



**UNIVERSIDAD
DE GRANADA**

**Variación morfológica y corpus lingüístico:
los diminutivos -ICO, -ITO, -ILLO en la provincia
de Almería (siglos XVIII y XIX)**

TESIS DOCTORAL

Pilar Arrabal Rodríguez

Dirigida por
Dr. Miguel Calderón Campos

GRANADA, 2023



**UNIVERSIDAD
DE GRANADA**

Variación morfológica y corpus lingüístico:
los diminutivos -ICO, -ITO, -ILLO en la provincia de
Almería (siglos XVIII y XIX)

TESIS DOCTORAL

Pilar Arrabal Rodríguez

Dirigida por
Dr. Miguel Calderón Campos

Programa de Doctorado en Lenguas, Textos y Contextos

Departamento de Lengua Española

Granada, 2023

Editor: Universidad de Granada. Tesis Doctorales
Autor: Pilar Arrabal Rodríguez
ISBN: 978-84-1195-165-4
URI: <https://hdl.handle.net/10481/89697>

Esta tesis doctoral se ha realizado gracias al Proyecto ALEA XVIII «Atlas lingüístico y etnográfico de Andalucía, s. XVIII. Patrimonio documental y Humanidades Digitales», financiado por FEDER / Junta de Andalucía – Consejería de Transformación Económica, Industria, Conocimiento y Universidades. Proyecto de referencia P18-FR-695. Asimismo, se enmarca en el Proyecto ALEA oriental-XVIII, financiado por FEDER / Junta de Andalucía – Consejería de Transformación Económica, Industria, Conocimiento y Universidades. Proyecto de referencia A-HUM-116-UGR20; y en el Proyecto HISPATESD «Hispanae Testium Depositiones. Las declaraciones de testigo en la historia de la lengua española (1492-1833)», financiado por el Ministerio de Economía, Industria y Competitividad. Referencia FFI2017-83400-P.

Los tres proyectos han sido desarrollados por el grupo de investigación «Diacronía de la lengua española» (DiLEs) (UGR-HUM-278), de la Universidad de Granada.

*No creo haber agotado los problemas del diminutivo.
Al revés: nunca hubiera sospechado la existencia de
tantos problemas, en paciente espera del investigador,
como los que se levantan en cada una de las soluciones
que aquí presento.*

Amado Alonso, 1961

*The beginning of any corpus study is the
creation of the corpus itself. The decisions that
are taken about what is to be in the corpus,
and how the selection is to be organized,
control almost everything that happens
subsequently. The results are only as good as
the corpus.*

John Sinclair, 1991

Agradecimientos

A quienes me han acompañado a lo largo de estos años debo reconocerles la infinita gratitud que les tengo por haberme ayudado, de un modo u otro, a completar esta etapa.

A mi director y tutor, Miguel Calderón Campos, por enseñarme mucho más de lo que ha implicado redactar estas páginas y desarrollar esta investigación. Mi más sincero agradecimiento por guiarme, asesorarme y dedicarme su tiempo desde el primer día que me abrió las puertas de su despacho y me propuso hacer la tesis. Por su confianza, paciencia y generosidad. Me ha transmitido su afán de superación, el rigor científico y otros tantos valores que son ya de por vida y que le agradeceré siempre.

En general, a todo el equipo que integra el grupo DiLEs, en el que he aprendido tanto y en el que espero seguir haciéndolo. A Gael Vaamonde, por su ayuda con la estadística y sus explicaciones una y otra vez cuando las he necesitado. Siempre dispuesto a enseñarme y a resolver las dudas más técnicas. A Mayte García-Godoy, por sus sabios consejos, ejemplo para mí de constancia y excelencia. A Inma González Sopena, por su bondad y predisposición. A Antonio Martín-Rubio, compañero en todo. A Yisela Ortiz, por su afecto cuando más me ha hecho falta. Y a Diego Reinaldos, por sus ánimos y sus diminutivos en *-ico* (mis preferidos).

Mi agradecimiento también a los profesores, miembros del departamento y compañeros de profesión que me han manifestado su apoyo y con los que me he cruzado durante estos años. Particularmente a la profesora Miriam Bouzouita, quien me ha brindado oportunidades muy enriquecedoras para mí y a quien tengo gran estima.

A mis compañeros de carrera, Rubén, Victoria, Ana Eva y Ana. La Filología me puso delante a personas de cuya motivación y constancia era imposible no contagiarse. Y en especial a Lurdes, quien me ha inspirado para muchas cosas de las que ella probablemente no sea consciente.

No puedo olvidarme de mis familiares, primas y amigos. Aquellos que han estado ahí cuanto tuve que dejar de estar y a quien también quiero dar las gracias.

A mis abuelos paternos. De ellos escuché por primera vez muchas de las palabras que me he ido encontrando en los textos de esta tesis y que he leído con sus voces. En mi cabeza se iban apareciendo los objetos que han formado parte del mobiliario de la casa del pueblo durante toda mi infancia y adolescencia, sin llegar a imaginar el valor que posteriormente tendrían para esta investigación. Y a mi abuela Pilar, mariposa curiosa; sé que este momento la llena de orgullo.

A mis padres, Miguel y Mara, por hacerme llegar hasta aquí y transmitirme la importancia del esfuerzo y la sensibilidad por los libros. Este logro es vuestro. A mi hermana Esther, por completarme con las cualidades que a mí me faltan. De ellos he recibido el apoyo discreto, pero el que se sabe que es incondicional.

Y a Juan Carlos. Siempre a mi lado, más aún en los momentos difíciles de estos años. Me ha hecho seguir adelante gracias a su cariño, comprensión y la admiración depositada en mí. Un día, mucho antes incluso de que este día se encontrara entre mis aspiraciones y se iniciara este camino, me dijo que yo tenía que ser doctora. Sé que sin él no habría sido posible acabar esta tesis, ni siquiera empezarla. Gracias, por tanto.

Resumen

Esta tesis doctoral tiene como objetivo reexaminar desde el punto de vista histórico la distribución y frecuencia de los sufijos diminutivos, particularmente en el español almeriense de los siglos XVIII y XIX a partir de inventarios de bienes. La idoneidad de esta fuente textual se asienta en dos aspectos fundamentales: la datación y localización geográfica inequívoca que caracteriza a este tipo de documentos y la ausencia del tono solemne con el que fueron redactados, tendentes a la practicidad y a la muestra de rasgos cercanos a la oralidad, entre los que destaca la presencia de diminutivos.

El periodo analizado se considera fundamental para la constitución del denominado español moderno en la provincia de Almería, una región poco atendida a pesar de su localización geográfica como frontera entre dos variedades lingüísticas fuertemente caracterizadas: el habla andaluza y el habla murciana.

Para ello se ha elaborado un corpus diacrónico especializado con un volumen de 53 300 palabras. Al mismo tiempo, se incluye un corpus de control con el que poder analizar en términos contrastivos el uso de los diminutivos en la provincia andaluza. De forma paralela al corpus de estudio, se proyecta un subcorpus con inventarios de Madrid de estas dos centurias. En total, el corpus documental se aproxima a las 119 000 palabras y está compuesto por 97 inventarios de bienes de Almería y Madrid pertenecientes a los siglos XVIII y XIX.

Los textos se integran en el corpus *Oralia Diacrónica del Español (ODE)*, que se caracteriza por seguir una metodología específica para su elaboración con el uso de la herramienta TEITOK (Janssen, 2016). Este sistema precisa la transcripción de los textos según el estándar XML-TEI y permite la publicación en línea de una triple edición digital (imagen del facsímil, edición semipaleográfica y normalización ortográfica), así como el procesamiento lingüístico y aprovechamiento del corpus. De especial relevancia para este estudio es la adjudicación de etiquetas morfológicas (POS) y lemas. Los problemas específicos que conlleva la anotación lingüística se tratan en el caso concreto del léxico de los inventarios de un corpus especializado como *ODE*.

El estudio de los diminutivos parte de la distinción metodológica entre diminutivos lexicalizados y diminutivos transparentes. Con respecto a los primeros, se proporciona un glosario que recoge 100 entradas en las que se incluyen las definiciones, acepciones y ejemplos que apoyan los significados manifestados en el corpus. En muchas ocasiones, *ODE* ha servido para mejorar la información lexicográfica disponible de determinadas voces, algunas de ellas escasamente documentadas. En cuanto a los diminutivos transparentes, se realiza un análisis estadístico que, partiendo de un contraste de hipótesis, establece diferencias significativas en el uso y distribución de los diminutivos en los periodos y lugares estudiados. Se confirman tendencias opuestas en el uso de los diminutivos, tanto desde el eje diatópico entre Almería y Madrid, como desde el eje cronológico entre los siglos XVIII y XIX, siendo las diferencias más destacadas en el caso de la provincia andaluza.

Palabras clave: diminutivos, inventarios de bienes, siglo XVIII y XIX, lingüística de corpus, Almería, Madrid.

Abstract

This thesis aims to re-examine the historical distribution and frequency of diminutive suffixes in the Spanish of Almería in the 18th and 19th centuries through inventories of goods. The suitability of this textual source is based on two main aspects: the clear dating and geographical location which characterises this type of documents and the absence of the formal style, which tends towards functionality and the use of features close to the oral language, especially the presence of diminutives.

The period analysed is considered essential for the constitution of the modern Spanish in the province of Almería, a region which has received little attention despite its geographical location as a frontier between two strongly characterised linguistic varieties: the Andalusian and Murcian modalities.

For this purpose, a specialised diachronic corpus with a volume of 53 300 words has been compiled. At the same time, it includes a control corpus with which to contrast the use of diminutives in the Andalusian province. This contains inventories of Madrid from the same two centuries. In total, the whole corpus is close to 119 000 words and is made up of 97 inventories of goods in Almería and Madrid from the 18th and 19th centuries.

The texts are integrated into the corpus *Oralia Diacrónica del Español (ODE)*, which is characterised by a specific methodology for its elaboration using the TEITOK tool (Janssen, 2016). This system requires the transcription of the texts according to the XML-TEI standard and allows the online publication of a triple digital edition (facsimile image, semi-palaeographic edition and standard edition), as well as the linguistic processing and use of the corpus. Of relevance to this study is the allocation of morphological tags (POS) and lemmas. The specific problems involved in linguistic annotation are dealt with in the specific case of the lexicon of the inventories of a specialised corpus such as ODE.

The study of diminutives is based on the methodological distinction between lexicalised diminutives and transparent diminutives. Regarding lexicalised forms, a glossary is provided with 100 entries including definitions, meanings and examples that support the meanings expressed in the corpus. On many occasions, ODE has served to improve the lexicographic information available on certain voices, some of which are scarcely documented. With regard to transparent diminutives, based on a hypothesis test, a statistical analysis establishes significant differences in the use and distribution of diminutives in the periods and places studied. Opposing trends in the use of diminutives are confirmed, from the point of view of their diachronic and diatopic evolution between Almería and Madrid and between the 18th and 19th centuries. Differences are more outstanding in the case of the Andalusian province.

Keywords: diminutives, inventories of goods, 18th and 19th century, corpus linguistics, Almería, Madrid

ÍNDICE

1. INTRODUCCIÓN.....	21
1.1. Presentación.....	21
1.2. Objetivos y metodología.....	22
1.3. Estructura de la tesis	25
2. CORPUS DOCUMENTAL	27
2.1. Antecedentes.....	27
2.2. Panorama actual de los corpus históricos del español desde la perspectiva de la lingüística de corpus	30
2.3. Criterios de selección.....	35
2.3.1. Límites geográficos	35
2.3.2. Marco cronológico.....	38
2.3.3. Tipología textual.....	40
2.4. Archivos consultados y recopilación del material	43
2.5. Descripción del corpus y características generales.....	45
2.5.1. Acceso al corpus	47
2.5.2. Corpus de estudio de Almería	48
2.5.3. Corpus de control de Madrid	53
2.6. Corpus complementarios	58
3. EDICIÓN DIGITAL DEL CORPUS EN XML-TEI	61
3.1. Criterios de transcripción.....	61
3.2. Transcripción en lenguaje XML	63
3.2.1. Conceptos básicos	64
3.2.2. Reglas de buena formación	66
3.2.3. Editor de texto XML	69
3.3. Codificación del manuscrito en XML-TEI.....	70
3.3.1. Cabecera <teiHeader>	72
3.3.1.1. Descripción bibliográfica del archivo <fileDesc>	72
3.3.1.2. La tipología textual <encodingDesc>	76
3.3.1.3. Algunos aspectos no bibliográficos <profileDesc>	76
3.3.1.4. Historial de revisiones del documento <revisionDesc>	78
3.3.2. Texto <text>	78
3.3.2.1. Estructura básica.....	79

3.3.2.2. Abreviaturas	81
3.3.2.3. Características del texto.....	83
3.3.2.4. Intervenciones del escribano	88
3.3.2.5. Conjeturas del editor.....	91
4. CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK.....	95
4.1. Empleo de la herramienta TEITOK.....	95
4.2. Tokenización	98
4.3. Normalización ortográfica.....	102
4.4 Funcionamiento del anotador lingüístico en TEITOK	105
4.5. Anotación morfosintáctica (POS) en el corpus <i>ODE</i>	110
4.5.1. Etiquetario	110
4.5.2. Casos especiales en la anotación morfosintáctica	112
4.5.2.1. Participios y adjetivos.....	112
4.5.2.2. Formas verbales.....	113
4.5.2.3. Adverbios	114
4.5.2.4. Locuciones.....	115
4.5.2.5. Palabras en lengua extranjera	115
4.6. Lematización	115
4.6.1. El concepto de lematización frente al de normalización	116
4.6.2. Decisiones metodológicas para la adjudicación de lemas	119
4.6.2.1. Sustantivos.....	119
4.6.2.2. Derivados nominales	121
4.6.2.3. Plurales inherentes.....	127
4.6.2.4. Adjetivos.....	134
4.6.2.5. Determinantes.....	135
4.6.2.6. Pronombres.....	136
4.6.2.7. Verbos.....	136
4.6.2.8. Adverbios	137
4.6.2.9. Numerales.....	138
4.6.2.10. Palabras invariables	138
4.6.2.11. Palabras con sufijación apreciativa.....	139
4.7. Triple visualización de los textos	139
4.8. Búsquedas en el corpus basadas en lenguaje CQL.....	142

4.8.1. Principios básicos de formación CQL	143
4.8.2. Uso de expresiones regulares como parte de la sintaxis CQL.....	146
4.8.3. Casos prácticos de búsquedas en el corpus ODE	151
4.8.3.1. Distinguir entre mayúsculas y minúsculas	152
4.8.3.2. Combinar varios criterios en una búsqueda.....	153
4.8.3.3. Especificar el contexto adyacente.....	153
4.8.3.4. Mostrar distintas opciones en una sola búsqueda	155
4.8.3.5. Excluir elementos de una búsqueda.....	157
4.8.3.6. Dejar huecos entre palabras	159
4.8.3.7. Indicar elementos opcionales	161
4.8.3.8. Calcular el número de tokens en el corpus	161
4.9. Interfaz de consulta del corpus <i>ODE</i>	162
4.9.1. Búsqueda simple.....	163
4.9.2. Búsqueda avanzada.....	164
4.9.3. Búsqueda comparada.....	169
4.9.4. Búsqueda en el mapa	170
4.9.5. Búsqueda genérica.....	171
4.9.6. Búsqueda con XPath.....	172
4.9.7. Distribución de palabras en el corpus.....	173
4.9.8. Otras formas de recuperación de datos.....	174
4.10. Revisión manual de errores en el corpus	175
4.10.1. Modificar la información asociada a una palabra.....	177
4.10.2. Corregir errores en bloque mediante búsquedas en el corpus	178
4.10.3. Verificar la anotación y corregir etiquetas mal construidas	181
4.10.4. Comprobar errores en la asignación de lemas	182
5. ESTADO DE LA CUESTIÓN	187
5.1. Caracterización del diminutivo.....	187
5.2. Periodización y extensión geográfica de <i>-ico</i> , <i>-ito</i> e <i>-illo</i>	192
5.3. El diminutivo en la actualidad: frecuencia y distribución	196
5.4. Fuentes para el estudio de los diminutivos.....	198
6. LOS DIMINUTIVOS <i>-ICO</i>, <i>-ITO</i>, <i>-ILLO</i> EN EL CORPUS	203
6.1. Consideraciones generales.....	203
6.2. Significado transparente y lexicalizado de los sufijos <i>-ico</i> , <i>-ito</i> , <i>-illo</i>	206
6.3. Criterios para la desambiguación de diminutivos transparentes y lexicalizados.....	214

6.3.1. El contexto como elemento desambiguador.....	219
6.3.2. Casos dudosos y problemas para la cuantificación	226
<i>Azadilla</i>	227
<i>Bacinilla</i>	228
<i>Banquillo</i>	229
<i>Braserilla</i>	233
<i>Cencerrilla</i>	234
<i>Cortinilla</i>	235
<i>Cucharilla</i>	236
<i>Mesilla</i>	237
<i>Platillo</i>	238
<i>Puntilla</i>	240
6.4. Diminutivos transparentes	241
6.4.1. Recuento de formas diminutivas en el corpus	241
6.4.2. Valor semántico del diminutivo en el corpus	244
6.5. Diminutivos lexicalizados	248
6.5.1. Recuento de diminutivos lexicalizados en el corpus	248
6.5.2. Características de los diminutivos lexicalizados	250
6.5.2.1. Regionalismos y voces poco documentadas.....	253
<i>Algodoncillo, albornocillo y borloncillo</i>	253
<i>Bastillo</i>	254
<i>Batistilla</i>	254
<i>Confitillo</i>	256
<i>Escamilla y raspilla</i>	257
<i>Francesita</i>	258
<i>(Lienzo) apretadillo y (lienzo) medianillo</i>	259
<i>Granillo</i>	261
<i>Horquillo</i>	263
<i>Manilla</i>	264
<i>Pajecillo</i>	265
6.5.2.2. La variación del diminutivo en voces opacas: el caso particular de «bufetillo».....	267
6.5.3. Glosario de lexicalizaciones	280
6.5.3.1. Estructura de las entradas	281

6.5.3.2. Fuentes utilizadas para la elaboración del glosario	283
6.5.3.3. Abreviaturas y signos empleados	289
6.5.3.4. Glosario	290
7. ANÁLISIS ESTADÍSTICO DE LOS DIMINUTIVOS -ICO, -ITO E -ILLO	313
7.1. Corpus y datos analizados	313
7.2. Variables e hipótesis de estudio.....	316
7.3. Modelo estadístico empleado y estructura del análisis.....	318
7.4. Resultados.....	322
7.4.1. Corpus de Almería.....	322
7.4.1.1. Dispersión de los datos en el corpus de Almería.....	322
7.4.1.2. Relación estadística entre diminutivo y siglo en el corpus de Almería.....	324
7.4.1.3. Disminución de -ico y aumento de -illo en el corpus de Almería	330
7.4.2. Corpus de Madrid	333
7.4.2.1. Dispersión de los datos en el corpus de Madrid	333
7.4.2.2. Relación estadística entre diminutivo y siglo en el corpus de Madrid	334
7.4.2.3. Alternancia de -ito/-illo en el corpus de Madrid.....	338
7.4.3. Análisis comparativo entre subcorpus	339
7.4.3.1. Contraste diatópico Almería vs. Madrid en el siglo XVIII	340
7.4.3.2. Contraste diatópico Almería vs. Madrid en el siglo XIX.....	343
7.4.3.3. Distribución geográfica	346
7.5. Discusión	349
8. CONCLUSIONES.....	355
8.1. Futuras vías de investigación.....	359
SUMMARY AND CONCLUSIONS.....	363
REFERENCIAS BIBLIOGRÁFICAS	369
Corpus.....	387
ÍNDICE DE TABLAS Y FIGURAS.....	389
ANEXOS	393
Anexo I. Ejemplo completo de cabecera etiquetada.....	395
Anexo II. Ejemplo completo de texto etiquetado	398
Anexo III. Catálogo de etiquetas XML-TEI utilizadas en <i>ODE</i>	404
Anexo IV. Relación de diminutivos transparentes incluidos en el análisis.....	408

Listado de abreviaturas y signos

acep.	Acepción
AHPAL	Archivo Histórico de Protocolos de Almería
AHPM	Archivo Histórico de Protocolos de Madrid
CQL	<i>Corpus Query Language</i>
CQP	<i>Corpus Query Processor</i>
FE	Frecuencia esperada
ff.	folios
FO	Frecuencia observada
FR	Frecuencia relativa
LC	Lingüística de corpus
POS	Etiqueta morfosintáctica (<i>part-of-speech</i>)
r.	recto
sf	sin foliación
s.v.	sub voce
v.	vuelto
XML	<i>Extensible Markup Language</i> : lenguaje de edición digital
*	Se antepone a construcciones agramaticales
@	Señala atributos en lenguaje XML

1. INTRODUCCIÓN

1.1. Presentación

Esta tesis doctoral surge del interés por reexaminar desde el punto de vista histórico la distribución y frecuencia de los sufijos diminutivos, particularmente en el español almeriense de los siglos XVIII y XIX, tomando como referencia documentación notarial. Este periodo, fundamental para la constitución del denominado español moderno, se caracteriza por asistir a un gran cambio en el paradigma de los diminutivos que sentaría las bases para su actual distribución. La bibliografía previa suele subrayar la prevalencia de *-ito* sobre *-illo* por primera vez en el siglo XVIII, basándose principalmente en el estudio de fuentes literarias para hacer recuentos de frecuencias y cálculos de porcentajes. Frente a los textos literarios o cultos, esta investigación parte de la tipología textual “inventario de bienes” para el estudio de los diminutivos. Su idoneidad para este fin se asienta en dos aspectos considerados fundamentales: la datación y localización geográfica inequívoca que caracteriza a este tipo de documentos, por un lado, y el carácter utilitario con el que fueron redactados, por otro, desprovisto de intención retórica e influenciado a veces por rasgos cercanos a la oralidad, entre los que se encuentra la abundancia en el uso de diminutivos.

Por otra parte, los estudios históricos sobre la materia han abordado mayoritariamente este aspecto desde una perspectiva general, mientras que la distribución de los diminutivos se complica a medida que se desciende a analizar más minuciosamente variedades particulares del español (Náñez Fernández, 2006, p. 33). Con esta tesis se pretende contribuir al estudio de la periodización del fenómeno a partir del análisis dialectal de la provincia de Almería, una región poco atendida y que sin embargo destaca por su localización geográfica como frontera entre dos variedades lingüísticas fuertemente caracterizadas: el habla andaluza y el habla murciana.

Para ello se ha elaborado un corpus diacrónico especializado con inventarios almerienses de los siglos XVIII y XIX con un volumen de 53 300 palabras. Al mismo tiempo, se incluye un corpus de control con el que poder contrastar el uso de los diminutivos en la provincia andaluza. De forma paralela al corpus de estudio, se proyecta un subcorpus con inventarios de Madrid de estas dos centurias. En total, el corpus documental en su conjunto se aproxima a las 119 000 palabras y está compuesto por 97 inventarios de bienes de Almería y Madrid pertenecientes a los siglos XVIII y XIX.

Los documentos se integran en *Oralia Diacrónica del Español (ODE)*¹, un corpus que se caracteriza por seguir una metodología específica para su elaboración de acuerdo con los últimos avances informáticos en la configuración de corpus históricos. La transcripción de los textos se ha desarrollado siguiendo el estándar XML-TEI con el objetivo de poder ofrecer una triple edición digital en línea (imagen del facsímil, edición semipaleográfica y normalización ortográfica) por medio de la herramienta TEITOK

¹ <http://corpora.ugr.es/ode/>

(Janssen, 2016). Este sistema es el que se ha utilizado para el procesamiento de los textos, así como para la anotación lingüística y aprovechamiento del corpus a partir de la interfaz de búsquedas.

La elaboración del corpus constituye uno de los ejes sobre los que se vertebra la tesis doctoral. La mayor atención se ha centrado en las tareas de diseño, procesamiento y posibilidades de explotación de los datos como una parte importante y necesaria de la que deben partir las investigaciones en lingüística histórica. De entre las funcionalidades de *ODE* destaca la anotación morfosintáctica mediante la adjudicación de etiquetas POS y la lematización de los textos. Los procesos automáticos de anotación lingüística se han desarrollado teniendo en cuenta el léxico específico de los inventarios de bienes y han estado ligadas a una revisión manual posterior para garantizar la creación de un corpus útil y fiable en el que puedan basarse los estudios diacrónicos.

La tesis se centra en el estudio de todas las voces con sufijación diminutiva en el corpus. Estas se han clasificado en dos grandes grupos: 1) aquellas en las que el sufijo ha adquirido un significado específico fruto de un proceso de lexicalización (diminutivos lexicalizados) y 2) aquellas en las que el sufijo conlleva un auténtico valor de disminución (diminutivos transparentes). Con respecto a las primeras, se proporciona un glosario que recoge 100 entradas en las que se incluyen las definiciones, acepciones y ejemplos que apoyan los significados manifestados en el corpus. En muchas ocasiones, *ODE* ha servido para mejorar la información lexicográfica disponible de determinadas voces, algunas de ellas escasamente documentadas.

En cuanto a los diminutivos transparentes, gracias al etiquetado morfosintáctico de *ODE* el listado completo de todos ellos se ha obtenido automáticamente a partir de una sola búsqueda. El recuento ha permitido realizar un análisis estadístico que, partiendo de un contraste de hipótesis, trata de establecer diferencias significativas en el uso y distribución de los diminutivos en Almería y Madrid durante los siglos XVIII y XIX. Las pruebas estadísticas determinan el tipo de relación que existe entre las distintas variables de estudio y confirman tendencias opuestas en el uso de los diminutivos, tanto desde el eje diatópico, como desde el eje cronológico, siendo las diferencias más destacadas en el caso de la provincia andaluza. El corpus utilizado asegura la comparabilidad de los textos y permite la cuantificación de los datos para el estudio de fenómenos morfológicos desde un enfoque variacionista.

1.2. Objetivos y metodología

Esta investigación tiene como principal objetivo examinar la evolución y frecuencia de distribución de los diminutivos *-ico*, *-ito* e *-illo* en la provincia de Almería durante los siglos XVIII y XIX a partir de un corpus conformado por inventarios de bienes. Se pretende analizar la cronología de los diminutivos en esta provincia, pero también abordar la cuestión desde una perspectiva diatópica, para lo que se incluye en el corpus documentación procedente de Madrid que sirva de contraste. Para alcanzar este propósito, se establecen unos objetivos previos enfocados, por un lado, a la elaboración

del corpus que servirá de estudio, por otro, al análisis de los diminutivos a la luz de los nuevos datos. Este objetivo general se desglosa a su vez en otros más específicos:

a) Objetivos de carácter tecnológico:

1. Seleccionar documentación manuscrita inédita y de interés lingüístico para crear un corpus de inventarios de bienes de los siglos XVIII y XIX procedentes de las provincias de Almería (corpus de estudio) y de Madrid (corpus de control).
2. Realizar una edición digital del corpus textual siguiendo los criterios propuestos por el estándar XML-TEI.
3. Normalizar, etiquetar morfosintácticamente y lematizar el corpus con la herramienta TEITOK.
4. Describir todo el proceso de transcripción y anotación del corpus, de manera que pueda servir como guía de trabajo para la elaboración de futuras ediciones digitales que utilicen TEITOK y, más concretamente, para los editores de nuevos textos que se integren en ODE.
5. Configurar un corpus documental de acceso libre y formato electrónico que facilite la posterior recuperación de los datos de acuerdo con los últimos avances de la lingüística computacional (CQL) y con el que se asegure la fiabilidad del etiquetado morfosintáctico y lematización, especialmente para las palabras con sufijación diminutiva.

b) Objetivos de carácter filológico:

6. Establecer los criterios lingüísticos que permiten diferenciar entre diminutivos transparentes y diminutivos lexicalizados y aplicar esos criterios al vocabulario de los inventarios de bienes del corpus.
7. Determinar el valor semántico del sufijo en las voces lexicalizadas en el corpus que se han creado mediante sufijación diminutiva.
8. Elaborar un glosario de los diminutivos lexicalizados del corpus para complementar la información lexicográfica disponible.
9. Cuantificar y analizar en términos estadísticos la distribución cronológica (ss. XVIII y XIX) y diatópica (Almería y Madrid) de los diminutivos transparentes en *-ico*, *-ito* e *-illo*.

Los estudios previos relativos a la cronología y distribución diatópica de los diminutivos desde una perspectiva histórica se han basado fundamentalmente en textos literarios. Las principales limitaciones que conlleva el manejo de este tipo de textos se relacionan con su dudosa datación y con las reconstrucciones estereotipadas de la lengua oral que en ellos se hace (§ 5.4. *Fuentes para el estudio de los diminutivos*). Por ello, para la recopilación del corpus de esta tesis se ha optado por la selección de documentos que puedan ser datados y localizados geográficamente de manera inequívoca y que hayan sido elaborados en situaciones naturales. La documentación de

archivo cumple sobradamente con estas características, por lo que se considera una fuente fiable cuando se trata de vincular con certeza fenómenos lingüísticos a determinados lugares o épocas. En concreto, los inventarios de bienes se prestan de manera excepcional al estudio de los diminutivos, pues era una práctica habitual de los tasadores de bienes matizar el tamaño de los objetos por medio del uso de diminutivos o de la adjetivación. Así, se ha creado un corpus conformado únicamente por inventarios de bienes con el que se asegure además la comparabilidad y el equilibrio que proporciona un corpus elaborado en torno a un solo tipo textual.

En primer lugar, se ha hecho una edición digital de los textos siguiendo el estándar XML-TEI. En segundo lugar, los textos se han incorporado a la plataforma informática TEITOK, donde han sido normalizados y anotados morfosintácticamente siguiendo los últimos avances en lingüística computacional y se encuentran disponibles para su consulta (§ 2.5.1. *Acceso al corpus*). Esta herramienta permite además la explotación del corpus gracias a la triple visualización de los textos y a la recuperación de la información a partir de la interfaz de búsquedas.

Durante el proceso de creación del corpus ha tenido especial importancia la anotación morfosintáctica de los textos. De manera concreta para los objetivos de esta tesis, el etiquetado de los diminutivos se ha basado en la aplicación de unos criterios previamente establecidos para determinar el grado de transparencia del diminutivo y asegurar que solo los considerados transparentes son etiquetados como auténticos diminutivos (§ 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*). Al proceso automático de anotación le ha seguido el de revisado manual, necesario para garantizar la fiabilidad del etiquetado y de los resultados obtenidos a partir de las búsquedas basadas en características morfosintácticas.

Para el estudio lingüístico de los diminutivos ha sido fundamental la anotación morfosintáctica del corpus y su posterior revisión. La recuperación selectiva de información en el corpus ha permitido obtener de forma automática el cómputo de todos ellos. Partiendo de este recuento, se realiza un análisis cuantitativo preciso que examine en términos estadísticos la evolución de los diminutivos en Almería y arroje nuevos datos sobre el habla andaluza oriental en esta provincia.

A las diferencias cronológicas se añade además el interés por examinar los datos desde una perspectiva diatópica. Los resultados obtenidos para la provincia de Almería se comparan con un subcorpus de la provincia de Madrid, también conformado por inventarios pertenecientes al mismo marco temporal (ss. XVIII y XIX). De esta manera, junto al corpus de estudio de la provincia de Almería, se proyecta además la creación de un corpus de control que permita completar el análisis cronológico con los datos que proporciona la variable geográfica en cada caso. Las hipótesis y variables de estudio se exponen detenidamente en § 7.2. *Variables e hipótesis de estudio* y la metodología específica del análisis cuantitativo de los diminutivos se explica en § 7.3. *Modelo estadístico empleado y estructura del análisis*.

En definitiva, esta investigación pretende aportar nuevos datos al estudio de los diminutivos en una zona relevante para el estudio de las hablas andaluzas orientales como es Almería; y lo hace a partir de un análisis variacionista en el que los datos obtenidos se contrastan con los de Madrid, provincia alejada geográficamente lo

suficiente como para esperar obtener diferencias significativas. El estudio aplica nuevas técnicas de análisis con las que ofrecer una visión más detallada y completa de la que disponemos hasta ahora sobre el uso de los diminutivos durante los siglos XVIII y XIX en estas provincias. La selección documental y la elaboración del corpus ocupan en esta tesis una parte importante, pues las tareas de diseño, procesamiento y explotación de los datos juegan un papel fundamental y necesario en los trabajos que pretenden contribuir al estudio de la historia del español valiéndose de las avanzadas técnicas informáticas que ofrece hoy en día la lingüística de corpus.

Esta tesis se alza por tanto con el propósito de revisar la evolución de los diminutivos utilizando un método de análisis estadístico adecuado a la cuantificación de los resultados obtenidos y partiendo de la creación de un corpus avanzado tecnológicamente.

1.3. Estructura de la tesis

Esta tesis se organiza en ocho capítulos. El primero de ellos es una introducción en la que se presenta el tema de estudio y se exponen los objetivos y la metodología empleada. Los capítulos restantes se dedican a la configuración del corpus y al análisis detallado del fenómeno lingüístico de interés.

El desarrollo de la investigación se apoya en buena parte en la elaboración del corpus que ha servido de base para el estudio de los diminutivos *-ico*, *-ito* e *-illo* en la provincia de Almería. Para reexaminar la evolución que afecta al paradigma de estos sufijos durante los siglos XVIII y XIX, se ha prestado especial atención a la creación de un corpus fiable y representativo que permita responder a los objetivos filológicos que nos hemos propuesto. Así, esta tesis se estructura en dos grandes bloques: el primero de ellos se dedica a la configuración del corpus, mientras que el segundo está destinado al análisis de los resultados derivados de este mismo corpus.

El primer bloque lo conforman los capítulos 2, 3 y 4 de la tesis, dedicados a la elaboración del corpus. El capítulo 2 (§ 2. *CORPUS DOCUMENTAL*) presenta una descripción de las características generales del corpus. En él se establecen los criterios para la selección de los textos y se expone la necesaria distinción entre corpus de estudio y corpus de control que articulará el posterior análisis cuantitativo. Para cada uno de los subcorpus se señalan los puntos geográficos representados y se presenta la nómina de documentos que lo integran.

Los capítulos 3 (§ 3. *EDICIÓN DIGITAL DEL CORPUS EN XML-TEI*) y 4 (§ 4. *CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK*) ahondan en la metodología con la que se ha elaborado el corpus. El tercer capítulo está dedicado a la edición digital, para la que se ha utilizado el estándar XML-TEI. Se exponen los criterios de transcripción y las normas que se han seguido para la codificación de los textos. Por último, el cuarto capítulo expone las distintas fases que se han seguido hasta llegar a conseguir un corpus normalizado, lematizado y anotado morfosintácticamente de acuerdo con los últimos avances de la lingüística computacional. El procesamiento informático se ha hecho a partir de la herramienta TEITOK, desde la que se realiza

además la búsqueda y recuperación de los datos. En este capítulo se dedica una parte importante también a esta cuestión, indispensable para poder explotar al máximo todas las posibilidades que ofrece un corpus avanzado tecnológicamente.

El segundo bloque se propone cumplir con los objetivos de carácter filológico formulados en esta tesis doctoral y por los cuales se acomete el estudio morfológico de los diminutivos *-ico*, *-ito*, *-illo* con dos claros propósitos: por un lado, diferenciar entre el valor transparente y lexicalizado del diminutivo en las palabras con sufijación diminutiva en el corpus; por otro, estudiar desde una perspectiva diacrónica y diatópica la variación que actúa sobre el uso de estos sufijos en los siglos XVIII y XIX en las provincias de Almería y Madrid. Este bloque está compuesto por los capítulos 5, 6 y 7.

En el capítulo 5 (§ 5. *ESTADO DE LA CUESTIÓN*) se incluye una revisión de los diferentes aspectos que han centrado el estudio de los diminutivos *-ico*, *-ito* e *-illo* en español. En él se presenta el estado de la cuestión principalmente desde las perspectivas diacrónica y diatópica del fenómeno, tanto de los estudios dedicados a épocas pasadas como los que se centran en el uso actual del diminutivo. El capítulo 6 (§ 6. *LOS DIMINUTIVOS -ICO, -ITO, -ILLO EN EL CORPUS*) contiene el estudio completo de los mencionados sufijos en el corpus de inventarios de la tesis. En él se aborda la distinción entre diminutivos lexicalizados y diminutivos transparentes, para la que ha sido necesario establecer unos sólidos criterios de clasificación. Aquellos lexicalizados, los cuales quedan fuera del análisis cuantitativo, se han incluido en un glosario en el que, junto a la definición del término, se ofrecen los ejemplos reales documentados en el corpus.

En el capítulo 7 (§ 7. *ANÁLISIS ESTADÍSTICO DE LOS DIMINUTIVOS -ICO, -ITO E -ILLO*) se realiza el análisis cuantitativo de los diminutivos. En él se desarrollan de manera más específica las preguntas e hipótesis de investigación y se lleva a cabo el análisis estadístico y la discusión de los datos. Este análisis es fruto del estudio de todos los términos con sufijación diminutiva (*-ico*, *-ito* e *-illo*) presentes en el corpus e incluye únicamente aquellos en los que se considera que el sufijo cumple con un valor semántico de aminoración del tamaño del objeto. El análisis de los resultados obtenidos se vertebra en base a la separación de los datos entre corpus base (Almería) y corpus de control (Madrid) para poder ofrecer un estudio diacrónico y diatópico de los diminutivos.

Al capítulo de análisis le sigue el capítulo 8 (§ 8. *CONCLUSIONES*), en el que se presentan las conclusiones derivadas de todo el proceso de estudio, así como otras posibles vías de investigación que quedan abiertas y que podrían complementar los resultados que aquí se exponen. Seguidamente se incluyen las conclusiones en inglés conforme a los criterios establecidos para la obtención de Doctorado Internacional.

Se cierra esta tesis con las referencias bibliográficas, el índice de tablas y figuras utilizadas en los distintos capítulos y los anexos.

2. CORPUS DOCUMENTAL

El corpus documental de esta investigación se compone íntegramente de inventarios de bienes procedentes de las provincias de Almería y Madrid de los siglos XVIII y XIX. Los documentos han sido seleccionados y editados *ad hoc* para el análisis que se desarrolla y constituyen un corpus especializado y avanzado tecnológicamente con el que poder extraer resultados concluyentes acerca de la evolución de los diminutivos en las provincias y siglos mencionados.

La proyección del corpus de esta tesis no se entiende sin aludir necesariamente a *Oralia diacrónica del español (ODE)*, el corpus del que forma parte (§ 2.1. *Antecedentes*) y que se inserta en la línea de los corpus especializados que se han ido desarrollando en las últimas décadas en el ámbito de los estudios diacrónicos. Las características de este tipo de corpus contrastan con las de los llamados corpus generales o de referencia, las cuales se desarrollan en § 2.2. *Panorama actual de los corpus históricos del español desde la perspectiva de la lingüística de corpus*.

Los criterios de selección que justifican la creación de un corpus como el que se utiliza en esta tesis se detallan en § 2.3. *Criterios de selección*. Para su elaboración, se han consultado los archivos históricos de las dos ciudades de interés y la selección del material ha dependido en algunos casos de las características de conservación de los documentos en estos archivos, tal y como se explica en § 2.4. *Archivos consultados y recopilación del material*.

Este corpus plantea a su vez la diferenciación entre corpus de estudio y corpus de control, necesaria para el contraste de los resultados. La composición y características de cada uno de estos subcorpus en base a esta distinción se aborda en § 2.5. *Descripción del corpus y características generales*. Por último, para el análisis de tipo cualitativo, se ha recurrido en ocasiones a la consulta de materiales adicionales externos a los que se aportan en esta tesis, los cuales se enumeran en § 2.6. *Corpus complementarios*.

2.1. Antecedentes

El corpus documental de esta tesis forma parte del corpus histórico *Oralia Diacrónica del Español (ODE)*. Este corpus, a su vez, es continuación del *Corpus diacrónico del español del reino de Granada (CORDEREGRA)*, nacido con el objetivo de estudiar fenómenos relacionados con la oralidad en declaraciones de testigos e inventarios de bienes comprendidos entre 1492 y 1833 del antiguo reino de Granada, formado por las actuales provincias de Málaga, Granada y Almería (Calderón Campos y García-Godoy, 2009; Calderón Campos, 2015). El *CORDEREGRA* llegó a alcanzar un volumen considerable cercano al medio millón de palabras, pero se centraba sobre todo en documentos granadinos. A pesar de constituir ya de por sí una interesante fuente de valor para los estudios históricos, el corpus no incluía documentación de otras zonas o regiones más alejadas con las que poder contrastar los fenómenos observados, y tampoco contaba con algunos de los adelantos informáticos que se han ido incorporando

en los últimos años y que permiten explotar al máximo los datos que ofrece. Como propósito de este afán de mejora, el corpus *ODE* surge con la intención de suplir las carencias que presentaba su predecesor y que estaban relacionadas principalmente con la representatividad geográfica del corpus y la metodología empleada para su elaboración². Para una muestra de las ventajas que aún *ODE*, así como de sus características tecnológicas, son esenciales los trabajos de Calderón Campos (2019a), Calderón Campos y Vaamonde (2020), Arrabal Rodríguez (2020) y Calderón Campos y Díaz-Bravo (2021).

ODE es actualmente un corpus mucho más amplio fruto de posteriores proyectos de investigación con los que se han ido incorporando textos procedentes de casi todas las provincias de Andalucía y de diferentes tipologías textuales: inventarios de bienes (Calderón Campos, 2018b; González Sopeña, 2022; Arrabal Rodríguez, 2023), declaraciones de testigo (García-Godoy, 2018) y certificaciones médicas (Calderón Campos, 2018a; Martín Rubio, 2023).

Los documentos seleccionados para esta tesis han sido las primeras aportaciones a *ODE* como un corpus ya modernizado tecnológicamente y cuyo objetivo es la ampliación geográfica con textos de nuevas regiones hispánicas. Con esta tesis, por tanto, se inició el proceso de incorporación de nuevos documentos, así como la puesta en marcha de un avanzado modelo tecnológico que permite mejorar el corpus ya existente. La creación del corpus documental que conforma esta tesis supone la aplicación de estas dos notables novedades:

En primer lugar, se mejora la representatividad de Andalucía oriental con la inclusión de documentos de Almería para ofrecer así un corpus completo, más rico y equilibrado entre las provincias orientales. Adicionalmente, la provincia almeriense ha tenido muy poca representación en los principales corpus históricos del español (véase más adelante § 2.3.1. *Límites geográficos*), por lo que se contribuye ahora a mejorar una situación bastante desfavorable para Almería y que sin embargo resulta imprescindible para completar los estudios del español de Andalucía. El aumento del corpus conlleva también la incorporación a *ODE* de documentos no andaluces con el objetivo de conseguir un subcorpus de control para poder contrastar la variedad andaluza. Está formado por inventarios de bienes de la Comunidad de Madrid del mismo periodo con los que establecer diferencias o similitudes comparativas entre ambas regiones. Con respecto al marco cronológico, se representan dos centurias clave para la configuración del español moderno: los siglos XVIII y XIX.

En segundo lugar, la ampliación del corpus ha supuesto un importante cambio en la metodología basado en la utilización del sistema TEITOK, una de las tecnologías más punteras en lingüística de corpus a cuyas características se dedica el Capítulo 4 de esta tesis (§ 4. *CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK*). El objetivo principal es el de poder ofrecer un corpus enteramente digital, lematizado y anotado

² Calderón Campos (2019a) expone las limitaciones que presenta el corpus *CORDEREGRA*, editado con un procesador de texto convencional como es *Microsoft Word*. Este método está en la base de la edición de numerosos corpus que hoy en día se ofrecen en línea, pero que no permiten aprovechar las ventajas de una edición digitalizada entendida en su sentido estricto.

lingüísticamente con el que realizar búsquedas complejas. Los avances tecnológicos permitirán además hacer un uso interdisciplinar del corpus debido a las diferentes ediciones que se ofrecen de un mismo texto. Este nuevo modelo tecnológico ha sido probado satisfactoriamente en otros corpus históricos como *Post Scriptum* (Vaamonde, 2015a).

El corpus pretende incorporar los últimos avances tecnológicos sin olvidar el rigor filológico de una documentación cuidadosamente seleccionada y que conserva en su edición las características paleográficas del manuscrito original, en consonancia con estándares previos de calidad como los propuestos por la Red CHARTA³ para la transcripción de textos antiguos.

La creación del corpus documental de esta tesis se hace coincidir con el diseño de otros corpus lingüísticos creados también a partir de algunas de las características ya mencionadas. De este modo, *ODE* se inspira en tres aspectos destacados de los siguientes corpus especializados:

Del *Corpus Léxico de Inventarios (CorLexIn)* se toma la idea de elaborar un corpus muy cerrado temáticamente en torno a inventarios de bienes. La principal ventaja de esta opción es la comparabilidad que ofrecen este tipo de documentos. En ellos se anotan casi siempre objetos muy similares, por lo que disponer de un corpus de inventarios de bienes de distintas áreas y de distintas épocas permitirá comparar el uso de los diminutivos, tanto diatópica como diacrónicamente (Morala, 2012b, p. 421; Calderón Campos, 2019c, p. 112; Calderón Campos y Vaamonde, 2020, p. 171)

Del corpus *CHARTA (Corpus hispánico y americano en la red: textos antiguos)* se toma el estándar de edición de la red internacional. Para la elaboración del corpus de esta tesis se tomarán estos criterios con ligeras modificaciones manteniendo la triple presentación de los documentos (§ 3.1. *Criterios de transcripción*).

Del corpus *Post Scriptum (Archivo Digital de Escritura Cotidiana en Portugal y España en la Edad Moderna)* se toma el modelo tecnológico. Este corpus aún, por primera vez, la edición digital de manuscritos con la elaboración de un corpus lematizado y anotado morfológica y sintácticamente, lo que facilita las búsquedas complejas gracias a la herramienta TEITOK.

De acuerdo con lo ya expuesto, el corpus de esta tesis se alza como un subcorpus integrante de *ODE* creado exprofeso para esta investigación. Nos servirá para analizar en términos comparativos la evolución de los diminutivos en la provincia de Almería con respecto a Madrid, por lo que este mismo conjunto de textos incluye ya en sí mismo dos subcorpus: por un lado, la documentación procedente de Almería conformará nuestro corpus de estudio; por otro, los textos de Madrid constituirán el corpus de control con el que comparar los datos de aquel en torno al uso de los diminutivos.

Es importante destacar esta distinción, pues cuando en esta tesis nos referimos de manera general al corpus, debe entenderse en todo momento el corpus que ha servido para nuestro análisis, esto es, únicamente el conformado por los inventarios de Almería y Madrid, aunque este a su vez forme parte de *ODE*. Esta diferenciación se debe a que

³ <https://www.redcharta.es/>

estos últimos han sido concebidos y editados específicamente para esta investigación. Esta tesis doctoral supuso el punto de inicio para la mejora desde el punto de vista temático y también tecnológico del corpus *ODE*: en el primero de los casos, implicó la inclusión de los primeros documentos y de la creación de un corpus de control madrileño que permitiese contrastar los resultados obtenidos a partir de la documentación andaluza; en segundo lugar, significó la implantación de una nueva metodología a la hora de crear, editar el corpus y explotar sus datos. Hoy en día *ODE* alcanza casi el millón de palabras con cerca de 900 documentos entre los que se incluyen también los del antiguo *CORDEREGRA*, adaptado a los nuevos criterios metodológicos. Es justo reconocer que la creación y edición de los documentos de esta tesis no habría sido posible sin los avances tecnológicos de *ODE*⁴.

2.2. Panorama actual de los corpus históricos del español desde la perspectiva de la lingüística de corpus

En un momento en el que cada vez son más los corpus disponibles del español y los estudios diacrónicos que se basan en ellos, podría parecer innecesario detenerse a analizar qué entendemos por *corpus*. Sin embargo, dado que los corpus han ido evolucionando con el paso del tiempo, es lógico pensar que también lo haya hecho la propia definición del término. El sentido tradicional de *corpus* como ‘recopilación de textos’ ha quedado anticuado y responde ya únicamente a un uso vago y coloquial. Este sentido laxo del término difiere del concepto actual en dos nuevos aspectos que resultan indispensables: el soporte electrónico y la codificación⁵.

De manera más precisa, un corpus se define por ser un conjunto de textos producidos en situaciones naturales y representativos de una lengua o variedad lingüística concreta que se almacenan en formato electrónico y se codifican para lograr una recuperación selectiva de la información⁶. Esta codificación supone incluir información extratextual con la adición de metadatos y mejorar las posibilidades de explotación de los textos mediante la anotación lingüística (Rojo, 2021, pp. 1-3). El componente informatizado es esencial en una nueva generación de corpus en los que, de otra manera, sería imposible trabajar con grandes cantidades de datos, pero también la única opción posible para realizar los procesos de anotación lingüística, que, gracias a los avances informáticos en procesamiento del lenguaje, pueden hacerse de manera automática.

⁴ Los textos que conforman el corpus documental de esta tesis suponen el 15 % del total de los inventarios incluidos en *ODE* hoy en día (cifras actualizadas con fecha 20/06/2023).

⁵ Sobre los inicios de los corpus textuales informatizados véase Enrique-Arias (2009, pp. 11-15) y Torruella (2017, pp. 31-39).

⁶ De forma más precisa, Torruella afirma que solo los corpus informatizados pueden llamarse *corpus textuales* o *corpus lingüísticos*, ya que el término *corpus* en sí mismo es bastante ambiguo y general (2017, p. 33, nota 23).

La tipología de los corpus hoy en día es muy variada y depende tanto de las funcionalidades que ofrece el corpus como de los objetivos con los que se diseña⁷. Centrándonos sobre todo en los de tipo diacrónico, nos referiremos a sus características partiendo de la habitual distinción entre corpus generales o de referencia y corpus específicos.

Los corpus generales comprenden los grandes corpus del español compuestos por al menos 100 millones de palabras. Se consideran “de referencia” porque pretenden ser representativos de toda una lengua, incluidas sus variedades, y reflejar diferentes modalidades de textos. Al estar diseñados sin objetivos lingüísticos específicos, permiten su uso para muchos tipos de investigaciones en el estudio de variados fenómenos, y la anotación morfosintáctica y lematización de que disponen permiten multiplicar sus posibilidades. Algunos ejemplos de este tipo son el *Corpus del español* (CdE) de M. Davies o los corpus de la Real Academia, tanto diacrónicos como sincrónicos: *Corpus diacrónico del español* (CORDE), *Corpus del Diccionario histórico de la lengua española* (CDH), *Corpus de Referencia del Español Actual* (CREA), *Corpus del Español del Siglo XXI* (CORPES XXI). En el extremo de los corpus de gran tamaño se encuentran los llamados corpus masivos, mucho más alejados cuantitativamente de los corpus de referencia y sin una cuidada selección textual, pero muy avanzados tecnológicamente; un ejemplo es el corpus *esTenTen18* de *Sketch Engine* o aquellos que se basan en búsquedas de textos en la web (Rojo, 2021, p. 79).

Disponer de grandes corpus implica también tener que lidiar con ciertas limitaciones en torno a la calidad de los datos. Varios trabajos han hecho una revisión sobre las características y posibilidades de explotación de estos corpus sin olvidar algunos de sus puntos más débiles, que principalmente se centran en la dudosa datación de los textos y en la fiabilidad de los anotadores automáticos (Rojo, 2010; Vaamonde, 2015b; Rodríguez Molina y Octavio de Toledo y Huerta, 2017; Campos Souto, 2018; De Benito Moreno, 2019; Davies y Parodi, 2022)⁸.

Aunque lo hace desde el ámbito de la lexicografía, Campos Souto alude a algunas de las características básicas que debería poseer cualquier corpus representativo que aspire a ser una herramienta fiable para el estudio de la lengua en su vertiente diacrónica:

- a. Los textos del corpus han de estar lematizados.
- b. Los textos deben someterse a un proceso de codificación textual.
- c. La interfaz de consulta debe permitir una variada gama de búsquedas.
- d. El corpus debe ser amplio y representativo.
- e. Los corpus deben ser filológicamente fiables en cuanto a la edición y selección.
- f. La copia digital de los testimonios debe ser accesible (Campos Souto, 2018, p. 25).

⁷ Un amplio panorama de los diferentes tipos de corpus en español puede verse en Parodi et al. (2022).

⁸ De manera más específica, para las posibilidades de explotación y limitaciones del CDH se puede consultar el trabajo de Campos Souto (2018) y De Benito Moreno (2019), este último también para el resto de los corpus académicos. En Vaamonde (2015b) se analizan las principales desventajas del CdE y del CORDE, y cómo estas se resuelven en el caso de un corpus especializado.

Los corpus generales a los que se ha hecho referencia más arriba difícilmente pueden cumplir con todos estos aspectos debido a cuestiones fundamentalmente prácticas relacionadas con el vasto volumen de los datos que contienen. Los avances tecnológicos permiten que los textos sean anotados y lematizados cada vez con un mayor porcentaje de acierto, pero rara vez estos corpus están revisados y depurados de errores, algunos de ellos significativos y que interfieren notablemente en los resultados de las búsquedas. Como señala la autora, el desafío ahora está en mejorar los programas y estrategias de anotación morfosintáctica para que los corpus generales sean más fiables (Campos Souto, 2016, p. 41).

Por otro lado, no se hace referencia en estos corpus a los criterios gráficos de edición de los textos y tampoco se tiene acceso a la imagen del manuscrito, lo que imposibilita comprobar la fidelidad de las transcripciones. Especialmente relevante para los estudios diacrónicos, entre los que el corpus de esta tesis se inserta, es poder contar con textos fielmente datados cronológica y geográficamente que puedan situar en el tiempo y el espacio la existencia o ausencia de determinados fenómenos lingüísticos. Es ya un lugar común señalar que el interés prioritario de los grandes corpus se ha centrado en la cantidad y no en la calidad de los textos que ofrece. Por ello, frente a estos, en las últimas décadas se viene desarrollando otro tipo de corpus más reducidos, pero más especializados, con los que se intenta suplir las carencias de los corpus de referencia del español.

Los corpus específicos o especializados se diseñan con un objetivo concreto y aportan datos para describir un tipo particular de lengua. Independientemente del modelo tecnológico que empleen, más o menos avanzado según el caso, de ellos se destaca su compromiso con la calidad filológica de las transcripciones, así como la rigurosa selección y datación de los documentos, lo que los convierte en herramientas fiables para los estudios históricos. Constituyen corpus de este tipo algunos como *CHARTA*, *Post Scriptum*, *CorLexIn*, *Corpus diacrónico y diatópico del español de América (CORDIAM)*, *Corpus de documentos Españoles Anteriores a 1900 (CODEA+ 2022)*, *Biblia Medieval*, *Old Spanish Textual Archive (OSTA)* y *ODE*⁹.

Otro de los aspectos en los que se centran los corpus especializados es el de la preocupación por el equilibrio y la representatividad. Una de las limitaciones de los corpus generales se basa en el problema de la diversidad textual, el cual ha supuesto un verdadero obstáculo para la lingüística histórica, que se ha visto forzada a la adopción de una u otra vía de análisis a la hora de afrontar el estudio de la materia: por un lado, la que opta por los análisis de tipo cualitativo y centrada en el estudio de las tradiciones discursivas; por otro, la de los que abogan por la cuantificación de grandes cantidades de datos mediante la selección de diferentes tipos textuales. En los últimos años, esta última vía se ha erigido como la más fiable para el estudio histórico del español (Kabatek, 2013, p. 15). Actualmente, son más los corpus históricos que se erigen sobre

⁹ De manera conjunta, algunos trabajos han hecho una comparación de las características y posibilidades que ofrecen los principales corpus diacrónicos del español: Calderón Campos (2019b), Sánchez Lancis (2022), Bertolotti y Company (2022). El portal *CORHIBER (Corpus Históricos Iberorrománicos)* ofrece un elenco de los principales corpus iberorrománicos además de información básica de cada uno de ellos.

diferentes tipologías textuales (*CORDE*, *CORDIAM*, *CHARTA*, *ODE*, *CODEA*, etc.) que aquellos que se limitan temáticamente a una sola (*CorLexIn*, *Biblia Medieval*, *Post Scriptum*). La codificación de los textos en los corpus lingüísticos actuales, cada vez más fidedignos y avanzados tecnológicamente, permite filtrar, en el caso de los primeros, por tipo textual. Esta opción facilita los estudios de tipo contrastivo en los que la variable ‘tipo textual’ puede interferir en la interpretación de los resultados asegurando así el criterio de comparabilidad¹⁰.

La diversidad textual es solo un ejemplo más de la necesaria codificación a la que deben someterse los corpus diacrónicos y que se encuentra estrechamente ligada con el soporte informático empleado para albergar los textos. Sin olvidar la calidad filológica, los corpus deben desarrollar una serie de avances tecnológicos que permitan mejorar la búsqueda y explotación de la información. El corpus *ODE*, en el que se integra el corpus de esta tesis, forma parte de esos corpus especializados, no solo en la temática, sino también en lo tecnológico, “más centrados en la calidad de los datos que en la cantidad, pero sin excluir la representatividad cuantitativa basada en la buena selección de los textos” (Calderón Campos, 2018b, p. 318). El corpus muestra también la imagen de los documentos originales y distintas ediciones del texto. Asimismo, se pretende ofrecer un corpus lingüístico anotado y lematizado que permita obtener información a partir de búsquedas avanzadas en las que además se ofrecen diferentes visualizaciones de los datos (§ 4.9. *Interfaz de consulta del corpus ODE*). En definitiva, un corpus acorde a la definición actual del término “corpus” con la que se iniciaba este apartado.

La aparición de corpus de semejantes características ha hecho que cambie también la forma en la que se trabaja en los estudios lingüísticos. Los corpus actuales permiten recuperar cantidades ingentes de datos de manera rápida y precisa, por lo que la noción actual de *corpus* a la que hemos aludido conlleva también una nueva interpretación del término *lingüística de corpus* (LC). Con él se alude a la metodología en la que se integran principalmente las tareas relativas a las fases de 1) diseño y compilación del corpus, 2) procesamiento, búsqueda y explotación de los datos y 3) análisis mediante técnicas estadísticas e interpretación de los resultados (Paquot y Gries, 2020). La estrecha relación entre todas estas fases, así como las múltiples y complejas tareas que conlleva cada una explica que existan discrepancias entre aquellos que la consideran como un método más (enfoque, técnica, etc.) entre todos los disponibles, una herramienta o una disciplina en sí misma (Torruella, 2017, p. 25). Algunos autores han llegado incluso a precisar el término y hacer una distinción entre *lingüística de corpus*, relacionada con la creación y uso de los corpus, y *lingüística con corpus* (Kabatek, 2016) o *lingüística del corpus* (Caravedo, 1995, p. 18), para aludir al análisis lingüístico que se desprende de los datos obtenidos a partir de corpus. Sin embargo, hoy en día parece haber cierto consenso en considerarla como una metodología transversal a muchas disciplinas (Parodi et al., 2022).

¹⁰ Para Kabatek (2013) habría siempre que tener en cuenta que la sola clasificación de géneros discursivos y tipologías textuales no es homogénea, ya que dentro de cada uno de estos grupos se incluyen textos muy diferentes entre sí.

De lo que no cabe duda es de que la LC pretende aproximarse a los fenómenos lingüísticos a partir de datos basados en hechos reales, empíricos, y cuantos más, mejor, por lo que se hace imprescindible contar con corpus textuales amplios y representativos, cuya gestión solo es posible a partir de sistemas informáticos que lo permitan. Asimismo, el enfoque bajo la LC lleva intrínseco el realizar análisis sistemáticos y exhaustivos, de lo que se espera que el investigador sea capaz de procesar la información obtenida mediante técnicas estadísticas refinadas con las que extraer conclusiones y probar nuevas hipótesis (Rojo, 2021, p. 49). La tendencia hacia la aplicación de modelos estadísticos es ya una realidad en auge, puesto que han demostrado aportar resultados más precisos sobre los fenómenos analizados y matizar conclusiones basadas únicamente en los recuentos de frecuencias absolutas (Nieuwenhuijsen, 2016, p. 217)¹¹.

Tradicionalmente, todo el prestigio filológico y el valor de la investigación han sido atribuidos a la interpretación conceptual de los análisis derivados de corpus. La construcción del corpus y los métodos cuantitativos han sido consideradas cuestiones marginales y la implicación del investigador en la creación de herramientas de calidad, labores muy poco reconocidas. Afortunadamente, cada vez son más los proyectos que se afanan por mejorar y crear corpus lingüísticos del español con las características anteriormente descritas, trabajos que aportan una visión crítica de lo que ofrecen y deberían ofrecer los corpus actuales. Pero queda mucho camino todavía por recorrer hasta alcanzar la plena convicción de que cada una de estas ramas son parte imprescindible y complementarias de la LC y que los estudios diacrónicos actuales no son posibles sin todas y cada una de las tres fases a las que se ha hecho referencia más arriba.

Creemos firmemente que la noción de LC conlleva un cambio no solo en las características conceptuales del análisis filológico a partir de textos, sino también en las herramientas con las que se aborda dicho análisis, las cuales deben cumplir con una serie de requisitos instrumentales básicos¹². Ambos procesos son intrínsecos al concepto actual de LC y no pueden entenderse el uno sin el otro. No se hace lingüística de corpus sin la utilización de corpus avanzados tecnológicamente y de calidad de la misma manera que no se hace lingüística de corpus sin una interpretación cuantitativa y cualitativa de los datos del corpus, por muy avanzado que este sea.

La investigación planteada en esta tesis se desarrolla, por tanto, en consonancia con los nuevos preceptos de la LC, entendida no solo como una metodología que mejora el procesamiento y acceso a los datos, sino también como un enfoque que posibilita cuantificar y aplicar técnicas estadísticas como antes no era posible y no se tenía conocimiento. El diseño, edición y explotación del corpus adquieren entonces un valor

¹¹ Para un conocimiento más profundo acerca de los métodos de análisis estadísticos empleados en lingüística de corpus véase Desagulier (2017), Brezina (2018), Paquot y Gries (2020) (especialmente las partes IV-V) y Serigos (2022).

¹² Esta idea es la que subyace en los principales manuales sobre lingüística de corpus en español en los que se incide en que el uso de nuevas herramientas exige también nuevos análisis de los datos, diferentes a los que se han venido haciendo hasta ahora (Enrique-Arias, 2009; Kabatek, 2016; Torruella, 2017; Parodi et al., 2022).

trascendental y se convierten en cuestiones fundamentales de toda investigación en lingüística diacrónica (Sinclair, 1991; Enrique-Arias, 2009, p. 16). La construcción de un corpus de calidad no es el fin mismo de la investigación, pero no debe olvidarse que es una de las tareas que integran hoy en día la metodología de la LC y la definen en sí misma. Por esta razón, esta tesis concede un peso importante a la edición (Capítulo 3) y procesamiento lingüístico (Capítulo 4) del corpus que ha servido como base al estudio que se presenta y con el que se asegura el respeto por el interés histórico-filológico. Reivindicar el papel que juegan los corpus específicos de calidad es algo necesario si se quiere contribuir al estudio de la variación y el cambio lingüístico desde una perspectiva diacrónica.

2.3. Criterios de selección

Teniendo en cuenta el carácter de la investigación, el diseño del corpus de esta tesis se asienta sobre los principios de comparabilidad y equilibrio. Se pretende conseguir un corpus lo más fielmente comparable de acuerdo con los parámetros geográfico, cronológico y tipológico. Para ello, se hace necesario un equilibrio en la selección de materiales del corpus de tal manera que la muestra documental represente el fenómeno de estudio de forma equilibrada en base a diferentes criterios extralingüísticos. Estos se describen detalladamente a continuación:

- a. El origen geográfico ligado a las provincias de Almería y Madrid.
- b. El marco cronológico establecido en los siglos XVIII y XIX.
- c. La tipología textual que conforman los inventarios de bienes.

2.3.1. Límites geográficos

Con el corpus de esta tesis se pretende estudiar la modalidad andaluza oriental en el espacio geográfico de Almería. Además del interés lingüístico que posee el oriente de Andalucía en su conjunto –formado, de oeste a este, por parte de Málaga, Granada, Jaén y Almería–, la situación particular de esta última le otorga especial importancia; en primer lugar, por su delimitación geográfica, en segundo lugar, por la poca presencia que hasta el momento se le ha dado en los corpus del español. En base a estos dos argumentos justificaremos la necesidad de elaborar un corpus representativo de esta provincia.

En cuanto a su localización, Almería se encuentra emplazada en el punto más al sureste de la península ibérica ocupando el extremo más oriental de Andalucía. Abierta en gran parte al mar Mediterráneo, la rodean por el sur y levante más de doscientos kilómetros de costa. En los límites opuestos, se halla flanqueada por Granada, al oeste, y la comunidad de Murcia, al noreste. Forma parte de la llamada Andalucía oriental, según la tradicional división lingüística que de esta comunidad se ha hecho. De Almería se ha destacado en varias ocasiones su personalidad bien definida como reflejo de

rasgos dialectales tanto propios como influidos por su frontera con Murcia (Narbona Jiménez et al., 2011). Precisamente su proximidad con Murcia le ha otorgado un carácter de puente entre ambas comunidades, convirtiéndola en una zona clave para el estudio de la configuración de la variedad castellana meridional en general, y como diferenciación entre la modalidad murciana y andaluza, en particular.

A pesar de esta estratégica situación geográfica como puente entre las variedades lingüísticas andaluza y murciana, en los corpus históricos del español se detecta una carencia alarmante de documentación almeriense si se la compara con la situación privilegiada de otras provincias andaluzas mucho más representadas. En los corpus académicos (*CORDE* y *CDH*) la presencia de Almería es nula y la situación tampoco mejora en exceso en otros más especializados como *CHARTA* y *CODEA* (Fernández Alcaide, 2021). Cabe destacar que la misma situación desfavorable podría aplicarse también a otras provincias andaluzas como Huelva y Jaén, por oposición a Sevilla y Córdoba, en el extremo más aventajado. La situación varía dependiendo también del siglo que se estudie: por ejemplo, en el *CDH*, la falta de textos andaluces se extiende a todo el siglo XVIII, con excepción de Cádiz (Fernández Alcaide, 2021, p. 10). En cualquier caso, se destaca en estos corpus el desequilibrio documental entre provincias andaluzas. Por tanto, urge ampliar la representatividad de textos andaluces para poder obtener una visión conjunta y equilibrada de la evolución del español en esta región.

En este sentido, el corpus de esta tesis pretende cubrir un vacío documental necesario y esencial para el estudio de una provincia que actúa como enlace entre dos variedades meridionales fuertemente caracterizadas como son las variedades andaluza y murciana. Por su ubicación geográfica, se trata de una zona de marcado interés lingüístico para el estudio de las hablas andaluzas orientales y, más particularmente, para el estudio de la evolución de los diminutivos, ya que especialmente en esta zona el uso del sufijo *-ico* es un rasgo caracterizador del español (Lapesa, 2014, p. 334) del que se pretende analizar su *continuum* evolutivo. La creación de un corpus de inventarios de Almería se ve justificada por las razones mencionadas.

Con la finalidad de realizar estudios de tipo contrastivo es imprescindible contar con corpus de control que permitan comparar los datos obtenidos con los de otras zonas y épocas. Por ello, unido a la creación del corpus almeriense, se ha proyectado la elaboración de un corpus de la Comunidad de Madrid que sirva como referencia a la hora de contrastar la modalidad lingüística de Almería con la de esta otra provincia. Los datos de ambos corpus brindarán los resultados necesarios para posibilitar el análisis en el plano diatópico. Con ello se espera poder establecer diferencias o similitudes comparativas concluyentes en cuanto al estudio de los diminutivos en estas regiones, puesto que se consideran dos zonas de interés lingüístico y que, en lo geográfico, se hallan alejadas entre sí.

De la Comunidad de Madrid cabe destacar su enclave geográfico en el centro de la península ibérica. Cercada por las dos Castillas, pasó a ser la capital del reino desde que Felipe II decidiera trasladar aquí la corte de manera permanente en 1561, momento a partir del cual la ciudad se convertiría en símbolo de poder y núcleo cultural.

En el panorama lingüístico, el habla de esta zona forma parte de la modalidad castellana, la cual se extiende por el centro y norte peninsular. Exceptuando los rasgos

propios de cada una de las regiones en las que se habla, el castellano, entendido como una de las modalidades que integran la variedad lingüística del español, es la más ampliamente extendida de las utilizadas en la Península. Históricamente, la modalidad castellana se ha identificado por factores extralingüísticos con el español modélico y ha servido como prototipo para la creación del español estándar (Moreno Fernández, 2020, p. 58). Por ende, Madrid, como centro neurálgico y capital administrativa, económica y cultural, gozaba del prestigio suficiente para ser también considerada un referente en lo relativo al idioma, convirtiéndose en el foco de irradiación del español general. Desde este punto comenzaría la estandarización a partir de la inauguración de instituciones lingüísticas y con la variedad castellana como referente normativo para la creación de un español general y homogéneo.

Madrid tuvo una clara influencia en la configuración del español moderno favorecida por el gran volumen de documentación que la ciudad generaba. Muchos de los usos lingüísticos de la documentación de corte administrativo que perviven hoy en día se consolidaron precisamente en los siglos XVIII-XIX, llegando incluso a penetrar en la lengua de uso (Sánchez-Prieto Borja, 2019, p. 106, Vázquez Balonga, 2019). Los rasgos pertenecientes a la norma castellana nortea, algunos de ellos generalizados desde las clases más humildes, se irradiarían desde Madrid hacia otros lugares empujados por el prestigio y la fuerza demográfica de la ciudad (Moreno Fernández, 2015, p. 3).

De manera inversa, los usos lingüísticos de las clases populares urbanas se reflejan también en la escritura, de manera que afloran en los documentos fenómenos orales de la época. Este hecho quedaría evidenciado en los inventarios de Madrid, a partir de los cuales se espera recabar datos interesantes en lo relativo al uso de los diminutivos por contraste con la región de Almería, más aislada lingüísticamente del foco normativo. Se espera que la variable geográfica juegue en este caso un papel decisivo en el mantenimiento de dialectalismos en unas y otras zonas. De confirmarse esta hipótesis, la documentación almeriense reflejaría un uso de los diminutivos opuesto al registrado en Madrid.

Para la creación del corpus se ha intentado seleccionar documentos procedentes de diferentes puntos geográficos de las provincias de Almería y Madrid. Se incluyen las capitales, pero también otros municipios más alejados con el fin de representar la región al completo. Para el caso de Almería, se ha recopilado documentación procedente tanto del interior como cercana a la costa, así como de localidades próximas a las provincias colindantes de Granada y Murcia. En total, los documentos almerienses proceden de 18 puntos geográficos repartidos por toda la provincia (§ 2.5.2. *Corpus de estudio de Almería*). Por su parte, de Madrid se han seleccionado documentos de la capital y de algunos municipios cercanos a esta, además de diferentes localidades que bordean el territorio perimetralmente para abarcar toda la provincia en su conjunto. Los municipios seleccionados suman 13 en total (§ 2.5.3. *Corpus de control de Madrid*).

De esta manera, se amplía considerablemente la red de puntos que conforma el corpus con el objetivo de incrementar el volumen de textos, pero también con la finalidad de no seleccionar solo aquellos que se han producido en las capitales de provincia. La importancia de representar el habla de los pueblos ha sido especialmente

señalada en el caso de Madrid, ya que, a la hora de describir la extensión de algunos usos lingüísticos, son los pueblos los que seguramente pueden aportar “claves geográficas más claras que las de la capital, por la afluencia a esta de gentes de todos los lugares” (Sánchez-Prieto Borja, 2019, p. 90). Sin embargo, aunque el estudio desarrollado en esta tesis parte de una muestra más amplia para completar lagunas documentales en lo que a extensión geográfica se refiere en Almería y Madrid, no persigue establecer áreas dialectales o locales dentro de estas provincias. La evolución de los diminutivos es por tanto provincial y los resultados obtenidos deben ser entendidos de manera general para ambas regiones sin que se establezcan diferenciaciones por áreas o zonas geográficas.

Conviene añadir que la selección de municipios incluidos en el corpus no ha sido del todo aleatoria. Se ha intentado, como ya se ha dicho, que estos se encuentren repartidos por toda la provincia para cubrir el mayor territorio posible, pero también se han tenido en cuenta las localizaciones que de Almería y Madrid se encuentran en otros corpus con documentación similar como el *CorLexIn* o *CODEA+ 2022* con el objetivo de recabar datos lo más fielmente comparables. Con respecto a estos corpus, en *ODE* se han mantenido los municipios que de Almería se localizan en ambos corpus¹³ y se ha ampliado la red con la inclusión de siete nuevos enclaves hasta alcanzar los 18 puntos geográficos de que consta el corpus de inventarios almeriense. En el caso de Madrid, se han seleccionado intencionadamente los tres municipios que incluye *CorLexIn* (Madrid, Pinto y Torrelaguna). En *CODEA*, donde la documentación madrileña es proporcionalmente mayor que la del resto de regiones, se han mantenido muchos de los municipios, aunque algunos de ellos se han descartado cuando el trabajo de archivo ha presentado problemas para acceder a los fondos de estas localidades. En cualquier caso, en ambas provincias se ha aumentado el número de lugares de procedencia de los documentos con respecto a los que se registran en estos corpus.

En definitiva, la selección de documentos se ha hecho en base a los criterios geográficos mencionados: el corpus se fundamenta en inventarios de bienes producidos en las actuales provincias de Almería y Madrid, procedentes tanto de la capital como de diferentes municipios repartidos por todo el territorio. Esta distribución no pretende establecer tendencias dialectales, sino que surge con el propósito de comparar los textos de ambas provincias desde una perspectiva diatópica y cuantificar los resultados obtenidos en torno a la evolución de los diminutivos en estas regiones.

2.3.2. Marco cronológico

Los textos que conforman el corpus documental se circunscriben cronológicamente al periodo comprendido entre 1700 y 1896. Este lapso temporal abarca por tanto dos siglos

¹³ Se trata de los municipios de Almería, Vera y Laujar de Andarax en el corpus *CODEA*, por un lado; y de Canjáyar, Cuevas de Almanzora, Dalías, Fiñana, María, Serón, Tabernas, Vélez Blanco y Zurgena, en *CorLexIn*, por otro. Todos estos municipios se incluyen también en *ODE*. La nómina completa puede verse en § 2.5.2. *Corpus de estudio de Almería*.

completos (XVIII y XIX) considerados como el inicio del denominado español moderno según la periodización que establecen los principales estudios (Eberenz, 1991, 2009; Cano Aguilar, 2004; Echenique Elizondo y Martínez Alcalde, 2005; Pharies, 2007; Sánchez Lancis, 2009; Lapesa, 2014).

En lo que respecta al fenómeno de los diminutivos en el español, el análisis del corpus pretende revelar si esta etapa histórica, y más concretamente el siglo XVIII, puede ser o no considerado como una frontera temporal del español en los mismos términos que plantea Sánchez Lancis (2012). Para este autor, el siglo XVIII no alberga un gran número de cambios lingüísticos relevantes, sino que más bien se caracteriza por pertenecer a una época de estabilidad de la lengua que no justifica la importancia que se le ha dado a esta centuria en las gramáticas históricas, las cuales se estarían basando principalmente en criterios externos para hablar del siglo XVIII como inicio del español moderno (2012, p. 46). Estas afirmaciones se basan en los estudios del español general, por lo que ahora, partiendo de un corpus regional representativo de Almería y Madrid, veremos si los siglos XVIII y XIX suponen momentos clave para la periodización de los diminutivos en estas dos variedades concretas.

Por el contrario, otros autores sí que han destacado el papel que juega el siglo XVIII, decisivo para la diferenciación de algunas variedades del español peninsular y americano. Company Company (2007, p. 35) demuestra que es en el XVIII cuando en el español mexicano se concentran varios cambios que suponen la definitiva configuración de esta variedad lingüística. De forma equivalente, también Mondéjar (2006) considera que “es en el XVIII cuando con toda propiedad se puede hablar de «habla andaluza», o de «andaluz» como un todo constituido, de cuya existencia se tiene conciencia” (2006, p. 5). A partir de este siglo la variedad lingüística andaluza comienza a adquirir la dialectalización plena.

La importancia de esta centuria en la historia del español se manifiesta además en el aumento que han experimentado en las últimas décadas los estudios diacrónicos focalizados en este periodo (Gómez Seibane y Ramírez Luengo, 2007; García-Godoy, 2012; Ramírez Luengo, 2013; Calderón Campos, 2014; Guzmán Riverón y Sáez Rivera, 2016), del que ya no se puede decir que esté desatendido.

En la selección del eje temporal conviene además señalar que en el Madrid de los siglos XVIII y XIX se constituye una nueva clase urbana enriquecida que contribuye a la extensión y difusión de determinados rasgos lingüísticos (Sánchez-Prieto Borja, 2019, p. 106) y que empujados por su prestigio se irradiarían desde este punto hacia otras variedades peninsulares. En este sentido, se pretende analizar para el caso concreto de los diminutivos el alcance que el uso de los sufijos en esta región tiene sobre otras en cuanto a la estandarización. De esta época se destaca la fundación de la Real Academia Española y sus esfuerzos por conseguir la homogeneidad lingüística, por lo que se trata de determinar la influencia que el empleo de los diminutivos en la variedad centro peninsular tiene sobre otras variedades periféricas, las cuales podrían tender hacia la nivelación con aquella durante este periodo o posteriores. Así, el análisis de este marco cronológico servirá también para evaluar cómo de decisivos son para los cambios lingüísticos algunos de los factores externos que se han mencionado, pues podría confirmarse que el peso que se ha otorgado al influjo normativo de algunas instituciones

como la Academia es excesivo para la evolución de algunos fenómenos del español (Sánchez Lancis, 2012).

De esta manera, reflejar la lengua de los siglos XVIII y XIX en las variedades almeriense y madrileña se considera esencial para aportar nuevos datos sobre la configuración de estas modalidades en el uso de los diminutivos. Por un lado, el corpus sirve de espejo del XVIII, siglo que se ha revelado decisivo en la configuración del español de algunas variedades; por otro lado, arroja datos del siglo XIX, necesarios para evaluar los efectos de la normalización, analizar la evolución lingüística a lo largo de este periodo y determinar si se trata de un periodo de cambios o estabilidad.

Por otra parte, la elección del marco cronológico se considera imprescindible ante la escasez de corpus representativos de este periodo. El siglo XVIII ha sido considerado como el gran olvidado (Company Company, 2012, p. 255), aunque en la actualidad cada vez son más las fuentes que incluyen documentación propia de esta etapa histórica fundamental para la configuración actual del español. De esta manera, el corpus *ODE* contribuye a los estudios históricos de este marco temporal con un corpus de 47 inventarios de bienes para el siglo XVIII y 50 para el XIX con un total de 118 817 palabras (§ 2.5. *Descripción del corpus y características generales*).

Con el fin de alcanzar la representatividad, para cada siglo se han establecido tres cortes cronológicos, de modo que cada siglo quedaría dividido en cuatro bloques de veinticinco años cada uno que se han completado con la selección de documentación de manera equitativa. Aunque esta división posibilitaría un estudio más exhaustivo basado en divisiones temporales más reducidas, para esta tesis el análisis se realizará por siglos completos, considerados cada uno de ellos como un conjunto cerrado de datos. Esta forma de proceder se ha considerado adecuada, dado que el análisis ya aporta diferencias significativas entre uno y otro siglo y que, por otro lado, la división del corpus en periodos temporales inferiores al siglo fragmentaría demasiado los datos quedando para algunos casos frecuencias muy reducidas difícilmente significativas en términos estadísticos.

2.3.3. Tipología textual

El corpus de esta tesis está constituido exclusivamente por inventarios de bienes. Dentro de esta tipología textual se incluye “cualquier texto hecho con la finalidad de enumerar, de la forma más minuciosa posible, los bienes de una persona o una institución” (Morala, 2012a, p. 200). La principal ventaja de un corpus compuesto por inventarios reside en la homogeneidad del tipo textual. Esta permite la comparación regional y cronológica de los documentos, pero a ello se suman una serie de no pocas características que por su idoneidad han hecho de los inventarios una valiosa fuente para la investigación en diversas áreas. De ellos se ha destacado la perfecta e inequívoca datación y localización geográfica de los documentos. Este tipo de documentos proporciona una información muy precisa en torno a la ubicación, fecha, identificación de los propietarios, profesión, etc. (Pedraza Gracia, 1999, p. 140), que para los estudios lingüísticos es un aspecto fundamental a la hora de poder adscribir determinados

fenómenos a zonas y épocas concretas. Además, la abundancia de esta documentación en los archivos de protocolos histórico-provinciales asegura una buena y abundante muestra representativa con la que elaborar el corpus.

La confección de los inventarios respondía a la necesidad de dejar constancia de los bienes de una persona por muy diversas causas, la más habitual por el fallecimiento de alguien, pero también, como señala Quirós García (2011, p. 427), por contratos matrimoniales, dotes, testamentos, impagos, compraventas, almonedas o particiones, entre otras muchas causas que eran bastante habituales en la época y de las que se dejaba testimonio de manera jurídica¹⁴. Todo este flujo de acciones cotidianas ha proporcionado una cantidad ingente de documentación especialmente útil para el investigador.

No es nuestro propósito principal realizar un análisis exhaustivo de la caracterización textual de este tipo de documentos. Al respecto pueden verse los trabajos ya citados de Pedraza Gracia (1999) y Quirós García (2011), así como las tesis doctorales de Rodríguez Manzano (2019) y Laguna Álvarez (2017), en los que se destaca el valor de los inventarios para los estudios lingüísticos. En el plano filológico, un corpus de inventarios como el que se propone basado en documentación notarial compensa en cierta medida la carencia de fuentes orales y permite analizar fenómenos relacionados con la oralidad con más fiabilidad que otros corpus de naturaleza culta. Este hecho se constata en el auge exponencial que han tenido los corpus y trabajos que se han basado en el estudio de textos de archivo frente a los textos literarios (Castillo Lluch y Díez del Corral Areta, 2019). Aunque los inventarios de bienes son textos de carácter notarial con abundantes fórmulas fraseológicas y tratamientos protocolarios, responden a un tono de formalidad neutra (Torruella, 2017, p. 112). Estos textos ofrecen datos léxicos de enorme valor para el estudio de la vida cotidiana, al mismo tiempo que abren la posibilidad de estudiar variedades morfológicas del español, debido en muchos casos al reflejo de una escritura muy cercana a la oralidad de los escribanos, quienes, dado el fin utilitario de los inventarios, no se preocupaban demasiado por la corrección formal (Calderón Campos y Vaamonde, 2020, p. 170).

Tradicionalmente, el análisis de los inventarios se ha centrado sobre todo en el estudio del léxico. Un claro ejemplo son las numerosas publicaciones derivadas del *CorLexIn* (Morala, 2012a; Perdiguero, 2012; Egido, 2016; Pérez Toral, 2017; Junquera Martínez, 2020), pero también de corpus propios de inventarios procedente de diferentes puntos de la geografía (Bastardín Candón, 2015; Puche Lorenzo, 2018; Torres Martínez, 2019; entre muchos otros), pero también se ha demostrado su valía para otro tipo de estudios relacionados con la variación gráfica, fonética o morfosintáctica (Puche Lorenzo, 2012; Calderón Campos, 2015, 2019c; Egido, 2019;

¹⁴ Esta diversidad en los tipos de documentos significa también la presencia de fórmulas propias para cada uno de ellos. La falta de homogeneidad en los inventarios ha sido puesta en evidencia por algunos autores, quienes han apuntado discordancias formales importantes entre este tipo de documentos “ya sea porque no existía para los inventarios una estructura diplomática tan estandarizada como para otro tipo de actas notariales, ya por las diversas causas que motivaban su elaboración, ya debido a los distintos niveles de pericia de los propios escribanos, o incluso por las diferentes épocas en que fueron redactados” (Quirós García, 2011, p. 431).

Arrabal Rodríguez y Calderón Campos, 2022). Esta tesis doctoral se inserta en la línea de los estudios históricos morfosintácticos a partir de inventarios de bienes con los que poder analizar el uso de los diminutivos.

En cuanto a la selección documental, esta se ha guiado simplemente por el requisito principal de que los textos contengan listas de bienes. Sin embargo, son muchos y muy variados los documentos en los que se incluyen bienes inventariados. Durante la búsqueda de archivo se han localizado declaraciones de pobre, poderes, venta de propiedades, cartas de dote y arras, testamentos, fianzas de cárcel, codicilos, arrendamientos, particiones de bienes, almonedas y subastas, repartos de herencia, redención de censos, permutas, etc. Así, bajo esta tipología se incluye un conjunto heterogéneo de documentos con diferencias formales en cuanto a su estructura y contenido que nos hizo reflexionar sobre cuáles eran los más adecuados para su inclusión en el corpus. En la práctica, en todos estos documentos confeccionados ante notario se registran bienes, sin embargo, no son muy abundantes en muchos de ellos. Ante tal acopio y disparidad de documentos, se optó por elegir aquellos en los que el listado de bienes era más extenso y variado. En este sentido, las cartas de dote y los inventarios por fin y muerte suelen ofrecer con una probabilidad mucho más alta largas listas de objetos que otro tipo de inventarios, por lo que se priorizaron estos frente al resto. Sin embargo, es conveniente hacer una aclaración al respecto, y es que la selección de los textos se ha hecho siempre en función de la riqueza léxica que poseen y no por que se haya detectado la mayor presencia de diminutivos en ellos. Esto último falsearía los resultados que aquí se presentan sobre el uso real de los diminutivos en los textos, por lo que es importante destacar que la selección se ha hecho sin sesgo alguno y se ha guiado principalmente por el valor de representatividad para toda clase de estudios, incluidos también los de tipo léxico¹⁵.

En consecuencia, los inventarios *post mortem* y las cartas de dote son los documentos con un peso mucho más considerable en el corpus. En algunas ocasiones el inventario podía realizarse sin que el difunto haya dejado testamento, por lo que este se realizaba tras el fallecimiento con los familiares presentes como testigos:

- (1) Nos constituimos en la casa mortuoria para practicar el inbentario de los bienes quedados por fallecimiento del hijo del primero Jose Diaz Plaza, becino que fue de esta villa [...] a la edad de treinta y un año, el cual no pudo practicar disposicion testamentaria por no haberle permitido su enfermedad [...] Estando todos presentes se requirio a la viuda que se allaba al cuidado de la casa, con el fin de que pusiese de manifiesto todos los vienes, efectos y papeles que ubiera sin ocultar ninguno (*ODE* – Almería, Tabernas, 1870).

¹⁵ Este hecho implica también una limitación; Pedraza Gracia advierte que las investigaciones basadas en inventarios dejarían fuera a las clases sociales más humildes, pues la mayoría de los inventarios eran realizados por personas de cierto nivel económico que poseían bienes de valor que declarar. “La realización del inventario responde a una necesidad jurídica y no a una obligación legislativa” (1999, p. 139).

Las cartas de dote se caracterizan por ser documentos regularizados y muy tipificados en cuanto a su estructura y formulación. Todas ellas tienen una disposición muy similar de la información y contienen las mismas expresiones protocolarias repetitivas en todos ellos, aunque también existen partes de texto libre. Por lo general, el marido es quien solicita al notario realizar el inventario de los bienes que recibe tras el casamiento. Así, los recibos de dote servían para atestiguar los bienes que eran llevados al matrimonio por parte de la mujer, aunque también es posible encontrar cartas de dote con los bienes que lleva el marido, normalmente cuando este se ha casado en segundas nupcias y lo que se pretende es evitar litigios entre los herederos de una y otra parte:

- (2) Trajo el susod[ic]ho ziertos vienes muebles, raizes y semovientes, asi suyos como de los hijos de la d[ic]ha su primera mujer, y para que en lo suvzesivo se sepa los que son y q[ue] entre sus hijos y los de la d[ic]ha Ysael Lopez no aya pleitos ni desazones por no hauer echo falta ahora ynspeccion de ellos, quiere hacer ymventario de los que trajo a d[ic]ho matrim[oni]o (ODE – Almería, Serón, 1754).
- (3) Y es así que huiendo llegado el caso de pasar el d[ic]ho Pedro Fer[nánde]z Grandizo a contraer matrimonio de segundas numpzias con Marta Lozano, vez[in]a de esta d[ic]ha villa; y allegado el de saber que vienes y haz[ien]da quedaron al tiempo del fallezim[ien]to de la d[ic]ha Luzia Fer[nánde]z y allarse el susod[ic]ho padre y lex[í]timo adm[inistrad]or de las personas y vienes de los d[ic]hos sus tres hijos menores y de la susod[ic]ha, y para sauer q[u]e gananziales de vienes lleva a d[ic]ho en seg[un]do matrimonio y para que en todo tiempo conste los que son (ODE – Madrid, 1747).

En resumen, la elección de inventarios de bienes como tipología documental para la elaboración del corpus se ha regido por el criterio de comparabilidad. El corpus favorece el contraste de los resultados al estar conformado por muchos textos de iguales características que además se encuentran bien identificados en cuanto a la fecha y el lugar de origen. Para los objetivos de esta tesis, permite además cuantificar el número de diminutivos empleados y cotejarlo con el de otras regiones y épocas en los mismos contextos de aparición. Para tal efecto se han reunido en mayor medida cartas de dote e inventarios *post mortem*, aunque también se recogen almonedas, repartos de herencia o tasaciones. En cualquier caso, para la selección de los inventarios se ha optado por incluir en el corpus aquellos que presentan mayor número de bienes, ya que serán con toda probabilidad los que aporten mayor riqueza léxica.

2.4. Archivos consultados y recopilación del material

Para la recopilación de los inventarios se visitaron los archivos históricos de las dos ciudades capitalinas con representación en el corpus: el Archivo Histórico Provincial de Almería (AHPAL) y el Archivo Histórico de Protocolos de Madrid (AHPM). El conjunto final de documentos fue el resultado de varias visitas a ambos archivos en los

que se empezó por seleccionar aquellos pertenecientes a las capitales, donde el volumen de documentación era mayor y mejor catalogado que para el resto de los municipios.

El proceso de selección se llevó a cabo *in situ* de acuerdo con la documentación que el archivo ofrecía y se iba encontrando, lo que también implicaba la digitalización de los documentos, realizada personalmente mediante cámara fotográfica. La reproducción del manuscrito forma parte de la edición digital de cada texto del corpus, por lo que ha sido imprescindible poder conseguir fotos de los originales en todos los casos. La selección de los documentos no ha supuesto grandes restricciones más allá del cumplimiento de los tres criterios principales mencionados hasta ahora: documentos de las provincias de Almería y Madrid producidos en los siglos XVIII y XIX y que incluyan listas de bienes. Para ello ha sido imprescindible que en cada inventario el escribano haya indicado de manera inequívoca el lugar y la fecha en la que se creó, de tal forma que cada documento esté debidamente identificado.

Los inventarios se localizaron en la sección de Protocolos de ambos archivos. Las signaturas que forman parte de estos fondos documentales incluyen tomos con una numerosa y heterogénea documentación reflejo de las variadas transacciones que tenían lugar en la vida cotidiana de la época. Pero no toda esta documentación incluye necesariamente inventarios de bienes; el proceso de selección ha exigido la consulta de cada tomo en busca de aquellos documentos que sí presentan listas de bienes.

Afortunadamente, la gran cantidad de inventarios de bienes en los archivos hizo que se pudieran priorizar unos documentos frente a otros. En este sentido, el trabajo de selección dependió en algunos casos de las características particulares que se iban encontrando en la documentación de cada archivo. En primer lugar, se descartaron aquellos manuscritos que plantearon alguna duda sobre su certera localización o datación en favor de aquellos de indiscutible origen y fecha, condición indispensable para la inclusión en el corpus.

En segundo lugar, el estado de conservación de los documentos fue también un aspecto clave en el proceso de selección. En general, la documentación de ambos archivos se conserva en buen estado, pero sobre todo en el archivo almeriense muchos de los legajos consultados mostraban un deficiente estado de conservación debido a la humedad, lo que se manifestaba con manchas en los folios y, en algunos casos, también rotos en los bordes que imposibilitaban la lectura. Otro inconveniente fueron los numerosos documentos que se han visto afectados por la acción de las polillas, lo que se dejaba ver en el deterioro de algunos legajos con los característicos rotos alargados causados por el arrastre de estos insectos. De esta manera, ante los daños insalvables en algunos de los legajos, y siempre en la medida de lo posible, se han priorizado los que gozan de un mejor estado de conservación con el fin de facilitar la digitalización del manuscrito, así como la transcripción y edición posterior. Sin embargo, dado que muchos de los tomos de algunos periodos estaban en iguales condiciones, hubo que incluir en el corpus inventarios visiblemente afectados¹⁶.

¹⁶ El deterioro por las causas mencionadas es notable en algunos inventarios. Algunos ejemplos son el “Inventario de los bienes que quedaron por fin y muerte de Antonio de Acosta y Macías” (AL1806I0023) (<http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/AL1806I0023.xml>) o el

En el archivo de Madrid la selección de los manuscritos supuso una dificultad añadida como consecuencia de la política de consulta de los documentos. La gran mayoría de los protocolos pertenecientes a la capital del marco cronológico que el corpus abarca se encuentra microfilmada, pero el sistema de consulta de los documentos en soporte microfilm es lento y tedioso y esto supuso una inversión de tiempo mayor para la digitalización de un número considerablemente menor de manuscritos. El archivo no permite la consulta directa de estos tomos en papel, por lo que la selección de los inventarios microfilmados se ralentizó mucho más que en el caso del archivo almeriense. Si bien en este último ha sido posible obtener todos los documentos en soporte fotográfico, en el caso de los inventarios madrileños microfilmados solo ha sido posible su recuperación a partir de fotocopias impresas¹⁷. Por este motivo, en el AHPM la selección documental estuvo determinada en buena medida por el soporte. Así, se prefirieron aquellos tomos cuya consulta en papel fuera posible siempre que el marco cronológico se viera representado en su totalidad.

En resumen, se han seleccionado textos que cumplen con los criterios extralingüísticos establecidos y que permiten ofrecer una muestra representativa. Las características particulares de conservación y gestión de la documentación en cada archivo nos llevaron a tomar algunas decisiones externas al diseño de la investigación sin que ello haya interferido en los principios de representatividad del corpus: en Madrid, se han priorizado los tomos en los que estaba permitido fotografiar personalmente los manuscritos; en Almería, hubo que descartar los inventarios que se encontraban en legajos muy afectados por la humedad o insectos y se optó por otros de mejor conservación.

El resultado de la labor de selección es la relación de inventarios que finalmente forman parte del corpus documental de esta tesis y que se encuentran disponibles para su consulta en *ODE*, cuyas características se describen en el siguiente apartado.

2.5. Descripción del corpus y características generales

El corpus documental de esta tesis plantea la creación de dos subcorpus compuestos por inventarios de bienes de los siglos XVIII y XIX a fin de poder examinar en términos comparativos la evolución de los diminutivos en dos variedades del español peninsular: 1) un corpus de estudio representativo de la provincia de Almería y 2) un corpus de control de la Comunidad de Madrid. La documentación madrileña será tomada como referencia a la hora de analizar los datos de Almería, de manera que el corpus en su conjunto está formado por dos subcorpus:

“Inventario de los bienes que quedaron por fin y muerte de Juan Yanguas Calvache” (AL1896I0045) (<http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/AL1896I0045.xml>).

¹⁷ De la consulta de tomos en soporte microfilm se seleccionaron para su incorporación al corpus de control un total de seis inventarios procedentes de cuatro tomos diferentes (P-15766, P-19452, P-22426, P-21694), todos ellos de la capital. Un ejemplo es el documento “Bienes que Tomás Muñoz recibe como dote por su casamiento con Catalina García” (MD1765I0112), cuyo facsímil en *ODE* son las copias escaneadas de los manuscritos impresos.

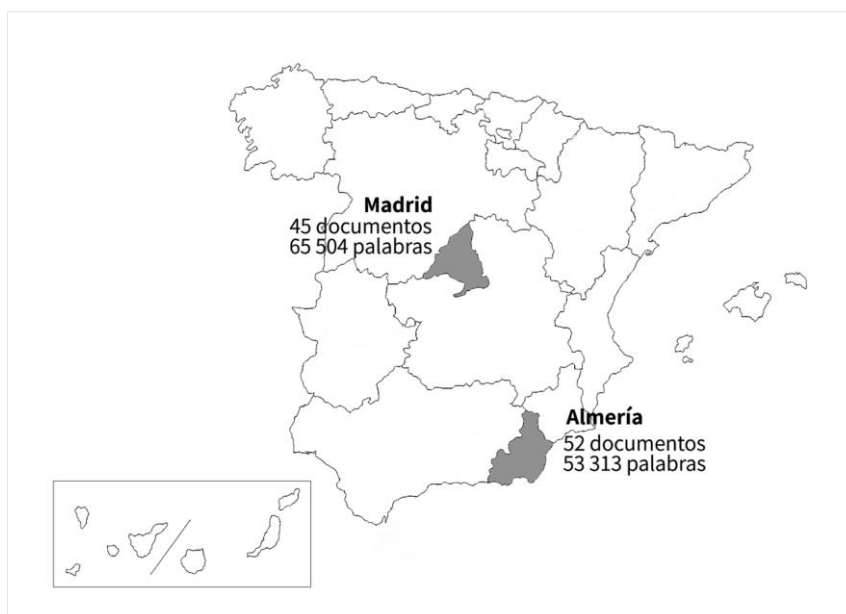


Figura 1: Delimitación geográfica y tamaño del corpus documental

La distinción dentro del corpus documental entre corpus base o de estudio y corpus de control es necesaria para la cuantificación de los resultados de acuerdo con el análisis estadístico que se lleva a cabo en esta tesis (Capítulo 7). Sin embargo, de manera general se alude al corpus de esta tesis como uno solo en su conjunto, del que debe entenderse que está compuesto a su vez por documentación procedente de las provincias de Almería y Madrid.

El propósito inicial fue alcanzar las 100 000 palabras: 50 000 palabras para cada provincia, 25 000 para cada centuria. Esta aproximación se basa en el modelo previo establecido para la elaboración del *CorLexIn*, en el que se fija que cada provincia esté representada por entre 20 000 y 30 000 palabras (Morala, 2012b, p. 422). En base a estas cifras, se espera que un corpus de un tamaño semejante sea lo suficientemente representativo para ofrecer datos concluyentes en torno a la evolución de los diminutivos en las provincias y los siglos mencionados. Finalmente, el volumen del corpus sobrepasa las 118 000 palabras conforme a la siguiente distribución:

	Corpus de estudio Almería	Corpus de control Madrid	
S. XVIII	27 544	29 980	
S. XIX	25 769	35 524	
Total	53 313	65 504	118 817

Tabla 1: Número de palabras en el corpus

Para calcular el número de palabras o tokens en el corpus se ha utilizado el software *LancsBox* (Brezina et al., 2020). Esta herramienta entiende por token cada una de las formas gráficas presentes en el corpus, entre las que se incluyen palabras y números.

Conviene por tanto destacar que las cifras de la Tabla 1 comprenden también los números en los textos, pero no los signos de puntuación¹⁸.

A pesar de las ligeras diferencias en el tamaño de los subcorpus, –destaca sobre todo el corpus de Madrid del siglo XIX con un mayor número de palabras–, se ha intentado alcanzar una representatividad homogénea entre las diferentes calas de la muestra documental que permita la comparación de los resultados también en lo que respecta al volumen de los corpus. En lo que sigue, haremos referencia a las características de cada uno de ellos de manera independiente. Por un lado, los 52 inventarios pertenecientes a la provincia andaluza de Almería conformarán el llamado corpus de estudio y permitirán evaluar la distribución de los diminutivos en una región considerada típicamente dialectal; por otro lado, los 45 documentos restantes son originarios de la Comunidad de Madrid y servirán como corpus de control a la hora de contrastar la documentación andaluza con esta otra procedente de una zona lingüísticamente diferenciada de la de aquella. En total, 97 inventarios de bienes producidos a lo largo de los siglos XVIII y XIX.

La distinción entre corpus de estudio y corpus de control, motivada por el carácter propio de la investigación, se fundamenta en el análisis contrastivo sobre los diminutivos. Sin embargo, debemos aclarar que, excepto que se especifique lo contrario, en esta tesis se alude de manera general al corpus documental como aquel que se ha creado específicamente para esta investigación y que integra todos los documentos procedentes de los siglos XVIII y XIX de Almería y Madrid.

2.5.1. Acceso al corpus

El corpus de esta tesis es parte integrante de *ODE* (§ 2.1. *Antecedentes*), por lo que los documentos de Almería y Madrid analizados se pueden consultar directamente en la página web del corpus, de acceso libre para cualquier usuario: <http://corpora.ugr.es/ode/>.

Para acceder al listado de documentos del corpus bastará con realizar una consulta a través de la propia interfaz de búsqueda avanzada del corpus señalando en el desplegable el tipo textual (inventario de bienes), la provincia (Almería o Madrid) y el siglo (XVIII o XIX) según interese. La búsqueda devuelve la lista completa de documentos ordenada cronológicamente del inventario más antiguo al más reciente en la que se indica el código de identificación del texto, el título, el año, el municipio y el tipo textual. En la imagen se muestran los inventarios de Almería del siglo XVIII:

¹⁸ El concepto de token puede variar considerablemente según la herramienta que se utilice. En TEITOK, por ejemplo, incluye también los signos de puntuación (§ 4.2. *Tokenización*), por lo que se ha preferido no incrementar artificiosamente estas cifras, ya que el número de palabras del corpus repercute de manera muy directa en el análisis estadístico para hallar las frecuencias relativas de los diminutivos.

Búsqueda en CQL:

<text> [] :: match.text_province = "Almería" & match.text_cat = "inv" & match.text_century = "XVIII"

24 resultados

ID	Título	Año	Lugar	Provincia	Tipo textual
AL1704I0001.xml	Inventario de los bienes que quedaron por fin y muerte de Pedro Rodríguez	1704	España, Almería, Cuevas de Almanzora	Almería	inventario de bienes
AL1709I0002.xml	Bienes que Pedro Carmona recibe como dote por su casamiento con Catalina García	1709	España, Almería, Tabernas	Almería	inventario de bienes
AL1714I0003.xml	Inventario de los bienes que quedaron por fin y muerte de Juan González Grano de Oro	1714	España, Almería, Cuevas de Almanzora	Almería	inventario de bienes
AL1715I0004.xml	Bienes que Leandro Guevara lleva a su segundo matrimonio con Margarita	1715	España, Almería, Serón	Almería	inventario de bienes
AL1721I0005.xml	Inventario de los bienes que quedaron por fin y muerte de D. Manuel Herrán	1721	España, Almería, Almería	Almería	inventario de bienes
AL1726I0006.xml	Inventario de bienes	1726	España, Almería, Abila	Almería	inventario de bienes
AL1730I0050.xml	Inventario de los bienes que quedaron por fin y muerte de Bernardo Carmona	1730	España, Almería, Vera	Almería	inventario de bienes

Figura 2: Acceso a los documentos a partir de la búsqueda en ODE

En caso de que se quiera obtener el conjunto completo de textos de una sola vez (todos los inventarios de Almería y Madrid de los siglos XVIII y XIX), bastará con incluir en el cajón de búsqueda ambas opciones separadas por una barra lateral (arriba en la Figura 2), quedando como sigue:

```
<text> [] :: match.text_province = "Almería|Madrid" & match.text_cat = "inv" &
match.text_century = "XVIII|XIX"
```

Este tipo de búsqueda utiliza el lenguaje CQL, una sintaxis específica para recuperar la información en TEITOK. Su funcionamiento y los distintos tipos de búsquedas en el corpus se detallan más adelante en los apartados § 4.8. *Búsquedas en el corpus basadas en lenguaje CQL* y § 4.9. *Interfaz de consulta del corpus ODE*.

2.5.2. Corpus de estudio de Almería

El corpus de Almería incluye inventarios de bienes procedentes de 18 municipios distribuidos por diferentes puntos de la provincia:



Figura 3: Mapa de municipios almerienses incluidos en el corpus

Forman parte del corpus almeriense un total de 52 textos, 24 de ellos fechados en el siglo XVIII y 28 en el XIX. Todos los documentos pertenecen al Archivo Histórico Provincial de Almería (AHPAL), que custodia documentación de la capital, así como de distintas comarcas de la provincia. Los inventarios de bienes forman parte de la sección de Protocolos notariales, concretamente de los fondos públicos judiciales de ámbito local, los cuales se organizan según las Alcaldías Mayores de las jurisdicciones correspondientes. Para la recopilación del corpus se consultaron los fondos de las Alcaldías Mayores de Almería, Canjáyar, Cuevas de Almanzora, Fiñana y Vera, cuyas firmas de los documentos seleccionados figuran en la siguiente tabla:

<i>Título</i>	<i>Signatura</i>	<i>Año</i>	<i>Municipio</i>
Inventario de los bienes que quedaron por fin y muerte de Pedro Rodríguez	AHPAL, P-3584 ff. 4r-5r	1704	Cuevas de Almanzora
Bienes que Pedro Carmona recibe como dote por su casamiento con Catalina García	AHPAL, P-1062 sf.	1709	Tabernas
Inventario de los bienes que quedaron por fin y muerte de Juan González Grano de Oro	AHPAL, P-3595 ff. 1v-5v	1714	Cuevas de Almanzora
Bienes que Leandro Guevara lleva a su segundo matrimonio con Margarita	AHPAL, P-4199 ff. 192r-193r	1715	Serón
Inventario de los bienes que	AHPAL, P-306	1721	Almería

quedaron por fin y muerte de D. Manuel Herrán	ff. 11r-17v		
Inventario de bienes	AHPAL, P-446 ff. 6r-8r	1726	Abla
Inventario de los bienes que quedaron por fin y muerte de Bernardo Carmona	AHPAL, P-43401 ff. 3v-4v	1730	Vera
Inventario de los bienes que quedaron por fin y muerte de D. Juan Mauricio Salido y Granados	AHPAL, P-1109 sf.	1736	Almería
Inventario de los bienes que quedaron por fin y muerte de José Fontana	AHPAL, P-457 ff. 2r-4r	1745	Fiñana
Bienes del caudal	AHPAL, P-457 ff. 11v-12r, 22r	1748	Fiñana
Inventario de los bienes que quedaron por fin y muerte de Antonio Gutiérrez	AHPAL, P-1136 sf.	1748	Roquetas de Mar
Inventario y tasación de los bienes que quedaron por fin y muerte de Francisca de Asién	AHPAL, P-1176 ff. 55r-62v	1750	Dalías
Bienes que Juan García Morales lleva a su segundo matrimonio con Isabel López	AHPAL, P-4235 ff. 7r-8r	1754	Serón
Bienes que Isabel López lleva a su segundo matrimonio con Juan García Morales	AHPAL, P-4235 ff. 9r-10r	1754	Serón
Inventario y tasación de los bienes que quedaron por fin y muerte de Josefa María González Alarcón	AHPAL, P-3848 ff. 9r-20r	1766	Cuevas de Almanzora
Inventario y tasación de los bienes que quedaron por fin y muerte de José Yepes	AHPAL, P-610 sf.	1774	Bacares
Inventario de los bienes que quedaron por fin y muerte D. ^a Antonia Navarro Segura	AHPAL, P-3848 ff. 8r-8v	1775	Cuevas de Almanzora
Inventario y tasación de los bienes que quedaron por fin y muerte de Pedro de Santander	AHPAL, P-474 ff. 97v-102v	1775	Escúllar
Inventario de los bienes que quedaron por fin y muerte de D. Fernando Gallego	AHPAL, P-474 sf.	1776	Fiñana
Inventario que hace Roque Alcalde	AHPAL, P-478	1784	Abrucena

de sus bienes propios y de los de su esposa, María Ana Gaona, tras el fin y muerte de esta	ff. 6r-7v		
Inventario, cuenta y partición de los bienes que quedaron por fin y muerte de Teresa García Rivera	AHPAL, P-478 sf.	1785	Fiñana
Inventario de los bienes que quedaron por fin y muerte de Pedro Sevilla Cortinas	AHPAL, P-3850 ff. 3v-5v	1786	Cuevas de Almanzora
Inventario de los bienes que han quedado por fin y muerte de Antonio Fernández y Guevara	AHPAL, P-43405 ff. 8r-11v	1791	Vera
Inventario de los bienes que quedaron por fin y muerte de D ^a Juana Pérez Ávalos	AHPAL, P-3853 ff.14v-17v	1792	Cuevas de Almanzora
Inventario de los bienes que quedaron por fin y muerte de Gerónimo Domínguez García	AHPAL, P-4951 ff. 7r-10v	1803	Zurgena
Inventario de los bienes que quedaron por fin y muerte de Antonio de Acosta y Macías	AHPAL, P-1298 sf.	1806	Almería
Inventario de los bienes que quedaron por fin y muerte de Juana González	AHPAL, P-4951 ff. 10r-11r	1809	Zurgena
Inventario de los bienes que quedaron por fin y muerte de José López García, correspondientes a su primera mujer	AHPAL, P-4951 ff. 17r-18r	1814	Zurgena
Inventario, cuenta y partición de los bienes que quedaron por fin y muerte de Joaquina Lao Rada	AHPAL, P-374 ff. [3v]-[6r]	1819	Alboloduy
Inventario de los bienes que quedaron por fin y muerte de Diego Rodríguez y Agustina de Vargas	AHPAL, P-496 sf.	1822	Fiñana
Bienes que José Plaza Hernández recibe como dote por su casamiento con Ana María Guirado Plaza	AHPAL, P-603 ff. 12r-14r	1822	Tabernas
Inventario de los bienes que quedaron por fin y muerte de M. ^a Antonia López	AHPAL, P-496 sf.	1824	Fiñana
Inventario de los bienes que quedaron por fin y muerte de Teresa Domínguez	AHPAL, P-4959 ff. 12r-17r	1825	Zurgena

Inventario de los bienes que quedaron por fin y muerte de Antonio de Latorre a sus dos hijas	AHPAL, P-496 ff. 28r-29v	1827	María
Inventario y tasación de los bienes que quedaron por fin y muerte de D. Agustín Domínguez García	AHPAL, P-4959 ff. 23r-25r	1829	Zurgena
Inventario de los bienes que quedaron por fin y muerte de D. Tomás Gómez	AHPAL, P-496 ff. 14r-20r	1832	Fiñana
Inventario de los bienes que quedaron por fin y muerte de María	AHPAL, P-496 sf.	1833	Abla
Inventario de los bienes que quedaron por fin y muerte de D. Francisco Pastor	AHPAL, P-496 sf.	1834	Escúllar
Inventario de bienes que quedaron por fin y muerte de Francisco Santander	AHPAL, P-496 sf.	1834	María
Inventario de los bienes que quedaron por fin y muerte de Juan Aparicio a sus hijos y herederos	AHPAL, P-496 ff. [2r-6r]	1835	Fiñana
Inventario de los bienes que quedaron por fin y muerte de D. José M. ^a Castreño	AHPAL, P-496 sf.	1837	Escúllar
Inventario de los bienes que quedaron por fin y muerte de Josefa Cabrerizo	AHPAL, P-496 sf.	1840	Fiñana
Inventario de los bienes que quedaron por fin y muerte de Tomás Navarro y Juana Paula	AHPAL, P-5189 ff. 19r-23r	1844	Cuevas de Almanzora
Bienes que recibe Joaquín de Vilches y Baeza como dote por su casamiento con Ana Martínez Vilches	AHPAL, P-5294 ff. 1r-5v	1844	Almería
Inventario de los bienes que quedaron por fin y muerte de Ana Josefa García Castillo	AHPAL, P-5189 ff. 1r-3v	1845	Cuevas de Almanzora
Inventario de los bienes embargados a Juan José de Soria y Francisco de Soria	AHPAL, P-5297 ff.	1848	Almería
Inventario de los bienes quedados por fin y muerte de Antonio de Oña Oña para proceder a su repartición	AHPAL, P-7711 sf.	1868	Tabernas
Inventario de los bienes que	AHPAL, P-7711	1870	Tabernas

quedaron por fin y muerte de José Díaz Plaza	sf.		
Inventario de los bienes que quedaron por fin y muerte de D. Santiago Morales Nieto	AHPAL P-8483	1882	Fiñana
Inventario de los bienes que quedaron en herencia por fin y muerte de D. Juan Martínez Valera	AHPAL, P-10673 ff. 43r-45r	1895	Vélez Blanco
Inventario de los bienes que quedaron por fin y muerte de Juan Yanguas Calvache	AHPAL, P-10187 ff. 6r-6v	1896	Laujar de Andarax
Inventario de los bienes que quedaron en herencia por fin y muerte de Isabel Moreno López	AHPAL, P-10187 sf.	1896	Canjáyar

Tabla 2: Relación de documentos de Almería con título, signatura, fecha y municipio

2.5.3. *Corpus de control de Madrid*

El corpus de Madrid está constituido por 45 documentos, 23 pertenecientes al siglo XVIII y 22 al XIX. Todos ellos proceden del Archivo Histórico de Protocolos de Madrid (AHPM), aunque para recopilar el corpus fue necesario consultar diferentes fondos documentales. Los inventarios de la capital se hallan en la sección de Protocolos Notariales, mientras que los inventarios de los municipios se encuentran recogidos en las papeletas del fichero de Protocolos Notariales de las localidades de la provincia de Madrid. En este fondo, constituido por 7.124 tomos, los protocolos se agrupan por distritos, posteriormente por localidades y a su vez por notarios, de manera que los protocolos reciben el nombre de la escribanía en la que se produjeron, es decir, de los notarios que los realizaron, procedentes de los distintos distritos notariales de la provincia.

En su conjunto, el corpus madrileño incluye inventarios de los 13 municipios señalados en el mapa:



Figura 4: Mapa de municipios de Madrid representados en el corpus

El listado completo de los documentos que han sido seleccionados para el corpus de Madrid se recoge en la tabla de abajo:

<i>Documento</i>	<i>Signatura</i>	<i>Año</i>	<i>Municipio</i>
Inventario de los bienes que quedaron por fin y muerte de Juan Fernández Vizcaíno de las Peñas	AHPM, T-11306 ff. 112v-114r	1700	Madrid
Inventario y tasación de los bienes que quedaron por fin y muerte de Bartolomé González	AHPM, T-11049 ff. 464v-466v ff. 471r-472r	1700	Madrid
Inventario de los bienes que quedaron por fin y muerte de Lucía González en herencia a Martín García	AHPM, T-11306 ff. 189r-193r	1701	Madrid
Inventario de los bienes que quedaron por fin y muerte de Alonso López	AHPM, T-11306 ff. 1033r-1033v ff. 1035r-1039v	1703	Madrid
Almoneda de los bienes de Francisco Ramírez realizada por Juan Montes como testamentario	AHPM, T-13966 ff. 1151v-1154r	1705	Madrid

Inventario de los bienes que quedaron tras fin y muerte de María Sánchez de Lázaro	AHPM, T-14527 ff. 323r-323v ff. 324v-325v ff. 326v-327r ff. 329r-329v f. 330v	1711	Madrid
Inventario de los bienes que quedaron por fin y muerte de Diego Espino	AHPM, T-14623 ff. 367r-369v	1715	Madrid
Inventario de los bienes que recibe Antonio de la Vega como dote por su casamiento con María Salazar	AHPM, T-42221 ff. 5r-6r	1717	Torrelaguna
Inventario de los bienes que quedaron por fin y muerte de Andrés Cacho de Herrera	AHPM, T-32930 ff. 72r-73r	1720	Pinto
Bienes que Manuel Sánchez recibe como dote por su casamiento con Paz Rodríguez	AHPM, T-29684 ff. 49r-50r	1722	Colmenar de Oreja
Bienes que Andrés de Vargas recibe como dote por su casamiento con Teresa de Mejorada	AHPM, T-14328 ff. 47v-51r	1724	Madrid
Inventario de los bienes y hacienda que quedaron por fin y muerte de María Fernández Calderilla	AHPM, T-14527 ff. 726r-732v	1728	Madrid
Bienes que Ángel Casanova recibe como dote por su casamiento con Francisca Caballero	AHPM, T-14833 ff. 26r-28r	1745	Madrid
Bienes que Pedro Fernández lleva a su segundo matrimonio tras la muerte de Lucía Fernández, su primera mujer	AHPM, T-14833 ff. 212r-212v	1747	Madrid
Carta de pago y recibo de dote a favor de D. ^a Paula Antonia Capellán	AHPM, T-15766 ff. 190r-208v	1750	Madrid
Bienes que Marcos Rodríguez recibe como dote por su casamiento con Leonor Rodríguez	AHPM, T-30640 ff. 159r-161v	1751	Brunete
Bienes que Tomás Muñoz recibe como dote por su casamiento con Catalina García	AHPM, T-19452 ff. 26v-29r	1765	Madrid
Inventario y división de los bienes que quedaron por fin y muerte de	AHPM, T-29776	1770	Estremera

Teresa Sarrín a sus herederos	ff. 438r-440v		
Inventario de los bienes Antonio Hernández recibe como dote por su casamiento con Francisca Aita Díaz de Aguilar	AHPM, T-33148 ff. 297r-299r	1779	San Lorenzo de El Escorial
Bienes que Matías Galiano aporta a su matrimonio con Manuela González de Lobera	AHPM, T-22426 ff. 55v-57v	1782	Madrid
Bienes que Gabriel del Peral recibe como dote por su casamiento con María Dominica Aguado	AHPM, T-32946 ff. 104r-108r ff. 109v-110r	1784	Pinto
Bienes que Pablo Ramón Domínguez recibe como dote por su casamiento con Francisca Miguel Rodríguez	AHPM, T-33152 ff. 20r-22v	1789	San Lorenzo de El Escorial
Inventario de los bienes que quedaron por fin y muerte de Diego Martín	AHPM, T-22426 ff. 280v-282v	1793	Madrid
Bienes que Santiago Gavilanes recibe como dote por su casamiento con Justa Varea	AHPM, T-30651 ff. 47r-50r	1800	Brunete
Bienes que Francisco Rodríguez recibe como dote por su casamiento con Vicenta Barrios	AHPM, T-23021 ff. 21v-23r	1801	Madrid
Bienes que María Josefa Delgado lleva a su matrimonio como dote	AHPM, T-23021 ff. 109v-111r	1805	Madrid
Bienes que Isidro Ferrer recibe como dote por su casamiento con Antonia de la Vega	AHPM, T-23021 ff. 247v-249v	1806	Madrid
Bienes que Eusebio de la Fuente aporta a su matrimonio con Anastasia Pérez	AHPM, T-41841 ff. 8r-12r	1809	Navalafuente
Inventario y tasación de los bienes quedados por fin y muerte de José Martínez Balledor	AHPM, T-21694 ff. 13r-17v	1815	Madrid
Bienes que Sebastián Antonio de la Brena recibe como dote por su casamiento con Martina de Ocaña y Peñalver	AHPM, T-21694 sf.	1818	Madrid
Inventario de los bienes que quedaron por fin y muerte de D.	AHPM, T-23219	1832	Madrid

Esteban Peyron y Merino	ff. 341r-341v f. 342v		
Bienes que Rufino Rey recibe como dote por su casamiento con Bibiana Gómez	AHPM, T-38423 ff. 37r-39r	1832	Navas del Rey
Bienes que Carlos Amallo recibe como dote por su casamiento con María Dolores Manget	AHPM, T-23524 ff. 11v-15r	1834	Madrid
Bienes que Celedonio López Alonso recibe como dote por su casamiento con Juana Hebrero	AHPM, T-44431 ff. 32r-33v	1834	Algete
Inventario de los bienes que quedaron por la fin y muerte de D. José Antonio Morales	AHPM, T-29864 sf.	1836	Estremera
Inventario de los bienes que quedaron por la fin y muerte de D. Antonio Álvarez	AHPM, T-25053 ff. 65r-68r	1840	Madrid
Inventario de los bienes que quedaron por la fin y muerte de D. José Laviña y Prat	AHPM, T-25211 ff. 57v-65r	1843	Madrid
Inventario de los bienes que quedaron por la fin y muerte de Josefa de Cuevas	AHPM, T-44479 ff. 161r-168r	1845	Ambite
Bienes que Bernardo Fernández recibe como dote por su casamiento con su futura esposa Mónica Jenara García	AHPM, T-26498 ff. 30r-33v	1852	Madrid
Bienes que Telesforo Simón recibe como dote por su casamiento con Dolores Raposo	AHPM, T-44436 ff. 15r-18r	1855	Algete
Bienes que Pedro González recibe como dote por su casamiento con Petra Almoguera	AHPM, T-33170 ff. 118r-121r	1858	San Lorenzo de El Escorial
Inventario y tasación de los bienes que quedaron por la fin y muerte de D. Antonio Carrascosa	AHPM, T-30174 ff. 103r-116r	1870	Aranjuez
Inventario de los bienes que quedaron por la fin y muerte de D. José M. ^a Fernández Crespo, vecino de la corte	AHPM, T-34560 ff. 51r-54v ff. 67r-73v	1877	Madrid
Inventario de los bienes que quedaron por la fin y muerte de Policarpo Cortecedo y Ayuso	AHPM, T-45918 ff. 854r-858r	1890	Alcalá de Henares

Bienes que Antonio Rodríguez recibe como dote por su casamiento con Margarita Casado	AHPM, T-37569 ff. 109r-111v	1893	Aranjuez

Tabla 3: Relación de documentos del corpus de Madrid

2.6. Corpus complementarios

El análisis estadístico sobre el uso de los diminutivos en los siglos XVIII y XIX se fundamenta en el corpus documental de esta tesis, compuesto por inventarios de bienes de Almería. Como ya se ha mencionado, el propio corpus de esta tesis incluye un subcorpus de Madrid con el que contrastar la información extraída, por lo que no se han tomado fuentes de control externas que sirvan de referencia al corpus de esta tesis.

El corpus en su conjunto ha servido también para desarrollar un análisis de tipo cualitativo a partir del cual confirmar la lexicalización del diminutivo en muchas de las palabras en los textos (§ 6.5.2. *Características de los diminutivos lexicalizados*), poder documentar voces escasamente documentadas (§ 6.5.2.1. *Regionalismos y voces poco documentadas*) o incluso aportar una definición de los diminutivos lexicalizados incluidos en él (§ 6.5.3.4. *Glosario*). Sin embargo, en algunas ocasiones, los ejemplos del corpus no han bastado para determinar con certeza algunos de estos aspectos y poder determinar con exactitud el valor semántico con el que se emplea el diminutivo, cuestión fundamental para este estudio¹⁹. Para ello ha sido necesario un análisis ulterior que va más allá de las zonas y periodos que abarca la documentación seleccionada. En dicho análisis cualitativo se han tenido en cuenta otras fuentes de datos ajenas a esta investigación y cuyos textos se han utilizado como referencia para el contraste de resultados. Para asegurar el criterio de comparabilidad textual, se han utilizado corpus históricos diseñados también en torno a inventarios de bienes o que contengan este tipo de textos. En este sentido, dos han sido los corpus principales que se han tomado como referencia y de los que se incluyen ejemplos en esta tesis:

a. *Oralia Diacrónica del Español (ODE)*

ODE se constituye como un “corpus de corpus” en el que se encuentran representadas diferentes zonas geográficas a lo largo de cuatro centurias (XVI-XIX). Concebido en sus orígenes como un corpus regional del reino nazarí de Granada (extendido a las actuales Málaga, Granada y Almería), ha ido ampliando su ámbito geográfico con la inclusión de documentos del resto de Andalucía y de otras regiones, y también de tipos textuales gracias al desarrollo de diferentes proyectos de investigación a la par que ha visto mejorado tecnológicamente el soporte informático en el que se ofrecen sus materiales (§ 2.1. *Antecedentes*). En lo que respecta a los inventarios de

¹⁹ Algunos ejemplos de la dificultad que ha supuesto en varios casos determinar el valor transparente o lexicalizado del diminutivo se analizan en § 6.3.2. *Casos dudosos y problemas para la cuantificación*.

bienes, los que se incluyen en *ODE* pertenecen a Andalucía (Huelva, Cádiz, Sevilla, Málaga, Granada y Jaén) y Extremadura (Cáceres y Badajoz) entre el espacio temporal comprendido entre los siglos XVI y XVIII²⁰.

Como se mencionó al inicio, si bien el corpus de esta tesis –conformado por los inventarios de Almería y Madrid– es un subcorpus dentro de *ODE*, en esta investigación ambos son considerados fuentes independientes: por un lado, el corpus *ODE* con su documentación andaluza y extremeña; por otro, el corpus documental de esta tesis, compuesto a su vez por un corpus de estudio y otro de referencia, de Almería y Madrid, respectivamente. Por ello, a la hora de incluir ejemplos procedentes de *ODE* se hace una distinción entre *ODE* – Almería y *ODE* – Madrid (corpus documental de esta tesis) y el resto de las provincias de *ODE* entendido como corpus complementario.

Del corpus *ODE* se toma únicamente la documentación relativa a inventarios de bienes, dejando a un lado cualquier otro documento que interfiera en la homogeneidad del tipo textual. Gran parte de los inventarios pertenecen al siglo XVIII, por lo que han servido para analizar las características de determinadas voces en el mismo espacio cronológico que se incluye en esta tesis. Los textos de *ODE* han servido para precisar la extensión geográfica de algunas palabras o comprobar sus contextos de uso con el objetivo de ayudar a definir o identificar el objeto al que aluden.

b. *Corpus léxico de inventarios (CorLexIn)*

Por su diseño específico, cerrado temáticamente en torno a inventarios de bienes, el *CorLexIn* ha sido también una fuente fundamental para el estudio cualitativo de esta tesis. Este corpus compila documentación notarial inédita del siglo XVII de prácticamente todas las provincias de España, pero también de otras partes de América, con el que se pretende dar a conocer el léxico hispánico del Siglo de Oro en el campo de los objetos de la vida cotidiana. Entre sus características más reseñables, destacamos su calidad filológica evidenciada en la cuidada selección y transcripción de los documentos, así como su tamaño, nada desdeñable para un corpus de características tan específicas como las que se han señalado. En el momento actual, su volumen supera las 1 600 000 palabras según se puede leer en su página web²¹. Precisamente su tamaño y su ambición por representar la mayor parte del territorio hispánico lo convierten en una fuente muy precisa para estudiar la variación diatópica y diacrónica del léxico de la vida cotidiana de la época (Morala, 2012a, 2014, 2015).

Para la investigación planteada en esta tesis, el empleo del *CorLexIn* como corpus complementario implica poder obtener una pieza previa en el estudio diacrónico de las voces con diminutivo. La información de nuestro corpus de inventarios de los siglos XVIII y XIX podrá complementarse con la relativa al XVII en el *CorLexIn*, que servirá para enriquecer con otros ejemplos los datos reflejados en nuestro corpus, especialmente en aquellos casos en los que nuestros inventarios no aportan información suficiente para el estudio de determinadas palabras.

²⁰ Según el estado actual del corpus *ODE* con fecha 20/06/2023.

²¹ <http://corlexin.unileon.es/> (fecha de consulta: 20/06/2023).

Teniendo en cuenta la diferencia en el marco temporal representado, del *CorLexIn* interesan no solo los inventarios procedentes de Almería y Madrid, con los que poder completar el recorrido de algunas voces, sino también los del resto de puntos, para comprobar la extensión geográfica o localizar ejemplos que ayuden a definir o aclarar el uso de determinados objetos en periodos anteriores al analizado en esta tesis.

c. Otros corpus complementarios

De manera ocasional, para algunos casos concretos se han consultado también como referencia otras fuentes basadas en textos archivísticos entre los que se documentan inventarios de bienes. Nos referimos a *CORDIAM*, *CODEA+ 2022*, *CORDE*, *CDH* y el *Léxico hispanoamericano de Peter Boyd-Bowman*, aunque este último no sea un corpus propiamente dicho.

Todos estos corpus han servido en algún momento para consultar voces con sufijación diminutiva de las que tanto su presencia como ausencia en ellos ha sido significativa para rastrear su difusión, tanto en el tiempo como en el espacio geográfico. El estudio cualitativo llevado a cabo acerca de los diminutivos en *-ico*, *-ito* e *-illo* se ha visto enriquecido con las aportaciones que ofrecen estas fuentes y de las más arriba descritas.

3. EDICIÓN DIGITAL DEL CORPUS EN XML-TEI

El punto de partida para la edición del corpus han sido las cerca de setecientas imágenes de los inventarios manuscritos tomadas en los archivos de Almería y Madrid. El objetivo principal fue el de conservar lo más fielmente posible toda la información textual y extratextual que proporcionan las imágenes a partir de una metodología adecuada que lo permitiera y que fuese más allá de una simple digitalización del texto.

En este punto, conviene distinguir entre una edición digitalizada, definida más bien por el formato, y generalmente con vistas a una impresión en libro; y una edición digital, definida por la metodología y enfocada en aprovechar los últimos avances tecnológicos en investigación lingüística con el objetivo final de explotar al máximo todas las posibilidades que ofrecen los datos (Pierazzo, 2016, p. 16). Esta última, por tanto, perdería gran parte de su potencial si se ofreciera en papel, pero su publicación en línea permite conservar tanto sus características paleográficas como extratextuales con rigor y fidelidad. Una edición de este tipo encaja con el verdadero significado de las denominadas Humanidades Digitales y supera el paradigma clásico según el cual el simple uso instrumental de lo informático justifica la denominación de “corpus digital”.

En este capítulo se desarrolla la metodología empleada para la edición digital del corpus de esta tesis. En primer lugar, se describen los criterios de transcripción que se han seguido (§ 3.1. *Criterios de transcripción*). La transcripción se ha realizado en el lenguaje informático XML (§ 3.2. *Transcripción en lenguaje XML*) adaptado al estándar TEI, un esquema de codificación especializado y de alcance internacional empleado comúnmente para la codificación de textos en Humanidades (§ 3.3. *Codificación del manuscrito en XML-TEI*). En las páginas que siguen se da respuesta a estos dos conceptos básicos: qué es XML y qué es TEI, y cómo se han utilizado ambas herramientas para la elaboración de *Oralia Diacrónica del Español (ODE)*.

3.1. Criterios de transcripción

La transcripción de los textos que integran el corpus de esta tesis se hace coincidir en lo fundamental con los criterios propuestos por la Red *CHARTA* (2013) para la edición paleográfica estrecha de fuentes archivísticas hispánicas. El objetivo es lograr una edición que asegure el rigor filológico de las transcripciones y con la que conservar las principales características del manuscrito original que puedan ser de interés para los usuarios del corpus que acceden a él con distintos fines investigadores.

Los criterios de *CHARTA* se han adaptado para conseguir una edición digital que permita un mejor aprovechamiento del corpus. Se obtiene así una edición semipaleográfica que, con respecto a los criterios originales de *CHARTA*, únicamente difiere de la edición paleográfica estrecha en los aspectos relativos a:

- a. El uso de mayúsculas y minúsculas.
- b. La puntuación.
- c. La separación de palabras.

El uso de las mayúsculas y minúsculas en el texto se ha modificado según la norma ortográfica actual, eliminando así cualquier mayúscula inicial que no coincida con los criterios modernos o añadiéndola según corresponda, como ocurre en el caso de muchos nombres propios de persona o municipio en los inventarios (i.e. *juan* o *Crucifijo* en el original, transcritas como *Juan* y *crucifijo*). Este criterio supuso además la regularización de mayúsculas en mitad de palabra, habitual en el *usus scribendi* de algunos escribanos, incluidos también aquellos casos en los que la *r* mayúscula aparece con valor de vibrante múltiple en palabras como *aRoio* (arroyo), *caRo* (carro) y transcritas como *aroio* y *caro*. Puesto que se ha intervenido el uso de mayúsculas y minúsculas, aquellos interesados en el estudio de este aspecto deben consultar la edición facsímil que se aporta (§ 4.7. *Triple visualización de los textos*).

La puntuación del manuscrito también ha sido modernizada según la ortografía vigente. Esto implica tener en cuenta dos aspectos: por un lado, significa que se ha eliminado cualquier signo de puntuación que aparezca en el original cuando este no se corresponda con el uso actual; por otro, debe entenderse que la puntuación ha sido intervenida y que si esta coincide con el original no podrá saberse a no ser que se consulte el facsímil del documento, disponible siempre junto a la transcripción.

Por último, se hace una separación modernizada de todas las palabras en el texto aun cuando estas aparecen ligadas en el original. Este criterio implica también la separación de contracciones no estándares como *deste* o *queste* –transcritas como *d este* y *qu este*–. Este criterio, facilita la lectura de los textos y simplifica el procesamiento posterior del texto, esencial para la identificación de tokens en el corpus (§ 4.2. *Tokenización*).

Se considera que estos cambios no tienen trascendencia lingüística ni interfieren en la calidad filológica de las transcripciones, pero favorecen que cualquier usuario haga uso del corpus aun cuando su interés no sea puramente lingüístico. La edición refleja una fiel transcripción del texto con la conservación de las grafías originales y con ello se asegura la viabilidad del corpus para el estudio de fenómenos fonéticos dialectales como la neutralización *r/l* (*almadura*, *arjofar*), cambios en el punto de articulación (*abuja*), pérdida de *-d-* intervocálica (*arao*, *abrioros*), seseo (*almires*, *alasena*) o ceceo (*arteza*, *bracero*). El uso de otras grafías, tengan o no interés fonético, se han mantenido igualmente para asegurar la fiabilidad con respecto a la fuente original. De esta manera, se respeta la variación grafemática de los siguientes casos:

- Grafías *j/y* con valor vocálico /i/: *ynventario*, *beynte*.
- Variación *j/x* y *z/c*, *ç*: *alaxas*, *biexo*, *açeite*, *lienço*, *bronze*...
- Grafía *n* cuando esta tiene valor de /ɲ/: *pequeno*, *barreno* (por *barreño*)...
- Duplicaciones gráficas al inicio o mitad de palabra: *arttessa*, *cassa*...
- Uso del signo @ como abreviatura de *arroba* para la unidad de medida en casos como “cuatro @ de tocino” o “una tenaga (*tinaja*) de beinte @”.
- Grafías *v/u* para /u/: *uino*, *vno*...

Entre las grafías difíciles de interpretar, la variación *v/u* ha planteado muchos casos en los que no ha sido posible establecer patrones caligráficos claros. En los casos dudosos

sin resolver se ha tendido siempre hacia la normalización, esto es, a la transcripción de *v* para la consonante y *u* para la vocal.

3.2. Transcripción en lenguaje XML

XML (*eXtensible Markup Language*) es un lenguaje informático empleado para anotar textos de manera que puedan ser procesados automáticamente. Forma parte de los llamados lenguajes de marcado, es decir, aquellos que añaden marcas o etiquetas a un texto para señalar una u otra característica presente en la fuente original (una abreviatura, un roto en la página, una palabra que se encuentra tachada, una adición posterior, etc.). Estas etiquetas dotan de significado al texto para que posteriormente sean interpretadas por un programa informático, el cual transforma la información en un tipo concreto de visualización (unas letras voladas, un corchete con puntos, un fragmento de texto tachado, etc.). Por consiguiente, XML permite describir las características que están presentes en un texto, tanto aquellas relacionadas con el formato como las relativas al contenido semántico del texto, por ejemplo, lo que es un título, una fecha o una firma. La marcación de todas estas cuestiones permite organizar y describir los datos basados en texto haciendo explícita la información para su procesamiento informático.

XML conforma un lenguaje internacionalmente conocido y útil para la codificación de textos. Permite crear cualquier etiqueta personalizada a la que se le otorga el valor que se desee para estructurar los datos, pero no existe un conjunto cerrado de etiquetas con un significado preestablecido, sino que estas se definen para crear los documentos. Ahora bien, basados en XML se han creado estándares para la edición de textos. Tal es el caso de las directrices propuestas por el consorcio TEI, que propone una metodología específica de codificación en el ámbito de las Humanidades que ha servido de base para multitud de proyectos internacionales, así como para la edición del corpus de esta tesis (§ 3.3. *Codificación del manuscrito en XML-TEI*).

Por su fácil procesamiento, ya que admite múltiples salidas posteriores de visualización e impresión, XML se encuentra en la base de la gran mayoría de las ediciones digitales académicas que se realizan actualmente (Pierazzo, 2016, p. 25). Uno de sus principales reclamos radica en las múltiples posibilidades orientadas a la visualización del texto. Las ediciones digitales procuran una publicación cuidada y atractiva que permita al usuario navegar por los textos de forma dinámica e interactiva: acceder al facsímil, contrastar diferentes versiones, omitir en la lectura las intervenciones editoriales o mostrar el desarrollo de las abreviaturas. En este sentido, este tipo de ediciones posibilitan un acercamiento a los textos como antes no era posible, desarrollando al máximo las posibilidades de visualización y edición. Un fantástico ejemplo de ello es el corpus *Exploring Medieval Mary Magdalene*²², en el

²² <https://maria-magdalena.ub.uni-freiburg.de/>

que se recopilan manuscritos medievales creando un recurso digital en el que se da solución a las limitaciones tradicionales de las ediciones manuscritas impresas.

Frente a estos objetivos de tipo presentacional, la marcación de textos cumple además un papel fundamental a la hora de recuperar información a partir de búsquedas selectivas. Por ejemplo, rescatar en una edición todos los nombres propios que aparecen en un documento o buscar en una colección de textos todos aquellos pertenecientes a un determinado año solo es posible si previamente se han marcado estos aspectos. La edición digital permite aunar información textual y metatextual y marcar todo aquello que se desea recuperar con posterioridad. Se trata de poder aprovechar al máximo todas las opciones que ofrece un texto marcado cuya información puede ser recuperada tras el debido tratamiento informático.

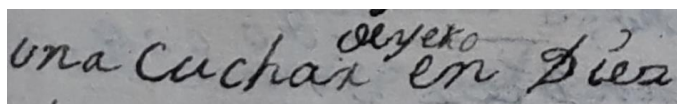
En definitiva, el anotado de textos en Humanidades, y de manera particular el de los textos manuscritos, ha permitido superar los inconvenientes que presentaban las ediciones impresas: en primer lugar, la visualización del formato y de algunas características escritas difícilmente trasladables al papel; en segundo lugar, la búsqueda y selección de información concreta que resulta imposible de recuperar de forma automática en una obra impresa. Las ventajas de la marcación en XML van más allá de la posterior visualización de los textos. Tanto o más relevante que esta, son para las investigaciones lingüísticas las posibilidades de explotación y recuperación de datos previamente marcados.

Con el objetivo de ofrecer una edición digital visualmente atractiva y que a la vez permita su posterior aprovechamiento lingüístico, el corpus *ODE* basa su metodología en el uso de XML para la transcripción de los textos. La sintaxis específica que requiere y el procedimiento de edición se detalla en las siguientes páginas.

3.2.1. Conceptos básicos

La sintaxis de XML se fundamenta en el uso de etiquetas que se representan entre caracteres angulares < >. Estas permiten marcar la información y organizarla siguiendo una estructura arbórea, ya que una etiqueta puede contener a su vez otras distintas. De esta manera, todos los elementos de un archivo XML se disponen de manera jerarquizada con elementos que incluyen otros elementos y que se encuentran correctamente anidados.

Para entender mejor el funcionamiento del lenguaje XML, explicaremos sus componentes básicos a partir del siguiente caso concreto de edición:



En el texto encontramos la frase “una cuchar de yero en diez”, pero se quiere dejar constancia en la transcripción de que las palabras “de yero” han sido añadidas posteriormente, concretamente se encuentran escritas por encima de la línea. Esta

característica presente en el manuscrito original se señala en XML por medio de etiquetas específicas que aportan esa información.

Para marcar el texto se utilizan etiquetas que comienzan con la forma <etiqueta> y terminan con la forma </etiqueta>. Estas últimas incluyen una barra inclinada después del ángulo de apertura y entre ambas se sitúa el texto que se quiere marcar. En nuestro ejemplo, se quiere señalar que las palabras “de yero” presentan una particularidad frente al resto, por lo que serán solo estas las que se marcan. Utilizaremos la etiqueta <add> (*addition*) para señalar que “de yero” es una adición por encima de la línea:

una cuchar <add>de yero</add> en diez

La suma de la etiqueta de inicio, la etiqueta de cierre y el texto marcado entre ambas conforman lo que se denomina un *elemento*.

Hasta ahora, se ha señalado que las palabras marcadas con <add> son una adición, pero esta etiqueta no añade información respecto a cómo han sido añadidas o el lugar que ocupan en el documento. Es posible precisar más la información que aportan las etiquetas completándolas con características nuevas que se incluyen en cada elemento en forma de *atributos*, los cuales se añaden siempre en la etiqueta de inicio. Por ejemplo, en nuestro caso, podemos especificar el lugar en el que el escribano ha escrito la adición. Los atributos (en naranja abajo) son las propiedades acerca de los elementos y su aparición dentro de una etiqueta es opcional. La información complementaria que se proporciona al atributo recibe el nombre de *valor* (en marrón) y se marca entre comillas. Atributo y valor se separan mediante el signo igual:

una cuchar <add place= "above">de yero</add> en diez

En resumen, se ha marcado solo una parte del texto con etiquetas <add> </add> que informan de que el fragmento ha sido añadido. Además, se completa esta información mediante un atributo para indicar el lugar (*place*) de la adición en el documento. El valor señala la propiedad del atributo, en este caso, por encima de la línea (*above*). De este ejemplo se deduce que la sintaxis de XML sigue la siguiente estructura:



Figura 5: Sintaxis básica de XML

La marcación de texto permite, además de representar visualmente las características textuales de la edición digital (Figura 6), recuperar posteriormente todas las palabras o

fragmentos en el corpus que se hayan marcado con una misma etiqueta, en este caso `<add>` (para ello véase § 4.9.5. *Búsqueda genérica*).

una cuchar^{de yero} en diez

Figura 6: Visualización en *ODE* del texto marcado con el elemento `<add>` `</add>`

Además de los elementos señalados, puede haber elementos vacíos que no precisan de contenido textual, como son las órdenes referidas a cuestiones estructurales. Por ejemplo, para señalar un cambio de línea no es necesario marcar todo el texto que incluye esa línea, sino que bastará con indicar en qué parte del texto se produce el cambio de línea. Este tipo de cuestiones, como los cambios de párrafo o el comienzo de una nueva página en el documento, pueden indicarse con un elemento que no necesita una etiqueta de inicio y otra de cierre, basta con una única etiqueta de apertura que termina con una barra inclinada y ángulo de cierre: `<etiqueta/>`. Si se quiere indicar el inicio de una nueva línea se marcará con el elemento vacío `<lb/>` (*line beginning*):

- (1) `<lb/>` Dos colchas de lana y una de zaraza. `<lb/>` Tres paños de lana. `<lb/>` Dos camas de madera. `<lb/>` Cuatro fundas pobladas (*ODE* – Almería, Fiñana, 1840).

Los elementos vacíos señalan una característica textual que se cumple hasta que se añade de nuevo el mismo elemento.

3.2.2. Reglas de buena formación

XML no se considera un lenguaje excesivamente complejo de producir ni de leer por un humano, pero sí atiende a una sintaxis específica que exige una estricta formación: solo si el documento está bien elaborado podrá ser procesado después. De modo que es importante comprobar que no existen errores de formación mediante la validación automática del archivo en cualquier procesador de texto. A continuación, se señalan algunas de las normas básicas que debe cumplir todo documento XML bien construido:

- a. Todas las etiquetas deben ir en pares (inicial y final). Cada elemento debe tener su correspondiente marca de apertura y de cierre, a excepción de los elementos vacíos, tal como sigue:

`<etiqueta>texto</etiqueta>`
`<etiqueta/> texto`

- b. Los elementos pueden contener texto, pueden contener a su vez otros elementos o pueden contener ambas cosas, texto y otros elementos:

`<etiqueta1>`

```
<etiqueta2> texto </etiqueta2>
</etiqueta1>
```

- c. El lenguaje XML se estructura de manera jerárquica, por lo que en cualquier documento debe haber siempre un elemento raíz que contenga a todos los demás, como el elemento <text> aquí abajo:

```
<text>
    <lb/>
    <lb/>
    <lb/>
    <lb/>
</text>
```

- d. XML permite el uso de comentarios internos que no serán procesados siempre que se delimiten entre los caracteres correspondientes: <!-- -->. Los comentarios no permiten la secuencia de guiones en su interior a excepción de los necesarios al inicio y final y no pueden introducirse dentro de las etiquetas, pero sí de los elementos, tal y como se muestra a continuación:

```
<etiqueta> texto <!-- esto es un comentario --> </etiqueta>
```

- e. Los caracteres angulares < > y las comillas " " tienen significado específico en el lenguaje de marcación, pues sirven para delimitar instrucciones y son interpretadas como parte de la sintaxis de XML. Por tanto, si en el documento original aparece algunos de estos caracteres de manera literal se debe escapar su uso en la transcripción mediante las siguientes estrategias:

<i>Carácter</i>	<i>Sustitución</i>	<i>Estrategia</i>
< (menor que)	lt (less than)	<
> (mayor que)	gt (greater than)	>
' (comilla simple)	apos (apostrophe)	'
" (comillas dobles)	quot (quotation mark)	"
& (et)	amp (ampersand)	&

Tabla 4: Relación de caracteres reservados en XML

XML exige una sintaxis de formación específica, pero no establece ningún conjunto cerrado de etiquetas, de manera que cada usuario puede crear sus propias etiquetas de acuerdo con sus intereses particulares. Sin embargo, deben tenerse en cuenta algunas consideraciones básicas en cuanto a la nomenclatura de las etiquetas y la formación de estas:

- f. XML es un lenguaje sensible a la detección de mayúsculas y minúsculas, lo que en la jerga informática se conoce como *case-sensitive*. Esto significa que los pares de etiquetas deberán coincidir también en este aspecto:

```
<etiqueta> </etiqueta>
<Etiqueta> </Etiqueta>
*<etiqueta> </Etiqueta>
```

- g. Los nombres de las etiquetas deben comenzar con una letra o guion bajo. El resto de los caracteres sí podrán ser letras, dígitos, guion bajo, punto o guion, exclusivamente:

```
<nombre_propio> </nombre_propio>
<ciudad-capital > </ciudad-capital >
<nombre.mujer> </nombre.mujer>
<autor> </autor>
```

- h. Convencionalmente, las etiquetas se escribirán siempre en minúscula, a excepción de aquellas compuestas por dos palabras, en las que la segunda de estas se escribirá con mayúscula inicial:

```
<nombrePropio>Pilar</nombrePropio>
```

- i. Los espacios en blanco son irrelevantes en XML, pero estos no se permiten dentro de las etiquetas después de los caracteres angulares:

```
<nombre >Pilar</nombre >
*< nombre>Pilar</ nombre>
```

- j. La nomenclatura de los atributos sigue las mismas reglas que los nombres de las etiquetas. Estos se colocan siempre en la etiqueta inicial y deben escribirse entre comillas inglesas (dobles o simples):

```
<add place= "above"> </add>
<add place= 'above' /> </add>
```

- k. Un elemento puede tener varios atributos y el orden de estos es irrelevante, pero no puede tener dos atributos iguales aun con distintos valores. El siguiente ejemplo indica el inicio de una nueva página con los atributos referidos al número (*n*) y a la imagen del facsímil con la que se vincula el texto (*facs*):

```
<pb n= "3v" facs= "foto006.jpg"/>
```

3.2.3. Editor de texto XML

La dificultad que supone el uso y aprendizaje del lenguaje XML está hoy superada gracias a la multitud de procesadores de texto que asisten durante el proceso de escritura y facilitan la tarea del etiquetado. La transcripción de los documentos que conforman el corpus de esta tesis se ha llevado a cabo íntegramente con el programa informático *Oxygen XML Editor*²³, uno de los muchos disponibles para facilitar el procedimiento de marcación de texto en lenguaje XML. Este programa es el de uso más extendido en el campo de las Humanidades Digitales debido a que es compatible con los principales sistemas operativos y al amplio conjunto de herramientas de creación y desarrollo de XML que proporciona. *Oxygen* incorpora multitud de plantillas para trabajar con los distintos esquemas existentes de codificación, entre los que se encuentra el propuesto por el consorcio TEI (*Text Encoding Initiative*)²⁴, al que se alude en el siguiente apartado.

Debido a que se requiere que los documentos XML sean válidos para poder ser procesados, es imprescindible que el documento esté correctamente elaborado de acuerdo con su sintaxis específica. Cuanto más extenso es un documento y mayor número de etiquetas contiene, más complejo y difícil de leer resulta para el investigador, por lo que aumentan las probabilidades de cometer errores de sintaxis. El formato XML exige que cada elemento se abra y se cierre conforme a las reglas de buena formación anteriormente descritas. Además, su estricta jerarquía requiere que todo elemento esté incluido a su vez en otro superior sin que se produzcan solapamientos. Para evitar errores de este tipo, el programa resulta de gran ayuda, pues asiste al usuario en la codificación al incorporar indicadores de corrección que señalan dónde se encuentra el error y cómo debe solucionarse cuando el documento no está bien formado sintácticamente:

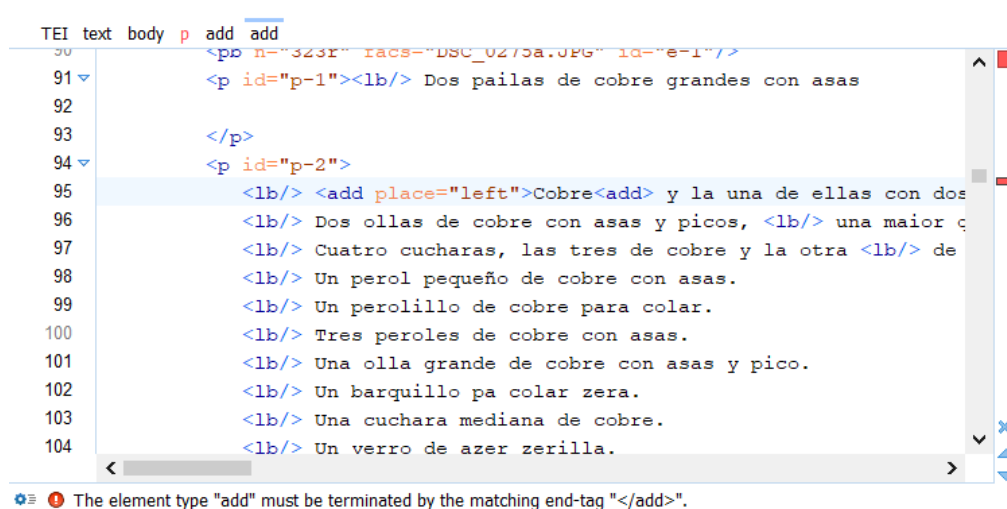


Figura 7: Aviso de error en el programa *Oxygen*. Elemento `<add>` `</add>` mal cerrado

²³ Licencia de uso previo pago y descargable en <https://www.oxygenxml.com/>. La versión utilizada ha sido la 19.1.

²⁴ <https://tei-c.org/>.

Para la transcripción de cada documento en *Oxygen* se ha seguido el estándar propuesto por el consorcio TEI (2018). El programa permite incorporar la plantilla modelo que se asocia con el esquema internacional XML-TEI y validar automáticamente los documentos de acuerdo con este estándar de edición. La validación consiste en comprobar que las etiquetas empleadas son las exigidas y que el documento cumple con las características necesarias y la estructura básica que debe tener cualquier documento asociado a TEI. Entre otras facilidades, el programa permite además realizar tareas de manera semiautomática, como autocompletar etiquetas, cerrar elementos abiertos o insertar atributos mediante menús desplegables en los que se muestran todos los valores posibles asociados a una misma etiqueta.

En definitiva, el programa proporciona numerosas utilidades que han amenizado y facilitado la tarea de la transcripción en XML. Las pautas seguidas para la codificación de los textos se detallan a continuación.

3.3. Codificación del manuscrito en XML-TEI

El lenguaje XML establece los requisitos obligatorios para transmitir información y que esta pueda ser procesada automáticamente y entendida por un ordenador, pero no proporciona el contenido exacto de las etiquetas XML que sirve para expresar las determinadas características que presenta la fuente original. Para ello existen estándares internacionales que fijan unas instrucciones precisas a la hora de codificar los textos. Estas directrices surgen con el propósito de alcanzar un método homogéneo en la codificación de textos que pueda ser utilizado de manera común y favorezca el intercambio de la información entre investigadores de todo el mundo. En el ámbito de las Humanidades, el estándar más conocido y extendido es el propuesto por el consorcio TEI (*Text Encoding Initiative*).

TEI ha desarrollado unas directrices basadas en la sintaxis XML para la representación digital de textos pertenecientes a muy diferentes géneros y tipologías. Para los estudios históricos, interesan especialmente los módulos dedicados a la codificación de textos manuscritos²⁵. En ellos se definen los elementos, etiquetas, atributos y valores que deben utilizarse para cada una de las características reflejadas en el manuscrito original que sirven para la representación de las características estructurales, físicas, visuales y conceptuales de este tipo de documentos. Por ejemplo, ya se señaló anteriormente que el elemento que indica una adición en el texto es <add> </add>. El etiquetado siguiendo la estrategia de marcado propuesta por TEI posibilita un análisis y explotación de los textos más allá de la simple versión digitalizada del manuscrito, ya que incluye no solo la codificación de las características formales, sino también de las de contenido.

La comunidad TEI proporciona además una amplia documentación sobre el uso de las guías con constantes modificaciones y actualizaciones según las necesidades de

²⁵ Concretamente los módulos 10-12 (“Manuscript Description”, “Representation of Primary Sources” y “Critical Apparatus”) de la guía P5 (Text Encoding Initiative Consortium, 2018).

los distintos grupos de trabajo que conforman la comunidad. El consorcio se mantiene continuamente activo y sus novedades se dan a conocer a través de conferencias anuales y listas de distribución. También en el sitio web es posible encontrar y descargar las distintas versiones de las directrices TEI. En el comienzo de elaboración del corpus de esta tesis, las pautas que se han seguido son las establecidas en la versión 3.4. de la guía TEI P5 (Text Encoding Initiative Consortium, 2018).

La principal ventaja de TEI radica en que sus directrices conforman ya un estándar para la edición digital sobradamente consolidado y desarrollado que ha servido como guía para multitud de proyectos de muy diversa naturaleza. Las pautas se publican en código abierto, lo que significa que no dependen del uso de ningún programa informático específico y las guías, ampliamente ejemplificadas, dan respuesta a casi cualquier caso de edición que se encuentre en una fuente manuscrita.

En el ámbito de la historia de la lengua española, la Red *CHARTA*, dedicada a la edición de documentación archivística hispánica, se ha interesado por adaptar la transcripción de manuscritos a los estándares TEI. El resultado tras varios años de trabajo ha sido la publicación de una guía de marcación XML-TEI basada en los antiguos criterios de edición *CHARTA* (Isasi Martínez et al., 2020) que se espera implementar próximamente y a la que se acogerán todos los grupos que conforman la Red. Por otro lado, de manera ya finalizada, la transcripción de la totalidad de los documentos que se incluyen en el proyecto *Post Scriptum* se realizó íntegramente y desde sus comienzos siguiendo el estándar XML-TEI (Vaamonde, 2016).

Las guías TEI permiten la selección de módulos optativos en función de los intereses particulares de cada investigación. Algunas características son personalizables, como por ejemplo los valores de un atributo o la posibilidad de elegir entre distintas soluciones de codificación para un mismo fenómeno. Al no tratarse de esquemas rígidos, cada grupo o proyecto puede adaptar las directrices generales a sus criterios de transcripción según sus intereses. Esta posibilidad, si bien puede ser una ventaja en varios aspectos, también puede convertirse en el origen de inconsistencias entre los proyectos que usan TEI. En este punto, es necesario hacer algunas puntualizaciones motivadas por la propia metodología de edición de nuestro corpus de inventarios, y de manera general del corpus *ODE*, en el que se integra.

En primer lugar, para la edición digital de los documentos se ha partido del modelo tecnológico propuesto por el proyecto europeo *Post Scriptum*, por el cual el procesamiento lingüístico de los datos se lleva a cabo con el sistema web TEITOK. Este sistema permite dar salida a la visualización y aprovechamiento de ediciones basadas en XML-TEI, pero su utilización implicó a hacer algunas modificaciones con respecto al uso estricto de las directrices TEI. Ello ha supuesto emplear algunas estrategias en la transcripción de los textos que nada tienen que ver con el estándar TEI, como por ejemplo la marcación de las abreviaturas (§ 3.3.2.2. *Abreviaturas*), pero que facilitan el posterior procesamiento de los documentos en TEITOK, bien para agilizar los procesos de tratamiento lingüístico, bien por el propio funcionamiento del sistema. No obstante, esto no ha sido impedimento para crear documentos igualmente válidos de acuerdo con el esquema TEI. Las directrices se han seguido lo más fielmente posible aun en el uso

de los valores de los atributos en lengua inglesa, atendiendo al carácter de internacionalización del corpus.

En segundo lugar, la estructura de los textos de *Post Scriptum*, conformado por cartas, difiere sustancialmente de la que aparece en los inventarios de bienes. Con respecto a *Post Scriptum*, para la edición de *ODE* se han realizado algunas alteraciones en el marcado de los textos, sobre todo en cuanto a las divisiones estructurales que caracterizan un texto epistolar. De esta manera, el uso de las etiquetas ha variado sustancialmente en uno y otro corpus.

El cumplimiento del estándar TEI en los documentos lo marca la existencia de un elemento raíz <TEI> </TEI>. El uso de este estándar determina qué elementos son obligatorios y optativos, los elementos que se pueden utilizar o no dentro de otros elementos, así como el nombre de los atributos utilizados. La estructura de un documento TEI, así como el marcado de cuestiones editoriales, rasgos físicos, información semántica u otros aspectos, puede ser todo lo simple o compleja que se desee, dado el amplio número de casos de edición que se plantean en sus directrices y las variadas tipologías textuales a las que se intenta dar cabida. A continuación, siguiendo el estándar TEI, se detallan las pautas adoptadas para la edición del corpus de inventarios de esta tesis, de manera particular, y para *ODE*, en general.

Cada texto transcrito en el corpus contiene dos grandes elementos <teiHeader> y <text> que se abordarán respectivamente en las siguientes secciones.

3.3.1. Cabecera <teiHeader>

El elemento <teiHeader> contiene los metadatos relevantes que se asocian a cada documento, es decir, toda la información complementaria al texto (año de creación, lugar, archivo que lo custodia, signatura, etc.). Constituye la cabecera del documento y consta de cuatro elementos principales que se detallan a continuación. Para cada uno de estos elementos se añaden ejemplos reales de casos documentados en los inventarios del corpus. Un ejemplo de la estructura que posee la cabecera de cualquier documento en *ODE* puede verse en § Anexo I. *Ejemplo completo de cabecera etiquetada*.

3.3.1.1. Descripción bibliográfica del archivo <fileDesc>

De los cinco elementos que puede incluir la cabecera, el elemento <fileDesc> (*file description*) es el único de presencia obligada según las reglas de buena formación de un documento TEI. En él se incluye la descripción bibliográfica completa del archivo electrónico y se compone de tres elementos también obligatorios para el cumplimiento del estándar TEI:

- a. <titleStmnt> (*title statement*) agrupa información sobre el título del archivo electrónico y los responsables de su contenido en las distintas tareas y fases de edición. Contiene a su vez los siguientes elementos:

- `<title>` proporciona el título del archivo electrónico. En el título se hará constar para cada documento su tipología dentro de los inventarios: carta de dote, testamento, partición de bienes, almoneda, reparto de bienes tras la muerte, etc.; así como el beneficiario de los bienes o el fallecido en el caso de testamentos y repartos.

(2) `<title>`Bienes que Pedro Carmona recibe como dote por su casamiento con Catalina García`</title>` (*ODE* – Almería, Tabernas, 1709).

(3) `<title>`Inventario de los bienes que quedaron por fin y muerte de Diego Martín`</title>` (*ODE* – Madrid, 1793).

- `<editor>` hace referencia al proyecto en el que se inscribe el documento. La realización de esta tesis se inserta en el proyecto *CORDEREGRA*, por lo que la totalidad de los documentos se asociarán a este mismo proyecto de investigación.

(4) `<editor>`CORDEREGRA: Corpus diacrónico del español del reino de Granada. 1492-1833`</editor>`

- `<funder>` alude a la fuente de financiación con la que fue posible realizar la edición digital.

(5) `<funder>`Junta de Andalucía/FEDER: P09-HUM 4466, 2009-2014`</funder>`

- `<respStmt>` informa sobre los responsables de distintos aspectos de la edición. Aparecerá tantas veces como sea necesario e incluirá un elemento `<resp>` con el nombre completo del responsable y un atributo *id* donde se señala la tarea llevada a cabo:

(6) `<respStmt>`
`<resp id="transcription">`Pilar Arrabal Rodríguez`</resp>`
`</respStmt>`
`<respStmt>`
`<resp id="standardization">`Pilar Arrabal Rodríguez`</resp>`
`</respStmt>`
`<respStmt>`
`<resp id="annotation">`Pilar Arrabal Rodríguez`</resp>`
`</respStmt>`

- b. `<publicationStmt>` (*publication statement*) contiene información sobre la institución, el proyecto o el grupo de investigación responsables de la

publicación y distribución de la edición digital. De ello se deja constancia en los siguientes elementos:

- <publisher> señala la institución responsable de la publicación.
- <pubPlace> indica el lugar donde se realiza la publicación.
- <distributor> proporciona el nombre de la persona, grupo o institución responsable de la distribución del texto.

(7) <publicationStmt>

```
<publisher>UGR, Universidad de Granada</publisher>  
<pubPlace>Granada</pubPlace>  
<distributor>HUM-278. Grupo de Investigaciones Histórico  
Lingüísticas y Dialectales. UGR Junta de Andalucía</distributor>  
</publicationStmt>
```

- c. <sourceDesc> (*source description*) describe la fuente a partir de la cual se ha derivado la edición digital e incluye el elemento <msDesc> (*manuscript description*) con toda la información referente al manuscrito único con el que se identifica la edición digital. Este elemento contendrá a su vez los siguientes:

- <msIdentifier> facilita la identificación del manuscrito en el archivo histórico del que procede mediante los siguientes elementos:
 - <country> para el país donde se ubica el archivo.
 - <settlement> para la ciudad a la que pertenece el archivo.
 - <institution> para el nombre completo del archivo.
 - <repository> para el fondo documental.
 - <idno> para la signatura que indica la localización del documento dentro del archivo. Irá siempre seguida de las siglas oficiales del archivo al que pertenece y separadas sigla y signatura por un espacio: AHPAL para el archivo almeriense y AHPM para el madrileño.

(8) <msIdentifier>

```
<country>España</country>  
<settlement>Almería</settlement>  
<institution>Archivo Histórico Provincial de Almería</institution>  
<repository>Fondo público judicial de ámbito local</repository>  
<idno>AHPAL P-478</idno>  
</msIdentifier>
```

- <msContents> describe el contenido intelectual del manuscrito, así como los elementos que los estructuran.

- `<summary>` incluye un resumen opcional del texto. En otros casos, cuando este se omite, aparecerá la abreviatura NA (no aplica).
- `<msItem>` describe elementos dentro del contenido intelectual del manuscrito. Incluye un atributo *class* donde se especifica el tipo de manuscrito fuente a partir del cual se ha realizado la edición (original, microfilm, fotocopia). Este elemento contiene a su vez la etiqueta `<locus>` con la foliación del documento. Se indicará el número de folio inicial y final separados por guion. En caso de que no exista foliación se indicará con la abreviatura *sf.* (sin foliación). En cambio, si la foliación está conjeturada se indicará entre corchetes, por ejemplo: [37r-40v].

(9) `<msContents>`

```

<summary>NA</summary>
<msItem class="original">
  <p><locus>sf</locus></p>
</msItem>
</msContents>

```

- `<additional>` incluye información adicional sobre el documento, como lo referente a permisos de publicación o su reproducción fotográfica. Contiene a su vez los elementos:

- `<adminInfo>` con un elemento `<availability>` dentro de `<p>` en el que se alude a los permisos de reproducción de las fotografías del manuscrito para cada archivo.
- `<surrogates>` con un elemento `<bibl>` dentro de `<cit>` en el que se indica el formato en el que se ofrecen las reproducciones de los manuscritos.

(10) `<additional>`

```

<adminInfo>
  <availability>
    <p>Para hacer uso de las imágenes es necesaria la
    autorización del Archivo</p>
  </availability>
</adminInfo>
<surrogates>
  <cit><bibl>facsimil    digital    guardado    en    formato
  JPEG</bibl></cit>
</surrogates>
</additional>

```

3.3.1.2. La tipología textual <encodingDesc>

En el elemento <encodingDesc> (*encoding description*) se documenta la codificación del texto electrónico. En este apartado se ha incluido la información relacionada con la tipología textual de los documentos que conforman el corpus *ODE*²⁶. Para ello, dentro del elemento <classDecl> se ha creado una taxonomía que define los códigos de clasificación utilizados en el corpus. En el caso de los inventarios de bienes el código que los identifica será siempre el de *inv* (inventario). Dentro de este elemento se deja constancia de la clasificación completa utilizada en *ODE*, que contiene cuatro categorías con su correspondiente código declarado en un atributo *id*, tal y como se detalla a continuación:

```
(11)  <encodingDesc>
      <classDecl>
      <taxonomy>
        <category id="inv">
          <catDesc>inventarios de bienes</catDesc>
        </category>
        <category id="cer">
          <catDesc>certificados médicos</catDesc>
        </category>
        <category id="dec">
          <catDesc>declaraciones de testigos</catDesc>
        </category>
        <category id="oth">
          <catDesc>otros</catDesc>
        </category>
      </taxonomy>
    </classDecl>
  </encodingDesc>
```

3.3.1.3. Algunos aspectos no bibliográficos <profileDesc>

El elemento <profileDesc> (*profile description*) proporciona una descripción detallada de los aspectos no bibliográficos del texto, como pueden ser los idiomas utilizados en el texto, las circunstancias de su producción, los contribuyentes o su estado de conservación. En él se han incluido las siguientes etiquetas:

- a. <settingDesc> incluye la información específica de la localización y datación a la que se hace referencia en el texto:

²⁶ Sobre los diferentes tipos textuales incluidos en *ODE*, véase § 2.1. *Antecedentes*.

- `<name>` contiene la localización geográfica de forma exhaustiva en el formato país, provincia, municipio. Esta etiqueta general contempla varios atributos:
 - *type* para señalar que el nombre hace referencia a un lugar de procedencia del manuscrito.
 - *subtype* para señalar la provincia.
 - *geo* para las coordenadas geográficas del lugar. TEITOK puede incluir automáticamente la latitud y longitud a partir de la información aportada en `<name>`.
- `<date>` incluye el año de creación del documento con los atributos *when*, donde se vuelve a indicar el año, y *when-custom* para el siglo en numeración romana.

```
(12) <settingDesc>
      <setting>
        <name type="place" subtype="Almería" geo="37.1709721 -
        2.8397107">España, Almería, Fiñana</name>
        <date when="1785" when-custom="XVIII">1785</date>
      </setting>
    </settingDesc>
```

- b. `<textClass>` agrupa la información que describe la naturaleza del texto en términos de una clasificación previa. Incluye a su vez la etiqueta `<catRef>`, donde se especifica mediante el atributo *target* la categoría concreta del documento mediante la remisión al código correspondiente definido en la taxonomía ya declarada anteriormente en el elemento `<classDecl>` (véase más arriba).

```
(13) <textClass>
      <catRef target="inv"/>
    </textClass>
```

- c. `<creation>` añade información sobre la creación del manuscrito original. Incluye un elemento `<name>` con el nombre del escribano cuando se conoce esta información. Si no es posible conocer este dato, en su lugar se indicará NA (no aplica).

```
(14) <creation>
      <name>NA</name>
    </creation>
```

3.3.1.4. Historial de revisiones del documento <revisionDesc>

El elemento <revisionDesc> (*revision description*) resume el historial de revisiones del archivo electrónico. En él se incluye el elemento <change>, donde se deja constancia de las modificaciones realizadas sobre el archivo con algunas especificaciones señaladas en los atributos *who*, para la tarea realizada, y *when*, para la fecha en la que tuvo lugar. Este elemento de la cabecera se genera automáticamente una vez que el texto se ha tokenizado en TEITOK (§ 4.2. *Tokenización*).

En el ejemplo siguiente el elemento <change> alude a la tarea de tokenización:

```
(15) <revisionDesc>  
      <change who="xmltokenize" when="2019-08-05">tokenized using  
      xmltokenize.pl</change>  
    </revisionDesc>
```

3.3.2. Texto <text>

El texto de cada documento en el corpus se incluye en la etiqueta <text> y consta mínimo de dos elementos obligatorios: <body> (cuerpo del texto) y <p> (párrafo). Cada documento en el corpus constituye un archivo XML con un único elemento <text> que incluye la transcripción estructurada tal y como se detallará en los subapartados siguientes. Un ejemplo completo de texto en el corpus puede verse en § *Anexo II. Ejemplo completo de texto etiquetado*. Asimismo, para facilitar la consulta de cualquier etiqueta utilizada en la edición, el conjunto completo de las etiquetas que se irán describiendo a continuación se recogen en § *Anexo III. Catálogo de etiquetas XML-TEI utilizadas en ODE*.

Para la edición de otra tipología de documentos, <text> podría componerse, además de <body>, de otros elementos que no conforman el cuerpo del texto como tal, sino que pertenecen a un prólogo, índice, sobrescritos, etc., o simplemente son anotaciones ajenas al texto. No obstante, en la codificación de inventarios de bienes, los elementos <text> y <body> coincidirán siempre en su contenido y ambos elementos abarcarán el conjunto de textos que conforman el corpus en su totalidad. De esta forma, el esquema del cuerpo del texto será siempre el que se muestra a continuación y que además coincide con la estructura mínima obligatoria que debe cumplir el elemento <text> en un documento TEI bien formado:

```
(16) <text>  
      <body>  
        <p>Texto del documento</p>  
      </body>  
    </text>
```

3.3.2.1. Estructura básica

El elemento `<body>` es de presencia obligada dentro de `<text>` y contiene la transcripción íntegra de un solo texto unitario en el que se excluye cualquier referencia o información ajena al propio cuerpo del texto.

Inicio de folio y enlace al facsímil

El inicio de un nuevo folio se indica con el elemento vacío `<pb/>` (*page beginning*) que debe contener a su vez los atributos *n* y *facs*. El atributo *n* proporciona el número de folio dentro del conjunto del documento, mientras que *facs* permite vincular el texto con la imagen correspondiente. El elemento `<pb/>` debe cumplir con las siguientes observaciones:

- a. Junto al número de folio se añaden las iniciales *r* o *v* correspondientes al recto o vuelto de la hoja: `<pb n= "8r"/>`
- b. El número de folio coincidirá con la numeración original en caso de que esta aparezca en los márgenes superiores del manuscrito. Si por el contrario la foliación se desconoce se puede indicar con la abreviatura *sf*: `<pb n= "sf"/>`. Si por el contrario esta se conjetura se indica entre corchetes: `<pb n= "[8r]"/>`
- c. En *ODE*, la vinculación de texto e imagen es relevante, pues supone la posibilidad de ofrecer una visualización del facsímil original del manuscrito. El nombre interno de la imagen deberá coincidir con el indicado en el atributo, al que se le añade también la extensión del archivo: `<pb n= "[8r]" facs= "IMG.0001.jpg"/>`

Inicio de párrafo

Los inicios de párrafo deberán marcarse con los elementos de apertura y cierre `<p>` y `</p>` respectivamente y abarcarán todo el conjunto de líneas correspondiente al párrafo hasta un nuevo punto y aparte. La división de párrafos se respeta conforme al manuscrito original salvo en una excepción: se ha establecido que las listas de bienes en las que se diferencian distintas tipologías encabezadas por un título al margen conformen también párrafos separados.

- (17) `<p>`En la villa de Pinto [...] para que en todo tiempo se sepa y aya cuenta y razon de los vienes que en esta dha villa tenia, paso a las casas donde tenia su abitazion y hizo el imbentario de ellos en la forma siguiente`</p>`
`<p>`Primeramente dos cantimploras de cobre grandes con sus corchas. Otra cantimplora de lo mismo mas pequeña con su cubo `</p>` (*ODE* – Madrid, 1720).

Inicio de columna

No es frecuente encontrar en los inventarios de bienes una disposición del texto por columnas. No obstante, en aquellos casos en los que se diferencian varias columnas, estas se marcan con el elemento `<cb/>` (*column beginning*). El texto en la columna se sigue marcando entre párrafos con la etiqueta `<p>` mencionada arriba. Un ejemplo de texto dispuesto en columnas es el documento AL1832I0031: «Inventario de los bienes que quedaron por fin y muerte de D. Tomás Gómez²⁷».

El número de columna se indica mediante el atributo *n*. El valor de este atributo no se indica con dígitos sino con letras, tal y como se indica a continuación:

- (18) `<cb n= "a"/>` `<p>`Ymbentario, cuenta, particion, adjudicacion a los bienes que fincaron por el fallecimiento de Dn Tomas Gomez [...] `</p>`
`<cb n= "b"/>` `<p>`En la villa de Fiñana, á diez y nueve dias del mes de octubre de mil ochocientos treinta y dos [...] `</p>` (ODE – Almería, Fiñana, 1832).

Cambios de línea

Los cambios de línea se especifican con el elemento vacío `<lb/>` (*line beginning*) tantas veces como líneas incluya el folio.

- (19) `<lb/>` Un cofre de bara y media de largo cubierto `<lb/>` de enzerado viejo.
`<lb/>` Otro cofre chico, maleta, cubierto de enzerado.
`<lb/>` Otro pa malecta de tres quartas de largo de enze`<lb/>`rado viejo.
`<lb/>` Otro cofre viejo sin serradura.
`<lb/>` Una arca de pino de bara y quarta de largo.
`<lb/>` Una mesa de pino cubierta de badana (ODE – Madrid, 1711).

Es posible indicar con el atributo *n* el número de la línea en la que el texto se halla: `<lb n= "1"/>`. No obstante, en el caso de ODE no se ha llevado a cabo esta opción para agilizar las transcripciones. Puesto que el conteo de líneas no ofrece posteriormente un aprovechamiento significativo, se ha decidido prescindir de su marcación.

A la hora de indicar un cambio de línea pueden darse las siguientes posibilidades:

- a. Que el cambio de línea afecte a palabras enteras. En este caso el elemento `<lb/>` se delimita con un espacio a izquierda y derecha:

- (20) una mantilla de encaje de `<lb/>` blanca negra

²⁷ Enlace directo a la visualización del documento en el corpus:
<http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/AL1832I0031.xml>

- b. Que el cambio de línea coincida en mitad de una palabra. En este caso el elemento `<lb/>` se usa sin espacios a uno y otro lado:

(21) cubiertos de plata nue`<lb/>`vos

- c. Que el cambio de línea y la rotura de palabra coincidan con el inicio de un nuevo folio. Se utilizan los elementos `<pb/>` y `<lb/>` sin espacios entre la palabra afectada:

(22) pendientes de dia`<pb n= "13v" facs= "DSC_0371.jpg"/><lb/>`mantes

3.3.2.2. Abreviaturas

El desarrollo de abreviaturas puede marcarse según el estándar TEI siguiendo distintas estrategias.

Una de ellas es mediante los elementos `<expan>` y `<ex>`. La etiqueta `<ex>` (*editorial expansion*) recoge los caracteres abreviados en la fuente original y que son añadidos por el editor, mientras que `<expan>` comprende tanto la abreviatura como las letras añadidas. Utilizando un ejemplo, la abreviatura de *señora* como *sra* quedaría de la siguiente forma:

(23) `<expan>s<ex>eño</ex>ra</expan>`

Otra posibilidad más compleja de marcar las abreviaturas se obtiene con la combinación de las etiquetas `<choice>`, `<expan>` y `<abbr>`. Esta última recoge la abreviatura reflejada en el manuscrito, `<expan>` encierra el desarrollo de la abreviatura, y `<choice>` engloba ambas etiquetas para indicar que todo ello es una expansión. La ventaja con respecto a la propuesta anterior es que esta última permite recuperar tanto la parte abreviada como la expandida²⁸. Tomando de nuevo el ejemplo anterior, el resultado sería el siguiente:

(24) `<choice>
 <abbr>sra</abbr>
 <expan>eño</expan>
</choice>`

Para el desarrollo de las abreviaturas, en cambio, en el corpus *ODE* se ha seguido una estrategia interna más sencilla que simplifica las posibilidades señaladas. Esta nueva opción se aleja del estándar TEI y es semejante a la seguida en el proyecto *Post Scriptum* con el objetivo de agilizar la fase de transcripción (Vaamonde, 2016, p. 64). Este procedimiento suprime el uso de etiquetas XML sustituyéndolas por un atajo en el

²⁸ Una ampliación al tratamiento de las abreviaturas en documentos históricos con variaciones a las propuestas que aquí se exponen puede verse en Isasi Martínez et al. (2020, p. 75).

que se emplea la doble pleca ||. La forma expandida se transcribe seguida de la forma original abreviada, separadas ambas por el símbolo || y sin espacios:

vno||vecino
ciud||ciudad

Este sistema permitirá una conversión automática ulterior a un formato XML una vez que los documentos son procesados en la plataforma TEITOK gracias a una resta matemática entre la palabra desarrollada y la abreviatura de esta, dando como resultado los caracteres expandidos añadidos por el editor.

Para el tratamiento de las abreviaturas en el corpus se han tenido en cuenta los siguientes aspectos:

- a. Solo se ha atendido al contenido alfabético de la abreviatura sin incluir puntos, barras o caracteres de cualquier otro tipo que pueda incluir el original. Esto se extiende a la común aparición de caracteres en superíndice, los cuales se han transcrito en el mismo nivel de línea, o a las abreviaturas seguidas de punto en el original, los cuales no se han tenido en cuenta en la transcripción:

v ^{no}	vno vecino
r ^s .	rs reales

- b. La ortografía original del manuscrito se debe mantener en la expansión de las abreviaturas, de manera que las grafías presentes en el texto deben verse incluidas también en el desarrollo de las abreviaturas.

essno	essno esscribano
setez ^{os}	setezos setezientos

- c. El desarrollo de las abreviaturas se hará atendiendo a las normas ortográficas actuales, independientemente de cuáles sean los hábitos de escritura del autor del manuscrito en el resto del texto. El uso sistemático en un documento, por ejemplo, de la letra *x* como representación del fonema /x/, o de *n* para el fonema /n/, no se verá reflejado en la expansión de las abreviaturas.

lmo	lmo legítimo
aos	aos años

- d. Las abreviaturas de origen griego, monogramas o la combinación de palabras gráficas con grafías ya castellanizadas como en los casos de *xptiano* (cristiano) o del nombre propio *xptobal* (Cristóbal) o XPO (Cristo) no se desarrollan, puesto que mezclan grafías de distintos alfabetos. El desarrollo de estas abreviaturas incumpliría el caso descrito en b., por lo que su

adaptación se lleva a cabo más adelante, en la tarea de normalización (§ 4.3. *Normalización ortográfica*).

- e. Tampoco se desarrollan los símbolos especiales utilizados como abreviaturas. El símbolo @ frecuentemente empleado en los inventarios como abreviatura de *arroba* se transcribe de manera literal y no se expande. De lo contrario, expandir esta abreviatura conllevaría incumplir el caso expuesto en b.

3.3.2.3. Características del texto

Se mencionan aquí cómo se han resuelto los casos de edición relativos a algunas características particulares del texto, tales como repeticiones del escribano (texto redundante), los casos de *sic*, las reproducciones literales, los fragmentos señalados o resaltados, las palabras escritas en otra lengua distinta a la española o las omisiones en el texto por olvido del escribano o deterioros en la fuente.

Texto redundante o superfluo

Se marcan con el elemento <surplus> aquellas grafías o palabras repetidas por olvidos del escribano o errores no cancelados. Se interpreta por tanto que son repeticiones no intencionadas. El elemento marcado será siempre el que aparece en segundo lugar.

- (25) Dos sabanas de lienzo de lino y estopa nuebas de siete baras cada una que, a cinco rs||reales vara importan <surplus>importan</surplus> 70.

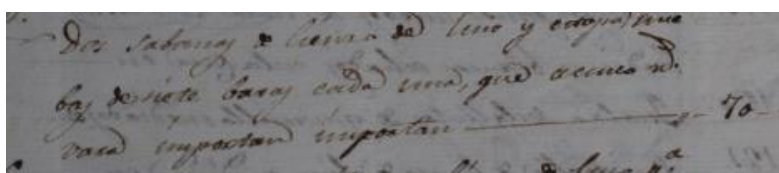


Figura 8: Texto superfluo por repetición

Sic

Los fragmentos de texto aparentemente erróneos o inexactos y que podrían interpretarse como errores de transcripción se marcan con el elemento <sic>. Esta etiqueta puede combinarse con el elemento <corr> (*correction*), que indica la corrección del elemento <sic> por parte del editor. El uso de ambas etiquetas se debe agrupar bajo el elemento <choice> para señalar las diferentes interpretaciones o alternativas de un mismo fragmento en el texto. Cabe destacar que el elemento <corr> proporciona una intervención del editor y que por tanto no se encuentra en el manuscrito original. Las

correcciones proporcionadas por el escribano y que sí forman parte de la fuente original se marcan con el conjunto de etiquetas <del>, <add> y <subst> descritas anteriormente.

No obstante, para la edición de este corpus se ha decidido no intervenir los elementos marcados con <sic>, pues una propuesta de enmienda editorial ya se proporciona en la posterior versión normalizada de los documentos (§ 4.3. *Normalización ortográfica*). A continuación, se muestran distintos ejemplos que representan una casuística de los segmentos que se han codificado con <sic> solo cuando estos pueden generar motivo de duda por distintas razones, entre las que se encuentran fallos de concordancia en el género o número, cambios por metátesis, disimilación o paragoge, fenómenos fonéticos reflejados en la escritura como rotacismo y seseo, o simplemente despistes del escribano.

Los fragmentos donde estos casos son recursivos, como lo pueden ser los fenómenos fonéticos, solo se han marcado en su primera aparición, entendiendo que son recurrentes en el escribano a lo largo de todo el texto.

- (26) dos <sic>grande</sic> y dos pequeñas

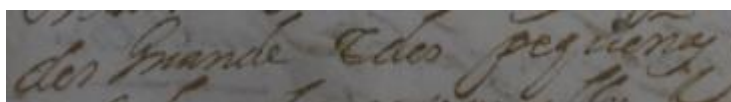


Figura 9: Elemento sic por falta de concordancia gramatical

- (27) con dos <sic>abriores</sic> de ojales

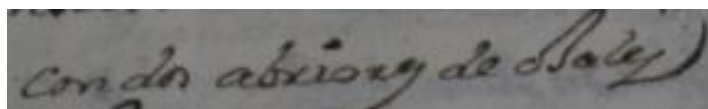


Figura 10: Elemento sic por pérdida de -d- intervocálica

- (28) ocho <sic>quados</sic> pintura basta

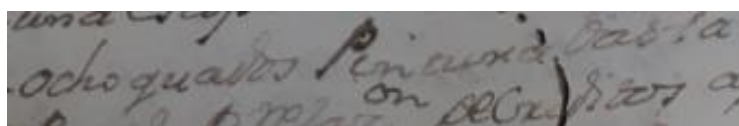


Figura 11: Elemento sic por omisión del escribano

Texto literal

Los fragmentos de texto literal que según las normas ortográficas actuales deben ir entre comillas se marcan con la etiqueta <q> (*quoted*). Se marcarán por tanto dentro de este elemento los términos técnicos o de jerga, referencias y citas externas o reproducciones literales.

- (29) candeleros antiguos marcados <q>A</q> con 11 marcos

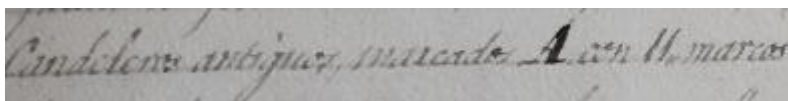


Figura 12: Fragmento de texto con reproducción literal

Texto resaltado

Los fragmentos de texto gráficamente diferentes al texto que los rodea se etiquetan con el elemento <hi> (*highlighted*), que podrá contener un atributo *rend* para indicar el tipo de resaltado: subrayado (con el valor *underlined*), negrita (con el valor *bold*), cursiva (con el valor *italic*), etc.

- (30) <hi rend= "underlined">Muebles</hi>

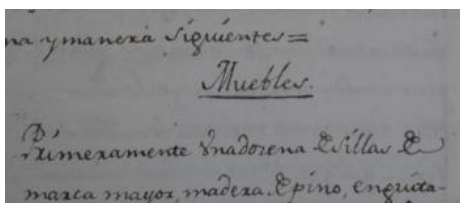


Figura 13: Texto subrayado

- (31) <hi rend= "bold">Ynventario</hi>

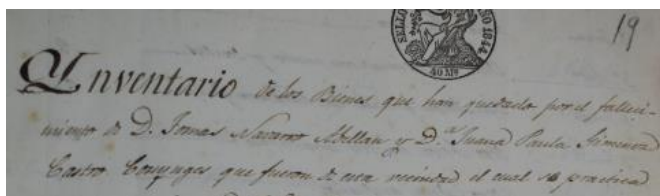


Figura 14: Texto marcado en negrita

Palabras o expresiones en lengua extranjera

Las palabras o frases en un idioma distinto al español se señalan con la etiqueta <foreign> que contendrá el atributo *lang* para indicar la lengua utilizada. En el corpus solo se ha registrado el uso del latín como lengua extranjera, por lo que el valor de esta etiqueta ha sido “*la*” en todos los casos tal y como corresponde para el latín.

En los inventarios de bienes, el uso del latín se reduce a expresiones fijas utilizadas de manera recurrente:

- (32) <foreign lang= "la"/>In dey nomine</foreign>, amen. Sepan cuantos esta carta de dote y arras bieren, como yo, Josef de Plaza Hernandez (*ODE* – Almería, Tabernas, 1822).

- (33) que el dia diez y nuebe de agosto contrajo matrimonio `<foreign lang= "la"/>`in facie ecclesia`</foreign>` con Franca Miguel Rodriguez (ODE – Madrid, San Lorenzo de El Escorial, 1789).
- (34) estando presentte y con asistencia de Antto Perez, curador `<foreign lang= "la"/>`ad litem`</foreign>` de los hijos menores q quedaron por fin y muerte del dho Alonso Lopez (ODE – Madrid, 1703).

Omisiones y deterioros del original

Las omisiones en la transcripción, ya sea por que el texto resulta ilegible por daños en el soporte, las realizadas por el propio escribano o por razones editoriales, se marcan con la etiqueta `<gap/>`. Esta etiqueta señala un punto en el documento donde el editor no es capaz de ofrecer una transcripción de lo que se ha escrito, por lo que se tratará siempre de un elemento vacío.

La razón por la que el texto se omite se indica con el atributo *reason*, cuyos valores serán alguno de los tres siguientes: con el valor *illegible* se señalan los vacíos en el texto causados por daños en el soporte original (Figura 15), por una grafía dificultosa, por un tachado del escribano imposible de leer (Figura 16) o cuando la encuadernación impide la lectura. Con el valor *omitted* se señalan los huecos presentes en el manuscrito original por despistes u olvidos del propio escribano (Figura 17). Por último, con el valor *editorial* se señala un fragmento de texto que se suprime en la transcripción por ser considerado irrelevante lingüísticamente por el editor (Figura 18). En los inventarios, se omiten con frecuencia algunos fragmentos previos al listado de bienes donde se incluyen expresiones formularias, extensas enumeraciones de personas que participan en una partición de bienes o los listados de objetos repetitivos considerados sin trascendencia o interés léxico.

El elemento `<gap/>` no señala segmentos inferiores a la palabra, por lo que la etiqueta se aplicará a palabras enteras independientemente de si el segmento dañado afecta solo a algunas grafías.

- (35) Yten una delantera de cama, de lienço, en diez `<lb/>` y seis rs. `<lb/>` Yten bara y media de manteles en dies rs. `<lb/>` Yten, de lana del colchon algo traida, en diez `<gap reason= "illegible"/>`



Figura 15: Texto omitido en la transcripción por deterioro del manuscrito

- (36) vn cofre encorado con su llave y zeradura; y en el se allo <lb/> <gap reason="illegible"> lo sigiente: dos esteras de junco mediadas; dos son<lb/>breros negros forados; y no se allaron mas vienes y se [...]

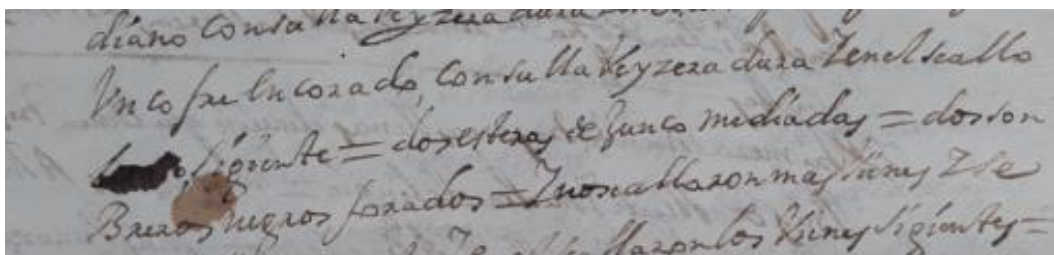


Figura 16: Texto tachado por el escribano imposible de leer

- (37) Seis mantas de cama chica á veinte rs cada una. <lb/> Una cortina de lana bieja.
<lb/> Tres cortinas de <gap reason="omitted"> para balcon á diez rs cada una.
<lb/> Una colcha de cama grande de percal frances encarnado [...]

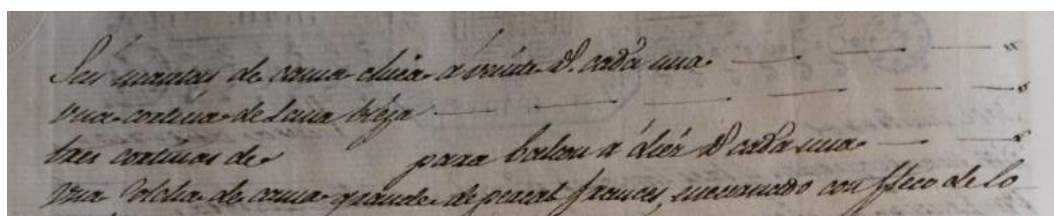


Figura 17: Texto omitido por el escribano

- (38) <lb/> <gap reason="editorial"/>
<lb/> <hi rend="underlined">Pago</hi>
<lb/> Primte se le adjudica en el valor de las <lb/> casas mortuorias setecientos rs. <lb/> Ytt un marjal de viña en la Acifria, lin<lb/>da levante [...]

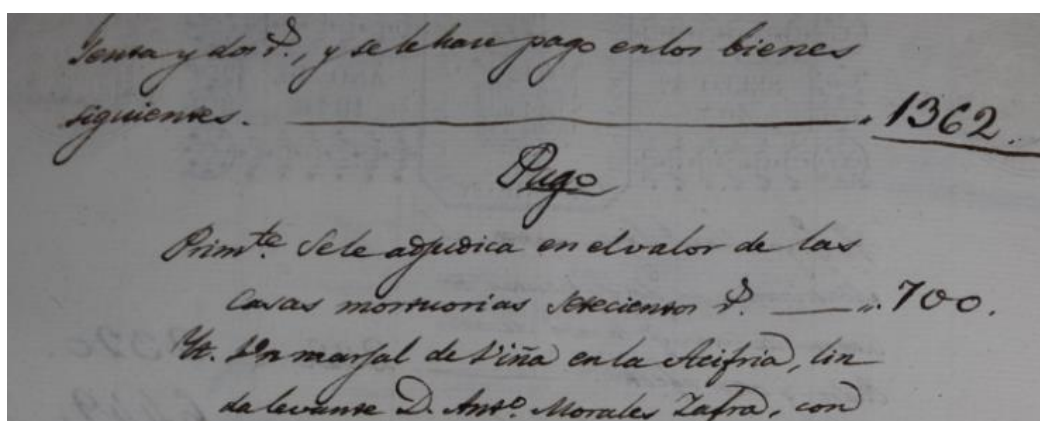


Figura 18: Texto omitido por razones editoriales

3.3.2.4. Intervenciones del escribano

Adiciones

Las adiciones en el original por parte del escribano se indican con la etiqueta <add> (*addition*). Al igual que con las cancelaciones, este elemento podrá aplicarse a letras, palabras o frases que hayan sido añadidas por el escribano o por una segunda mano, ya sea fuera del nivel de línea o en los márgenes de la caja de escritura. Cada elemento <add> deberá contener un atributo *place* donde se indica el lugar específico de la adición en el documento.

Las adiciones en los márgenes izquierdo, derecho, superior o inferior se indican con los valores *left*, *right*, *top* y *bottom* respectivamente (Figura 19), mientras que las adiciones por encima de la línea de escritura o interlineados se marcan con el valor *above* (Figura 20), o *below* si en cambio se hacen por debajo de la línea (Figura 21).

(39) <add place= "left">Se dio de limosna</add>

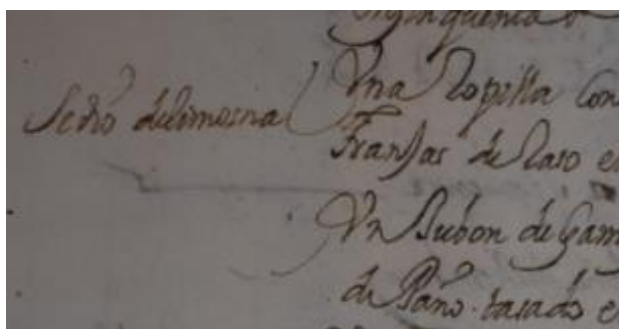


Figura 19: Adición en el margen izquierdo

(40) Un arado<add place= "above">con su reja</add> en veynte y quatro rs

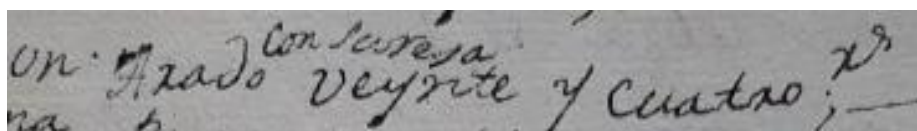


Figura 20: Adición por encima de la línea

(41) Vn corchon <add place= "below">poblado de lana</add> color negro y pajizo

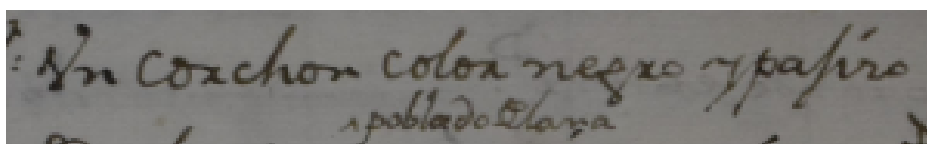


Figura 21: Adición por debajo de la línea

Cancelaciones

Cuando hay texto en el manuscrito cancelado por el propio escribano mediante tachado, raspado o cualquier otro modo se aplica el elemento `<del>` (*deletion*) siempre y cuando sea posible leer el fragmento de texto eliminado. Esta etiqueta puede aplicarse a grafías, palabras enteras o a partes de palabra. La etiqueta llevará un atributo *rend* en el que se indica que el texto ha sido cancelado con el valor *cancelled*:

- (42) Otro cofre `<del rend= "cancelled">` mui`</del>` tunbado de baquetta de dos barras de largo `<lb/>` claueteado con sus pies.

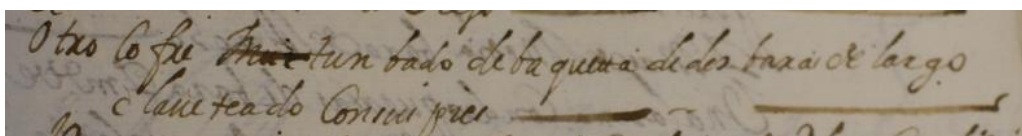


Figura 22: Texto tachado por el escribano

- (43) dhos nuestros padres Diego Rodri`<lb/>`gues y Agustina de Vargas `<del rend= "cancelled">` los dos`</del>` `<lb/>` el primero, en estado de matrimonio, la segunda en el de don`<lb/>`cella

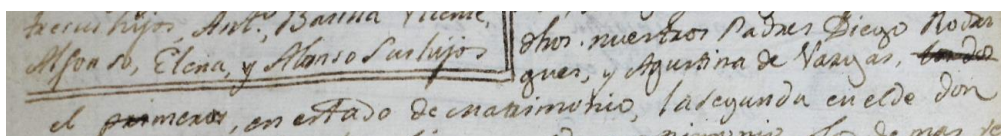


Figura 23: Texto tachado

Restituciones

En muchos casos, las cancelaciones implican una adición posterior como resultado de alguna enmienda por parte del escribano. De este modo, ambas acciones están estrechamente relacionadas y, combinadas, constituyen un único elemento: la corrección de un error. Las modificaciones realizadas en el manuscrito que implican un tachado, sobrescrito o una adición con la intención de subsanar algún fallo en la escritura se agrupan bajo el elemento `<subst>` (*substitution*). Esta etiqueta combina las ya mencionadas anteriormente para las cancelaciones y adiciones, por lo que contendrá a su vez `<del>` (cuyo valor del atributo será *overwritten*) y `<add>` (con valor *inLine*). En caso de que no fuese posible descifrar el elemento cancelado no se utilizará la etiqueta `<del>` sino `<gap/>` (Figura 26). Se muestran a continuación algunos ejemplos de sustituciones en el original:

- (44) ba`<subst>`
`<del rend= "overwritten">`ll`</del>``<add place= "inLine">`y`</add>`
`</subst>`eta

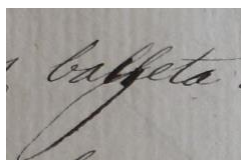


Figura 24: Restitución con sobrescrito

- (45) `<subst>`
`<del rend= "cancelled">arcas</del><add place= "above">onzas</add>`
`</subst>`

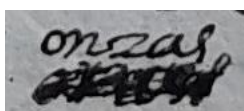


Figura 25: Restitución con tachado y adición por encima de la línea

- (46) uno que a dho precio ynportan
`<subst>`
`<gap reason= "illegible"/><add place= "inLine">beynte y quatro rs</add>`
`</subst>`

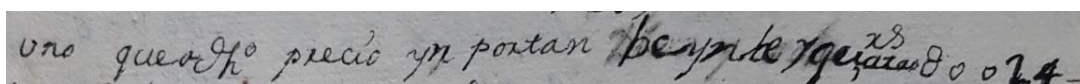


Figura 26: Restitución con raspado y sobrescrito

Cambios de mano

Las secuencias de texto escritas por una nueva mano se etiquetan con el elemento vacío `<handShift/>` al inicio del fragmento producido por la nueva mano. Se entenderá que todo lo que se transcribe a continuación corresponde a otro autor. En consecuencia, si hay otro cambio de mano deberá indicarse de nuevo con el elemento `<handShift/>`. Este podrá llevar un atributo *new* para indicar el número de manos que intervienen en el texto. Véase con el siguiente ejemplo:

- (47) `<handShift new= "2"/> <add place= "left"/>Sala del valcon</add> Primeramte`
`una mesa redonda a mas de me<lb/>dio servir en quarentta y zinco rs charolada.`
`<lb/> Ytt seis sillas pequeñas de estrado con asientos.`

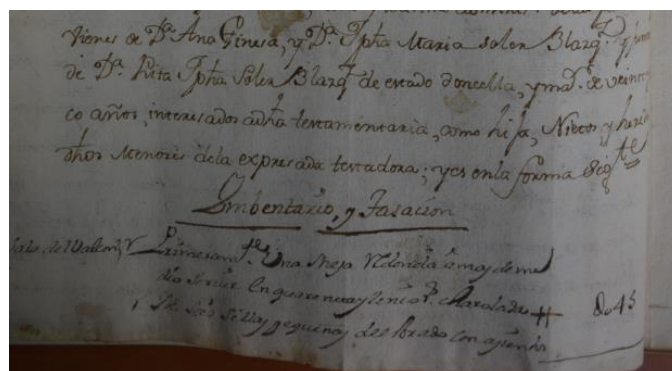


Figura 27: Cambio de mano

Firmas y rúbricas

Las firmas se marcan con la etiqueta <seg> (*arbitrary segment*), que en TEI indica cualquier segmentación por debajo del nivel de un fragmento de texto. La etiqueta llevará un atributo *type* con el valor *signature*. Al carecer de interés lingüístico, no se marcan las rúbricas, garabatos, figuras o símbolos sin interpretación lingüística que puedan aparecer acompañando a las firmas o en cualquier otra parte del manuscrito.

(48) <seg type="signature"/>Martinez</seg>

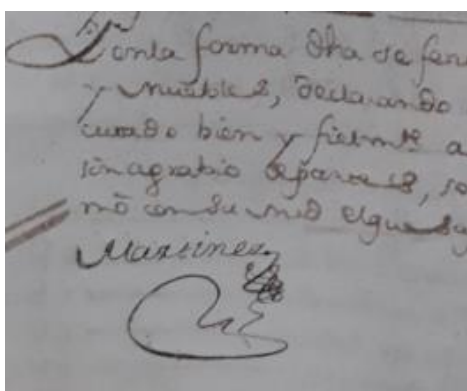


Figura 28: Firma y rúbrica

3.3.2.5. Conjeturas del editor

Los casos de conjeturas se señalan con el elemento <supplied> para indicar que algún fragmento del texto no puede transcribirse con total seguridad. A esta etiqueta se le añade un atributo *cert* donde se indica el grado de certeza con el que se transcribe el elemento dudoso. Con los valores *high* y *low* se señala en este caso el alto o bajo grado de verosimilitud con el que se transcribe el elemento conjeturado. Las restituciones editoriales se harán siguiendo las normas ortográficas actuales, independientemente de los hábitos escriturales del escribano. Por otro lado, nótese que cuando no es posible suministrar algún contenido, la etiqueta será <gap/>.

Las razones por las que el texto se conjetura pueden ser variadas. La gran mayoría de las ocasiones no es posible transcribir el segmento con seguridad debido a que la fuente resulta dañada, manchada o deteriorada por agentes externos que dificultan la lectura (Figura 29). En otras ocasiones, es la encuadernación la que impide leer los márgenes del manuscrito, pero se pueden ver algunas grafías que permiten restituirlo (Figura 30).

Cuando la fuente no es comprensible debido a una caligrafía dificultosa del escribano se señala la propuesta más viable (Figura 31). En otros casos, a pesar de que la fuente es legible, la caligrafía utilizada no permite dotar de sentido al segmento, lo que impide hacer de él una lectura comprensible (Figura 32).

- (49) cama de madera de pino a <supplied cert= "high">medio servir</supplied>²⁹

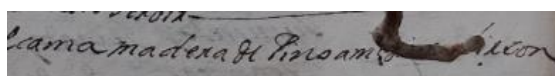


Figura 29: Conjetura segura por daños físicos del soporte

- (50) <supplied cert= "high">Tienda</supplied> de pretinero

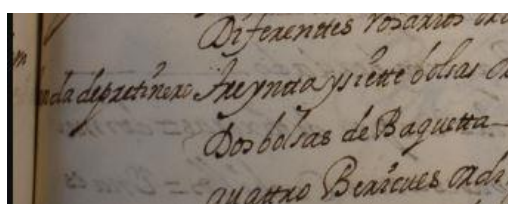


Figura 30: Conjetura segura por la encuadernación del documento

- (51) una cortina mui bieja y sin <supplied cert= "low">uso</supplied>

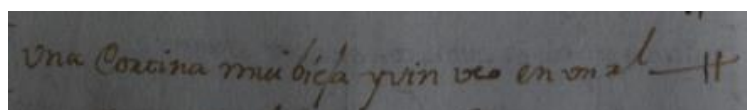


Figura 31: Conjetura no segura por grafía dudosa

- (52) a medio serbir, cabida de dos <supplied cert= "low">cartaros</supplied>

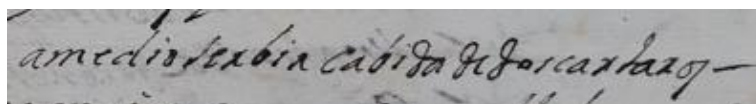


Figura 32: Conjetura no segura de palabra incomprensible

²⁹ En este inventario se alude recursivamente al grado de uso de los objetos con la expresión “a medio servir”, por lo que es el propio contexto el que facilita la conjetura en este caso (ODE – Almería, Cuevas de Almanzora, 1704).

Conviene destacar que la etiqueta <supplied> solo se aplica a palabras o a conjuntos de palabras y nunca a partes o grafías individuales. En la Figura 32 el contexto hace pensar que *cántaros* sería la interpretación más factible. A pesar de ser una única grafía la que plantea dudas, la etiqueta se aplica a la palabra entera por cuestiones relacionadas con la recuperación de información: si solo se marcaran los caracteres comprometidos únicamente sería posible la recuperación de estos caracteres sueltos y no de la palabra entera que ha planteado dudas.

4. CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK

La transcripción de los textos en XML-TEI tiene como objetivo conseguir un *corpus lingüístico* según las nociones con las que actualmente se entiende este término: un conjunto de textos en formato electrónico, codificado y preparado para la recuperación selectiva de información (Rojo, 2021, p. 2). Tras la codificación de los textos comienza una nueva fase en la edición del corpus encaminada a la publicación de los textos y a su procesamiento lingüístico considerada meramente computacional (Calderón Campos, 2019a, p. 29). La herramienta que se ha utilizado para ello es TEITOK.

En este capítulo se analiza este sistema y su viabilidad para la creación de un corpus histórico especializado, lematizado y anotado morfosintácticamente de acuerdo con los últimos avances de la lingüística computacional (§ 4.1. *Empleo de la herramienta TEITOK*). El procesamiento informático del corpus ha constado de las fases automáticas de tokenización (§ 4.2. *Tokenización*), normalización (§ 4.3. *Normalización ortográfica*) y anotación lingüística, esta última realizada con el programa informático NeoTag (§ 4.4 *Funcionamiento del anotador lingüístico en TEITOK*), que se ha ocupado de la asignación de etiquetas gramaticales (§ 4.5. *Anotación morfosintáctica (POS) en el corpus ODE*) y lemas (§ 4.6. *Lematización*).

Los procedimientos de tratamiento lingüístico a los que se ha sometido el corpus, detallados extensamente en estos apartados, han permitido ofrecer distintas versiones de un mismo texto (§ 4.7. *Triple visualización de los textos*). Tras todo este proceso, el objetivo es poder recuperar la información contenida en el corpus de una manera rápida y precisa, especialmente la que afecta a la marcación de los diminutivos. Para ello, TEITOK proporciona un motor de búsquedas especializado con el que poder explotar al máximo todas las posibilidades que ofrecen los documentos. Su funcionamiento requiere de una sintaxis propia basada en el lenguaje CQL (*Corpus Query Language*) que se detalla en § 4.8. *Búsquedas en el corpus basadas en lenguaje CQL*. De manera más específica, ODE dispone de una interfaz adaptada a la información que contiene el corpus y que integra el lenguaje CQL en los distintos tipos de búsquedas que ofrece (§ 4.9. *Interfaz de consulta del corpus ODE*).

Por último, con el propósito de crear un corpus que garantice la fiabilidad de los datos para los estudios diacrónicos, todos los textos se han sometido a una revisión manual (§ 4.10. *Revisión manual de errores en el corpus*). Puesto que el método de revisión ha supuesto en algunos casos depurar errores a partir del sistema de búsquedas, el procedimiento que se ha seguido se describe tras aludir al funcionamiento del buscador.

4.1. Empleo de la herramienta TEITOK

La marcación de textos en lenguaje XML-TEI se presta a múltiples salidas para la explotación de los datos, pero en sí misma supone trabajar con archivos difíciles de leer y complejos cuanto más etiquetas se añaden. Este lenguaje ha sido diseñado para que

sea entendido y procesado por un ordenador, pero no para que sea leído por un humano. Así, la edición realizada hasta ahora no es válida para una publicación directa (ni en papel ni en línea) de las transcripciones de los textos que integran el corpus, pues esta carece de formato y resulta ilegible y nada conveniente para la investigación:

```
<lb/> Yt treynta y seys cabras, ha veynte y cinco rs ca<lb/>da una, <sic>iportan</s
nuebecientos rs. <lb/> Yt hotra cabra en diez y hocho rs. <lb/> Yt un macho en trey
rs. <lb/> Yt unas catorce hobejas, a treynta rs, inportam cu<lb/>atrocientos y veyn
rs. <lb/> Yt dos cerdas en nobenta rs. <lb/> Yt una mula en mil rs. <lb/> Yt una mu
en seyscientos y setenta y cinco rs. <lb/> Yt una bura en trecientos rs. <lb/> Yt d
<sic>nube</sic> fanegas y media de centeno, a <lb/> veynte y seys rs la fanega,
ynportan quyni<lb/>entos y siete rs. <lb/> Yt veynte fanegas de panizo, ha beynte y
cua<lb/>tro rs, <sic>yportan</sic> cuatrocientos y hochenta rs. <lb/> Yt tres faneg
media de cebada ha ve<subst><del rend="overwritten">n</del><add>y</add>nte</subst>
<lb/> y cuatro rs. <lb/> Yt de tres cuartillas de bas, treynta rs. <lb/> Yt de una
fanega y tres cuartillas de <add place="above">a</add>billas, <lb/> ciento y cinco
<lb/> Yt de un celemin de garbanzos, 5. <lb/> Yt de una caldera en <subst><gap
reason="illegible" id="e-41"/><add>cinc</add></subst>uenta y cuatro. <lb/> Yt
una arca, cincuenta y cinco rs. <pb n="sf" facs="20180601_120600.jpg" id="e-43"/>
<lb/> Yt una tarima en diez y seys rs. <lb/> Yt una romana en beynte y cinco rs. <l
Yt una manta de lana en seys rs. <lb/> Yt de dos guardapies, cien rs. <lb/> Yt de d
sabanas, setenta y seys rs y medio. <lb/> Yt una basquiña en cincuenta rs. <lb/> Yt
```

Figura 33: Transcripción en XML-TEI de un inventario del corpus

Las transcripciones en XML-TEI son el primer paso necesario para una publicación digital, pero no el fin último. Trabajar en este código requiere disponer posteriormente de un método informático capaz de procesar la información con los fines deseados o, dicho de otro modo, disponer de una herramienta que permita “traducir” los textos de entrada a un formato más accesible. En nuestro caso, se pretendía la publicación en línea de los inventarios acompañada de un mayor aprovechamiento lingüístico de los textos que permitiera recuperar toda esa información. El objetivo era ofrecer los textos en un formato digital que trascendiera la simple visualización. Ambas cuestiones, publicación en línea y procesamiento automático de los textos para su conversión en un corpus buscable, se resuelven gracias al empleo de la herramienta TEITOK.

TEITOK es un sistema informático creado por Maarten Janssen (2016) con el que es posible visualizar, anotar, editar y explotar corpus, todo ello en línea y en una única plataforma o entorno web. En primera instancia, TEITOK ofrece una solución a la visualización de los datos en XML-TEI, pero permite además otras posibilidades encaminadas a la anotación lingüística de los textos y su posterior explotación³⁰. Su funcionamiento se sustenta en el proceso de tokenización (§ 4.2. *Tokenización*), esto es, la creación de tokens, para poder añadir a un texto base distintos niveles de edición sin que sea necesario crear versiones separadas de un mismo texto.

TEITOK permite anotar y añadir al texto información de muy diverso tipo de acuerdo con los intereses de cada corpus (morfológica, sintáctica, léxica, metatextual...), para lo que utiliza una estructura modular en la que se pueden

³⁰ Información más específica sobre el funcionamiento de TEITOK y todas sus posibilidades puede verse en la web <http://www.teitok.org/> y en los trabajos de Janssen (2016, 2018), Janssen, Ausensi y Fontana (2017), Calderón Campos (2019a), Calderón Campos y Vaamonde (2020), Arrabal Rodríguez (2022).

desarrollar distintas funciones. El sistema incorpora un normalizador, con el que ofrecer una versión modernizada del texto; y un anotador, con el que es posible añadir también información lingüística a los textos (etiqueta morfosintáctica POS y lema). Recientemente se está trabajando en *ODE* en la marcación de fenómenos fonéticos y en la anotación semántica del corpus, aunque sobre estos procesos no se hará mención en esta tesis.

El sistema incorpora a su vez un avanzado motor de búsquedas basado en lenguaje CQL (§ 4.8. *Búsquedas en el corpus basadas en lenguaje CQL*) que permite conjugar toda la información –textual y extratextual– contenida en el corpus para aprovechar al máximo todas sus posibilidades de explotación. De manera que la misma herramienta para crear, editar y alojar el corpus, sirve a su vez para recuperar información.

Otra de las características que impulsan el éxito de TEITOK radica en que su interfaz permite a los investigadores editar y gestionar su propio corpus de una manera fácil e intuitiva, de modo que no son necesarios conocimientos informáticos específicos para su uso. Numerosos proyectos lo han empleado para la creación de corpus de muy variadas características y tipologías. Entre los de tipo histórico, se ha mencionado en reiteradas ocasiones el caso de *Post Scriptum*, cuyo modelo se ha tomado como referencia para la elaboración de *ODE*, pero el número de proyectos que utilizan TEITOK no cesa de aumentar y actualmente está en vías de ser implantado para la elaboración de otros corpus históricos del español como el *CORDICan* (Viña Brito y Corbella, 2018), el *COSUIZA* (Castillo Lluch y Díez del Corral Areta, 2018) o el *CODHECUN* (Bravo-García, Mancera Rueda y Martín Aizpuru, 2021).

Así pues, TEITOK ha sido la herramienta con la que se han llevado a cabo todas las tareas necesarias para el procesamiento informático de *ODE* (tokenización, normalización y anotación lingüística). Todas ellas se ejecutan automáticamente, pero es posible realizar cambios y modificaciones posteriores de forma sencilla.

En el caso concreto del corpus de inventarios de esta tesis, el procesamiento lingüístico de los textos se ha realizado de manera completa e individualizada para cada uno de ellos y no en distintas fases para todos los textos en bloque. Esto quiere decir que todo el proceso que comprende las tareas automáticas de tokenización, normalización ortográfica, anotación lingüística (POS y lema) y revisión manual posterior se ha realizado de principio a fin documento a documento. Solo una vez completadas todas las fases para un mismo texto se ha iniciado con el procesamiento del siguiente, y así con los 97 inventarios que integran el corpus documental. Esta práctica, que sin duda resulta lenta si lo que se quiere es incorporar de manera inmediata documentación al corpus para aumentar su volumen, se ha considerado la más apropiada por dos razones principales.

En primer lugar, ha permitido disponer de un corpus con textos enteramente normalizados y anotados en todo momento, desde la inclusión del primero de ellos hasta el último. Así, se han podido realizar búsquedas complejas a cualquier nivel de edición con el objetivo de obtener resultados preliminares que permitieran establecer patrones para la confirmación o no de las hipótesis planteadas en esta tesis doctoral, aun cuando el corpus iba creciendo gradualmente. En segundo lugar, la revisión manual y paulatina

de los textos que se iban agregando, con la corrección de errores derivados de los procesos automáticos de normalización y anotación y la posterior actualización del anotador, ha supuesto evitar errores que hubieran sido recurrentes en el procesamiento automático de cada nuevo texto. Por ello, la tarea de revisión ha desempeñado un papel trascendental para la elaboración de este corpus (§ 4.10. *Revisión manual de errores en el corpus*). Se ha intentado de esta manera “entrenar” al anotador con el vocabulario presente en los inventarios y mejorar progresivamente el porcentaje de acierto. Esta cuestión ha sido indispensable para agilizar la tarea de revisión en una fase en la que el corpus tenía ya un número considerable de textos.

En los apartados siguientes se exponen de manera pormenorizada las distintas fases de las que ha constado el procesamiento del corpus una vez que los archivos en XML-TEI han sido importados a TEITOK. Continúa el capítulo con la descripción del formato de visualización de los textos, el funcionamiento de las búsquedas y la tarea de revisión de los procesos automáticos.

4.2. Tokenización

La tokenización es el proceso por el cual se crea un token por cada palabra o forma ortográfica en el corpus. Se trata, en términos informáticos, de la unidad mínima de que consta un corpus. Siguiendo la sintaxis XML, se añade a cada palabra el elemento `<tok> </tok>`, de manera que en una frase en el texto como “vn arconzillo pequeño viexo” se reconocen los siguientes tokens:

(1) `<tok>vn</tok> <tok>arconzillo</tok> <tok>pequeño</tok> <tok>viexo</tok>`

La tokenización en TEITOK es automática gracias a un programa informático que accede al archivo XML e inserta cada palabra dentro del elemento `<tok>`. Cada forma ortográfica es considerada un token independiente al que se le asigna también un número único de identificación dentro del texto. Tras este proceso, el archivo de entrada queda modificado con esta nueva información, que permitirá ir incorporando posteriormente nuevas características para cada una de las palabras gracias a la creación de distintos atributos. Por ejemplo, la palabra *arconzillo* de la frase anterior es la forma original que aparece en el manuscrito (atributo *@form*), su correspondiente forma normalizada es *azadoncillo* (atributo *@nform*), morfológicamente se trata de un sustantivo masculino singular con presencia de un sufijo diminutivo (atributo *@pos*) y se engloba bajo el lema *arcón* (atributo *@lemma*) (ejemplo 2). Si además la palabra se viera afectada por un deterioro del original y esta fuera una conjetura del editor con un alto grado de fiabilidad, la etiqueta que en la transcripción se ha utilizado para indicar esta característica envolverá al token en su totalidad (ejemplo 3). Si la palabra fuera una abreviatura, como por ejemplo lo es *ziud* de *ciudad*, el token podría incluir también información relativa a su expansión (atributo *@fform*) (ejemplo 4). Así, los tokens de

las palabras que acabamos de mencionar recogerían como sigue toda esta información una vez que se ha completado la tokenización automática del texto³¹:

- (2) <tok id="w-171" form="arconzillo" nform="arconcillo" pos="NCMS00D" lemma="arcón">arconzillo</tok>
- (3) <supplied cert="high"><tok id="w-171" form="arconzillo" nform="arconcillo" pos="NCMS00D" lemma="arcón">arconzillo</tok></supplied>
- (4) <tok id="w-78" form="ziud" fform="ziudad" nform="ciudad" pos="NCFS000" lemma="ciudad">ziud</tok>

El token, por tanto, es algo así como una especie de cajón en el que se recogen distintos aspectos en torno a una misma palabra. En el caso concreto de *ODE*, la información que se añade abarca los distintos niveles que se han señalado, pero de acuerdo con los intereses de cada proyecto, podrían incluirse tantos atributos como distintos tipos de información se deseen añadir al corpus³².

La tokenización supone uno de los primeros pasos para el tratamiento automático de los textos, ya que implica preparar los datos para una nueva anotación más allá de la realizada a nivel ortográfico durante la fase de transcripción. Asimismo, gracias a ella será posible recuperar los datos a partir de búsquedas posteriores (Kilgarriff y Kosem, 2012). TEITOK se sustenta en la división por tokens para conseguir aunar en un único archivo diferentes ediciones de un mismo texto, de modo que un token contiene toda la información disponible en torno a la palabra en cuestión. Por ello, es un procedimiento generalizado en la gran mayoría de los corpus digitales actuales³³ que, sin embargo, difiere considerablemente de uno a otro en su realización. Conocer qué se considera como *token* en un corpus es esencial para interpretar los resultados obtenidos:

differences in the way tools are designed will have an impact on almost all corpus analyses. For example, differences in the definition of a word will impact not only on word frequency counts, but also on the values of type-token ratios, the strengths of collocation between words, and the N-grams and keywords produced for a corpus (Anthony, 2015, p. 151).

En TEITOK el concepto de token incluye palabras, dígitos y signos ortográficos de puntuación, por lo que un recuento de tokens en el corpus incluye estos tres aspectos.

³¹ Por razones explicativas, se mantiene en estos ejemplos y en los siguientes la *@form*, pero este atributo no es necesario a efectos de procesamiento del corpus, ya que queda sobrentendido al coincidir con la forma de entrada en la transcripción (véase más abajo Figura 34). La *@form* sí es obligatoria en aquellos casos como los de (3), en los que la palabra forma parte de una etiqueta XML.

³² Recientemente se está trabajando en *ODE* en la adición de dos nuevas capas de anotación lingüística (*@ltags* y *@tags*) para marcar fenómenos fonéticos y campos onomasiológicos respectivamente.

³³ En la amplia mayoría de los corpus actuales más relevantes del español, la posibilidad de realizar búsquedas simples y combinadas radica en el proceso previo de tokenización (De Benito Moreno, 2019).

Para la identificación automática de tokens el programa se basa en el reconocimiento de caracteres alfabéticos, números y signos de puntuación. La frontera de palabra queda delimitada por los espacios en blanco, los cuales sirven como distintivo para tokenizar, así como en el caso de los números de una o más cifras, también delimitados por espacios. Una excepción es el tratamiento de la puntuación, inmediata a las palabras.

La tokenización automática en TEITOK rara vez suele plantear problemas. Posibles fallos en la identificación de tokens se deben en la mayoría de las ocasiones a errores humanos cometidos durante el proceso de transcripción o a archivos no válidos que incumplen la sintaxis XML-TEI y deben ser revisados. Una vez realizada la fase de tokenización, el corpus lingüístico lo conforma todo lo que se incluye en las etiquetas <tok>. Este proceso permite aunar la información textual, previamente señalada en la fase de edición digital, con la información lingüística, añadida en las fases que siguen de procesamiento automático (Janssen, 2016, p. 4038).

La tokenización del texto es esencial para la distinción entre los diferentes niveles de edición de los que constará el corpus (edición semipaleográfica y edición normalizada, por un lado, y anotación morfosintáctica y lema, por otro), todos asociados a una palabra y gestionados desde un único archivo XML. Cada uno de estos niveles lo conforman los distintos atributos creados dentro de <tok> que se han especificado anteriormente. Sin embargo, no todos los tokens en el corpus constarán de todos estos atributos o niveles de edición. Este puede ser el caso de palabras que no poseen *@fform* porque no se trata de abreviaturas o bien de palabras que no poseen *@nform* porque su normalización coincide con la forma original del manuscrito. En definitiva, algunas palabras podrán tener todos los atributos posibles, mientras que otras solo los correspondientes a la anotación lingüística (POS y lema), ya que la forma original es la misma en todos los niveles de edición. Véase la codificación de las palabras *brasero* y *bracero* (variante ceceante):

(5) <tok id="w-482" form="brasero" pos="NCMS000"
lemma="brasero">brasero</tok>

(6) <tok id="w-695" form="bracero" nform="brasero" pos="NCMS000"
lemma="brasero">bracero</tok>

Sin embargo, para que una palabra como *brasero* (ejemplo 5), pueda ser recuperada también como *bracero* (ejemplo 6) a partir de una búsqueda modernizada, TEITOK supera estas diferencias mediante un sistema de herencias en el que partiendo de la forma que aparece en la transcripción, el resto de los niveles asumirá esta forma si no se encuentra antes ninguna otra:

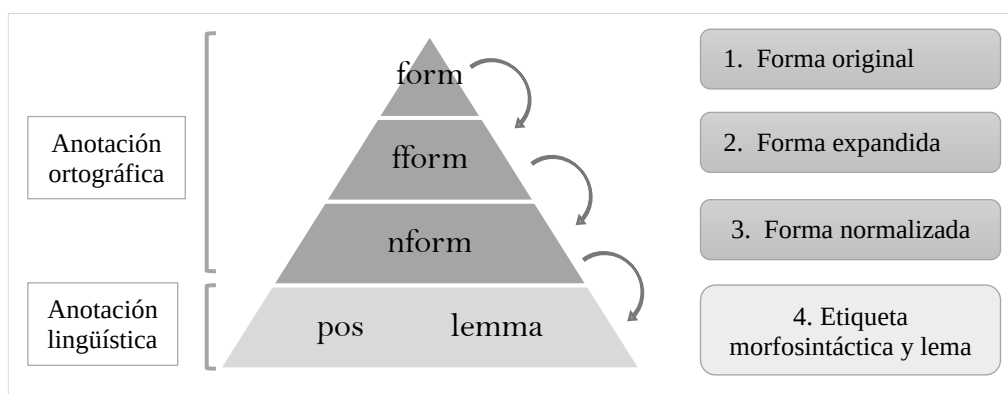


Figura 34: Funcionamiento de TEITOK mediante el sistema de herencias

La forma normalizada es la misma que la forma original si no se especifica ninguna otra. Por otra parte, la anotación lingüística utiliza siempre la palabra normalizada para la adjudicación de etiquetas morfosintácticas y lemas, por lo que en caso de que no haya forma normalizada, el sistema asumirá que esta es la misma que la forma original. Este método permite que no sea necesario replicar información que, además de ser una tarea improductiva, podría favorecer la presencia de errores que habría que corregir manualmente en cada uno de los niveles (Janssen, 2016, p. 4038).

Este sistema jerárquico es esencial para entender el funcionamiento de TEITOK y la superposición de capas o niveles de edición que se añaden al corpus y que permiten disponer de diferentes versiones de cada texto aglutinadas en un único archivo fuente gracias a la tokenización. Cada una de estas capas (normalización, anotado morfosintáctico y lema) son el resultado de las tareas de procesamiento informático a las que se ha sometido el corpus y que se desarrollan en los apartados siguientes. Tras cada una de ellas el archivo XML de partida se modifica automáticamente con la inclusión de nueva información que ha sido revisada manualmente (§ 4.10. *Revisión manual de errores en el corpus*).

Como es lógico pensar, todo este procedimiento supone añadir al archivo inicial cada vez más atributos a las palabras, de manera que se obtiene un archivo con elementos cada vez más largos que hacen imposible la lectura de los textos, la corrección de errores directamente sobre ellos o realizar cualquier búsqueda. A la versión inicial de la transcripción mostrada páginas atrás en la Figura 33 puede compararse esta otra versión del mismo documento una vez realizadas todas las fases del procesamiento lingüístico del texto:

```

<lb id="e-23"/> <tok form="Yt" fform="Ytem" id="w-156" nform="item" lemma=
>unas</tok> <tok id="w-158" lemma="14" pos="Z">catorce</tok> <tok id="w-1
emma=", " pos="Fc">,</tok> <tok id="w-161" lemma="a" pos="SPS00">a</tok> <
orm="reales" id="w-163" lemma="real" pos="NCMP000" stags="I1">rs</tok><to
tan" lemma="importar" pos="VMIP3P0">inportam</tok> <tok form="cuatrocient
trocientos</tok> <tok id="w-167" lemma="y" pos="CC">y</tok> <tok id="w-16
s" id="w-169" lemma="real" pos="NCMP000" stags="I1">rs</tok><tok id="w-17
<lb id="e-25"/> <tok form="Yt" fform="Ytem" id="w-171" nform="item" lemma
73" lemma="cerdo" pos="NCFP000">cerdas</tok> <tok id="w-174" lemma="en" p
nta</tok> <tok form="rs" fform="reales" id="w-176" lemma="real" pos="NCMP
<lb id="e-26"/> <tok form="Yt" fform="Ytem" id="w-178" nform="item" lemma
>una</tok> <tok id="w-180" lemma="mulo" pos="NCFS000">mula</tok> <tok id=
/tok> <tok form="rs" fform="reales" id="w-183" lemma="real" pos="NCMP000'
<lb id="e-27"/> <tok form="Yt" fform="Ytem" id="w-185" nform="item" lemma
>una</tok> <tok id="w-187" lemma="muleto" pos="NCFS000">muleta</tok> <tok
m="seiscientos">seyscientos</tok> <tok id="w-190" lemma="y" pos="CC">y</t
s="CC">y</tok> <tok id="w-193" lemma="5" pos="Z">cinco</tok> <tok form="i

```

Figura 35: Resultado del archivo XML tras la tokenización y procesamiento lingüístico completo (normalización, anotación morfosintáctica y lematización)

TEITOK facilita el manejo de los documentos y su modificación gracias a una interfaz que permite acceder a los archivos sin necesidad de trabajar directamente sobre el código XML. Este sistema de codificación de los textos es el que permite también recuperar la información en cualquiera de los niveles de anotación a partir de búsquedas en el corpus en CQL (§ 4.8.1. *Principios básicos de formación CQL*).

4.3. Normalización ortográfica

La normalización ofrece una versión modernizada del texto original según las normas ortográficas actuales. Esta opción permite disponer de una edición estandarizada que facilite la lectura de cada texto en el corpus. Por un lado, favorece la consulta del corpus para un público más amplio no necesariamente interesado en cuestiones paleográficas desde el punto de vista filológico; por otro, para el procesamiento informático del corpus, la normalización supone reducir los errores de los anotadores automáticos y posibilita la recuperación posterior de información (Calderón Campos y Vaamonde, 2020, p. 181).

Con la normalización todas las variantes ortográficas de una misma palabra quedan vinculadas a una forma estándar. Este procedimiento es especialmente útil en el caso de la documentación manuscrita antigua, caracterizada por la amplia variación ortográfica y la diversidad de formas en las que puede aparecer escrita una palabra incluso en el mismo texto. La normalización en los corpus históricos tiene la ventaja de poder recuperar a partir de una sola búsqueda todas las apariciones de un término, lo que es bastante eficaz, sobre todo para aquellas palabras que presentan una gran variabilidad de formas ortográficas, muchas de ellas con repercusión fonética:

- (7) ‘enaguas’: *henaguas, naguas, senaguas, sanaguas, enaguas...*
- ‘muaré’: *moer, moe, mue, muer...*
- ‘tinaja’: *tenaja, tenaxa, tenaga, tinaja, tinaxa, tynaja, ttinaja...*
- ‘muselina’: *moselina, mosolina, mosulina, mozolina, murselina, burselina...*
- ‘trébedes’: *trebes, extrevedes, estrebedes, estreves, trevedes, trebedes...*

Entre las variantes ortográficas de una misma palabra pueden aparecer algunas poco previsibles, por lo que la normalización se convierte en estos casos en una herramienta especialmente útil para recuperar variantes infrecuentes que raramente se podrían encontrar a partir de búsquedas aisladas (Calderón Campos, 2019b, p. 45). Por esta razón, en *ODE* la búsqueda por defecto está configurada para acceder a la forma normalizada de cada palabra y facilitar al usuario la consulta de las formas originales, entre las que se localizan algunas difíciles de suponer como las siguientes:

- (8) *omenaje* por ‘menaje’
- samucas* por ‘jamugas’
- cofaina* por ‘jofaina’
- sanaguas* por ‘enaguas’
- henero* por ‘enero’
- jugon* por ‘jubón’

La normalización afecta únicamente al nivel gráfico de las palabras y no se vincula con su paradigma flexivo (i.e. *niño*, *niña*, *niños*, *niñas*). Esta cuestión corresponde a la asignación de lemas y pertenece a una fase posterior de configuración del corpus (§ 4.6. *Lematización*). Si bien la lematización es más frecuente en los corpus actuales, no así la normalización ortográfica, algo que resulta vital para, por un lado, salvar el obstáculo que supone la enorme variedad gráfica que suelen presentar los textos antiguos; y por otro, para que el proceso automático de lematización funcione correctamente. *ODE*, por tanto, representa un corpus normalizado y lematizado al mismo tiempo. Sin embargo, en muchos corpus que no presentan ediciones normalizadas de los textos, la lematización se relaciona en cierto modo con la normalización ortográfica. En *ODE* se trata de dos procesos diferenciados con implicaciones también diferentes en la extracción de los datos. Esta cuestión se trata con mayor profundidad en § 4.6.1. *El concepto de lematización frente al de normalización*.

La consulta del *Diccionario de la lengua española* (*DLE*, 2019) ha sido determinante para la normalización de variantes fonéticas minoritarias. Algunas de estas variantes de una misma palabra aparecen recogidas en el diccionario formando entradas lexicográficas distintas. En estos casos, la variante menos extendida suele remitir por definición sinónímica a la forma más general, la cual es considerada la forma de referencia. Esta última, tomando como base el *DLE*, es la que se ha escogido para la normalización de aquellas voces que presentan diversidad fonética en el corpus. Pueden mencionarse como ejemplos las alternancias *frezada-frazada*, *zafrán-azafrán*, *jamuga-samuga*, *almario-armario*, *chamelote-camelote*, *macerina-mancerina*, *aljofaina-jofaina*, todas ellas variantes recogidas en el diccionario para las que se ha utilizado como voz normalizada la segunda de cada pareja.

En otros casos, la normalización ortográfica ha servido también para agrupar bajo la forma estándar las variantes desusadas de una misma palabra. Así ha ocurrido con los términos que en el *DLE* señalan su vigencia de uso con alguna de las distintas

marcas de variación diacrónica –*ant.*, *desus.* y *p. us.*³⁴–. En el corpus, esta circunstancia se aplica, por mencionar solo algunos ejemplos, a los pares *alatón-latón*, *flueco-fleco*, *defunto-difunto*, *mesmo-mismo*, *alcabuz-arcabuz* o *cuchar-cuchara*. Las formas anticuadas de cada par se encuentran registradas en el *DLE* y remiten a la variante actual (segunda de cada par), la cual se ha utilizado para la normalización.

Es conveniente destacar que la normalización en el corpus afecta únicamente al nivel ortográfico del texto, por lo que no se han modernizado otros aspectos morfosintácticos como el de las palabras que se emplean con género distinto al recogido en el *DLE* (**una almirez*, **la fin y muerte*, **hintera-hintero*, **legona-legón*) o la fusión gráfica de algunos compuestos (*hoja de lata* en lugar de *hojalata*). En estos casos la normalización conserva las características originales del manuscrito. Por otra parte, cabe recordar que la modernización ortográfica de mayúsculas y minúsculas, así como de la puntuación, es un aspecto que ya se tuvo en cuenta en la fase inicial de transcripción de los documentos (§ 3.1. *Criterios de transcripción*), por lo que estas cuestiones ya fueron intervenidas previamente y no se han modificado en esta fase. En cambio, sí que se han normalizado aquí las abreviaturas de origen griego (*xptobal* → *Cristóbal*, *XPO* → *Cristo*) y los signos no alfabetizables (@ → *arroba*) que fueron transcritos conforme el original durante la edición digital de los textos (véase sobre este último aspecto el apartado § 3.3.2.2. *Abreviaturas*).

En TEITOK, el proceso de normalización es automático y consiste en la adición de la forma gráfica actual mediante la incorporación de un atributo *@nform* (*normalized form*) a cada token en el corpus. El resultado para la forma *alaxas* sería el siguiente:

(9) <tok id="w-196" nform="alhajas">alaxas</tok>

El proceso se realiza gracias a un programa informático que permite importar parámetros de normalización de un corpus previo y aplicarlos a otro corpus. Por ello, lo más apropiado es que los datos importados se asemejen en la medida de lo posible al tipo de léxico que contienen los textos del nuevo corpus. En el caso de *ODE*, los primeros documentos fueron normalizados tomando como base la información del corpus *Post Scriptum*. Esto supuso algunos errores automáticos recurrentes en el caso de palabras que eran nuevas en *ODE*. No obstante, conforme iba aumentando el volumen del corpus, estos parámetros se actualizaron periódicamente a fin de ampliar los datos de los que consta el normalizador.

Tras la normalización automática se ha realizado una lectura íntegra del texto para localizar posibles errores y evitar que estos se sucedan en fases siguientes. Durante esta revisión se han corregido manualmente fallos automáticos aplicados a un solo token o a palabras en bloque mediante el sistema de edición que ofrece TEITOK; este se describe con detalle en § 4.10. *Revisión manual de errores en el corpus*.

³⁴ *Anticuada* para aquellas palabras o acepciones propias de la Edad Media o desechadas ya en el español moderno; *desusado* para las vigentes en la Edad Moderna pero no de uso actual; y *poco usado* para palabras cuyo uso está prácticamente extinguido (Puente Gutiérrez, 2000, p. 315).

El proceso completo de normalización prepara y facilita sobremanera la tarea posterior de anotación lingüística de los textos y resulta indispensable para mejorar la eficacia del anotador en cuando a la adjudicación automática de etiquetas morfosintácticas y lemas en el corpus.

4.4 Funcionamiento del anotador lingüístico en TEITOK

La anotación lingüística del corpus consiste en añadir a cada palabra información relativa a la función morfosintáctica y al lema. Con respecto a la primera, esta se hace por medio de una etiqueta POS (*part-of-speech*) con la que se indica la categoría gramatical y los rasgos morfológicos de la palabra en su contexto. En cuanto al lema, este señala la forma bajo la que se engloban todas las variantes flexivas de la palabra. Aunque etiqueta y lema se añaden automáticamente al mismo tiempo, a algunas de sus características se alude en los siguientes apartados de forma separada.

La asignación de etiquetas y lemas es automática en TEITOK y se realiza con el programa informático NeoTag (Janssen, 2012). Su ejecución incorpora de manera simultánea dos nuevos atributos: *@pos* y *@lemma*, para la etiqueta morfosintáctica y el lema, respectivamente. El resultado es un token con una nueva anotación lingüística.

NeoTag funciona utilizando datos de entrada importados de un corpus previamente anotado, o bien puede tomar como referencia los datos del propio corpus cuando este ya alcanza un tamaño considerable. Al igual que para la normalización, en la fase inicial del corpus los parámetros que se utilizaron fueron los de *Post Scriptum*. Una vez alcanzadas las 100 000 palabras, el anotador se desligó de estos parámetros para entrenarse con los datos de nuestro corpus de inventarios alojado en TEITOK. Esto es posible porque NeoTag puede, además de anotar corpus, utilizar el corpus en sí mismo como datos de entrada para su entrenamiento. Estos se actualizan constantemente al mismo tiempo que el corpus aumenta su tamaño con la anotación de nuevos textos (Janssen, 2016, p. 4039).

Se muestra a continuación el resultado del proceso automático de anotación para varios ejemplos de palabras en el corpus:

- (10) *vurras*
<tok id="w-657" nform="burras" pos="NCFP000" lemma="burro">vurras</tok>
- (11) *enbudico*
<tok id="w-536" nform="embudico" pos="NCMS00D" lemma="embudo">enbudico</tok>
- (12) 75
<tok id="w-1794" pos="Z" lemma="75">75</tok>
- (13) *ymbentareados*
<tok id="w-1473" nform="inventariados" pos="VMP00PM" lemma="inventariar">ymbentareados</tok>

(14) *vino*
`<tok id="w-1607" pos="NCMS000" lemma="vino">vino</tok>`
`<tok id="w-624" pos="VMIS3S0" lemma="venir">vino</tok>`

NeoTag adjudica a cada forma ortográfica una etiqueta en función de la frecuencia de aparición que esta tiene en los parámetros. Para cada uno de los ejemplos anteriores el programa utiliza diferentes reglas basadas en patrones de análisis lingüístico para reconocer la presencia de morfemas flexivos (10), sufijos diminutivos (11), caracteres numéricos (12) o partículas enclíticas (como se explicará más adelante). Puede además aplicar modelos probabilísticos según la frecuencia de aparición que una palabra con su respectiva etiqueta tiene en el corpus cuando esta tiene una anotación diferente dependiendo del contexto, como ocurre en (14) con la palabra *vino*, la cual puede ser un sustantivo o un verbo.

En cuanto a la asignación de lema, este se extrae a partir de la forma normalizada de la palabra de acuerdo con el sistema de herencias en TEITOK (§ 4.2. *Tokenización*, véase Figura 34), por lo que el anotador modificará cuando sea necesario esta forma estandarizada para crear el lema. Aplica para ello reglas de segmentación textual basadas en la cadena final de caracteres de la palabra (Janssen, 2012, p. 2). El anotador automático busca en el corpus de entrenamiento el lema de otras palabras con igual terminación y lematizará la nueva forma según la información que ya tiene para la mayoría del resto de formas similares.

La fase automática de anotación contempla también el tratamiento de las contracciones, esto es: dos o más palabras gramaticales que conforman una sola forma ortográfica. Es el caso de las partículas *al* y *del* o de formas verbales como *habiéndolos* o *haberme*. Si se considera que cada token corresponde a una sola forma ortográfica, no habría posibilidad de anotar morfosintácticamente cada lexema de manera independiente. TEITOK resuelve este problema con la adición de elementos `<dtok>` dentro de la etiqueta `<tok>`. El resultado tras la anotación del ejemplo previo *haberme* –*averme* en el original– sería el que sigue:

(15) *averme*
`<tok id="w-44" form="averme" nform="haberme">averme<dtok id="d-44-1 form="aver" nform="haber" pos="VAN0000" lemma="haber"/><dtok id="d-44-2" form="me" pos="PP1CS000" lemma="me"/></tok>`

En palabras nuevas o desconocidas de las que no tiene datos, NeoTag posee estadísticas de gramática universal y sabe aplicar *n*-gramas³⁵ en base a factores como la terminación de la palabra, la combinación con otras palabras u otras características de similitud. Por ejemplo, si consideramos que *inventariados* (13) es una nueva palabra desconocida para el anotador de la que no tiene registros previos, a la hora de adjudicarle una etiqueta POS y un lema, el programa buscará en sus parámetros todas aquellas palabras terminadas en *-dos*. Con respecto a la etiqueta morfosintáctica, los datos del corpus de

³⁵ Secuencia de palabras que suele ser frecuente en un corpus.

entrenamiento reflejan que las palabras con esta terminación suelen corresponder a las formas plurales de sustantivos, adjetivos o –lo más frecuente– participios, por lo que lo anotará como un participio más. Esta elección se fundamenta en que el sistema, a partir de la terminación de otras palabras, utiliza la frecuencia de distribución de las etiquetas de esas otras palabras en el corpus para calcular la probabilidad léxica del nuevo término (Janssen, 2012, p.2). Cuando el anotador automático ha establecido, basándose en la probabilidad, que se trata de un participio, tratará de obtener el lema segmentando la forma normalizada y añadiendo la terminación propia que corresponde a los lemas de las formas verbales.

La adjudicación automática de lemas se sustenta en un proceso prácticamente matemático en el que a cada palabra se le resta su desinencia y se le suma una nueva que se añade a la raíz. Siguiendo con el anterior ejemplo, el anotador modifica automáticamente la forma *inventariados* cambiando la desinencia. La nueva forma adopta la terminación del resto de palabras similares. El resultado de esta operación da lugar al lema, obtenido a partir de la supresión de la terminación en *-dos* y la adición de una nueva en su lugar, correspondiente a la forma del infinitivo en *-r*.

Mediante este sistema basado en la terminación de las palabras el programa acierta la anotación de numerosos casos. Por el contrario, en las palabras homógrafas, el anotador no puede elegir en base a la cadena final de caracteres y deberá elegir entre varias posibilidades, seleccionando siempre la opción más probable de entre las registradas en el corpus de entrenamiento.

Este sistema de anotación ha sido testado con un 95 % de eficacia en corpus de español moderno (Janssen, 2012, p. 3). No obstante, en un corpus de inventarios de los siglos XVIII y XIX como el que aquí se trata, la tasa de acierto ha sido menor. Las excepciones al cumplimiento de los patrones morfológicos mencionados harán al anotador fluctuar entre distintas posibilidades en las que a veces el resultado escogido no es el que se espera. Puesto que la asignación de etiqueta POS y lema son tareas simultáneas, fallos en la anotación morfosintáctica conllevan en muchos casos fallos también en la creación del lema (16). De manera similar, pero con distintas desinencias de las que se han venido ejemplificando, pueden explicarse también los casos de adjetivos o sustantivos anotados y lematizados como verbos por su analogía con los gerundios de las formas terminadas en *-ndo* o con las formas en plural del presente en primera persona (17), así como los errores cuando el significado de la palabra en su contexto no cumple con la etiqueta asignada (18). El sistema de funcionamiento de NeoTag explica errores en la anotación automática como los que se muestran a continuación:

(16)	<i>viudo</i>	→	POS: *VMP0000	→	lema: *viur
	<i>demasiado</i>	→	POS: *VMP0000	→	lema: *demasir
	<i>Quesada</i>	→	POS: *AQ0FS0	→	lema: *quesado
(17)	<i>hediondo</i>	→	POS: *VMG0000	→	lema: *hedior
	<i>orden</i>	→	POS: *VMIP3P0	→	lema: *order
	<i>céntimos</i>	→	POS: *VMIP1P0	→	lema: *centimar

(18) se *fueron* a la hacienda → POS: *VSIS3P0 → lema: *ser³⁶

El anotador automático no es del todo fiable a la hora de distinguir los diferentes usos de una misma palabra con respecto a la función que cumple en el texto. De esta manera, palabras coincidentes en su grafía como son los verbos *ir* y *ser* en las formas de pretérito de (18) tienen una mayor tendencia a fallar en la anotación automática. Por el momento, y con los datos que actualmente alberga el corpus, el anotador no es capaz de elegir correctamente en base a los datos que posee entre los distintos lemas para una misma forma ortográfica y esta revisión deberá ser necesariamente manual.

De manera estricta, los casos de lemas fallidos mostrados en los ejemplos (16-18) no deben ser considerados exactamente errores. Se trataría de la forma lematizada esperable que corresponde a cada palabra según la categoría gramatical que le ha sido asignada (etiqueta POS). De acuerdo con su etiqueta, la adjudicación del lema cumple a la perfección con las reglas de la morfología flexiva de las palabras y aplica con lógica las características funcionales del español, aunque es evidente que no acierta en todos los casos. Por esta razón, los ejemplos señalados son una prueba fehaciente de la necesidad de realizar una revisión manual de la anotación automática para evitar que estos errores pasen al corpus de entrenamiento y así evitar que se vuelvan a producir con la anotación de nuevos textos.

Como se ha explicado, los fallos en la adjudicación de lema suelen estar vinculados con una incorrecta etiqueta POS. Los errores cometidos por el anotador automático exclusivamente en la lematización son muy escasos y rara vez se producirán fallos en el lema que no deriven de inconsistencias previas presentes en el corpus de entrenamiento (Janssen, 2012, p. 2). Sin embargo, no son infrecuentes los casos de aquellas palabras que, aun siendo correctamente etiquetadas, han fallado en la lematización. La creación automática del lema suele ser exitosa para las palabras que flexionan según las tendencias o leyes generales, pero fracasa en otros muchos casos en los que se presentan irregularidades motivadas por cambios fonéticos que responden a cuestiones puramente de evolución histórica del español. Particularmente problemáticos son los fenómenos que afectan al paradigma verbal, como por ejemplo la diptongación en las formas de presente (*sentir* > *siento*), los casos de inflexión vocálica (*medir* > *midió*) o el perfecto etimológico (*querer* > *quise*, *saber* > *supe*). Este tipo de anomalías que están sujetas a cambios en el vocalismo y consonantismo, incluso a alternancias ortográficas (*hoz* / *hoces*), han motivado fallos automáticos en la lematización como los siguientes:

(19)	<i>vecas</i>	→	POS: NCFP000	→	lema: *vec
	<i>hoces</i>	→	POS: NCMP000	→	lema: *hoc
(20)	<i>sastres</i>	→	POS: NCMP000	→	lema: *sastr
	<i>catres</i>	→	POS: NCMP000	→	lema: *catr

³⁶ Por oposición a este otro ejemplo: *fueron* tasadas → POS: VSIS3P0 → lema: ser

(21)	<i>sienten</i>	→	POS: VMIP3P0	→	lema: *sienter
	<i>ríen</i>	→	POS: VMIP3P0	→	lema: *ríer
	<i>midió</i>	→	POS: VMIS3S0	→	lema: *midier
(22)	<i>quisieron</i>	→	POS: VMIS3P0	→	lema: *quiser
	<i>pusieron</i>	→	POS: VMIS3P0	→	lema: *puser
(23)	<i>valido</i>	→	POS: VMP0000	→	lema: *valir
	<i>tenidos</i>	→	POS: VMP00PM	→	lema: *tenir
(24)	<i>salen</i>	→	POS: VMIP300	→	lema: *saler
	<i>añaden</i>	→	POS: VMIP3P0	→	lema: *añader
	<i>produciendo</i>	→	POS: VMG0000	→	lema: *producer

Estos errores se explican por la técnica de segmentación de las palabras en el proceso automático de lematización que usa NeoTag (Arrabal Rodríguez, 2020, p. 72). Para crear el lema, el programa no es capaz de tener en cuenta las irregularidades morfológicas que responden a determinadas reglas ortográficas del sistema de la lengua (19), ni tampoco de considerar la *-e* final de los sustantivos como parte integrante de la raíz y no de la desinencia (20). Por otro lado, en lo que a la flexión verbal se refiere, el sistema también fallará en el lema de los verbos que no se ajustan a los paradigmas regulares de conjugación. Esto es, verbos con alternancia vocálica /e/~i/, /e/~ié/, irregularidades en el cierre vocálico (21) o verbos con las alternancias vocálicas y consonánticas (alternancias mixtas) como las que suelen presentar los conocidos como pretéritos fuertes (NGLE § 4.9c, p. 219) (22). La terminación en *-ido/a/os/as* en la que coinciden los participios de los verbos de la segunda y tercera conjugación también suele ofrecer dificultades al sistema a la hora de adjudicar un lema en infinitivo con forma en *-er* o *-ir* y lo mismo ocurre con otras terminaciones en las que coinciden vocal temática y desinencia, por lo que es imposible para el anotador reconocer automáticamente a qué conjugación pertenece el verbo y acaba seleccionando la que considera más probable (23 y 24). Otros ejemplos de lemas erróneos se han dado también con algunos participios irregulares.

La totalidad de ejemplos que se han mostrado conforman una casuística de los errores que con más frecuencia se han detectado tras realizar la anotación lingüística automática con el programa NeoTag. El proceso automático alcanza un porcentaje de error que exige necesariamente una revisión manual de las etiquetas morfosintácticas, así como del lema. No obstante, muchos de estos errores tuvieron una mayor presencia en las primeras fases de anotación del corpus, cuando el volumen de textos anotados era todavía considerablemente bajo. El porcentaje de fiabilidad mejoró considerablemente cuando el corpus aumentó su volumen y el anotador pasó a utilizar los propios parámetros de ODE. Aun así, una posterior revisión manual de todos los textos ha sido condición *sine qua non* para que estos pasaran a formar parte del corpus.

La corrección manual y la continua actualización del anotador resultaron claves para que las formas correctamente anotadas se incorporaran al corpus de entrenamiento

y así aumentar progresivamente la tasa de acierto. El método de corrección y revisión de errores se expone en § 4.10. *Revisión manual de errores en el corpus*.

4.5. Anotación morfosintáctica (POS) en el corpus ODE

La anotación morfosintáctica implica la adjudicación de una etiqueta morfológica para cada palabra del texto según la función gramatical que cumple en él. En TEITOK es posible anotar automáticamente los textos con el programa informático NeoTag (§ 4.4 *Funcionamiento del anotador lingüístico en TEITOK*). En este apartado se alude al conjunto de las etiquetas que se han utilizado para la anotación del corpus, así como a algunos criterios metodológicos que hubo que adoptar para la anotación de casos particulares.

Por las características de esta investigación, el corpus de esta tesis destaca por la anotación precisa de sufijos apreciativos, algo que contrasta con el anotado morfosintáctico mucho más superficial que presentan otros corpus académicos de referencia como el CDH y CORPES XXI (Calderón Campos, en prensa) con los que no es posible, o al menos, resulta bastante costoso, realizar estudios de este tipo.

La búsqueda de las palabras con presencia de un verdadero sufijo diminutivo ha sido trascendental para el análisis estadístico desarrollado en esta tesis. Su recuperación automática ha sido posible gracias a la anotación morfosintáctica, que ha permitido descartar el resto de las palabras en las que el diminutivo se ha lexicalizado³⁷:

(25)	Un <i>arillo</i> de yerro con quarenta y ocho escarpías	NCMS00D
	Vnos pendientes de aljofar con sus <i>arillos</i> de oro	NCMP000

La anotación constó de una primera fase automática y de una segunda fase de revisión manual con la que se asegurar la fiabilidad de este corpus para estudios basados en aspectos morfosintácticos.

4.5.1. Etiquetario

El etiquetario utilizado para anotar el corpus se ha basado en el utilizado por el proyecto *Post Scriptum* para la anotación de cartas manuscritas, aunque con algunas ligeras modificaciones. Este sigue las directrices del estándar propuesto por el grupo EAGLES³⁸ para la anotación de lexicones en lengua española (Vaamonde, 2015a).

Cada etiqueta se fundamenta en una cadena combinada de letras y números asociados a distintos rasgos morfosintácticos que dependen de la posición que ocupen dentro de la cadena. La primera letra de la etiqueta representa siempre la categoría gramatical a la que pertenece la palabra, y el resto de los caracteres adquieren su

³⁷ El distintivo ‘diminutivo’ se refleja en la etiqueta morfosintáctica con la presencia de una D (Tabla 5).

³⁸ <https://www.ilc.cnr.it/EAGLES96/home.html#ancora>

significado en relación con ella. Por ejemplo, la letra P indica el rasgo plural en la cuarta posición de los pronombres, sin embargo, en segunda posición señala el tipo de pronombre, personal en este caso. En los verbos, en cambio, el rasgo plural se define en la quinta posición y no antes. Por tanto, el significado de las letras dentro de la etiqueta está determinado por un sistema de posiciones más que por el significado en sí de los caracteres individualmente.

El etiquetario completo utilizado para la anotación del corpus se puede consultar en la web del corpus *ODE* (en la sección *Etiquetario* del menú principal). En la Tabla 5 se recoge únicamente el conjunto de las etiquetas que se aplican a adjetivos y nombres, por ser estas las únicas categorías gramaticales en el corpus a las que se adjuntan diminutivos. El rasgo ‘diminutivo’ se refleja en la etiqueta morfosintáctica con una D en la segunda posición de adjetivos y en última posición en el caso de los sustantivos.

	Posición	Rasgo		Ejemplos	
ADJETIVO (A)	1. Tipo	Q	Cualitativo	una caldera <u>rota</u>	AQ0FS0
		O	Ordinal	su <u>primera</u> mujer	AO0FS0
		X	Posesivo	dote y caudal <u>suyo</u>	AX0MS0
	2. Grado	0	No aplica	dos más <u>usadas</u>	AQ0FP0
		A	Aumentativo	cintas <u>anchonas</u>	AQAFP0
		D	Diminutivo	un cazo <u>chiquito</u>	AQDMS0
		C	Comparativo	en <u>peor</u> uso	AQCCS0
		S	Superlativo	una <u>purísima</u> Trinidad	AQSFS0
	3. Género	M	Masculino	los <u>infrascritos</u>	AQ0MP0
		F	Femenino	pintura <u>basta</u>	AQ0FS0
		C	Común	de lana <u>azul</u>	AQ0CS0
NOMBRE (N)	4. Número	S	Singular	<u>media</u> docena	AQ0FS0
		P	Plural	dos lebrillos <u>grandes</u>	AQ0CP0
		N	Invariable		
	5. Función	0	No aplica		
	1. Tipo	P	Propio	<u>Agustín Segura</u>	NP00000
		C	Común	<u>sombrilla</u> de seda	NCFS000
	2. Género	M	Masculino	arrobas de <u>aceite</u>	NCMS000
		F	Femenino	tinaja para <u>agua</u>	NCFS000
		C	Común	escribano y <u>testigos</u>	NCCP000
	3. Número	S	Singular	<u>abanico</u> de nácar	NCMS000
		P	Plural	<u>abanicos</u> para chimenea	NCMP000
		N	Invariable	<u>lunes</u> diez y ocho	NCMN000
	4.	0	No aplica		
	5.	0	No aplica		
	6. Grado	0	No aplica		
		D	Diminutivo	<u>pañuelitos</u> blancos	NCMP00D
		A	Aumentativo	<u>botijón</u> viejo	NCMS00A

Tabla 5: Etiquetado morfosintáctico de adjetivos y nombres en el corpus *ODE*

4.5.2. Casos especiales en la anotación morfosintáctica

En este apartado se detallan los criterios adoptados para la anotación de los casos que han presentado dudas o cuyo etiquetado ha sido objeto de mayor reflexión.

La anotación lingüística de *ODE*, a excepción de ligeras modificaciones en algunos aspectos, se basa en buena parte en las directrices seguidas para el etiquetado morfosintáctico del corpus *Post Scriptum* (Vaamonde y Magro, 2016). Las pautas aquí descritas no pretenden ser una guía exhaustiva de la anotación en *ODE*, sino una síntesis de algunos casos particulares que pueden ser de interés a la hora de recuperar información a partir de la anotación morfosintáctica de las palabras en el corpus. Para ello es relevante conocer cómo se han etiquetado determinadas categorías gramaticales a efectos de poder obtener los resultados deseados.

Cabe destacar que la anotación lingüística del corpus es morfosintáctica: la elección de la etiqueta se realiza en base a la función que la palabra cumple en la oración, de modo que una misma palabra puede aparecer varias veces en el corpus, pero tener distinta anotación (i.e. *dicho* como participio / *dicho* como determinante / *dicho* como sustantivo)³⁹. No obstante, la anotación sí es morfológica en el caso de las formas que actúan como unidades léxicas y que se anotan de manera independiente, palabra por palabra. Esta circunstancia afecta a numerosos casos en el corpus, como se expondrá más adelante a propósito de las unidades fraseológicas:

pavo real	NCMS000 + AQ0CS0
alguacil mayor	NCMS000 + AQCCS0

4.5.2.1. Participios y adjetivos

La afinidad sintáctica que en muchos casos une a participios y adjetivos ha llevado a considerarlos de manera conjunta. Para la anotación de los adjetivos que coinciden con los participios actuales de algunos verbos se han establecido los siguientes criterios:

- a. Se anotan como adjetivos cuando han adquirido un significado propio distinto al que aporta el verbo del que derivan. Se trata de los participios que funcionan como adjetivos en cuanto a modificadores del nombre:

sábanas ya <i>andadas</i>	AQ0FP0
marcos <i>cuadrados</i>	AQ0MP0

³⁹ La dificultad de asignar automáticamente etiquetas morfosintácticas, mayor si cabe en un corpus diacrónico, se hace patente en muchos casos y de manera especial en aquellos que han sufrido un cambio de categoría. Un ejemplo evidente es el caso de *dicho*. Sobre la anotación de esta palabra concreta en *ODE*, véase Calderón Campos y García-Godoy (2023).

- b. Se anotan como participios flexionados cuando el contenido semántico no sufre ninguna variación con respecto al significado del verbo, independientemente de que funcionen como adjetivos:

bienes <i>inventariados</i>	VMP00PM
silla baja <i>forrada</i>	VMP00SF
año <i>pasado</i>	VMP00SM
pañuelo <i>bordado</i>	VMP00SM

- c. Se anotan como nombres comunes las palabras que proceden de participios considerados léxicamente ya sustantivados:

lienzo con <i>deshilados</i>	NCMP000
calzones de <i>estezado</i>	NCMS000
crystal y <i>vidriado</i> fino	NCMS000

En cuanto al tratamiento de los participios, estos se dividen entre participios con flexión y participios sin flexión:

- a. Son participios no flexionados los integrantes de tiempos compuestos formados por el auxiliar *haber* + participio, puesto que no incluyen información de género y número:

había <i>mandado</i>	VMP0000

- b. Son participios flexionados los que no forman parte de tiempos verbales compuestos, así como los participios de las pasivas flexionadas con *ser*, los cuales se conjugan en género y número. Entran en esta casuística los participios que cumplen función adjetival ya mencionados arriba:

el <i>expresado</i> mi marido	VMP00SM
fue <i>tasada</i>	VMP00SF
sábanas <i>guarnecidas</i>	VMP00PF

4.5.2.2. Formas verbales

La anotación de formas verbales contempla en la segunda posición de la etiqueta morfosintáctica tres categorías diferentes según la función y la tipología del verbo:

- a. Verbo auxiliar (A). Todas las formas de *haber* cuando este funciona como auxiliar:

bienes que <i>ha</i> dejado el difunto	VAIP3S0

los bienes que <i>hayan</i> quedado	VASP3P0

- b. Verbo *ser* (S). Todas las ocurrencias del verbo *ser* incluidas aquellas usadas en la formación de la voz pasiva y de las perífrasis verbales:

vecinos que <i>fuleron</i> de la villa	VSIS3P0
el matrimonio <i>fuere</i> disuelto	VSSF3S0

- c. Verbo principal (M). Todos los verbos que gozan de pleno significado léxico, incluido también el verbo *haber* cuando no funciona como auxiliar:

<i>contrajo</i> matrimonio	VMIS3S0
dos orcas de <i>aventar</i>	VMN0000
en la forma que <i>haya</i> lugar	VMSP3S0
<i>habrá</i> más de cuatro años	VMIF3S0

4.5.2.3. Adverbios

Se reduce considerablemente la anotación de los adverbios y se agrupan todos bajo la etiqueta general de adverbio (R), en la que no se deja constancia de cualquier otro rasgo en el que se indique la pertenencia a alguna de las tradicionales categorías en las que se suelen clasificar este tipo de palabras en base a su significado (adverbios de lugar, de cantidad, de tiempo, de manera, de duda, de afirmación, de negación, etc.). Esta simplificación difiere del criterio adoptado en el corpus *Post Scriptum*, en el que los adverbios se dividen en cuatro tipos diferentes, y va encaminada a mejorar el porcentaje de acierto del anotador.

En el caso de las locuciones adverbiales, se anota de manera separada cada pieza de la unidad léxica.

<i>abajo</i> de dichas casas	R
unas enaguas <i>algo</i> traídas	R
<i>no obstante</i>	R + R

Los adverbios coordinados que derivan de adjetivos donde solo el segundo presenta la forma acabada en *-mente* se etiquetan ambos como adverbio:

<i>bien</i> y <i>fielmente</i>	R + R
<i>real</i> y <i>efectivamente</i>	R + R

4.5.2.4. Locuciones

Los grupos sintácticos formados por la agrupación de dos o más palabras como lo son las locuciones preposicionales o adjetivales no se consideran una única unidad o token a nivel de anotación morfosintáctica. Se ha conservado una etiqueta por cada una de las palabras independientes que forman parte de estas expresiones fijas. En estos casos, la anotación es de carácter morfológico e individual para cada forma:

tela de algodón <i>a rayas</i>	SPS00 + NCFP000
dinero <i>contante y sonante</i>	AQ0CS0 + CC + AQ0CS0

4.5.2.5. Palabras en lengua extranjera

Las palabras en lengua distinta al español, únicamente en latín para este corpus, no se han anotado morfosintácticamente. Se trata de las palabras integrantes de expresiones latinas, generalmente locuciones adverbiales, presentes en los inventarios, como *in solidum*, *in facie ecclesiae*, *ad litem* o *in dei nomine*. Estas formas no son recuperables a partir de la anotación morfosintáctica.

4.6. Lematización

La lematización consiste en asignar a cada palabra una forma base o lema bajo la que se adscriben sus diferentes formas flexionadas. El lema actúa como representante de todas las posibles variantes conjugadas que puede adoptar una palabra, por lo que constituye la forma con la que el término se incluye en los diccionarios. Tras una búsqueda en un corpus lematizado de la palabra *hacer*, el resultado podrá contener las formas *hecho*, *hagan*, *hizo*, *hará*, *hacer*, etc.

En el caso de *ODE*, la lematización forma parte de la fase de anotación lingüística del corpus y se desarrolla de forma automática con el programa NeoTag. El anotador asigna a cada palabra una etiqueta gramatical y un lema mediante la adición a cada token de un atributo denominado *@lemma* (§ 4.4 *Funcionamiento del anotador lingüístico en TEITOK*). El lema se obtiene mediante procedimientos informáticos de segmentación a partir de la forma normalizada de la palabra. A este proceso le ha seguido la revisión manual de lemas para detectar y corregir errores automáticos (§ 4.10. *Revisión manual de errores en el corpus*).

En las páginas que siguen se profundiza en las implicaciones que tiene el proceso de ‘lematización’ en *ODE* a la hora de extraer información, por contraposición a la concepción habitual del término en otros corpus del español. Asimismo, se precisan las decisiones metodológicas que se han seguido para la asignación de lemas en el caso concreto de este corpus y, más concretamente, en el léxico perteneciente a los inventarios de bienes.

4.6.1. El concepto de lematización frente al de normalización

El lema constituye la forma base bajo la que se agrupa el conjunto de variantes flexionadas de una palabra y que se corresponde con las características que afectan al género, número, persona, tiempo, modo y, con menos frecuencia, al caso (*NGLE* § 1.5b, p. 21). En las palabras con propiedades flexivas –nombres, adjetivos, determinantes, pronombres y verbos– se escoge la forma base que representa a todo el paradigma flexivo, mientras que en las palabras invariables que carecen de flexión –adverbios, preposiciones, conjunciones e interjecciones– el lema no se modifica y coincide con la forma invariable de la palabra.

La lematización es una característica cada vez más extendida de los corpus informatizados actuales teniendo en cuenta las numerosas ventajas que ofrece para la recuperación de información y el análisis de datos. De entre todas las funcionalidades que un ejemplo de corpus productivo y eficaz puede ofrecer, la lematización ha sido una de las técnicas más demandadas por los investigadores, quienes la consideran esencial para las tareas de análisis lingüístico (Díaz Bravo, 2015). Sin embargo, a pesar del creciente número de corpus especializados en lengua española, la lematización está todavía lejos de ser una práctica habitual en la mayoría de ellos. Tras un análisis de seis corpus significativos para el estudio de la historia de la lengua, solo dos –*CorLexIn* y *Post Scriptum*– resultan estar lematizados en su totalidad (Calderón Campos, 2019b, p. 56). Por su parte, De Benito Moreno (2019) examina diferentes aspectos en un total de trece corpus, de los cuales siete se encuentran lematizados⁴⁰.

De forma casi generalizada, en muchos de estos corpus la lematización está relacionada con la normalización ortográfica de las palabras, de modo que el lema contempla no solo las variantes flexivas de las palabras, sino también las variantes gráficas. La lematización exige por sí misma una estandarización previa de las formas gráficas, sin embargo, pocos corpus ofrecen una versión normalizada de los textos, lo que dificulta las búsquedas cuando lo que interesa es una única forma y no un lema entero. Por ejemplo, en el *CDH*, en el que es posible realizar búsquedas lematizadas, la búsqueda de *mujer* devuelve los resultados *mulier*, *mugier*, *muller*, *Mujer*, *muxer*, etc., pero también los de *muxeres*, *muglieres*, *muilleres*, *mujercillas*, etc., sin que sea posible descartar los casos en plural. Por el contrario, la búsqueda por ‘forma’ para rescatar solo casos en singular no contempla las distintas formas gráficas de la palabra en los textos. La cuestión es aún más problemática en el caso de los verbos, pues si lo que interesa es solo una forma verbal concreta, la única opción es recuperarla a partir de la búsqueda del infinitivo, en la que habrá que ir descartando todo el resto de las formas del paradigma que no interesan. Solo en un corpus normalizado y lematizado será posible hacer este tipo de búsquedas más precisas.

Por otra parte, en un corpus no normalizado, los errores en la lematización serán mucho más abundantes, puesto que la modernización ortográfica ya supone un control previo de la variabilidad gráfica asociada a una misma palabra. Volviendo al *CDH*, las

⁴⁰ *CORPES XXI*, *CDH*, *CdE*, *ESLORA*, *Val.Es.Co 2.0*, *COSER*, *Biblias Hispánicas* y *Post Scriptum*, a excepción de *CORDE*, *CREA* o *CORDIAM* (De Benito Moreno, 2019, p. 6).

formas *España* y *Espana* constituyen lemas distintos en el corpus, debido a un error de normalización.

Por lo tanto, normalización y lematización tienen implicaciones diferentes a la hora de recuperar información en un corpus, más concreta la primera, más amplia la segunda. *ODE*, siguiendo el modelo de *Post Scriptum*, sí contempla este nivel de marcación que concierne a la forma estándar de las palabras, la cual es accesible para el usuario desde el motor de búsquedas. Conviene precisar esta distinción, pues el concepto de ‘lematización’ en el sistema TEITOK no es sinónimo de ‘modernización ortográfica’, la cual corresponde a una fase previa de elaboración del corpus (§ 4.3. *Normalización ortográfica*) y cuyas ventajas se asumen en la fase de lematización.

El lema, por tanto, engloba aquellas palabras que en el corpus forman parte de un mismo paradigma flexivo, tanto para verbos (*inventariadas*, *inventarió*, *inventariamos*, *inventariando*, etc.) como para sustantivos o adjetivos (*blanco*, *blanca*, *blancos*, *blancas*). La lematización en *ODE* concierne únicamente a la agrupación por lemas y no a la sistematización de las variantes gráficas. Sin embargo, puesto que la lematización abarca el conjunto de todas las formas normalizadas de cada palabra, “arrastra” también las variantes gráficas que se asocian a cada una de esas formas. La diferencia entre normalización y lematización se entiende bien al comparar el resultado de las búsquedas de *tinaja* en el corpus como forma normalizada (a la izquierda) y como lema (a la derecha):

Search Query	[nform="tinaja"]
Group query	Transcribed form
Total	243
Reference size	1166498

Graph:

Table	▼

 | Count:

Group	Count	WPM	Percent
tinaja	128	109.73	52.67
tinaxa	46	39.43	18.93
tenaja	35	30	14.4
tynaja	15	12.86	6.17
tenaxa	9	7.72	3.7
ttinaja	5	4.29	2.06
tenaga	4	3.43	1.65
ttinaxa	1	0.86	0.41

Search Query	[lemma="tinaja"]
Group query	Transcribed form
Total	380
Reference size	1166498

Group	Count	WPM	Percent
tinaja	107	91.73	28.16
tinajas	97	83.15	25.53
tinaxas	47	40.29	12.37
tinaxa	43	36.86	11.32
tenaja	32	27.43	8.42
tenajas	13	11.14	3.42
tenaxa	9	7.72	2.37
tynaja	6	5.14	1.58
tenaga	4	3.43	1.05
tenaxas	4	3.43	1.05
tynajas	3	2.57	0.79
ttinaja	3	2.57	0.79
tinajillas	2	1.71	0.53
tinajitta	2	1.71	0.53
tinaxitas	1	0.86	0.26
tinajitas	1	0.86	0.26
tinajuelas	1	0.86	0.26
tinajita	1	0.86	0.26
tinagas	1	0.86	0.26
Tinajas	1	0.86	0.26
tenajita	1	0.86	0.26
tinajica	1	0.86	0.26

Figura 36: Resultado de las búsquedas normalizada (izquierda) y lematizada (derecha) de la forma *tinaja*. Concordancias ordenadas en base a la forma original transcrita

Como se puede comprobar, la consulta por forma normalizada es mucho más restrictiva. Esta búsqueda devuelve únicamente los casos asociados a una forma estandarizada concreta: *tinaja*, mientras que la consulta lematizada incluye además el resto de las formas de su paradigma flexivo: *tinaja*, *tinajas*, *tinajillas*, *tinajitas*, *tinajica*, etc. con sus correspondientes variantes ortográficas. La lematización vincula un mayor número de tokens, por lo que se trata de una marcación más amplia de los textos con la que se obtienen más resultados en una sola búsqueda. Su procesamiento concierne exclusivamente a la forma en que se constituyen los lemas en el corpus, ya que la estandarización gráfica tiene lugar previamente con la edición normalizada de cada texto. La normalización, por tanto, queda incluida también en la lematización y la diferencia entre ambas radica, para esta última, en los siguientes aspectos:

1. Vinculación de las variantes con flexión de género y número con la forma en masculino singular.
2. Vinculación del paradigma verbal bajo la forma en infinitivo.
3. Vinculación de las palabras con sufijos diminutivos de significado transparente con la forma base de la que derivan (*pequeñito* → lema ‘pequeño’).
4. Agrupación de las formas con distinción gráfica entre mayúsculas y minúsculas a favor de esta última (*inventario* e *Inventario* → lema ‘inventario’).

Una mención especial merece el tratamiento de las formas con sufijación diminutiva, de cuyo lema se ha excluido el sufijo en los casos en los que este tiene un significado transparente (*tinajilla* → lema ‘tinaja’). Por el contrario, y como es evidente, el lema asume el sufijo en aquellas palabras en las que el diminutivo se ha lexicalizado (*tornillo* → lema ‘tornillo’).

Por último, resta por mencionar que bajo el lema se reúnen exclusivamente las variantes morfológicas que comparten el mismo significado léxico. Por esta razón, este criterio ha sido decisivo en algunos casos para determinar el lema de palabras con variación de género o número. Así, las variantes formales de una palabra que no comparten significado léxico presentan lemas diferentes (*vidriero* → lema ‘vidriero’, *vidriera* → lema ‘vidriera’). Esta cuestión y otras en torno a la lematización del corpus se trata ampliamente más adelante (§ 4.6.2.2. *Derivados nominales*).

Como se ha explicado, el concepto de lema puede variar entre corpus e incluso entre diccionarios, los cuales adoptan diferentes criterios a la hora de establecer la forma base de las palabras. El estándar EAGLES difiere en algunos aspectos de los criterios comúnmente establecidos. Un ejemplo es la lematización de algunos pronombres (la forma *yo* se constituye como lema de los pronombres átonos *me*, *mí* y *nos*, mientras que los diccionarios suelen considerar estas cuatro formas como lemas en sí mismos). De este modo, resulta imprescindible conocer cómo se ha lematizado un corpus a la hora de interpretar los resultados tras una búsqueda lematizada.

La forma en la que se ha establecido el lema para cada una de las categorías gramaticales en este corpus de inventarios se aborda a continuación.

4.6.2. Decisiones metodológicas para la adjudicación de lemas

Se recogen en este apartado los criterios establecidos en el corpus *ODE* para determinar el lema de las palabras.

4.6.2.1. Sustantivos

Para los sustantivos que designan personas o seres animados (*alcalde, burro, hermano, hijo*) el lema se constituirá siempre con el género no marcado, esto es, la forma en masculino como uso genérico de ambos sexos (*NGLE* § 2.2a, p.85). Así, se elegirá la forma masculina singular para lematizar los sustantivos variables en género y número:

<i>ministro</i>	lema: ministro
<i>sobrinos</i>	lema: sobrino
<i>cerdas</i>	lema: cerdo

Los nombres que poseen un único género gramatical y que por tanto solo poseen variación en el rasgo morfológico referente al número, como los llamados epicenos (*gallina, oveja*), los comunes en cuanto al género (*testigo, cónyuge*) o los sustantivos heterónimos (*hombre / mujer, toro / vaca*) se lematizan bajo la forma en singular que indica su género gramatical inherente:

<i>alguacil</i>	lema: alguacil
<i>mercedes</i>	lema: merced
<i>mujeres</i>	lema: mujer

Mención aparte merecen aquellos sustantivos invariables que no poseen flexión en los rasgos de género y número, los cuales se lematizan bajo su misma forma:

<i>lunes</i>	lema: lunes

En el caso de los nombres propios, no se lleva a cabo la lematización por excelencia bajo la forma en masculino singular de los sustantivos. La variación de género que admiten algunos antropónimos se mantiene inalterada conformando lemas independientes. Por tanto, los nombres *Juan, Juana, Luis y Luisa* constituyen por sí mismos un lema propio cada uno. Ocurre lo mismo con aquellos nombres, principalmente femeninos, que refieren unicidad aun presentando forma plural (*Ángeles, Dolores, Nieves*); y lo mismo sucede con algunos apellidos (*Megías, Poyatos, Mateos*) y topónimos (*Indias*), los cuales se lematizan bajo su misma forma y no con el singular. Distinto es el caso de los antropónimos que se asimilan a los sustantivos comunes y se pluralizan cuando el mismo nombre es compartido por varios individuos o realidades (*NGLE* § 3.6a, p. 160), como ocurre en “los *Manueles*”, “las *Marías*”, “las *Américas*”. En estos casos, el lema será la forma en singular: ‘manuel’, ‘maría’, ‘américa’.

Como convención adoptada de forma unánime por los programas automáticos de lematización, el lema se indica siempre en minúscula, incluido el de las formas que constituyen nombres propios. Todo lo mencionado para los nombres propios se ejemplifica en los siguientes casos extraídos del corpus:

<i>Antonio</i>	lema: antonio
<i>Antonia</i>	lema: antonia
<i>Angustias</i>	lema: angustias
<i>Nieves</i>	lema: nieves
<i>los Medinas</i>	lema: medina

Aquellos nombres de personas o lugares que proceden de nombres comunes y que, por tanto, coinciden con ellos en la forma, coincidirán también en el lema:

- (26) a. Yt[em] otro de sermones de s[a]n Juan de la *Cruz* (*ODE* – Madrid, 1793).
 b. [...] la viuda Rafaela Guerrero *Cruz* (*ODE* – Almería, Tabernas, 1870).
 c. Una *cruz* de oro guarnecida de esmeraldas (*ODE* – Almería, Fiñana, 1824).
 d. Un peso de yer[r]o de *cruz* (*ODE* – Almería, Abia, 1726).

Así, las palabras señaladas en los ejemplos precedentes se englobarán bajo la forma ‘cruz’ en todos los casos. Recuérdese que el lema de los nombres propios va en minúscula, por lo que independientemente de que se aluda a realidades diferentes el lema será siempre el mismo. No ocurre así con los nombres propios en plural que del mismo modo coinciden con nombres comunes:

- (27) a. Estando en la cama mortuoria de Pedro Sevilla *Cortinas* (*ODE* – Almería, Cuevas de Almanzora, 1786). [lema → ‘cortinas’]
 b. Tres juegos de *cortinas* blancas adamascadas (*ODE* – Almería, 1848). [lema → ‘cortina’]

En estos casos, el apellido adquiere un lema propio, mientras que el sustantivo común seguirá las normas generales de lematización mencionadas anteriormente para los nombres comunes. Así, para estos casos se adjudicarán distintos lemas: ‘cortinas’ y ‘cortina’ respectivamente. Esta circunstancia se da repetidamente en el corpus con multitud de apellidos con formas en plural (*Paredes*, *Espadas*, *Ríos*, *Granados*) y topónimos (*Tabernas*, *Caniles*, *Lagos*), entre otros:

<i>Cuevas de Almanzora</i>	lema: cuevas
<i>Teresa Abad</i>	lema: abad
<i>varias flores y figuras</i>	lema: flor
<i>Miguel Flores</i>	lema: flores

Para la lematización de los derivados nominales y los plurales inherentes se ha seguido una metodología específica. Aunque también pertenecientes a la categoría de los sustantivos, estos casos se tratan de forma separada en los apartados que siguen.

4.6.2.2. Derivados nominales

Se desarrolla en este apartado cómo se ha llevado a cabo la lematización de algunos sustantivos derivados mediante sufijación. Debido a las características inherentes de un corpus conformado por inventarios de bienes, muchos de estos derivados designan objetos y utensilios, pero también nombres de oficios, lugares o estancias de la casa, prendas, telas, etc. tal y como cabe esperar de este tipo de documentación como reflejo de la vida cotidiana de la época. Es por esta razón por la que nos centraremos en algunos de los derivados que más productividad tienen en la formación de sustantivos para nombrar a este tipo de realidades de las que constan en gran medida los inventarios de bienes. Así, nos referimos principalmente a los derivados a partir de los sufijos *-dor / -dora*, *-dero / -dera*, *-ero / -era*, *-torio / -toria*, *-a / -o*, entre los más comunes para formar nombres de persona, instrumento o lugar (NGLE § 6.6, p. 450 y ss.).

De entre estos sustantivos, algunos permiten la variación de género sin que esto afecte a su significado (*gobernador* y *gobernadora*), mientras que en otros esta variación conlleva implicaciones semánticas (*vidriero* y *vidriera*). Para unos y otros la elección del lema se ha llevado a cabo de manera distinta, lo que ha permitido establecer dos grandes grupos: cuando la variación no implica cambio de significado, se elige en lema de la forma en masculino; por el contrario, cuando la variación de género sirve para designar realidades diferentes, ambas variantes constituirán lemas independientes. Este planteamiento se entenderá mejor a partir de los siguientes ejemplos:

- (28) a. Vecino de la misma y *curador* nombrado (ODE – Almería, Tabernas, 1870).
b. [...] madre tutora y *curadora* de d[ic]hos herederos (ODE – Madrid, Estremera, 1836).
- (29) a. Vna *cazadora* de color azul marino (ODE – Madrid, Alcalá de Henares, 1890).
b. Un reloj de sobremesa, bronce dorado, con una figura representando un *cazador* (ODE – Madrid, 1877).
- (30) a. [...] como *contador* y partidador nombrado por mí (ODE – Almería, Fiñana, 1835).
b. Vn *contadorcito* de una quarta de alta, concha y marfil (ODE – Madrid, 1715).

En los ejemplos incluidos en (28) el lema ‘*curador*’ abarcará las formas *curador* y *curadora*, pues ambos conforman las variantes morfológicas de *curador* como la persona que tiene a su cargo a otros individuos. El término, perteneciente al campo del derecho, puede aludir al sexo femenino o masculino sin que esta alternancia en la flexión implique cambios en el significado léxico y, por tanto, tampoco en el lema.

No ocurre lo mismo con los ejemplos que se muestran posteriormente: el uso del femenino en (29) es una característica meramente gramatical del sustantivo que designa ‘cierto tipo de prenda’, al contrario de lo que sucede en (29), donde el género morfológico sí coincide con el sexo. La diferenciación entre *cazador* y *cazadora* es semántica, pues la alternancia conlleva una diferenciación de significados que se

reflejará también en la elección del lema. Ambas palabras constarán de lemas independientes, ‘cazadora’ en el primer caso, ‘cazador’ en el segundo.

Por el contrario, esta diferenciación de lemas no tendrá lugar en casos como los que se indican en (30), donde distintos valores semánticos coinciden en la misma forma morfológica de masculino o femenino. El término *contador* puede nombrar realidades distintas y hacer referencia a una persona (30a) o bien a un tipo de mueble que sirve como mesa o escritorio (30b). Si bien existe un cambio en el significado, este solo es identificable por el contexto de aparición, ya que el morfema de género permanece invariable. Para ambos casos se atribuye un único lema: ‘contador’.

De esta manera, la lematización de sustantivos con variación de género ha dependido de las implicaciones semánticas que conlleva cada una de las variantes. Siguiendo este criterio, estos sustantivos se dividen en dos grupos: 1) aquellos en los que la alternancia masculino/femenino no supone un cambio de significado y 2) aquellos en los que sí. Para cada uno de estos grupos la adjudicación del lema se ha desarrollado de forma diferente.

a. Variación de género sin repercusión semántica

Se incluyen en este grupo aquellos sustantivos que permiten variación de género cuando esta no comporta un cambio de significado.

En un gran número de ocasiones, la derivación se aplica a nombres que designan personas: *herrero/-a*, *hojalatero/-a*, *vidriero/-a*. La variación existente entre *cerrajero* y *cerrajera* únicamente aporta información relevante a la hora de diferenciar el sexo del referente cuando este es un hombre o una mujer, pero el concepto semántico al que aluden no cambia: “persona que hace o repara cerraduras” (*DLE* s.v. CERRAJERO, RA). En otras ocasiones, los adjetivos se sustantivan para formar nombres de instrumentos (*sujetador*, *azucarero*), algunos de los cuales pueden estar también sujetos a variación de género. Es el caso de *cucharero* y *cucharera*, entendidos como el ‘utensilio donde se colocan las cucharas’. Esta alternancia entre el femenino y masculino registrada en el corpus responde a una variación de tipo estilística o individual que tampoco hace que se vea comprometido su significado.

Tal y como se observa a partir de los ejemplos señalados hasta ahora, algunos sustantivos que permiten alternancia de género no se ven afectados por cambios en su significado, independientemente de que esa variación esté motivada por distintos aspectos: la referencia al sexo, un uso estilístico o dialectal en los casos mencionados. La no repercusión semántica ha sido condición *sine qua non* para que el lema de estas palabras sea siempre su forma en masculino singular. No obstante, conviene hacer una serie de precisiones que se han tenido en cuenta para su lematización, las cuales se detallan a continuación.

Ciertos sustantivos en el corpus presentan alternancia de género cuando aluden a un mismo concepto (*azucarero/a*, *chocolatero/a*, *huevero/a*):

(31) a. Un juego de porcelana [...] jarra para leche y *azucarera* (*ODE* – Madrid, 1877).

- b. Ít[em] un *azucarero* y pimentero de plata blanca (ODE – Almería, 1721).
- (32) a. Una *chocolatera* de azófar (ODE – Almería, Fiñana, 1882).
b. Quatro *chocolateros* de cobre (ODE – Madrid, 1715).
- (33) a. Una *huevera* también de porcelana (ODE – Madrid, Alcalá de Henares, 1890).
b. Doce *hueberos* (ODE – Madrid, 1840).

Aunque en la mayoría de los casos una de las variantes suele ser mayoritaria, estas palabras se documentan indistintamente tanto en su forma masculina como femenina sin que esto trascienda en el significado, que siempre alude al utensilio que contiene azúcar, chocolate y huevos, en este caso. En algunas ocasiones, como ocurre con los ejemplos anteriores, esta variación de género que se atestigua en el corpus queda registrada también en los diccionarios⁴¹.

Sin embargo, en otros casos, el corpus permite constatar una alternancia de género que no se contempla en las fuentes lexicográficas. Un ejemplo de ello es el que ofrece la variación *frutero/a* para el significado de ‘recipiente’. En los diccionarios académicos, el uso del femenino únicamente tiene cabida como alusión a la persona que vende fruta, si esta es una mujer, y no como el utensilio utilizado para servir la fruta (cfr. *NTLLE*). La documentación del corpus, en cambio, permite confirmar casos de vacilación con este significado (34). Otros ejemplos similares son los que aporta la vacilación *cobertera/-o* (‘pieza para tapar ollas’), *cucharera/-o* (‘utensilio para colocar las cucharas’), *tortera/-o* (‘cazuela plana’) o *papelera/-o* (‘mueble para guardar papeles’) registrada igualmente en el corpus, pero no en las fuentes lexicográficas académicas. Del mismo modo que ocurrió con *frutera*, se considera que *cobertero*, *cucharera*, *tortera* y *papelero* fueron variantes morfológicas que alternaron en algún momento de la historia con sus equivalentes en masculino o femenino sin que ello tuviera implicaciones semánticas. El corpus resulta especialmente valioso para estos casos en los que se permite testimoniar la existencia de la vacilación de estos términos:

- (34) a. Una media fuente obalada con su tapa. Una *frutera* (ODE – Madrid, 1840).
b. Diez *fruter*as de china, largas unas y cuadradas otras (ODE – Madrid, 1843).
c. Cuatro *fruter*os de ped[erna]l en ochenta r[eale]s (ODE – Almería, 1806).
d. Un *frutero* de loza inglesa pint[a]da de azul (ODE – Madrid, 1840).
- (35) a. Vna *cucharera* en quatro rr[eale]s (ODE – Madrid, San Lorenzo de El Escorial, 1789).
b. Un *cucharero*, *coberterero* y tabla de labar (ODE – Madrid, Aranjuez, 1870).
c. Quinze *torteras* de cobre desiguales (ODE – Madrid, 1703).
d. Mas dos *torteras* con sus *coberteras* (ODE – Cáceres, 1675).
e. Vn *tortero* en vn real (ODE – Almería, Zurgena, 1825).

⁴¹ Así se refleja en el *DLE* para los casos mencionados. El *DEA* recoge la variación *azucarero/azucarera*, aunque considera como rara la variante en femenino.

- (36) a. Se hallo en un *papelero* vn legajo pequeño (*ODE* – Badajoz, 1685).
b. Una *papelera*, forma de escritorio [sic], de nogal con tapa y dentro doce gabectas y su puertezuela (*ODE* – Madrid, 1711).
c. Una *papelera* con su mesita de nogal (*ODE* – Madrid, 1765).

Los ejemplos precedentes constituyen casos de variación atestiguados en *ODE*. No obstante, puede ocurrir que para otras palabras el corpus solo documente una de las variantes morfológicas que realmente tuvieron vigencia. La alternancia *panero/a* no se manifiesta en los inventarios, donde solo es posible encontrar la forma *panera* con el significado de ‘cesta en la que se coloca el pan’. Sin embargo, su equivalente *panero* se documenta en otras fuentes con el mismo significado. Tanto en documentación de corpus, en inventarios del *CorLexIn*: “un *panero* con su asa de yerro” (Zamora, Villalobos, 1654); como en fuentes lexicográficas tales como el *DEA*, con la consideración de raro (s.v. PANERO, -RA, acep. 5). Lo mismo sucede con la pareja *arrendadero/a*; *ODE* únicamente registra la variante en femenino, sin embargo, aparece siempre en masculino en las obras lexicográficas⁴².

Se incluirían también en este grupo aquellas palabras que aun designando el mismo utensilio aceptan variación de género en función de preferencias geográficas, como ocurre con muchos de los derivados en *-dor / -dora*: *computador/a*, *aspirador/a* (*NGLE* § 6.7q, p. 461). No se encuentran en cambio en el corpus muchos ejemplos de este tipo. Las palabras presentes en el corpus mencionadas anteriormente, que alternan para referirse a un mismo utensilio, no muestran diferencias geográficas significativas.

Una mención aparte merece el caso de *chocolatero* y *chocolatera*. El *DLE* considera actualmente el uso del masculino como una variante morfológica adscrita dialectalmente a Andalucía (s.v. CHOCOLATERO, RA, acep. 3). No reflejan lo mismo los datos obtenidos en *ODE*, donde el masculino se documenta también en inventarios madrileños. En inventarios de provincias andaluzas como Jaén, Cádiz, Sevilla o Almería se documenta la alternancia *chocolatero/-a*. Se deba o no a fenómenos dialectales, la existencia de esta alternancia sin repercusión semántica se ha reflejado con un único lema en masculino [*chocolatero*, *chocolatera* → lema: ‘chocolatero’].

Por último, se lematizan también en masculino las palabras que aluden a personas y que se mencionaron al inicio, como *cerrajero/a*, en las que la alternancia de género se corresponde con el sexo. Se documentan en el corpus numerosos sustantivos de este tipo, en su mayoría formados a partir de la adición de los sufijos *-ero/a* y *-dor/a* que presentan en el español una gran productividad para formar nombres de oficios u ocupaciones que designan personas (*NGLE* § 6.8i, p. 466). Se lematizan por tanto en su forma masculina *administrador*, *alpargatero*, *apreciador*, *arrendador*, *boticario*, *carpintero*, *cazador*, *cerrajero*, *comendador*, *comisario*, *comprador*, *contador*, *curador*,

⁴² *Arrendadero*: “Anillo de hierro con una armella que se clava en madera o en la pared, y sirve para atar las caballerías en los pesebres por las riendas o por el ramal de la cabezada” (*DLE* s.v. ARRENDADERO). Los cuatro casos de la voz en el corpus pertenecen a un inventario madrileño de 1711 y se documentan siempre en femenino plural.

gobernador, hojalatero, labrador, morador, partidor, poblador, procurador, recaudador, tasador, testador, trabajador.

En suma, los casos de alternancia de género en los ejemplos tratados se consideran variantes morfológicas (no léxicas) en tanto a que el uso del masculino o el femenino no conlleva cambios semánticos. Todos ellos se lematizan bajo el lema en masculino singular.

b. Variación de género con trascendencia semántica

Distinto a los casos mencionados arriba es el constituido por aquellos en los que la variación morfológica ha comportado un cambio en el significado.

Muchos de los sustantivos derivados, además de denominar personas, han pasado a designar nuevas realidades. Por ejemplo, *vidriero/-a* designa al ‘hombre o la mujer que trabaja el vidrio’, pero *vidriera* ha adoptado también el significado de ‘ventana o escaparate de vidrio’, por lo que se trata de un concepto nuevo para el que se emplea únicamente la forma en femenino. De esta manera, para uno y otro caso el lema será distinto: [*vidriero/-a* ‘persona’ → lema: ‘vidriero’] [*vidriera* ‘ventana’ → lema: ‘vidriera’].

- (37) a. [...] maestro *vidriero* y ojalatero que fue del d[ic]ho Aranjuez (*ODE* – Madrid, Aranjuez, 1870).
b. Diez y ocho postigos de *vidrieras* grandes y pequeños (*ODE* – Madrid, 1715).
c. Otro relicario de platta con *vidriera* (*ODE* – Almería, Cuevas de Almanzora, 1766).
d. Seis cortinillas de *bidriera* de muselina estampada (*ODE* – Madrid, 1840).

Lo mismo puede aplicarse a otras muchas palabras que podrían designar tanto a la persona como al objeto: *aceitero/aceitera, quesero/quesera, lechero/lechera*. En todos estos casos, al lema en masculino se adscriben los sustantivos referidos a ‘persona’ o los adjetivos del sustantivo (i.e. “dos mantas *lecheras* (adj.) → lema: ‘lechero’) mientras que a las formas en femenino les corresponde el valor semántico ‘recipiente’ (i.e. “una *lechera* con llave (sust.) → lema: ‘lechera’).

Los casos de arriba presentan lemas distintos, en masculino y femenino, puesto que designan realidades diferentes: la variante en femenino se suele aplicar al concepto de ‘recipiente’. Sin embargo, en otros ejemplos, la denominación de dos conceptos diferentes coincide en la forma. Es el caso de los sustantivos polisémicos como *azucarero* o *costurero*, que pueden designar tanto a la ‘persona’ como al ‘recipiente’ (el cual puede también presentar alternancia de género), pero ambos se incluirán bajo el mismo lema con la forma en masculino singular. En estos casos se debe asumir que la lematización contempla las variantes flexionadas, pero no las variantes polisémicas cuando estas coinciden formalmente: *costurero* ‘persona’ y *costurero* ‘objeto de costura’ o *azucarero* ‘productor de azúcar’ y *azucarero* ‘recipiente’:

- (38) a. Un *azucarero* y un pimentero de plata (*ODE* – Almería, 1721).
 b. Un juego de café de china con [...] *azucarero* (*ODE* – Madrid, 1843).
 c. Un *azucarero* de loza (*ODE* – Madrid, 1877).
 d. Vna *azucarera* labrada (*ODE* – Cádiz, Puerto Real, 1780).
- (39) a. Josefa de Plaza Garcia, de cuarenta y seis años de edad, viuda de Antonio de Oña Oña, *costurera* (*ODE* – Almería, Tabernas, 1868).
 b. Un *costurero* redondo de caoba (*ODE* – Madrid, 1843).

En otros derivados la alternancia de género no sirve para distinguir entre personas y objetos, sino para aludir a objetos diferentes, los cuales se registran en el diccionario en entradas diferentes. Estos casos aceptan la alternancia morfológica de género para indicar diferencias en la forma, tamaño o función del utensilio designado: *cabecero* / *cabecera*, *gorro* / *gorra*, *mazo* / *maza*, *sementero* / *sementera*, *talego* / *talega*, *perol* / *perola*. Estos sustantivos se han empleado siempre bajo una única forma en femenino o masculino y el cambio de una u otra conlleva modificaciones semánticas, por lo que cada una de las palabras integrantes de estas parejas constituye un lema propio.

Se da un último caso de sustantivos que designan objetos o utensilios para los que no se manifiesta la alternancia de género y tampoco hay constancia de que se haya producido esta variación en algún momento de la historia del español. La designación de los objetos a los que aluden no admite una alternancia de género para nombrar la misma realidad y solo se documenta el uso de una de las variantes, femenina o masculina, según el caso: *espetera*, y no **espetero*; *tocador* y no **tocadora*; *tapadera* y no **tapadero*. La lematización de estos casos se ha hecho en base a su género gramatical [*espetera* → lema: ‘*espetera*’, *tocador* → lema: ‘*tocador*’].

Como recapitulación de todo lo expuesto, cabe destacar que la lematización de sustantivos con alternancia de género se ha llevado a cabo en masculino o femenino en base a las implicaciones semánticas de las formas designadas. En aquellos en los que la variación no comporta un cambio de significado, se lematiza en masculino o femenino, según la forma registrada en el diccionario (*DLE*) y mayoritaria en el corpus.

Se lematizan en masculino singular los siguientes casos:

- a. Los sustantivos o adjetivos en los que la variación -o/-a alude al género de la palabra: *costurero/-a*, *lechero/-a*, *vidriero/-a*, *herrero/-a*.
- b. Los sustantivos masculinos en -dor que designan instrumentos o lugares: *abridor*, *asador*, *bastidor*, *calentador*, *colador*, *cenador*, *cobertor*, *comedor*, *mostrador*, *pasador*, *peinador*, *rascador*, *tajador*, *tenedor*, *tocador*, *velador*.
- c. Los sustantivos masculinos derivados en -dero y -ero que designan recipientes o utensilios: *azucarero*, *costurero*, *frutero*, *platero*, *repostero*, *zafero*, así como las variantes minoritarias en femenino sin trascendencia semántica de algunos de ellos (*azucarera*, *frutera*).

- d. Los derivados mediante la adición de sufijos aumentativos ya lexicalizados: *almohadón, cambrayón, escopetazo, tazón*.

Por el contrario, se lematizan en femenino singular:

- a. Los sustantivos femeninos que designan seres inanimados en los que el género responde a una propiedad gramatical inherente de la palabra sin conexión con el sexo (NGLE § 2.3c, p. 90): *cazadora, plata, sábana, almohada, puerta*... Entre estos se encuentran muchos sustantivos derivados en *-dera* y *-era* que designan recipientes o utensilios: *abrazadera, aceitera, cafetera, cantarera, compotera, cuajadera, ensaladera, escupidera, espetera, espumadera, fiambarrera, fosforera, lechera, papelera, quesera, regadera, rinconera, salsera, salvadera, tapadera, tartera, tembladera, vidriera, vinagrera*.

4.6.2.3. Plurales inherentes

Se incluyen en este grupo los sustantivos que solo admiten su forma en plural, conocidos como *pluralia tantum* o plurales inherentes (i.e. *adentros, fauces, expensas*) (NGLE § 3.8f, p. 171), aquellos que prefieren el uso en plural, si bien la forma en singular también se admite (*tijera / tijeras*) y los nombres de objetos dobles en los que el plural no es informativo, pues designan a un solo objeto (*alicates, gafas*) (NGLE § 3.8r, p. 177). La lematización de estos casos se ha llevado a cabo en la forma singular o plural dependiendo de algunas de las características que se explican a continuación.

Determinados plurales inherentes no admiten el uso del singular en ninguno de sus posibles contextos de aparición. En el corpus, algunos ejemplos son *enseres, nupcias, extramuros* o *andas*. El fenómeno afecta también a otros compuestos como *contramangas, limpiadientes, cortaplumas, portavelas, portaviandas* o *portavinagreras*, cuyas variantes en singular no se documentan en el corpus. Este tipo de palabras se han lematizado con la forma invariable de plural, puesto que su uso se registra únicamente bajo esta forma. Por el contrario, otras como *portamonedas* o *guardapiés* (véase más abajo el estudio concreto de esta palabra) sí admiten su uso también en singular, el cual es posible documentar en algún momento de la historia del español, en *ODE* o en otros corpus complementarios, aunque la forma en plural sea la mayoritariamente aceptada y la que se integra en los diccionarios. Independientemente de su empleo en singular o en plural, este no parece haber afectado a su significado en ninguno de los casos, por lo que se lematiza en singular:

- (40) a. Vn *portamoneda* de rejilla de plata (*ODE* – Madrid, Alcalá, 1890).
b. Cogió la prudente señora el *portamonedas* que estaba aún bien repleto y se lo guardó (*CDH* – B. Pérez Galdós, *Fortunata y Jacinta*, 1887).

Otras palabras admiten la variación de número y se documenta la misma forma tanto en singular como en plural. En estos casos, la elección del lema en singular o plural se ha

visto supeditada a las implicaciones semánticas que pueda tener esta alternancia, como más adelante se explica para los términos *angarilla* / *angarillas* o *tijera* / *tijeras*.

Otro grupo de plurales inherentes es el asociado a los nombres de objetos dobles constituidos por dos partes (NGLE § 3.8r, p. 177). Suelen integrar este grupo aquellas palabras pertenecientes al campo semántico de la vestimenta como *calzas*, *calzones*, *calzoncillos*, *medias*, *pantalones*, *guardapiés*; pero también un gran número de objetos registrados en los inventarios y compuestos igualmente por varias piezas o partes: *lentes*, *anteojos*, *orejeras*, *angarillas*, *jamugas*, *tenazas*, *tijeras*. El DLE suele lematizar con su forma en singular las palabras pertenecientes a este grupo e indicar como observación junto a la definición correspondiente la marca gramatical “usado más en plural”. Estas palabras se caracterizan por permitir la elección en singular y plural en los mismos contextos de aparición sin mutar el significado, por lo que se lematizan en singular. El caso concreto de *angarillas*, *jamugas* y *tijeras* se analiza con mayor profundidad más abajo.

No forman parte del grupo de objetos dobles los que se denominan *duales léxicos*. A diferencia de los anteriores, estos objetos “se agrupan en pares formados por piezas o elementos independientes” (NGLE § 3.8u p.178). Aunque el uso más extendido es el plural, es posible el empleo del singular para aludir a cada uno de los componentes de manera individual sin que esto suponga cambios semánticos. Estas palabras se lematizan por lo tanto en singular, a pesar de que en el corpus siempre o casi siempre se documentan en plural. Estas son: *calcetines*, *zapatos*, *borceguíes*, *pantuflos*, *pendientes*, *guantes*.

Por último, existe otro grupo de palabras que no admiten variaciones de número sin una alteración en el significado. Se localizan en el corpus los vocablos *aguaderas*, *grillos* y *trébedes*. Estas palabras aceptan solo el uso del plural con el significado al que se alude en los inventarios en los ejemplos de abajo: (41) “armazón de madera, esparto, mimbre u otra materia semejante, con divisiones, que se coloca sobre las caballerías para llevar en cántaros o barriles agua u otras cosas” (s.v. AGUADERA acep. 8); (42) “conjunto de dos grilletes con un perno común, que se colocaban en los pies de los presos para impedirles andar” (DLE s.v. GRILLOS); y (43) “aro o triángulo de hierro con tres pies, que sirve para poner al fuego sartenes, peroles, etc.” (s.v. TRÉBEDE acep. 2):

- (41) a. Unas *aguaderas* grandes de minbre, viejas, quitadas las asas (ODE – Cáceres, 1621).
b. Unas *aguaderas* de cuatro cantaros (ODE – Almería, Cuevas de Almanzora, 1704).
c. Un borrico con su albarda, *aguaderas* y seron (ODE – Cádiz, Villamartín, 1748).
- (42) a. [...] lo puso en vn calabozo y en vn cepo con dos pares de *grillos* (ODE – Granada, 1615).
b. Vnos *grillos* viejos (ODE – Madrid, Navalafuente, 1809).
- (43) a. Unas *trebedes* de yerro grandes (ODE – Madrid, 1711).
b. Unas *treuedes* grandes redondas (ODE – Almería, Serón, 1754).

El empleo de estas formas en singular adquiriría nuevos significados distintos a los que se atestiguan en el corpus, recogidos igualmente en el diccionario en acepciones diferentes. Para estos casos en los que la variación singular / plural comporta un cambio en el significado, la lematización se produce separadamente⁴³. Así, las palabras mencionadas se lematizan en plural.

En definitiva, la elección del lema para los plurales inherentes se ha llevado a cabo siguiendo un criterio similar al ya mencionado para los derivados nominales tratados en el apartado anterior, esto es, basado en torno al cambio o no del significado. Así, se lematizan en singular aquellas palabras que, indistintamente de la variante empleada, no cambian su significado, como *portamoneda(s)*, *lente(s)* o *calzoncillo(s)*, con independencia de que la forma en plural sea la más usada. Por otra parte, la lematización se ha realizado en plural en los siguientes dos casos:

- a. Cuando la palabra únicamente permite este uso en español, sea cual sea su significado: *enserres* o *viveres*.
- b. Cuando el uso del plural implica un cambio en el significado con respecto al uso en singular, como en *aguaderas*, *trébedes* o *grillos*.

Este planteamiento es el que se ha seguido a la hora de lematizar algunas de las palabras que se analizan a continuación y que han presentado algún tipo de duda: *angarillas*, *jamugas*, *guardapiés* y *tijeras*. Se ha tratado de estudiar si la alternancia singular o plural documentada en el corpus adquiere matices semánticos distintos con el fin de proceder a su correcta lematización en base a los criterios expresados.

Angarilla(s)

En *ODE* el término aparece registrado con distintos significados. El más habitual es el que en los inventarios de bienes se relaciona con el de ‘andas para transportar materiales de distinto tipo’ (*DLE* s.v. ANGARILLA, acep. 5). El objeto podía servir para trasladar alimentos, líquidos, materiales o cualquier otro bulto, ya sea cargado por la fuerza humana de una o varias personas o bien a lomos de un animal. En varias ocasiones se relaciona con el trabajo de campo. Aunque señalado en plural en el diccionario, el corpus documenta ocurrencias tanto de la forma en singular como en plural:

- (44) a. Unos palos en forma de *angarillas* para sacar mies (*ODE* – Almería, Cuevas de Almanzora, 1704).

⁴³ En singular, el significado de *aguadera* se identifica como sinónimo de ‘abrevadero’ (*DLE* s.v. acep. 2). El de *trébede* como el de “habitación o parte de ella que, a modo de hipocausto, se calienta con paja” (*DLE* s.v. acep. 1). Por último, *grillo* hace referencia al insecto.

- b. Una *angarilla* para sacar basura en cinco reales (*ODE* – Burgos, 1706).
- c. Dos borricos con sus aparejos y *angarillas* para llevar piedra (*ODE* – Cádiz, Puerto Real, 1733).
- d. *Angarilla* y torba dies reales (*ODE* – Huelva, 1791).
- e. Una *angarilla* de llevar cristales (*ODE* – Madrid, Aranjuez, 1870).

Otro uso del término es el equivalente al de ‘camilla para transportar personas’. *ODE* recoge este sentido en una declaración de testigo y en su posterior certificado médico. En ambos documentos el término se emplea para aludir al objeto en el que fue transportado el cadáver, en este caso, de una mujer hallada muerta. Aunque el diccionario académico actual lo recoge con su uso en singular con el significado de ‘camilla’, en *ODE* el empleo del término se realiza en plural en las tres ocasiones:

- (45) a. dio orden [...] para que le diesen sepultura. Y, abriendoselo llebado en un lecho de *angarillas* que tienen para d[ic]ho efecto (*ODE* – Granada, 1699).
- b. [...] donde se alló en las *angarillas* de la hermandad de la Car[ida]d el querpo difunto de la muxer contenida en esta causa, el qual miraron y reconozieron los d[ic]hos dos zirujanos. Y auiéndolo hecho [...] dijeron an uisto, mirado y reconozido el querpo difunto que está en d[ic]has *angarillas* (*ODE* – Granada, 1699).

Por último, el corpus documenta en una ocasión el uso de *angarillas* como el equivalente al de *jamugas* (véase más abajo), esto es, el de ‘silla de tijera para montar a mujeriegas’ (*DLE* s.v. acep. 9). Prueba de esta sinonimia son los siguientes ejemplos localizados, uno de ellos en *ODE*. De nuevo, aunque el *DLE* lo recoja bajo la forma plural, la documentación de corpus atestigua también su uso en singular para este sentido:

- (46) Una *angarilla* para yr fuera, digo dos, la una ouo la dicha doña Jeronima de su tia doña Fran[cis]ca (*ODE* – Cáceres, 1621).
- (47) Dos *angarillas* de mugeres uiejas (Miguel Borge, 2016, p. 354).

En definitiva, a pesar de la información que proporciona el diccionario, el corpus documenta la alternancia del término en lo relativo al número en cualquiera de los significados con los que se le asocia. No obstante, la forma plural es la que se utiliza como lema en la mayoría de las fuentes lexicográficas. Ocurre así en el *DEA*, en los diccionarios académicos –hasta la edición de 2001, cuando el lema pasó a ser *angarilla*– y en una larga lista de diccionarios recogidos en el *NTLLE*. En *Autoridades*, de manera particular, se añade en la definición “usado siempre en plural” (s.v. ANGARILLAS).

Cabe destacar que en los corpus consultados el uso del plural está mucho más extendido que el empleo de la forma en singular, independientemente del significado. Así se manifiesta en el *CorLexIn*, el *Corpus del Español* de Mark Davies, el *CORDIAM* o el *CDH*. No obstante, aunque el uso del término en plural goza de mayor

representación tanto en repertorios lexicográficos como en los corpus trabajados, el uso del singular *angarilla* no deja de estar presente en dichas fuentes, por lo que para *ODE* se ha resuelto su lematización en singular como la forma base que englobaría al mismo tiempo también aquellos usos en plural, aunque estos conformen mayoría. De este modo, se ha querido cumplir con el criterio previamente establecido según el cual, si una palabra se encuentra representada en el corpus en sus formas plural y singular, se lematizará siempre bajo la forma en singular:

<i>angarilla</i>	lema: <i>angarilla</i>
<i>angarillas</i>	lema: <i>angarilla</i>

Jamuga(s)

Un caso similar y estrechamente vinculado con el anterior es el que ofrece la variación *jamuga / jamugas*. El término se refiere a una silla de tijera con respaldo que se coloca sobre la caballería y que es utilizado generalmente para el transporte de las mujeres. Así se explicita en los siguientes ejemplos extraídos del corpus:

- (48) a. Unas *amugas* para viaje (*ODE* – Almería, Fiñana, 1882).
 b. Ytt[em], las *jamuguillas* de madera para cavallería (*ODE* – Almería, 1750).
 c. Un par de *samucas* y unas sogas en cinco r[eale]s (*ODE* – Madrid, Algete, 1834).

El *DLE* establece una relación de sinonimia entre *jamugas* y *angarillas*, como ya se ha indicado más arriba a raíz del análisis de esta última. Los ejemplos que allí se señalaron avalan la estrecha relación semántica que vincula a ambos términos. *Jamugas* aparece en nuestro corpus en más ocasiones, aunque la falta de especificación nos impide conocer si su uso en estos casos se relaciona con algún otro significado distinto al mencionado. Por el contrario, en otros dos ejemplos se evidencia sin lugar a duda otro significado no contemplado en los diccionarios y que relaciona el término con un objeto empleado para las tareas de labranza, tal y como se muestra a continuación:

- (49) a. Unas *jamuguillas* para acarrear en un r[ea]l (*ODE* – Jaén, Villacarrillo, 1700).
 b. Dos pares *amuguillas* para barcinar⁴⁴ en cincuenta céntimos (*ODE* – Almería, Fiñana, 1882).

Estos ejemplos⁴⁵ atestiguarían la sinonimia *jamugas-angarilla* también en el sentido de ‘utensilio usado como transporte’ al mismo tiempo que suponen un valioso ejemplo de

⁴⁴ *Barcinar*: “coger las gavillas de mies, echarlas en el carro y conducirlas a la era” (*DLE*, s.v. *BARCINAR*). La voz se marca como un dialectalismo propio de Andalucía.

⁴⁵ En estos ejemplos y en los anteriores, el uso del diminutivo aplicado a *jamugas* no se considera lexicalizado, por lo que se ha incluido en el recuento de diminutivos transparentes analizados estadísticamente en esta tesis.

variación léxica en esta época. El vínculo entre ambas palabras se manifiesta en una relación sinonímica de ida y vuelta en la que ambos términos “se prestan” sus significados: *angarilla* equivale a *jamuga* como ‘silla de caballería’ y *angarilla* equivale a *jamuga* como ‘objeto de transporte’.

Con respecto a la variación entre el uso en singular y en plural, el *DLE* lematiza el término en singular, aunque señala su uso mayoritario en plural (s.v. JAMUGA), que también confirman los corpus consultados. El empleo del singular se reduce a casos aislados. Solo se ha localizado un ejemplo en *CorLexIn*: “Vna *jamuga* buena, con quatro mançanillas de açófar” (Toledo, Illescas, 1626), siendo *jamugas* la forma más difundida. En efecto, en *ODE* solo se encuentran ocurrencias de la forma en plural, por lo que su uso en singular podría haberse extinguido previamente. No obstante, puesto que esta variación de número se dio en algún punto de la historia, se opta por su lematización en singular. Adicionalmente, dadas las características de esta palabra como sustantivo perteneciente a los nombres de objetos dobles, el uso en singular tampoco alteraría el significado, por lo que también en este aspecto se ve justificada la elección del lema:

<i>jamugas</i>	lema: <i>jamuga</i>

Guardapié(s)

El uso en plural es el más extendido en el mundo hispánico y como tal aparece lematizada esta forma en el diccionario académico actual. *Guardapié* no parece tener cabida en los repertorios lexicográficos (no se encuentra en el *NTLLE* ni en las papeletas del *Fichero General*). Sin embargo, la forma singular sí se encuentra documentada en corpus como el *CorLexIn* (9 ocurrencias), el *CORDE* (1 caso) o en *ODE* (47 ocurrencias), del que se muestran a continuación algunos casos:

- (50) a. Un *guardapié* de nobleza, color de coral (*ODE* – Granada, 1727).
- b. Vn *guardapié* de senpiterna verde (*ODE* – Almería, Serón, 1754).
- c. Ytt[em] el *guardapié* de nobleza rrosada (*ODE* – Almería, Dalías, 1750).

El uso en singular, aunque minoritario, alcanza representación en documentos andaluces del siglo XVIII (Jaen, Granada, Cádiz o Almería). Aunque en el *CdE* el hallazgo más reciente data de 1850, *ODE* lo registra en 47 ocasiones de un total de 220 casos. No obstante, si bien *guardapiés* es la forma con mayor extensión por excelencia, es posible localizar *guardapié* en corpus diacrónicos sin que esto trascienda en el significado. Por esta razón se lematiza bajo la forma en singular, la cual englobaría también los usos en plural:

<i>guardapié</i>	lema: <i>guardapié</i>
<i>guardapiés</i>	lema: <i>guardapié</i>

Tijera(s)

El empleo del término se registra en los inventarios tanto en su forma en plural (51) como en singular (52):

- (51) a. Unas *tigeras* en medio real (*ODE* – Almería, Bcares, 1774).
b. Unas *tixeras* pequeñas con bayna (*ODE* – Almería, Cuevas de Almanzora, 1766).
c. Una almuadilla de coser con sus *tigeras* (*ODE* – Madrid, San Lorenzo de El Escorial, 1789).
d. Un estuche con cortaplumas, *tijeras* y otras frioleras de oro (*ODE* – Madrid, 1843).
e. Un par de *tijeras* de costura en un escudo (*ODE* – Madrid, Aranjuez, 1870).
- (52) a. Una mesa para comer de *tigera*, madera de pino (*ODE* – Madrid, 1832).
b. Dos escaleras de mano [...]. Una ídem de *tejera* [sic] (*ODE* – Madrid, Aranjuez, 1870).
c. Un catre de *tigera* con su tela (*ODE* – Madrid, 1877).
d. Un catre de los llamados de *tijera* (*ODE* – Almería, Canjáyar, 1896).

Aunque lematizado en singular, el *DLE* advierte de que el uso en plural está más extendido cuando se hace referencia al ‘objeto común compuesto de dos hojas y utilizado para cortar’. Este significado se corresponde con la primera acepción de esta entrada en el diccionario. Por otro lado, en la segunda acepción se alude con el término en singular a “cosa compuesta, como la tijera, de dos piezas cruzadas que giran alrededor de un eje” (*DLE* s.v. TIJERA, acep. 2). En los corpus, el empleo del plural queda ligado mayoritariamente a aquellos usos relacionados con la primera acepción, mientras que el empleo del singular se vincula con los contextos contemplados en la segunda definición. Esta distribución del singular y plural se puede observar claramente en los ejemplos mostrados arriba, donde el término en plural *tijeras* cita expresamente el objeto en sí (51) y el singular *tijera* queda restringido a un uso por analogía de dicho objeto (52). Por lo tanto, la variación de número sí adquiere en este caso matices semánticos diferentes.

Aunque menos extendido, el empleo del singular sin cambios semánticos también se documenta en los corpus del español:

vidrio prescian mucho, y esas cosas de Flandes, así como cuchillos, y espejos, y	tijeras	, y cosas de marfil, y cuentas, y corales.
Sedeño le dió de aquellas cosas que a los indios son gratas, así como cuchillos e	tijeras	, e hachas para cortar árboles, y les hizo buenas obras.
algodón picado (e primero hilado), que lo hacen quedar como la borra que dejan las	tijeras	de los tundidores, y era de cuantas colores puede haber, e aquéllas muy finas.
en el año que nunca las cubre pelo. Tienes tres trasquiladores, cada qual con su	tijera	, y dejas tales los cuerosque el ganado desespera. Y después que as trasquilado
van, y si les ladra algún can arrójasle tu cayado. Bastara que trasquilaras con tu	tijera	la vieja: cada año de cada ovejaun velloçino sacarás. Que lana te sobraríay el
tiranás, que sin jamás cansaros movéis las manos negras y ligeras! Maldigan las	tijeras	de riguroso filo, que de valor tan raro, de todo el mundo amparo, cortaron tan
troncos cada uno con un anillo al cabo, así que si como las partes por do toman las	tijeras	y de la parte contraria del fiel ay dos medias caxas cada una asida con el tronco
Vnas	tijeras	doradas, dies marauedis... x
E, asimesmo, mandamos que las personas que vendieren las lanas de	tijera	y añinos, o menudos de lanas lavadas, que las laven y vendan enxutas y según y
puntas del cabo de la muestra, en cantidad de media quarta de cada cabo, cortando con	tijeras	, porque parezca que fue desorejado. Y que, demás d'esto, executen las penas en
sellen, quiten a los tales paños las puntas del cabo de la muestra cortándolas con	tijeras	de cada cabo media quarta, de manera que el dicho paño quede desorejado. E, esto
podemos algo apañar. Nuestra fiesta es el interesse y la ganancia. Todos traemos	tijeras	que sotilmente puedan cortar, porque ninguno ay que no aprenda artes muy agudas
coles y cercenauanlas con vna	tijera	sin tocar en la buba: ui tan poco que lleuasse nada de
pueden trabajar curanse cortando los callos con vna	tijera	poco apoco y mundificado las con el poluo del
Otrosi atandolas con vn hilo de seda se caen. y la que se pudiere cortar con la	tijera	que tuuiere el pie delgado bien se puede hazer: y hechar encima de la cortadura
suceden deste muchas passiones y fluxos de sangre. E muchos determinan con vna	tijera	de abrir el prepucio de luengo aluengo de vna tigerada
las biiij Primera viceras que le haze abrir el capullo con la	tijera	se lo haze despues acabar decortar. E por estas el abrir delcapullo con la tigera
tigera se lo haze despues acabar de cortar. E por estas el abrir del capullo con la	tijera	segun es dicho digoque es vna obra muy falsa y de muchotrabajo al que la sufre
fuertemente; & con vna	tijera	sabiamentesea auuerto a la larga por el ?? ladoespecie
que son mas para abrir fasta que traygas el solapo a tanta delgadeza que con la	tijera	lopuedas cortar sin dr ningunapassional paciente & seras

Figura 37: Resultado de la consulta en el *CDH* del lema *tijera*. Ejemplos del s. XVI

Puesto que ambas variantes *tijera* / *tijeras* se emplean indistintamente con el mismo significado, estas se han lematizado en el corpus bajo la forma en singular, la cual englobaría el resto de las ocurrencias de la palabra también en plural:

<i>tijeras</i>	lema: tijera
<i>tijera</i>	lema: tijera

4.6.2.4. Adjetivos

Se agrupan bajo su forma en masculino singular:

- a. Las formas adjetivales con flexión de género y número:

<i>francés</i>	lema: francés
<i>sucias</i>	lema: sucio
<i>redondos</i>	lema: redondo

- b. Los adjetivos con afijación afectiva:

<i>pequeñicos</i>	lema: pequeño
<i>viejecilla</i>	lema: viejo
<i>picarón</i>	lema: pícaro

- c. Los adjetivos apocopados derivados de *bueno*, *malo*, *santo* y *grande* y los numerales ordinales derivados de *primero* y *tercero*. Nótese que estas formas se lematizan bajo su forma plena:

<i>gran</i>	lema: grande
<i>san</i>	lema: santo
<i>buen</i>	lema: bueno

<i>primer</i>	lema: primero
<i>tercer</i>	lema: tercero

- d. Los superlativos absolutos formados con el sufijo *-ísimo* y que en el corpus suelen corresponder a tratamientos nominales. Estas palabras se lematizarán bajo su forma base sin el afijo:

<i>santísimo</i>	lema: santo
<i>ilustrísimo</i>	lema: ilustre
<i>excelentísimo</i>	lema: excelente

Se considera una excepción el apelativo antonomástico *Purísima*, el cual funciona como un nombre propio: “Una *Purísima* con marco y cristal” (*ODE* – Almería, Cuevas de Almanzora, 1844). Su lematización por tanto se ajustará a los criterios correspondientes a los nombres propios:

<i>Purísima</i>	lema: purísima

- e. Las formas comparativas de los adjetivos *bueno*, *malo*, *grande* y *pequeño*, las cuales se agrupan bajo su lema propio en singular:

<i>mejor</i>	lema: mejor
<i>peor</i>	lema: peor
<i>mayores</i>	lema: mayor
<i>menores</i>	lema: menor

Difieren del criterio establecido por la Real Academia los casos mencionados en c. y d. Las formas apocopadas de los adjetivos que se han señalado constituyen un lema propio en el diccionario académico actual y lo mismo cabe decir de las formas superlativas, también lematizadas de manera independiente.

4.6.2.5. Determinantes

Del mismo modo que para los sustantivos y adjetivos, el lema de los determinantes está constituido por el masculino singular de aquellas formas variables en género y número, o por la forma en singular si solo poseen variación en cuanto al número:

<i>otro</i>	lema: otro
<i>nuestro</i>	lema: nuestro
<i>todos</i>	lema: todo
<i>varios</i>	lema: varios
<i>bastantes</i>	lema: bastantes

Las formas apocopadas de algunos determinantes indefinidos se lematizan bajo su forma extendida siguiendo el mismo procedimiento de lematización que el anteriormente mencionado para el caso de los adjetivos apocopados *primer*, *gran*, etc. (Criterio c. de los adjetivos. Véase arriba):

<i>algún</i>	lema: alguno
<i>ningún</i>	lema: ninguno
<i>un</i>	lema: uno

4.6.2.6. Pronombres

Los pronombres personales se lematizan bajo su forma propia independientemente de la persona a la que hagan referencia:

<i>yo</i>	lema: yo
<i>me</i>	lema: me
<i>mí</i>	lema: mí
<i>nosotros</i>	lema: nosotros
<i>nos</i>	lema: nos

Este criterio difiere del establecido por el grupo EAGLES, según el cual todos los pronombres personales se lematizan bajo las formas en singular *yo*, *tú* y *él* de acuerdo con los rasgos gramaticales de persona que presentan.

El resto de los pronombres de otras clases, demostrativos, posesivos, indefinidos, interrogativos y relativos, siguen los mismos criterios que los ya mencionados para el caso de los determinantes:

<i>cuántas</i>	lema: cuánto
<i>demás</i>	lema: demás
<i>nadie</i>	lema: nadie

4.6.2.7. Verbos

Todas las formas verbales se han lematizado bajo el infinitivo del verbo. De este modo, las formas *tienen*, *tendría*, *teniendo* y *tuviera* se agrupan bajo la forma *tener*.

En el corpus, cada token constituye una sola forma ortográfica (§ 4.2. *Tokenización*), por lo que la anotación lingüística se realiza por palabras gráficas y no por unidades sintácticas. Esto afecta a los tiempos compuestos formados con *haber*, a las construcciones pasivas con *ser* o a las formas perifrásticas, que se lematizan individualmente y no en bloque. Así, todas estas formas verbales se lematizarán como palabras independientes bajo el infinitivo de los verbos oportunos:

<i>desposarán</i>	lema: desposar
<i>valoradas</i>	lema: valorar
<i>habiendo hallado</i>	lema: haber + hallar
<i>fuleron tasados</i>	lema: ser + tasar
<i>había de otorgar</i>	lema: haber + de + otorgar

4.6.2.8. Adverbios

Los adverbios se lematizan bajo su forma invariable. También aquellos derivados con el sufijo *-mente* toman como lema esta forma y no la del adjetivo del que derivan:

<i>también</i>	lema: también
<i>primeramente</i>	lema: primeramente
<i>primero</i>	lema: primero

En lo que respecta a las formas apocopadas de ciertos adverbios, se han aplicado los siguientes criterios:

- a. La forma *tan* usada ante adjetivos o adverbios en las construcciones comparativas del tipo *tan ... como / cuanto / que ...* se lematizan con la forma extendida, de acuerdo con el criterio establecido para las formas apocopadas que permiten ciertas categorías de palabras (adjetivos: *gran, san*; determinantes: *algún, ningún*):

<i>tan</i>	lema: tanto
<i>tanto</i>	lema: tanto

- b. Los adverbios coordinados en *-mente* se lematizan ambos bajo la forma extendida:

<i>real y efectivamente</i>	lema: realmente + efectivamente
<i>clara y abiertamente</i>	lema: claramente + abiertamente

- c. Se lematizan como formas independientes los cuantificativos indefinidos *mucho* y *muy*, a pesar de que esta última sea la forma reducida de *mucho*. De esta manera, la lematización de esta pareja de adverbios es como sigue:

<i>muy</i>	lema: muy
<i>mucho</i>	lema: mucho

La anotación individual de las formas ortográficas en el corpus afecta en gran medida a las locuciones adverbiales en tanto que constituyen “expresiones fijas constituidas por varias palabras que equivalen a un adverbio” (NGLE § 30.15a, p. 2374). Estos grupos

de palabras que conforman una única unidad en el ámbito de la sintaxis (*al final, de prisa, a gusto*) no constituyen, en cambio, grupos fijos en el corpus, donde cada forma ortográfica constituye una unidad independiente por sí misma. En el caso de las locuciones, cada palabra integrante de la unidad se lematiza también de manera autónoma con su forma propia. El lema seguirá las pautas que le corresponden según la categorización morfológica de cada palabra de la expresión fija:

<i>de veras</i>	lema: de + veras
<i>a sabiendas</i>	lema: a + sabiendas
<i>de oídas</i>	lema: de + oír
<i>a medias</i>	lema: a + medio
<i>a gatas</i>	lema: a + gato

4.6.2.9. Numerales

El lema de los numerales ordinales será siempre la palabra en masculino singular:

<i>segunda</i>	lema: segundo
<i>primeras</i>	lema: primero

Por el contrario, el lema de los numerales cardinales se representa con las cifras en numeración arábiga, independientemente de que en el manuscrito original el número aparezca escrito con palabras. Al mismo tiempo, los números romanos se han lematizado con el equivalente a los dígitos arábigos:

<i>setecientas</i>	lema: 700
<i>10</i>	lema: 10
<i>VIII</i>	lema: 8

4.6.2.10. Palabras invariables

Las palabras invariables conformadas por las preposiciones, conjunciones o interjecciones se lematizan bajo su única forma invariable:

<i>con</i>	lema: con
<i>para</i>	lema: para
<i>pero</i>	lema: pero
<i>si</i>	lema: si
<i>ah</i>	lema: ah

4.6.2.11. Palabras con sufijación apreciativa

Debido a las características intrínsecas de esta tesis doctoral se ha considerado oportuno realizar una mención específica al tratamiento de aquellas palabras con sufijos de valor afectivo en cuanto a su lematización.

La consideración de las palabras con sufijación apreciativa (particularmente la de diminutivos), ha supuesto en algunos casos un complejo ejercicio de reflexión a la hora de dictaminar si los términos deben tratarse propiamente como diminutivos o bien como lexicalizaciones. De este modo de proceder ha dependido, en primer lugar, la anotación morfosintáctica de la palabra, pero afecta también directamente a su lematización en segundo término. Aquellos casos en los que se considera que la palabra ha adquirido ya un nuevo significado, el diminutivo se mantiene como parte integrante de la forma lematizada, de lo contrario, el lema se construirá sobre la base léxica de la palabra sin el sufijo correspondiente:

<i>molinillo</i>	lema: molinillo
<i>pajecillo</i>	lema: pajecillo
<i>manecilla</i>	lema: manecilla
<i>perolillo</i>	lema: perol
<i>contadorcito</i>	lema: contador
<i>jicarica</i>	lema: jícara

La distinción entre diminutivos transparentes y lexicalizados ha supuesto un profundo ejercicio de reflexión sobre el que se sustenta el análisis de esta tesis (§ 6. *LOS DIMINUTIVOS -ICO, -ITO, -ILLO EN EL CORPUS*). En algunos casos, gracias al contexto, ha sido fácil dictaminar si estamos ante verdaderos diminutivos o lexicalizaciones en las que se ha perdido la referencia al tamaño del objeto. Sin embargo, en muchas otras ocasiones resolver esta disyuntiva ha sido una tarea más compleja y en los casos en los que el contexto no es suficiente para resolver esta clasificación, el empleo del diminutivo bien podría estar indicando un auténtico valor de aminoración o un uso lexicalizado del término. Al tratamiento de las palabras con sufijación diminutiva en el corpus se dedica el Capítulo 6, por lo que no conviene extenderse más en este punto. No obstante, sí conviene mencionar que esta problemática y su consecuente resolución ha afectado de manera muy directa al proceso de lematización –junto al de anotación POS–, pues de con qué valor se consideran los sufijos diminutivos depende la adjudicación de uno u otro lema y su posterior recuperación a partir de búsquedas lematizadas en el corpus.

4.7. Triple visualización de los textos

TEITOK facilita no solo la edición y tratamiento informático de los textos, sino que funciona además como un sistema donde alojar los documentos para su publicación y

lectura en línea. En este sentido, actúa como una página web al uso que permite a cualquier usuario consultar y navegar por los textos de una manera fácil e intuitiva.

La tokenización permite que toda la información relacionada con un mismo texto quede almacenada en un único archivo (§ 4.2. *Tokenización*). Esto, que por un lado implica una larga y compleja cadena de código que hace imposible su lectura y manejo para la investigación, por otro lado, supone una solución viable al problema de la visualización de diferentes versiones de un mismo texto que de otro modo deberían ofrecerse de manera independiente.

La edición digital de los textos tiene como resultado ofrecer para cada uno de ellos las imágenes del manuscrito original, una edición semipaleográfica y una edición modernizada. TEITOK proporciona un entorno sencillo y atractivo para el usuario que da salida a la publicación de textos editados en XML-TEI gracias a una interfaz amigable. Este sistema permite además exhibir la fotografía del facsímil original, por lo que en todo momento es posible comprobar la veracidad de las transcripciones y disponer de las distintas versiones del texto, cambiando cómodamente de una a otra al seleccionar la opción deseada en el menú: la edición semipaleográfica –en la que se puede mostrar o no la expansión de las abreviaturas– (Figura 38) y la edición normalizada (Figura 39). Además, se puede incorporar al texto la información morfosintáctica y el lema de cada palabra (Figura 40).

Inventario de los bienes que quedaron tras fin y muerte de María Sánchez de Lázaro

Fecha	1711
Localidad	España, Madrid, Madrid
Proyecto	CORDEREGR: Corpus diacrónico del español del reino de Granada. 1492-1833
Financiación	Junta de Andalucía/FEDER: P09-HUM 4466, 2009-2014
Archivo	Archivo Histórico de Protocolos de Madrid
ID del manuscrito	AHPM T.14527
Transcripción	Pilar Arrabal Rodríguez

[less header data](#)

Opciones de visualización

Texto: **Transcripción** [Forma expandida](#) [Forma normalizada](#) -
 Mostrar: [Colores](#) [Formato](#) [<pb>](#) [<lb>](#) [Imágenes](#) -
 Etiquetas: [Etiqueta POS](#) [Lema](#)

Dos pailas de cobre grandes con asas

Cobre y la una de ellas con dos ganchos en y yerro ella para azer zerilla.

Dos ollas de cobre con asas y picos, una maior q otra.

Cuatro cucharas, las tres de cobre y la otra de yerro.

Un perol pequeño de cobre con asas.

Un perolillo de cobre para colar.

Tres peroles de cobre con asas.

Una olla grande de cobre con asas y pico.

Un barquillo pa colar zera.

Una cuchara mediana de cobre.

Un yerro de azer zerilla.

Un torno de madera de alamo y negro con sus manixas de yerro para blanquear zera.

Figura 38: Visualización en TEITOK de un documento XML-TEI con su cabecera, texto e imagen del facsímil (edición semipaleográfica)

Texto: Transcripción Forma expandida Forma normalizada - Mostrar: Colores Formato <pb> <lb> Imágenes - Etiquetas: Etiqueta POS Lema

Dos pailas de cobre grandes con asas

Cobre y la una de ellas con dos ganchos en y hierro ella para hacer cerilla.

Dos ollas de cobre con asas y picos, una mayor que otra.

Cuatro cucharas, las tres de cobre y la otra de hierro.

Un perol pequeño de cobre con asas.

Un perolillo de cobre para colar.

Tres peroles de cobre con asas.

Una olla grande de cobre con asas y pico.

Un barquillo para colar cera.

Una cuchara mediana de cobre.

Un hierro de hacer cerilla.

Un torno de madera de álamo y negro con sus manijas de hierro para blanquear cera.



Figura 39: Visualización en TEITOK de la edición normalizada e imagen del facsímil

Texto: Transcripción Forma expandida Forma normalizada - Mostrar: Colores Formato <tpb> <lb> Imágenes - Etiquetas: Etiqueta POS Lema

Dos	pailas	de	cobre	grandes	con	asas	
Num	Nombre C	Prep	Nombre C	Adj	Prep	Nombre C	
2	paila	de	cobre	grande	con	asa	

Cobre	y	la	una	de	ellas	con	dos	ganchos	en
Nombre C	Conj	Det	Det	Prep	Pron	Prep	Num	Nombre C	Prep
cobre	y	el	un	de	ellos	con	2	gancho	en

y	hierro	ella	para	hacer	cerilla	.
Conj	Nombre C	Pron	Prep	Vb	Nombre C	Punt
y	hierro	el	para	hacer	cerilla	.

Dos	ollas	de	cobre	con	asas	y	picos	.
Num	Nombre C	Prep	Nombre C	Prep	Nombre C	Conj	Nombre C	Punt
2	olla	de	cobre	con	asa	y	pico	.

una	mayor	que	otra	.
Det	Adj	Conj	Det	Punt
un	mayor	que	otro	.

Cuatro	cucharas	.	las	tres	de	cobre	y	la	otra



Figura 40: Visualización en TEITOK de la edición normalizada en la que se muestra la etiqueta morfosintáctica y lema de cada palabra

Detrás de esta visualización se encuentra el procesamiento informático en TEITOK del archivo XML correspondiente al documento de las figuras que se acaban de mostrar. La edición digital de partida es la del documento que se adjunta en los anexos para la

cabecera y texto (§ Anexo I. Ejemplo completo de cabecera etiquetada y § Anexo II. Ejemplo completo de texto etiquetado).

4.8. Búsquedas en el corpus basadas en lenguaje CQL

Todo el proceso de configuración del corpus descrito hasta ahora no tiene razón de ser si no se dispone de un sistema apropiado que permita seleccionar los datos y recuperar la información en base a los intereses que rigen cada investigación. Ello implica en muchas ocasiones poder cruzar variables de distinto tipo y precisar los criterios de una búsqueda para delimitar al máximo todas las opciones posibles y reunir solo aquellas que sean verdaderamente relevantes para nuestro estudio. Los textos del corpus han sido procesados a fin de poder tener distintas versiones de cada uno de ellos. Cada token en el corpus condensa información de diferente tipo gracias a la codificación en XML por medio de los atributos, pero para poder acceder a toda esa información se necesita un sistema que pueda leer e interpretar todos esos datos y la manera en la que están codificados. Para ello TEITOK utiliza un motor de búsquedas basado en lenguaje CQL.

CQL (*Corpus Query Language*) –en español, “lenguaje de consulta de corpus”– es un lenguaje especialmente diseñado para la recuperación de datos lingüísticos basado en una sintaxis propia que admite una enorme capacidad de concreción y de posibilidades combinatorias en las búsquedas de un corpus. Su empleo exige seguir una estructura específica y unos criterios muy bien definidos que se explican más adelante.

Este sistema fue diseñado por el Instituto para el Procesamiento del Lenguaje Natural (IMS) de la Universidad de Stuttgart⁴⁶ y establece unas características particulares no solo en cuanto a la sintaxis del lenguaje, sino también sobre cómo se visualizan los datos y cómo debe estar previamente configurado el corpus para que el sistema funcione. El lenguaje CQL supuso una solución al problema de los distintos niveles de anotación en corpus informatizados a la hora de gestionar y combinar toda la información que contienen:

When a corpus is thus regarded as a structured object composed of several different knowledge sources, a problem arises because different knowledge sources require possibly different access methods. [...] These different knowledge sources, access strategies and usage situations are best supported by a hierarchical, modularized system architecture where the single modules can be combined in different ways to adapt the system to various usage situations (Christ, 1994, p. 24).

Numerosos corpus se valen hoy día de este lenguaje como medio para rentabilizar al máximo los datos de que disponen. En este sentido, las guías que explican el funcionamiento de otros corpus, como las elaboradas para *Sketch Engine* (Cambridge Sketch Engine, 2012), *CICA* (Gago-Jover y Pueyo Mena, 2022) y *Post Scriptum*

⁴⁶ <https://www.ims.uni-stuttgart.de/en/research/projects/corpus-workbench/>

(Vaamonde, uso interno), han servido de gran apoyo para la confección de los apartados que siguen, los cuales se han visto enriquecidos por la existencia de tales manuales.

Las siguientes páginas tratan de explicar con mayor detalle qué es CQL y como funciona aplicado al corpus *ODE*. Este sistema, a su vez, se sirve del empleo de expresiones regulares como estrategia para aumentar considerablemente las posibilidades de recuperación de datos. La combinación de CQL con expresiones regulares convierte al buscador en una potente herramienta para explotar el corpus. Su verdadera utilidad se demuestra en § 4.8.3. *Casos prácticos de búsquedas en el corpus ODE*, donde se ofrecen ejemplos concretos de búsquedas reales en las que se combinan estos dos aspectos: lenguaje CQL y expresiones regulares.

4.8.1. Principios básicos de formación CQL

CQL se basa en el token como elemento básico para recuperar datos, de manera que una condición imprescindible para su empleo es que el corpus esté previamente tokenizado (Christ, 1944, p. 26). El sistema busca por tokens y accede a cada uno de ellos para seleccionar un tipo concreto de información según los criterios de consulta especificados en cada caso. La orden que emplea para ello se basa en el esquema básico:

[atributo = "valor"]

La sintaxis específica de CQL se fundamenta en el uso de los corchetes para representar cualquier token. Se sabe que la información que contiene cada token se codifica por medio de diferentes atributos, cada uno de los cuales se corresponde con un tipo distinto de información (§ 4.2. *Tokenización*) y que en el caso de *ODE* son los siguientes:

- a. *@form*: forma original de la palabra en el manuscrito

Búsqueda CQL	Resultados obtenidos
[form = "zerdo"]	zerdo
[form = "vacido"]	vacido
[form = "tinaxa"]	tinaxa
[form = "ynbentariado"]	ynbentariado

Tabla 6: Búsqueda CQL por forma original (*form*)

- b. *@fform*: forma expandida de una abreviatura

Búsqueda CQL	Resultados obtenidos
[fform = "vecino"]	vo, vno, veco, vzo
[fform = "escribano"]	esno, ssno, escno, scrio
[fform = "siguiente"]	sigte, sige, siguite

Tabla 7: Búsqueda CQL por forma expandida (*fform*)

c. *@nform*: forma modernizada según las normas ortográficas actuales

Búsqueda CQL	Resultados obtenidos
[nform = "cerdo"]	zerdo, serdo, cerdo
[nform = "vacío"]	vaçio, vacido, basido, vazio, vacio...
[nform = "tinaja"]	tinaja, tinaxa, tenaja, tenaxa, tenaga...
[nform = "inventariado"]	ynbentariado, imbentareado, inventariado

Tabla 8: Búsqueda CQL por forma normalizada (*nform*)

d. *@pos*: anotación morfosintáctica mediante etiqueta gramatical

Búsqueda CQL	Resultados obtenidos
[pos = "NCMS000"] (N= nombre, C= común, M= masc., S= sing.)	cerdo, retablo, señor, revólver...
[pos = "AQ0MS0"] (A= adjetivo, Q= calificativo, M= mas., S= sing.)	vacío, limpio, alto, grueso, inglés...
[pos = "NCFS000"] (N= nombre, C= común, F= fem., S= sing.)	tinaja, almohada, orza, mesa...
[pos = "VMP0000"] (V= verbo, M= principal (<i>main</i>), P= participio)	inventariado, tasado, heredado...

Tabla 9: Búsqueda CQL por etiqueta gramatical (*pos*)

e. *@lemma*: forma lematizada bajo la que se incluyen las variantes flexionadas

Búsqueda CQL	Resultados obtenidos
[lemma = "cerdo"]	cerdo, cerda, cerdos, cerdito, cerditillos...
[lemma = "vacío"]	vacío, vacía, vacíos, vacías
[lemma = "tinaja"]	tinajas, tenaxas, tinajica, tenaja...
[lemma = "inventariar"]	inventariado, inventariaron, inventarió, inventariar...

Tabla 10: Búsqueda CQL por lema (*lemma*)

En CQL es necesario hacer explícita el tipo de información que se desea recuperar por medio del atributo –que representa el nivel de anotación– dentro del token según el esquema mencionado. El valor, delimitado entre comillas, incluye los criterios específicos de la búsqueda. La elección de un atributo u otro supone acceder a un nivel concreto de anotación dentro del corpus y tiene implicaciones distintas a la hora de obtener resultados, tal y como se ha podido comprobar en los ejemplos de las tablas de arriba: un mismo valor tendrá resultados diferentes en búsquedas con atributos distintos. En definitiva, en el esquema básico de CQL [atributo = "valor"] el atributo es el criterio general de búsqueda, mientras que el valor es el criterio específico.

El buscador de TEITOK funciona escribiendo directamente en el cajón de búsqueda las instrucciones en CQL que se acaban de explicar, por lo que es necesario

conocer cómo se ha codificado la información dentro de cada token (atributos), que podrá variar según el corpus consultado.

Este sistema permite buscar por dos o más elementos a la vez si se repite el anterior esquema tantas veces como palabras o tokens se deseen recuperar. De manera que es posible delimitar y definir el contexto de aparición de una palabra si lo que interesa son unidades fraseológicas o expresiones fijas. En este sentido, las búsquedas de varios tokens pueden resultar muy útiles a la hora de establecer combinaciones léxicas en torno a una palabra para conocer qué palabras o tipos de ellas suelen aparecer junto a una forma determinada. Un caso práctico en un corpus de inventarios es recuperar la construcción *sustantivo + de* para conocer de un simple vistazo de qué material solía estar hecho un determinado objeto. Por ejemplo, si se aplica esta búsqueda al caso de *cama*, será necesario buscar dos tokens consecutivos en el corpus, para lo que debemos repetir el esquema ya conocido. En este caso, se ha optado por una búsqueda lematizada para así incluir los casos de *cama* tanto en singular como en plural:

[lemma = "cama"] [lemma = "de"]

rs. Vna armadura de cama de aya en cinco rs.	1706 Burgos
. <u>Lecho cotidiano</u> : Una cama de aya de medios mastiles encandelada	1706 Burgos
rs. Tres armaduras de cama de aya uiexas sin cordeles a	1706 Burgos
ducados. Vn paño de cama de frisa colorada con vnas trepas	1544 Córdoba
Vna cama de red de olanda y cortado	1592 Badajoz
rs. Unos paños de cama de caliqui tassados en seis duos	1592 Badajoz
en cient rreales. Vna cama de madera de canpo y una	1592 Badajoz
en quatrocientos rs. Dos camas de pino y una de nogal	1715 Almería
. n 20 ft cinco camas de pino, la una nueva	1721 Almería
y del susodho Dos camas de pino con su cuerda,	1754 Almería
en veinte reales. Dos camas de madera en doce rs.	1827 Almería
rs. 22 ... Dos camas de pino de marca mayor,	1829 Almería
valen 8. Yt dos camas de madera á 9 rs,	1832 Almería
vale 12. Yt tres camas de madera vale 27. Yt	1832 Almería

Figura 41: Resultado de la búsqueda *cama de* en el corpus *ODE*

En lenguaje CQL no es significativa la separación de tokens mediante espacios en blanco en la búsqueda de palabras consecutivas. Tampoco son trascendentes los posibles espacios que pueda haber dentro de los corchetes que delimitan cada token. No obstante, la sintaxis CQL se vuelve más estricta en lo relativo a la presencia de espacios dentro de los valores de un atributo, es decir, aquello que se incluye entre comillas. De esta manera, cualquiera de las siguientes búsquedas en CQL se consideran válidas, a excepción de la última, en la que el espacio en blanco se interpreta como otro carácter más y no obtendrá ningún resultado:

[lemma = "cama"] [lemma = "de"]

```
[lemma = "cama"] [lemma = "de"]  
[lemma = "cama"] [lemma = "de" ]  
*[lemma = " cama"] [lemma = "de "]
```

El esquema utilizado hasta ahora reproduce la sintaxis básica de CQL. Sin embargo, este sistema de búsquedas acepta como parte integrante de los valores de cada token el uso de expresiones regulares (véase el apartado abajo) para realizar búsquedas avanzadas con las que recuperar datos de manera muy selectiva. Las opciones de explotación del corpus aumentan considerablemente gracias a la compatibilidad de CQL con el uso de este tipo de comodines.

4.8.2. Uso de expresiones regulares como parte de la sintaxis CQL

Se conoce como expresiones regulares al conjunto de signos o símbolos que sirven para sustituir caracteres en una búsqueda. Su significado es convencional y se utilizan para buscar patrones comunes que pueden repetirse, como, por ejemplo, para recuperar todas las palabras que empiecen, contengan o terminen por alguna secuencia concreta (i.e. la búsqueda *des.** encontrará todas las palabras que empiecen por *des-* sin importar el número ilimitado de caracteres que vengan a continuación). Las expresiones regulares son usadas con bastante frecuencia en los buscadores para recuperar información de manera muy selectiva, ya que permiten incluir o excluir ciertos elementos haciendo más restrictivos los criterios de consulta. Comúnmente se las conoce como *regex*, por su contracción del término en inglés *regular expression*.

En este sentido, las expresiones regulares cumplen algunas de las funciones que en los corpus tradicionales desempeñan los comodines, más simples de utilizar, pero a los que superan con creces en cuanto a posibilidades de explotación se refiere. Los comodines difieren de las expresiones regulares en que estas están sujetas a convenciones internacionales, lo que significa que sus usos y funciones no varían y se ajustan siempre a las mismas pautas, sea cual sea la herramienta de búsqueda utilizada. Por el contrario, los comodines y las funciones que desempeñan pueden variar sustancialmente de un corpus a otro.

Las expresiones regulares se pueden utilizar en lenguaje CQL como parte de su sintaxis, casi siempre dentro de lo que se considera el valor del atributo dentro del esquema básico. La búsqueda del ejemplo que se ha mencionado anteriormente se representaría en CQL de la siguiente forma:

```
[nform = "des.*"]
```

El empleo de expresiones regulares añade a las búsquedas criterios específicos con un alto nivel de concreción. Algunas de estas expresiones, como se verá más adelante, son incluso más útiles cuando se combinan entre sí, y permiten realizar numerosas

combinaciones que pueden llegar a ser bastante complejas. Se incluyen aquí las más frecuentes y al final de este apartado se recopilan todas ellas (Tabla 11).

a. Punto .

Sustituye a un único carácter (letra o dígito). Se trata del comodín por excelencia en expresiones regulares y equivale al signo de interrogación en otros corpus que no usan CQL, como por ejemplo el *CORDE*:

Expresión regular	Resultados obtenidos
ca.o	cabo, caso, cazo, etc.
.aza	taza, maza, haza, raza, caza, Baza, etc.
.aber	caber, haber, saber

b. Asterisco *

Representa cero o más veces el carácter al que precede. La combinación del asterisco seguido del punto .* permite representar cualquier carácter cero o más veces. Esta combinación resulta muy útil y versátil para recuperar, por ejemplo, inicios y terminaciones de palabras:

Expresión regular	Resultados obtenidos
fe*	fe, fee, feee, feeee...
cal.*	cal, calabaza, calado, calavera, calzones, calzada, caldo, caldera...
.*tilla	beatilla, cotilla, cuartilla, Castilla, motilla, mantilla, paletilla...

c. Signo más +

Representa el carácter al que precede una o más veces. Pospuesto al punto .+ devolverá las ocurrencias de cualquier carácter una o más veces:

Expresión regular	Resultados obtenidos
oye+	oye, oyee, oyeee, oyeeee...
.+ción	acción, atención, bendición, habitación, nación, población, relación, velación...
pre.+	precio, preciso, preciosa, prenda, prevenir, pretil...

d. Cierre de interrogación ?

Representa el carácter al que precede cero o una vez. Pospuesto al punto .? devolverá las ocurrencias de cualquier carácter cero o una sola vez:

Expresión regular	Resultados obtenidos
des?cir	descir, decir
las?	las, la
mas.?	mas, masa

e. Corchetes []

Especifica un conjunto cerrado de caracteres posibles, ya sean letras o dígitos. El uso del guion (-) dentro de los corchetes permite establecer rangos para los caracteres establecidos:

Expresión regular	Resultados obtenidos
[blr]ata (la letra <i>b</i> , <i>l</i> o <i>r</i> antes de <i>-ata</i>)	bata, lata, rata
[a-z]aso (cualquier letra de la <i>a</i> a la <i>z</i> delante de la secuencia <i>-aso</i>)	caso, laso, paso, raso, taso, vaso
[1-5] (cualquier número del 1 al 5)	1, 2, 3, 4, 5

f. Paréntesis ()

Agrupar varios caracteres para tratarlos como una única unidad a la que se pueden aplicar a su vez otras expresiones regulares:

Expresión regular	Resultados obtenidos
(des)*compuesto	compuesto, descompuesto
un(a)	un, una

g. Llaves { }

Indica el número de veces que se debe repetir el carácter precedente. Se puede aplicar a caracteres o a tokens enteros (§ 4.8.3.6. *Dejar huecos entre palabras*):

Expresión regular	Resultados obtenidos
hola{3}	holaaa

.{4} (cualquier carácter 4 veces = palabras de 4 letras)	cinco, plata, medio, villa, negro, pares, viejo, decir, hecho, lugar...
[pos="AQ0MS0"]{2} (un adjetivo masculino singular dos veces seguidas)	amarillo pequeño, antiguo redondo, blanco viejo, buen cristiano...

Además de los signos ya mencionados, CQL admite el uso de otras expresiones que funcionan de manera similar a los operadores lógicos –con función Y, O y NO– para agrupar varios elementos en una misma consulta. Se consigue así una búsqueda combinada que permite conectar varios elementos mediante alguna de las condiciones que se detallan a continuación:

h. *Ampersand &*

Establece una relación copulativa entre dos elementos de una búsqueda (condición Y). Recupera resultados en los que se cumplen todos los criterios unidos mediante el signo & a fin de concretar los criterios de una búsqueda y reducir el número de posibles resultados. En lenguaje CQL se utiliza dentro de la estructura de un token para combinar atributos diferentes asociados entre sí:

[nform= "fueron" & lemma= "ser"]
Forma *fueron* perteneciente al lema *ser*

En esta búsqueda se especifican dos condiciones para una única palabra: la que corresponde a la forma normalizada (@*nform*) y la correspondiente al lema (@*lemma*). Los atributos unidos mediante el símbolo & devolverán solo aquellos casos en los que se cumplen las condiciones expresadas para cada atributo, esto es, las ocurrencias de *fueron* siempre y cuando pertenezcan al lema *ser*. Como resultado obtendremos aquellas formas en las que ambas condiciones se cumplen para todos los casos, descartándose por tanto las formas de *fueron* pertenecientes al verbo *ir*:

ya difunta, vecinos que **fueron** de la villa del
 que fue, cuios vienes **fueron** tasados por | personas inteligentes y
 , difunta, vezinos que **fueron** y es el dho su
 muger, difuntos vezinos | que **fueron** de dha villa, por
 ha mi ruego que lo **fueron** presentes | Cezilio Gimenez, Franco
 para sauer su balor, **fueron** tasadas | en zinco pesos,
 , vezinos y naturales que **fueron** | de esta dha va,
 los quales, regulados qe | **fueron** por personas practicas, se
 , difuntos y vecinos qe **fueron** de la misma villa;
 en la clausula precedente le **fueron** entregados al don | Antonio
 en el discurso d ella **fueron** los alcaldes de | dha billa

Figura 42: Búsqueda combinada. Resultado de la forma *fueron* del verbo ser

i. Pleca |

Conecta en una búsqueda varias alternativas mediante la función disyuntiva (condición O). Recupera resultados en los que se incluye cualquiera de las opciones indicadas. Funciona solo dentro del valor del token y puede aplicarse a palabras enteras o a fragmentos de ellas si se combina con la expresión regular ():

[nform= "pasa|tasa"]

Formas modernizadas *pasa* y *tasa*

[form = "azucarer(o|a)"]

Formas originales *azucarero* y *azucarera*

[lemma = "haber|tener|poseer"]

Formas flexionadas de los lemas *haber*, *tener* y *poseer*

j. Cierre de exclamación !

Excluye elementos en una búsqueda negando el criterio que le sigue (condición NO). En sintaxis CQL, se antepone siempre al signo igual para indicar que lo que sigue a continuación debe ser distinto a lo expresado. La expresión se puede emplear para negar un criterio dentro de un mismo token o para excluir tokens enteros:

[nform = ".+ndo" & pos! = "VMG0000"]

Palabras acabadas en *-ndo* que no sean gerundios:

segundo, *cuando*, *hediondo*, *hondo*, *Fernando*, *fondo*, *redondo*...

[nform= "mesa"] [nform!= "de"]

Secuencia *mesa* + cualquier forma que no sea la preposición *de*:

mesa chica, mesa camilla, mesa con, mesa bufete, mesa cajón, mesa altar...

Con el objetivo de facilitar el uso de todas las expresiones regulares, en la siguiente tabla se recopilan los signos mencionados y sus funciones:

<i>Signo</i>	<i>Valor</i>	<i>Función y significado</i>
.	Punto	Sustituye un único carácter.
*	Asterisco	Sustituye un carácter cero o más veces.
+	Más	Sustituye un carácter una o más veces.
?	Cierre de interrogación	Representa cero o una vez el carácter que le precede.
[]	Corchetes	Establece una lista de caracteres opcionales o rangos delimitados por guion. i.e. [bv], [a-z]
()	Paréntesis	Agrupar varios caracteres sobre los que aplicar otras expresiones regulares.
{n}	Llaves	Permite establecer el número de veces que puede repetirse el carácter o elemento precedente.
&	<i>Ampersand</i>	Añade una condición que debe cumplirse en todos los casos (función copulativa).
	Barra vertical o pleca	Combina varias opciones (función disyuntiva).
!	Cierre de exclamación	Niega o excluye el elemento que sigue a continuación (función negativa).

Tabla 11: Principales expresiones regulares empleadas en búsquedas CQL

4.8.3. Casos prácticos de búsquedas en el corpus ODE

Hasta ahora se ha ilustrado el funcionamiento del lenguaje CQL y el significado de las principales expresiones regulares, su uso y potencial, para la recuperación de datos lingüísticos. Estas instrucciones son aplicables a otros muchos corpus también basados en el mismo sistema y a otros buscadores que permiten el uso de expresiones regulares como parte de su sintaxis. En este apartado se combina el uso del lenguaje CQL con las ventajas que ofrecen las expresiones regulares en el caso concreto del corpus *ODE*.

Cabe destacar que algunos de los casos que se proponen solo son posibles gracias al nivel específico de anotación gramatical de este corpus en concreto. El etiquetado del corpus va más allá de la anotación superficial de las categorías principales *sustantivo*, *adjetivo*, *pronombre*, *verbo*, etc., y marca características específicas que posibilitan la recuperación precisa de la información. Por otro lado, ello se debe a la revisión manual a la que han sido sometidos los textos, lo que permite

ofrecer un corpus normalizado, anotado y lematizado con el menor número de errores posibles.

Los casos que se detallan a continuación ofrecen una idea de las posibilidades que brinda la combinación de lenguaje CQL con expresiones regulares para la explotación del corpus *ODE* a partir de búsquedas reales susceptibles de interés lingüístico.

4.8.3.1. Distinguir entre mayúsculas y minúsculas

Por defecto, el motor de búsqueda de *ODE* es *case-sensitive*. Esta expresión en inglés se emplea en términos de recuperación de datos para señalar que es relevante la distinción de caracteres entre mayúsculas y minúsculas a la hora de dictar órdenes en una consulta. Por lo tanto, cualquier búsqueda en el corpus tiene en cuenta la diferencia entre mayúsculas y minúsculas si no se especifica lo contrario.

Conocer este aspecto es significativo para un usuario que desea consultar el corpus, pues la búsqueda de una palabra solo en minúscula puede estar discriminando otras ocurrencias de la misma palabra escrita con todas o alguna de sus letras en mayúscula. Esto quiere decir que la búsqueda de “casa” y “Casa” obtendrá resultados diferentes, pues el sistema entiende que se trata de palabras distintas.

En aquellos casos en los que no interese distinguir entre mayúsculas y minúsculas se puede aplicar la expresión regular `%c` fuera del valor del atributo, pero dentro del token en sintaxis CQL, que contemple todas las ocurrencias de una palabra, incluidas también aquellas en las que al menos uno de sus caracteres está escrito con letras capitales:

```
[nform = "cocina"%c]
```

El uso de esta expresión permite especificar una búsqueda del tipo *case-insensitive*, es decir, insensible a mayúsculas y minúsculas. El resultado será una concordancia en la que el sistema no considere la diferencia entre ambos tipos de letra para una misma palabra. A continuación, se muestra la concordancia obtenida de la consulta anterior para la palabra *cocina* en el corpus:

```
de Ortega, alguazil mor | [...] | Cozina Primte un bail y vnas  
tasado | en una peseta. | Cocina | 374 Una mesa de pino  
cobertores blancos y negros. | Cosina Un plattero de madera.  
Y se paso a la cozina , a mano yzquierda,  
seruir. | Quatro sillas de cozina a medio seruir. | Vnos  
pequeno. | Quartto dormittorio de | cozina . | Vn quadro, pintura  
Que sirve para trasttos de conzina . | Y en ella solo  
mo. | Unas tenazas de cocina , real y mo.  
ochenta y quatro de la cocina ; y los trescientos y
```

Figura 43: Resultado de búsqueda *case-insensitive* para la voz *cocina*

4.8.3.2. Combinar varios criterios en una búsqueda

Una de las ventajas de CQL, favorecida también por la anotación múltiple a la que se ha sometido el corpus, es la de poder aplicar en una misma búsqueda distintos criterios que afecten a varios niveles de edición de una sola palabra. De esta manera, se puede ser más preciso en la búsqueda si se combinan varios atributos para un mismo token, es decir, si se define más de una condición combinando forma, etiqueta o lema, mediante el uso de &:

[nform = "vino" & pos = "NC.+"]

Forma *vino* como sustantivo común

Esta posibilidad es especialmente útil en los casos que se ven afectados por la ambigüedad gramatical. La combinación de palabras y etiquetas, por ejemplo, favorece la desambiguación de las palabras homógrafas cuando la búsqueda únicamente por forma normalizada no ayuda en la selección de los casos que interesan. Así, palabras y etiquetas pueden combinarse para obtener, por ejemplo, los casos de *cobre* como sustantivo o de *vino* únicamente como verbo. Se muestran a continuación otros ejemplos:

[nform= "como" & pos= "R"]

Forma *como* con función adverbial

[nform= ".+illo" & pos= "N.+D"]

Formas acabadas en *-illo* que sean diminutivos:

arconcillo, braserillo, conchoncillo, espejillo...

[form="a.+" & nform= "h.+"]

Formas originales que empiecen por *a-* y cuya normalización empiece por *h-*:

habas, haber, hacha, hacia, halló, harina, haza...

4.8.3.3. Especificar el contexto adyacente

La sintaxis CQL permite la búsqueda de palabras consecutivas para obtener dos o más elementos adyacentes y definir el contexto en el que queremos que aparezca una palabra. De esta manera, es posible buscar varios tokens de manera simultánea escribiendo en el cajón de búsqueda uno tras otro en su correspondiente sintaxis CQL:

[nform = "vecino"] [nform = "que"] [nform = "fue"]

La búsqueda de arriba pretende recuperar todas las variantes asociadas a las formas normalizadas de la cadena *vecino que fue*. Sin embargo, la consulta por palabras adyacentes no tiene por qué ser del mismo nivel para todos los tokens: se pueden

combinar distintos tipos de atributos para obtener, por ejemplo, un lema o una forma normalizada seguidas de una etiqueta morfosintáctica. Así, se podrían obtener todas las construcciones pasivas con *ser* más participio con una búsqueda como la siguiente, cuyo resultado se muestra debajo:

[lemma = "ser"] [pos = "VMP00SM"]

qual Dios Nro Señor a	cido seruido	de darne, y
y llaves, el que	fue apreciado	en settezos rrs. N
, q con la hechura	fue apreciado	en veinte y quatro rs
y en otra ocazion que	fue apresado	de los enemigos yngleses como
, que con la hechura	fue apreziado	en ziento y once rs
servizio de Dios Nro Señor	fue desposado	y velado, segun lo
a la tierra de q	fue formado	y quando muriere mi cuerpo
a la tierra de q	fue formado	y quando muriere quiero mi
a la tierra de q	fue formado	; y quando muriere,
de la ttierra de que	fue formado	y quando muriera mi cuerpo
virtud de su poder que	fue otorgado	en la villa de Madrid
por quanto la voluntad divina	fue servido	llevarse de esta presente
ellas delgadas, y todo	fue tasado	en quarenta. n 48
con todos sus peltrechos;	fue tasado	en ochocientos rrs de vn
, zerradura y llabe que	fue tasado	en cien rrs. N
en trescientos rrs en que	fue tasado	. N 21 Ytt dos

Figura 44: Resultado de la búsqueda *ser* + participio

O bien estudiar los casos en los que no se produce apócope de los adjetivos como *bueno* y *grande*. Para ello se efectuará una búsqueda en la que se combine la forma normalizada seguida de la etiqueta morfosintáctica correspondiente a nombre común:

[nform = "grande"] [pos = "NC.+"]

buscar a los susodhos con	grande alvoroito	y escandalo y ocaasionando a
y a menos preçio con	grande alibio	de los pobres, mas
por tener con el referido	grande amistad	, y le dixo:
, con el que profeso	grande amistad	, conoziendo igualmte a su
rs. Yd Vn cazo	grande azofar	, vsado, en quinze
en el suelo con muy	grande colera	y brio, diziendo que
que le haze adulterio con	grande escandalo	y mormuracion de los vzos
, de que se daba	grande escandalo	en la rrepuca. Y
villa y su comarca muy	grande escandalo	y mormuracion, de tal
dha herida fue dada con	grande fuerça	de pedrada, y que
porque le hizo vna muy	grande herida	en los cascos, de
Ysael de Escouar donçella de	grande onestidad	y rrecojimto y criada y
gran nota y escan dalo y	grande ynfamia	. Digan, etza.

Figura 45: Resultado de la búsqueda de la secuencia *grande* + sustantivo

4.8.3.4. Mostrar distintas opciones en una sola búsqueda

En otros casos, puede resultar útil obtener varias alternativas en una sola búsqueda. CQL permite seleccionar distintas opciones dentro de los criterios establecidos y obtener por tanto varios resultados en una misma concordancia. Este tipo de búsqueda combinada es posible con el uso de la pleca o barra vertical | delimitando ambos términos y puede ser especialmente útil cuando lo que se quiere es obtener una visión global que facilite el estudio de la distribución de distintos términos en el corpus. Por ejemplo, la búsqueda conjunta de los sinónimos *latón* y *azófar* devolverá un único listado en el que es posible ver todas las ocurrencias de ambas voces gracias a una consulta como la siguiente:

[nform= "latón|azófar"]

sarten grande y dos casos de **azofar** biejos con | sus cabos
gasta y recaton | Un almires de **azofar** con su mano de
que costo | Dos cazos de **azofar** con sus cabos de jerro
tres duos | Dos candeleros de **azofar** de pie alto en dos
un real | Un casito de **asofar** con su cabo pequeño en
blanco y dos candeleros de **açofar** | tasado en ocho reales | Una
Masianello, dos marcos | de **laton** dorado, otros dos id
milesimas. | Una cofaina de **laton** , dos medias fuentes | grandes
yd y una | ceytera de **laton** nueva en catorce rs.
rs. | Yd Otra de **laton** , encarnada, en doce
caja. | Una escribania de **laton** compuesta de tintero, obleadera
libros. | Una escrivania de **laton** compuesta de cuatro piezas.
rs. | Tres candeleros de **laton** á tres rs cada uno
rs. | Una espumadera de **latton** en dos rs. | Un
de parez, dorados de | **laton** con adornos de cristal,

Figura 46: Resultado de las alternativas *latón* / *azófar* en una sola consulta

Una búsqueda como esta puede darnos una visión de cuál de estas dos variantes léxicas procedentes del árabe gozaba de mayor presencia.

En lenguaje CQL, la opción de combinar varios resultados al mismo tiempo en una sola búsqueda no solo es especialmente útil para el estudio de variantes léxicas en palabras enteras, también para la distribución de variantes morfológicas o fonéticas, por lo que la alternancia deberá aplicarse en estos casos a fragmentos de palabras. Por ejemplo, una búsqueda como la siguiente nos ofrece una distribución de los diminutivos en *-ito* e *-illa* del lema ‘cuchara’:

[nform= ".+(ita.*|illa.*)" & lemma= "cuchara"]

rs vn. | Ytem una **cucharilla** grande en doce. | Ytem
unas | | un candelero y una **cucharilla** en veinte y seis rs
un | caliz con patena y **cucharilla** , dos vinageras con | su
con patena sobre ostia y **cucharilla** de plata con 17 onzas
rs. | 40. Seis **cucharillas** de yerro en cuatro rs
44. Diez y siete **cucharillas** para café, al
de lo mismo y una **cucharita** de metal. | Yt:
espadin. | Diez y ocho **cucharitas** de oro pa café con
cuchillos de plata. | Doce **cucharitas** pa café de plata con
negros. | Diez y siete **cucharitas** pa cafès y tres pinzas
. | Tres espumaderas. | Catorce **cucharitas** hechas en La Coruña.
. | Ytt dos dosenas de **cucharitas** de estrado de plata.
Ytem una espumadera y dos **cucharitas** de yerro en tres rs
onzas. | Vna docena de **cucharitas** de plata | para café pesan

Figura 47: Búsqueda combinada de variantes morfológicas del diminutivo

El resultado permite estudiar la distribución de los diferentes sufijos diminutivos en el caso de una palabra concreta, pero búsquedas similares pueden aplicarse a otros tantos fenómenos de alternancia morfológica, fonética o léxica en los textos. Más ejemplos de explotación del corpus gracias a la alternancia de resultados en una misma búsqueda se muestran abajo:

[nform= ".+er(o|a)" & pos= "AQ,+"]

Formas acabadas en *-ero/a* que sean adjetivos:

extranjero, entero, casera, espartera, verdadero, venidero, soltera, áspero...

[nform= "chocolater(o|a)"]

Variación de género de *chocolatero/-a*:

chocolatero, chocolateros, chocolatera, chocolateras

[lemma= "asno|burro"]

Ocurrencias del lema *asno* y *burro*:

Asnos, asnos, burro, burra, burros, burras

[form= "fr(a|e)zada.*"]

Alternancia vocálica e/a en las formas originales de *frazada* o *frazadas*

Frezada, frezadas, frazada, frazadas

4.8.3.5. Excluir elementos de una búsqueda

Se ha visto que combinar varios elementos en una búsqueda resulta extremadamente útil para agrupar datos en base a diferentes criterios. Por el contrario, igual de ventajoso puede ser excluir elementos con el fin de acotar la búsqueda y obtener resultados más precisos que cumplan solo con criterios o contextos de aparición más específicos. En CQL esta opción se permite con el uso del signo de exclamación ! antepuesto al signo = de cada atributo, para negar valores enteros, o la expresión [^] para negar uno o varios caracteres. Explicaremos esta función a partir de dos casos prácticos.

Por ejemplo, supongamos que interesa el estudio de las expresiones temporales construidas con las formas impersonales de *haber* del tipo “habrá mes y medio”, “ha ocho días que salió”. Para ello se puede realizar una búsqueda por lema (*haber*) en la que se contemplen todas las ocurrencias de este verbo en cualquiera de sus formas: [lemma= "haber"].

Sin embargo, el resultado de esta búsqueda probablemente sea demasiado amplio, ya que se incluirán también numerosas formas que no interesan, la mayoría de ellas, en las que *haber* actúa como verbo auxiliar (*ha traído, habiendo visto, haber recibido*, etc.). Se necesita entonces combinar la anotación sintáctica para seleccionar solo aquellos casos en los que *haber* goza de significado pleno. En este caso, necesitamos afinar la búsqueda negando uno de los elementos para que los resultados de *haber* no sean verbos auxiliares, tal y como se muestra a continuación:

```
[lemma= "haber" & pos!= "VA.+"]
```

Esta búsqueda se acerca más a nuestros intereses, pero podríamos filtrar aun más el resultado para eliminar las formas impersonales de *haber* con el sentido de ‘hallarse’, ‘estar’ o ‘existir’ que tampoco interesan en este caso. Se puede añadir a la consulta anterior una restricción más referida a las formas impersonales más frecuentes (*hay, había, haber, habido/habidos*):

```
[lemma= "haber" & nform!= "hay|había|haber|habido.*" & pos!= "VA.+"]
```

La lista de formas se ha reducido de nuevo considerablemente y un resultado como el siguiente reflejará más fielmente la estructura lingüística que se está buscando desde un principio:

diga y declare cuánto tiempo	ha	que enviudó, dijo que
de proceso, dijo que	ha	ocho días que salió de
mi marido y digo que	ha	tiempo de diez años que
esta, la dicha esclava	ha	justa e común estimación,
mejor forma que puedo y	ha	lugar en derecho, otorgo
más lugar haya y mejor	ha	para verse al
con más facultades ni	haberes	para pagarle su amor
contra el dicho don Gaspar	habiendo	escribano en dicha ciudad y
licenciado Pablos López, que	habrá	dos años poco más o
Miguel de Lara, no	habrá	hombre más desgraciado que yo
la voluntad de nuestro señor	habrá	efecto el dicho casamiento,
número de esta ciudad,	habrá	catorce o quince años,
enemigo, se cortó	habrá	un año árboles de él
ella, y dijo que	habrá	nueve días que él prestó
la dicha Águeda de Escol	habrá	veinte años, y en
4 Ítem si saben que	habrá	un mes poco más o
de la pregunta es que	habrá	quince días, poco más

Figura 48: Resultado de la búsqueda en la que se excluyen los casos de *haber* auxiliar

Para esta búsqueda la restricción se ha aplicado a todo el valor del atributo, pero se pueden aplicar restricciones a fragmentos de una palabra, uno o varios caracteres dentro de ella, con la expresión [^], en la que se incluye el carácter que se pretende negar:

[form="palan[^g]ana"]

Esta búsqueda rescatará todas las formas originales en las que la palabra *palangana* no se ha escrito con *g*. El resultado obtenido devuelve los casos de la variante fonética *palancana*.

Las búsquedas mostradas se basan en la especificación de diferentes criterios mediante la exclusión de elementos, algo especialmente útil cuando se trata de búsquedas como las que se han mostrado y en las que la combinación de distintos tipos de anotación no basta para limitar al máximo el número de resultados obtenidos. A continuación, se añaden otros ejemplos de búsquedas en las que podría aplicarse también esta función:

[form="a.+" & nform!="a.+"]

Formas originales con prótesis de *a-* cuya normalización no empieza por *a-*:
aforrado, atramado, atravesanos, alatón...

[lemma="."+ble" & pos!="AQ.+"]

Lemas acabados en *-ble* que no son adjetivos:
mueble, roble, sable

[nform= ".+ar" & pos!= "VMN.+"]
Formas normalizadas acabadas en *-ar* que no son verbos en infinitivo:
azúcar, altar, collar, titular, olivar, espaldar, pajar, nácar...

[form="^[^j]ofaina"]
Formas originales distintas a la variante normativa *jofaina*:
cofaina

4.8.3.6. Dejar huecos entre palabras

La búsqueda explícita de dos o más elementos en una sola consulta puede ayudar a delimitar el contexto en el que queremos que aparezca una palabra. Sin embargo, en ocasiones esos elementos pueden no estar próximos entre sí, sino que se encuentran separados por otras palabras formando expresiones más complejas. Por ejemplo, si interesa saber cómo eran o de qué material estaban hechos los bufetes antiguos (‘tipo de mesa’), una búsqueda en el corpus del tipo *bufete de* + sustantivo, nos dará una buena aproximación: [nform="bufete"] [nform="de"] [pos="NC.+"]⁴⁷

El resultado devolverá casos como *bufete de nogal, de caoba, de castaño, de herraje, de lino, de madera*, etc. Sin embargo, este resultado podría estar excluyendo otros muchos casos en los que al término *bufete* se le añade una valoración de cualquier otro tipo, como por ejemplo la construcción *bufete* + adj. Para ello, CQL permite hacer búsquedas dejando espacios entre palabras que representan tokens en los que no interesa especificar nada mediante el uso de corchetes vacíos:

[nform="bufete"] [] [nform="de"] [pos="NC.+"]

⁴⁷ Búsquedas de características similares son las que se han realizado para analizar el caso de los términos *bufete-bufetillo*, cuya evolución semántica y morfológica se analiza en § 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*.

en diez rreales. Vn	bufete alto de nogal	en quarenta reales. Çinco
en beinte ducados. Vn	bufete alto de nogal	en çinquenta reales. Vn
puerta christales, con su	bufete cubierto de concha	y ebano, en trescientos
. n 60 It un	bufete forrado de paño	azul en ciento y cincta
baqueta mediados. Yt un	bufete grande de caoba	. Yt un belon.
mrs. Yt. un	bufete grande de caoba	, bale ciento y cinco
taburetes de castaño, vn	bufete grande de caoua	, vn escritorio de madera
. n 2 It otro	bufete grande de caova	herrado, del mismo
baqueta en 0108. Un	bufete grande de gabeta	en 0024. Una mesa
y cinquenta reales. Vn	bufete grande de herrage	en ziento y veinte rs
, en tres ducados Otro	bufete grande de madera	en seis res Todos los
. n 82 Yt un	bufete grande de nogal	con sus hierros. n
quinçe rs. Ytt un	bufete grande de nogal	con sus hierros en setenta
en nobenta rs. Un	bufete grande de nogal	ochenta rrs. Una harca
rs. Ytten, un	bufete grande de nogal	con su erraxe apreziado en
y tres reales. Vn	bufete grande de pino	en quarentta reales. Vn
en ocho pessos. Vn	bufete grande de zedro	y con erraje en ocho

Figura 49: Resultado de una búsqueda con elementos opcionales en la estructura *bufete* + cualquier cosa + *de* + sustantivo

En efecto, la búsqueda devuelve un gran número de casos en los que a la construcción *bufete de* + sust. se le añade un adjetivo. La misma consulta podría hacerse añadiendo tantos huecos como se deseen. Se puede precisar el número exacto de huecos que se desean recuperar utilizando la expresión `[] {n}`. El uso de llaves indica la repetición del carácter precedente las veces que se indican:

```
[nform="bufete"] [] {3} [nform="de"] [pos="NC.+"]
bufete con sus cajones de pino, bufete de una tabla de nogal...
```

Conviene resaltar que este tipo de búsquedas devuelve solo las ocurrencias en las que hay el número exacto de elementos que se indica y no menos. Si, por el contrario, no interesa que este número sea fijo, sino que sea orientativo, pueden establecerse rangos. En este caso la búsqueda devolverá expresiones con un número variable de huecos entre palabras. Este rango se indica también entre llaves con los números separados por una coma. Esta búsqueda y otras similares a las mencionadas se muestran a continuación:

```
[nform="tanto"] [] {1,3} [nform="como"]
Construcción tanto / como de uno a tres elementos:
tanto raíces como, tanto de noche como, tanto vecinas de Vélez como...
```

```
[lemma= "ser"] [] [pos= "NCFS000"]
Búsqueda de la construcción ser + cualquier cosa + sust. femenino singular:
era una niña, es de latón, es la verdad, es de edad, es de madera...
```


[lemma= "ser"] [] [] [pos= "NCMS000"]

Búsqueda de la construcción *ser* + sust. masculino singular en la que se interponen dos elementos:

era de mi hermano, era el dicho hombre, es en su poder, es todo su valor...

[nform="más"] [] {2,4} [nform="que"]

Construcción comparativa *más que* formada de 2 a 4 elementos:

más de lo que, más bienes o deudas que, más de dos mil reales que...

4.8.3.7. Indicar elementos opcionales

En CQL es posible especificar elementos opcionales que pueden o no aparecer en una búsqueda con el uso de la interrogación ? como expresión regular justo después del elemento opcional. Este criterio puede aplicarse tanto a palabras enteras como a fragmentos de ellas o caracteres y puede ser útil cuando se quiere dejar abierto el contexto de aparición de una estructura determinada. Algunos casos prácticos se ejemplifican a continuación:

[nform="no"]? [lemma="ser"] [pos="NC.+"]

Construcción *ser* + sust. antepuesta o no al elemento “no”

fue alcalde, siendo testigos, no era dueño, no ser hora, no eran jueces...

[form="(.)tre.edes"]

Formas originales de la palabra *trébedes*, escrita con *b* o *v*, con prótesis opcional de algún elemento:

treuedes, estrebedes, extrevedes

[nform="igual"] [nform="de"]? [pos="NC.+"]

Construcción *igual* + sust. con adición opcional de la preposición:

igual madera, igual de raso, igual color, igual forma...

4.8.3.8. Calcular el número de tokens en el corpus

Saber el número de tokens en un corpus es importante para poder determinar el alcance de los resultados obtenidos en una búsqueda. Asimismo, es también decisivo a la hora de establecer frecuencias relativas en torno a la extensión de una voz o a la difusión de un fenómeno concreto, por lo que conocer este dato puede ser útil en algunos casos.

Como ya se sabe, en lenguaje CQL un token es todo aquello que se representa entre corchetes []. Una búsqueda de estas características, con corchetes vacíos sin nada en su interior, devolverá todos los tokens incluidos el corpus, independientemente de la información que contengan:

Búsqueda en el corpus	
Búsqueda en	
CQL:	

Figura 50: Búsqueda para rescatar todos los tokens en el corpus

De acuerdo con el concepto de token en TEITOK, es importante recordar que esta búsqueda no equivale al número total de palabras en el corpus (§ 4.2. *Tokenización*). Los signos de puntuación y los dígitos también conforman tokens en el corpus, por lo que la suma de todos los tokens también incluirá estos aspectos además de las palabras. Sin embargo, en el caso de que solo interese conocer el número real de palabras, se puede aplicar en CQL la expresión regular ya conocida para la negación de criterios en una búsqueda, y así excluir la puntuación y los números de esta cuantificación a partir de una búsqueda de características morfosintácticas:

[pos! = "F.+|Z"]

Tanto los signos de puntuación como los dígitos constituyen tokens con anotación gramatical (F para los signos ortográficos, Z para los números)⁴⁸, por lo que, excluyendo esta característica en el nivel de búsqueda morfosintáctica, todos los tokens que contengan alguna de estas etiquetas quedarán al margen del cómputo final. En la consulta señalada arriba se utiliza la expresión regular ! para negar lo que sigue a continuación, es decir, para excluir los tokens que cumplan con una u otra de las características mencionadas (función disyuntiva: |). Esta búsqueda equivale a pedirle al corpus todos los tokens que contienen cualquier anotación gramatical, a excepción de los que poseen las características que se especifican. El resultado será el número de palabras propiamente dichas que conforman el corpus en el momento de la consulta.

4.9. Interfaz de consulta del corpus ODE

El motor de búsqueda de ODE utiliza el lenguaje CQL para realizar distintos tipos de consultas atendiendo a diferentes niveles de complejidad e intereses. La interfaz de búsqueda avanzada no solo facilita las búsquedas cruzadas, sino que permite a usuarios no familiarizados con el funcionamiento del lenguaje CQL utilizar el buscador y realizar búsquedas precisas sin necesidad de conocer expresiones regulares. En este sentido, la interfaz hace accesible el corpus a todo tipo de público y lleva incorporado un sistema que permite “traducir” los criterios de búsqueda al lenguaje informático CQL sin tener

⁴⁸ Véase la sección “Etiquetario” en el menú principal de la web de ODE para la anotación morfológica del corpus: <http://corpora.ugr.es/ode/index.php?action=tagset>

todos los conocimientos previos sobre ello que se han mencionado en el apartado anterior.

Por defecto, la visualización de los resultados de una búsqueda se muestra en el conocido formato de concordancia KWIC (*Key Word in Context*). De forma adicional, para cada búsqueda TEITOK permite hacer recuento de frecuencias en base a distintos criterios (lema, forma original, etiqueta POS, etc.) y mostrar el resultado en diferentes formatos: tabla, gráfico circular, gráfico de barras, etc. De esta manera, TEITOK proporciona un motor de búsquedas eficaz al mismo tiempo que ofrece diferentes opciones útiles para la visualización de datos directamente obtenidos de una búsqueda en el corpus. Algunas de estas opciones de visualización se van mostrando en las páginas que siguen a medida que se detallan cada una de las opciones de consulta que permite el buscador.

4.9.1. Búsqueda simple

La consulta simple permite al usuario buscar por una o varias palabras en el cajón principal de búsqueda:



The image shows a web interface for a simple search. On the left, the text 'Búsqueda en CQL:' is displayed. To its right is a long, empty rectangular input field. Further to the right is a button labeled 'Buscar'. Below the 'Buscar' button, the word 'opciones' is written, likely serving as a link to more search options.

Figura 51: Interfaz de búsqueda simple en el corpus *ODE*

Esta consulta no admite buscar por fragmentos de palabras, por lo que la cadena de caracteres especificada en la consulta debe coincidir con palabras enteras para obtener un resultado, a no ser que en la búsqueda se emplee alguna expresión regular. Esto es así porque la búsqueda devuelve tokens, los cuales se corresponden siempre con palabras enteras y no con partes de ellas.

La búsqueda simple accede por defecto a la forma normalizada de cada token (vinculada con el atributo *@nform*), de manera que cualquier consulta debe escribirse conforme a las normas ortográficas actuales para obtener un resultado. La concordancia obtenida incluirá todas las formas ortográficas de la palabra en el corpus:

Búsqueda en

visualizar

CQL:

167 resultados • ipm: 142,94 • Showing 0 - 100 • siguientes

Text:

Etiquetas:

contexto	Lucas Perez, natural y becino de esta villa , de estado casado,	1829 Málaga
contexto	yo Ynatio de Mora, besino de esta billa , yjo de Baltasar de	1719 Huelva
contexto	dijo llamarse y ser bezino de esta billa , del qual rreziui	1764 Almería
contexto	espreso llamarse y ser bezino de esta villa , del qual,	1764 Almería
contexto	dijo llamarse y ser bezino de esta villa , del qual estand	1764 Almería
contexto	Diego Tostado, boticario, ueçino d esta villa , a quien doy fee	1687 Cáceres
contexto	, Juan Enrique Moreno, uezino de esta ua de Linares, hijo lexítimo	1765 Jaén
contexto	dha hermita Antonio Santiago, uezino de esta uilla , a qn por dho	1753 Granada
contexto	harando con Juan Relimpio, uezino de esta uilla , por uajo de las	1753 Granada
contexto	nunpzias de Barme Checa, uezino de esta uilla , otorgada entre este,	1778 Jaén
contexto	, pareçio Esteuan Saz, uezo de esta ua e natural del lugar	1608 Cáceres
contexto	Francisco Manjon, veçino d esta va de Vacarrillo, como mejor	1705 Jaén
contexto	efetto de Diego Galuan, veçino d esta villa , como depositario q es	1675 Cáceres

Figura 52: Búsqueda simple de la estructura “vecino de esta villa”

4.9.2. Búsqueda avanzada

Este tipo de búsqueda permite recuperar información a partir de cualquiera de los niveles de edición del corpus (formas originales, formas expandidas en el caso de abreviaturas, formas normalizadas, etiquetas gramaticales y lemas). Adquiere mayor complejidad, ya que posibilita multitud de combinaciones entre los distintos niveles permitiendo realizar búsquedas cada vez más sofisticadas.

La búsqueda avanzada permite además filtrar los datos de acuerdo con variables de carácter extratextual, es decir, aquella información que forma parte de los metadatos del documento como pueden ser la fecha o la localización geográfica del manuscrito. Su interfaz (Figura 53) está pensada para combinar en una sola búsqueda información de distinto tipo: por un lado, los datos puramente lingüísticos relacionados con el contenido de los textos (bloque izquierdo); por otro, las cuestiones extratextuales de cada documento alojado en el corpus (bloque derecho):

Figura 53: Interfaz de búsqueda avanzada del corpus *ODE*

El cruce de cuestiones textuales y extratextuales favorece el análisis multidisciplinar de los datos y enriquece enormemente las opciones de búsqueda, aumentando de manera exponencial las posibilidades de explotación del corpus. La interfaz facilita la búsqueda cruzada y permite filtrar los resultados en base a diferentes variables.

En cuanto a la información de carácter textual, la búsqueda avanzada permite buscar por cualquiera de las formas de una palabra, esto es, su forma ortográfica original, su forma expandida y su forma normalizada, así como por su etiqueta POS o su lema (vinculados respectivamente todos estos aspectos a los atributos *@form*, *@fform*, *@nform*, *@pos* y *@lemma* comprendidos en cada token).

El sistema permite buscar por palabras enteras o por fragmentos de ellas, lo que se puede indicar mediante el desplegable que aparece junto a cada uno de los campos de búsqueda (“igual a”, “empieza por”, “termina en”, “contiene”). Por defecto, el buscador marca la opción “igual a”, que tendrá resultado si el valor introducido coincide con una palabra real en el corpus. Los campos de búsqueda permiten recuperar la información lingüística asociada a cada uno de los elementos de que consta un token en el corpus, esto es:

- Forma transcrita: permite buscar por la forma exacta de la palabra tal y como aparece originalmente en el manuscrito: *cassa, yglesia, muxeres...*
- Forma expandida: permite buscar por la expansión de una palabra cuando esta ha sido abreviada por el escribano: *çiudad, principalmente, becino...*
- Forma normalizada: ofrece resultados por forma estandarizada conforme a las normas actuales de ortografía. El resultado obtenido devolverá todas las formas originales asociadas a la forma estándar: *casa, iglesia, mujeres, vecino...*
- Etiqueta POS (*part of speech*): permite la búsqueda por palabras gramaticales según la etiqueta morfosintáctica (§ 4.5.1. *Etiquetario*): *NCFS000, AQ0MS0, VMP0000...*

- e. Lema: permite la búsqueda por forma lematizada. Esta opción permite obtener como resultado todas las formas normalizadas asociadas a un mismo lema: *haber, mujer, vecino...*

En cuanto a la búsqueda por etiqueta POS, la propia interfaz de búsqueda facilita un constructor de etiquetas que sirve como asistente para la construcción de estas. Al clicar en esta opción se abre un desplegable que permite al usuario seleccionar los rasgos morfosintácticos correspondientes a la etiqueta deseada. En la imagen de abajo se muestra un ejemplo para obtener todos los casos del pronombre *la* en función de complemento indirecto. En este ejemplo, la categoría principal será la de pronombre (P), para la que deberán marcarse el resto de las características según como sigue (nótese que debe marcarse el caso dativo para obtener casos de laísmo):

Construcción de Etiquetas: Etiqueta POS

Etiqueta principal: Pronombre

tipo: Personal

persona: tercera

género: femenino

número: singular

caso: dativo

poseedor: [cualquiera]

polaridad: [cualquiera]

Insertar Agregar

Figura 54: Constructor de etiquetas en la interfaz de búsqueda

Al clicar en el botón inferior “Insertar”, el sistema generará automáticamente la etiqueta morfosintáctica apropiada (PP3FSD00). El resultado será una concordancia con los casos de laísmo en el corpus:

mill y quinientos ducados que	la	mando esten en poder de	1644 Badajoz
que era una desbergonzada y	la	cruzaria la cara, a	1726 Granada
Palomas y la envistio dicendola : <i>Mira picara, / que</i>			1784 Granada
no poder detenerse,	la	dixo espresase a su padre	1783 Málaga
aobadas que son las que	la	dejo legadas don Thomas Antonio	1750 Madrid
los bienes y hazienda que	la	dejo por la fin y	1751 Madrid
, por el amor que	la	tiene y por auer estado	1651 Badajoz
con el a lo que	la	contesto que si estaba asi	1818 Burgos
trumento contundente asismismo	la	allo otras dos leves contusiones	1809 Valladolid
el, con tal que	la	ottorgue carta de pago y	1712 Cádiz
de todos los bienes que	la	han pertenecido por la legitima	1783 Jaén
la dha mi esposa,	la	mando y dono en arras	1783 Jaén
. Y los bienes que	la	van adjudicados para su pago	1783 Jaén

Figura 55: Ejemplos de laísmo como resultado de búsqueda por etiqueta gramatical

Con respecto a la información de carácter extratextual, la búsqueda avanzada permite recuperar datos clasificados en torno a diferentes metadatos que forman parte de la cabecera de cada documento. En ella se incluye información referente al título, fecha y lugar de creación, tipología documental, localización archivística, descripción de la fuente o editores responsables de la edición, entre otros elementos.

El buscador de *ODE* permite al usuario filtrar los datos de acuerdo con los siguientes aspectos extratextuales:

a. *Título*

Permite buscar por una o varias palabras clave integrantes en el título de los documentos. Por ejemplo, la búsqueda ‘dote’ en este campo devolverá todos aquellos documentos en los que la palabra *dote* forma parte del título. Esta búsqueda posibilita saber cuántos de los inventarios alojados en el corpus se corresponden con dotes. Lo mismo podría aplicarse a otras palabras clave como los testamentos, embargos, almoneda o repartos de herencia. El resultado genera una tabla como la que se muestra abajo donde se pueden ver qué textos cumplen con el criterio establecido en la búsqueda:

Título	Año	Lugar	Provincia	Tipo textual
Carta de dote y arras de Catalina García	1709	España, Almería, Tabernas	Almería	inventario de bienes
Carta de dote de Ana María Guirado Plaza	1822	España, Almería, Tabernas	Almería	inventario de bienes
Capital y dote de Joaquín de Vilches en favor de Ana Martínez Vilches	1844	España, Almería	Almería	inventario de bienes
Dote de Leonor Rodríguez	1751	España, Madrid, Brunete	Madrid	inventario de bienes
Carta de dote de Paula Guerrero	1702	España, Málaga, Frigiliana	Málaga	inventario de bienes
Inventario de los bienes que recibe Antonio de la Vega como dote por su casamiento con María Salazar	1717	España, Madrid, Torrelaguna	Madrid	inventario de bienes
Carta de dote de D. ^a Francisca Caballero	1745	España, Madrid	Madrid	inventario de bienes

Figura 56: Documentos correspondientes a dotes. Búsqueda avanzada por ‘título’

b. *Año*

Permite establecer un rango cronológico para limitar las búsquedas de datos en el corpus. En este campo es obligatorio especificar una fecha concreta, de inicio y final, o bien solo de inicio o solo de final, para acotar la búsqueda.

Por ejemplo, la consulta 1) de la Figura 57 recupera documentos en un abanico de cien años, desde el 1750, hasta el 1850, incluidos. En el caso que solo interese

obtener datos para un año en concreto, se deberá escribir la misma fecha en las dos casillas separadas por guion, tal y como se muestra en 2).

Por otro lado, si lo que interesa es simplemente marcar un límite temporal a la hora de obtener resultados, esto puede hacerse proporcionando una sola fecha en cualquiera de las dos casillas del campo de búsqueda; esto tendrá efectos diferentes. En el caso de 3), la concordancia obtenida incluirá los resultados posteriores a la fecha indicada, es decir, desde el año 1750, incluido, hasta el último año del que se tienen datos en el corpus. En el caso 4), los datos resultantes pertenecerán al año 1750 y anteriores, hasta la fecha más temprana para la que el corpus registra datos.

1) Año 1750 - 1850

2) Año 1750 - 1750

3) Año 1750 -

4) Año - 1750

Figura 57: Búsqueda por rangos cronológicos, anteriores o posteriores a una fecha concreta

c. Lugar

Permite filtrar los resultados de acuerdo con la variable relativa al lugar de procedencia del texto. Este campo es de libre cumplimentación, lo que permite al usuario especificar un lugar concreto según sus preferencias, entre las que se pueden incluir provincias o también municipios. La búsqueda, por ejemplo, de Aranjuez, Pinto o de Navalafuente, todos ellos municipios de Madrid, devolverá los documentos registrados en el corpus originarios de estos lugares.

Sin embargo, conviene tener en cuenta que la búsqueda de ‘Madrid’ en este campo con el fin de filtrar solo documentos del municipio de Madrid, esto es, de su capital, no tendrá este resultado. En cambio, nos devolverá los resultados de toda la provincia incluidos también sus pueblos. Para obtener los textos solo pertenecientes al municipio de la capital deberá especificarse en la búsqueda la provincia y el municipio separados por una coma, tal y como se muestra a continuación:

Lugar Madrid, Madrid

Figura 58: Búsqueda de documentos pertenecientes al municipio de Madrid

d. Provincia

La búsqueda en este campo devuelve el total de documentos pertenecientes a una provincia, por lo que es una búsqueda más amplia que la anterior por ‘lugar’. Este tipo de búsqueda ofrece un desplegable de las provincias representadas en el corpus entre las que es posible escoger solo una de ellas.

e. Tipo textual

En este campo es posible elegir entre las diferentes tipologías que aloja el corpus, esto es, certificados médicos, declaración de testigos, inventarios de bienes o aquellos recogidos bajo la categoría “Otros”. Entre estos últimos se encuentran algunos documentos que no se incluyen en alguna de las categorías anteriores, como por ejemplo cartas.

La búsqueda por tipo de texto es especialmente útil si solo interesa el estudio de una única tipología documental, pero también si lo que se pretende es detectar diferencias o similitudes entre diferentes tipos de documentos. Filtrar los datos por medio de esta selección puede ser especialmente valioso para estudiar el posible cambio lingüístico sujeto a tradiciones discursivas diferentes. Así, una misma búsqueda en la que se incluyen varios tipos textuales puede arrojar datos reveladores en cuanto a la comparabilidad de los resultados.

f. Proyecto

ODE es el resultado del trabajo de diferentes proyectos de investigación entre los que es posible también filtrar por documentos adscritos a los mismos. Concretamente, los inventarios de bienes de esta tesis se vinculan con los proyectos *CORDEREGRA* y *ALEA XVIII*, por lo que filtrar por proyectos devolverá documentación diferenciada.

g. Siglo

Permite filtrar los datos mediante el criterio cronológico del siglo al que pertenecen. La búsqueda avanzada en este campo admite seleccionar uno de los siglos para los que hay datos en *ODE*, desde el XVI al XIX. En aquellos casos en los que interesa recabar datos de periodos mayores, la búsqueda por año, mencionada más arriba, resulta más adecuada y específica.

4.9.3. Búsqueda comparada

La búsqueda comparada permite realizar dos o más búsquedas simultáneas con las que comparar hasta tres formas diferentes. La interfaz de búsqueda ofrece una tabla en la que se debe proporcionar un nombre de consulta para cada una de las opciones y la búsqueda propiamente dicha formulada en sintaxis CQP. Esta opción es útil para analizar variación léxica en el corpus. Por ejemplo, se puede obtener la distribución en el corpus de las distintas denominaciones de un mismo objeto:

Nombre de la búsqueda	Búsqueda en CQL
Jofaina	[lemma="jofaina"]
Zafa	[lemma="zafa"]
Palangana	[lemma="palangana"]

Figura 59: Ejemplo de búsqueda comparada en el corpus ODE

El resultado de esta búsqueda ofrece un recuento estadístico para el que es posible elegir el tipo de visualización entre varias opciones (tabla, gráfico, barras, etc.). El usuario también puede descargar los datos obtenidos en diferentes tipos de archivos.

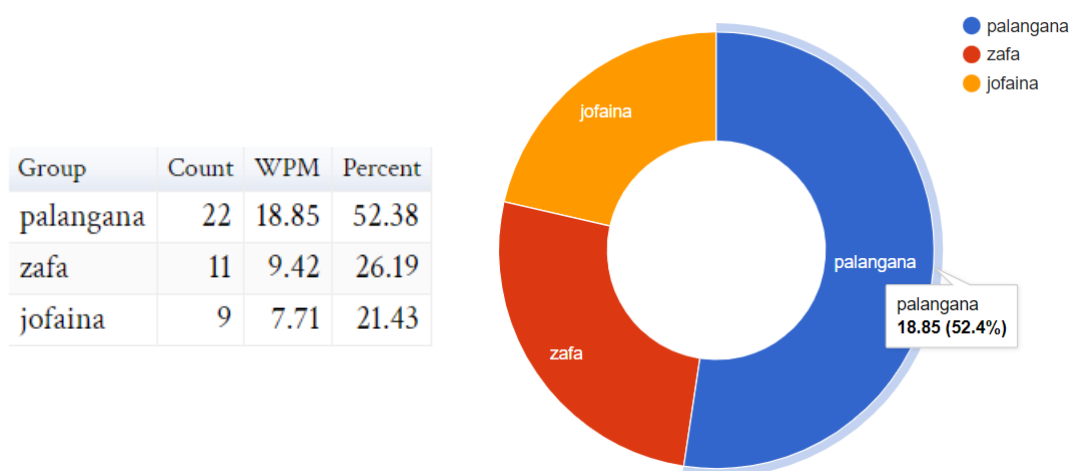


Figura 60: Distribución de las variantes *zafa*, *jofaina* y *palangana*. Búsqueda comparada

4.9.4. Búsqueda en el mapa

TEITOK incorpora la posibilidad de geolocalizar automáticamente los documentos del corpus añadiendo a cada uno de ellos las coordenadas geográficas del lugar del que proceden. Esta función permite visualizar sobre el mapa el número de documentos vinculados a un lugar, pero sirve también para realizar búsquedas cuyos resultados puedan verse reflejados en un mapa y no como una concordancia. El sistema facilita el cartografiado automático de los datos, lo que permite observar a golpe de vista la distribución de fenómenos lingüísticos en el espacio.

Este tipo de búsquedas sobre el mapa son especialmente útiles para visualizar fenómenos de variación diatópica y delimitar el uso de determinadas variantes ortográficas, fonéticas o léxicas, según interese. Por ejemplo, la palabra *delantal*, que aparece con cierta frecuencia en los inventarios, tuvo una variante ya en desuso (*devantal*) que el corpus permite delimitar geográficamente de manera muy clara.

Con una interfaz de búsqueda similar a la ya mostrada para las búsquedas comparadas (Figura 59), la consulta de las formas *devantal* / *delantal* ofrece la siguiente distribución diatópica:

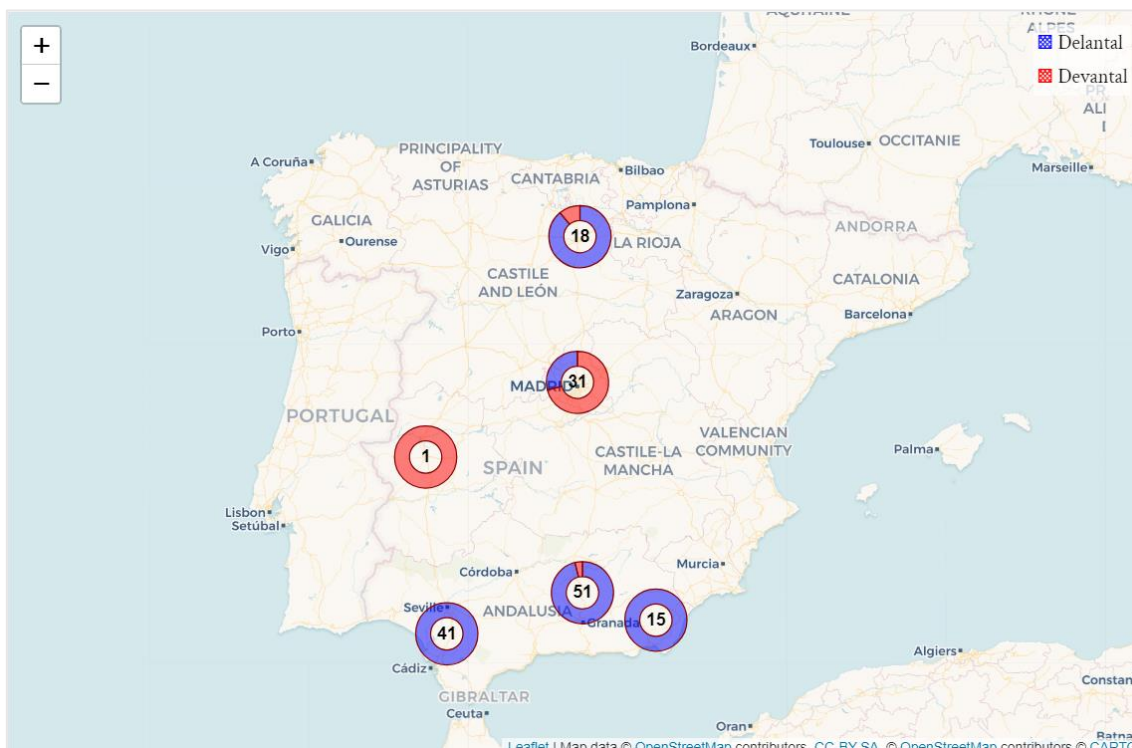


Figura 61: Búsqueda en el mapa. Distribución diatópica de las formas *delantal* y *devantal*

Bajo la forma *devantal* se han agrupado las variantes gráficas *avantal*, *devantal*, *avental*, *debantar*, *deuantal*, etc., mientras que bajo *delantal* se incluyen las variantes fonéticas *delantar*, *delantal*. El resultado sobre el mapa ofrece una clara distribución diatópica de la variante *devantal*, circunscrita mayoritariamente a inventarios madrileños y burgueses. De forma aislada se registra un caso en Cáceres (año 1613) y un par de ellos en Jaén⁴⁹.

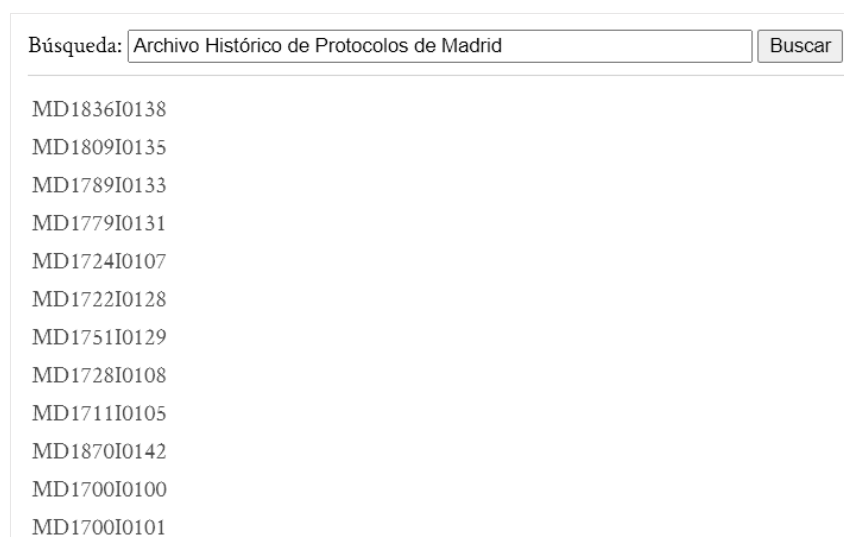
Con esta búsqueda se demuestra la utilidad de los mapas para representar variantes consideradas dialectales. El ejemplo en este caso refleja un fenómeno de variación fonética, pero la misma búsqueda comparada de *palangana*, *zafa* y *jofaina* mencionada más arriba podría plasmarse también sobre el mapa para delimitar el uso de diferentes variantes léxicas.

4.9.5. Búsqueda genérica

La búsqueda genérica permite recuperar cualquier tipo de información alojada en el corpus, ya sea sobre los datos que conforman la cabecera del documento o sobre el

⁴⁹ La variante *devantal* ha sido estudiada a raíz de documentos del *CorLexIn*, en los que el término se vincula geográficamente en el siglo XVII sobre todo con la franja nororiental de la Península (Miguel Borge, 2018, p. 83; Bartolomé y Junquera, 2018, p. 130). Los datos de *ODE* confirman en su mayoría esta distribución y añaden vestigios del uso del término también en otras zonas como Cáceres y Jaén.

cuerpo del texto. Resulta útil cuando se quieren obtener resultados en base a alguno de los criterios específicos que no se contemplan en la interfaz de búsqueda avanzada como, por ejemplo, el archivo al que pertenecen los manuscritos o el nombre del escribano como elementos extratextuales. Este tipo de búsqueda permite recuperar el conjunto de los documentos procedentes del Archivo Histórico de Protocolos de Madrid, ya que esta información se encuentra en todos los documentos que proceden de Madrid:



The image shows a web-based search interface. At the top, there is a search bar with the text "Búsqueda:" followed by a text input field containing "Archivo Histórico de Protocolos de Madrid". To the right of the input field is a button labeled "Buscar". Below the search bar, a list of document identifiers is displayed, each on a new line. The identifiers are: MD1836I0138, MD1809I0135, MD1789I0133, MD1779I0131, MD1724I0107, MD1722I0128, MD1751I0129, MD1728I0108, MD1711I0105, MD1870I0142, MD1700I0100, and MD1700I0101.

Búsqueda:	Buscar
MD1836I0138	
MD1809I0135	
MD1789I0133	
MD1779I0131	
MD1724I0107	
MD1722I0128	
MD1751I0129	
MD1728I0108	
MD1711I0105	
MD1870I0142	
MD1700I0100	
MD1700I0101	

Figura 62: Búsqueda genérica. Documentos del Archivo Histórico de Protocolos de Madrid

Además de buscar contenido de tipo extratextual, la búsqueda genérica permite obtener información relacionada con la marcación textual de los documentos. Por ejemplo, pueden recuperarse aquellos documentos en los que el escribano ha hecho alguna restitución en el texto. Esta característica se marca con la etiqueta `<subst>` (*substitution*), por lo que su búsqueda devolvería la lista de documentos en el corpus que incluyen una cancelación y una restitución por parte del escribano.

Cabe destacar que la búsqueda genérica devuelve solo el conjunto o listado de los documentos que incluyen la característica deseada, pero no los casos concretos dentro de estos documentos, para esto último resulta útil la búsqueda con *XPath* que se explica a continuación.

4.9.6. Búsqueda con *XPath*

Este tipo de búsqueda funciona utilizando el lenguaje informático *XPath* para recuperar información a partir de la estructura jerárquica de los documentos XML. A diferencia de la búsqueda genérica, recupera las palabras o fragmentos que han sido marcados con una etiqueta determinada. En lenguaje *XPath*, se emplea la doble barra inclinada `//` seguida de la marcación que se desea recuperar. Por ejemplo, para recuperar todos los casos de texto restituido, marcados con la etiqueta `<subst>`, bastará una búsqueda como la siguiente:

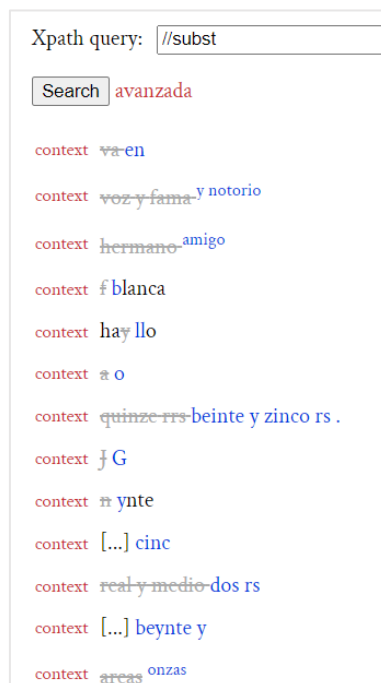


Figura 63: Búsqueda con lenguaje *XPath*. Casos de restituciones marcadas con <subst>

4.9.7. Distribución de palabras en el corpus

Esta opción permite conocer estadísticas actualizadas de manera automática de los datos que contiene el corpus en el momento de la consulta. El usuario puede obtener recuentos generales de palabras en función de diferentes criterios extralingüísticos. Así, es posible consultar el número de palabras contenidas en el corpus según el origen geográfico, el siglo, la tipología textual o el proyecto de investigación al que se adscriben. Por ejemplo, en el momento actual (agosto 2023), la distribución por tipo textual y fecha es como sigue:

Palabras por Siglo

XVIII	574,430
XVII	270,068
XVI	211,754
XIX	110,557
XV	421

Palabras por Tipo textual

inventario de bienes	828,536
declaración de testigos	250,146
certificado médico	72,782
otros	15,766

Figura 64: Distribución de palabras en el corpus por siglo y tipo textual

Conviene mencionar que la distribución de palabras hace en realidad una distribución por tokens. Los tokens en el corpus no tienen por qué coincidir únicamente con palabras reales, puesto que en TEITOK la consideración de token incluye además los signos de puntuación y los números (§ 4.2. *Tokenización*). En estas tablas, por tanto, se ofrece el número total de tokens en el corpus, los cuales pueden corresponderse tanto con

Nº	Trigrama
321	reales de vellón
207	forma de derecho
115	pares de medias
112	lienzo de lino
110	reales de plata
108	sábanas de lienzo
81	madera de pino
80	señal de cruz
73	tablas de manteles
69	pares de calcetas
68	medias de seda
67	tabla de manteles
64	almohadas de lienzo
61	capa de paño
56	varas de lienzo

Figura 66: Trigramas más frecuentes de la combinación *sustantivo + de + sustantivo*

4.10. Revisión manual de errores en el corpus

El estudio de los diminutivos llevado a cabo en esta tesis se sirve precisamente de la anotación morfosintáctica del corpus para la extracción de los datos analizados, por lo que ha sido indispensable comprobar la eficacia del etiquetado tras la ejecución automática en TEITOK del programa NeoTag. De manera general, la revisión se extiende también al resto de procesos informáticos descritos (tokenización, normalización, anotación morfosintáctica y lematización), de forma que se pueda garantizar la fiabilidad de la totalidad del corpus para su utilización en estudios históricos de distinta naturaleza.

La comprobación se ha basado en la lectura atenta de cada documento, el cual se ha sometido a una doble revisión desarrollada en dos fases: 1) revisión tras la tokenización y normalización ortográfica. En esta primera lectura se han comprobado además los posibles errores de transcripción una vez que el archivo XML-TEI ha sido ya importado a la plataforma TEITOK; y 2) una segunda revisión tras la asignación automática de etiqueta morfosintáctica y lema con el programa NeoTag. Puesto que el etiquetado y lematización son procesos simultáneos, su revisión se ha hecho también de manera conjunta.

Especialmente laboriosa ha sido la revisión de la anotación lingüística del corpus (etiqueta morfosintáctica y lema). La corrección de lemas en el corpus se ha realizado de manera paralela a la revisión del etiquetado morfosintáctico. Este modo de obrar se justifica en base a dos razones principales. La primera de ellas tiene que ver con la estrecha vinculación etiqueta-lema del anotador, como ya se explicó en § 4.4 *Funcionamiento del anotador lingüístico en TEITOK*. La adjudicación automática del lema depende de la anotación lingüística de la palabra, de manera que muchos de los

errores en la lematización se explican por una etiqueta morfosintáctica mal asignada. La segunda razón, muy en relación con la anterior, se explica por una cuestión de pragmatismo: un alto porcentaje de fallos en los lemas suponen también errores en la etiqueta, por lo que se ha considerado oportuno realizar la corrección a la vez. Esto permite modificar ambos aspectos mediante una sola acción, mientras que revisar etiquetas y lemas separadamente implica en numerosas ocasiones tener que modificar información de la misma palabra varias veces.

La eficacia del anotador y la revisión manual de los textos anotados automáticamente son aspectos estrechamente vinculados entre sí: la revisión permite mejorar el porcentaje de acierto del anotador y, a su vez, del éxito del anotador automático depende la buena anotación de los textos en los que cada vez sea menos necesaria una revisión exhaustiva. Por ello, la revisión manual del corpus es fundamental no solo para garantizar la ausencia de errores, sino para aumentar los datos de los corpus de entrenamiento con los que funciona tanto el normalizador como el anotador y mejorar el éxito de estos tras los procesos automáticos. Solo así será posible ir reduciendo progresivamente el porcentaje de error y, en consonancia, el tiempo dedicado a la revisión manual. Los fallos en la anotación fueron mucho más numerosos en los inicios de creación del corpus, pero disminuyeron en fases finales.

La eficacia del procesamiento informático del corpus está ligada a su tamaño, mejor cuanto mayor es su volumen, pero tan importante como aumentar el número de textos lo es la revisión de los mismos. En lo que respecta a la anotación morfosintáctica, aspecto clave para esta tesis, mejorar la calidad del corpus de entrenamiento es otro de los factores que se han tenido en cuenta para conseguir un corpus enteramente anotado y sin fallos en la anotación morfosintáctica de palabras con sufijación diminutiva. Errores en la adjudicación de etiquetas pasarían a formar parte de los parámetros del corpus de entrenamiento provocando fallos recursivos cada vez que se realiza la anotación automática de nuevos textos. Por esta razón, se exige necesariamente una revisión manual que garantice la calidad del etiquetado morfosintáctico por medio del programa NeoTag. Las revisiones se han realizado documento a documento tras la ejecución de la anotación automática y esto ha evitado acumular en el corpus mismos errores un alto número de veces. La depuración de errores en la anotación ha ido acompañada además de constantes actualizaciones del corpus de entrenamiento a fin de incorporar los cambios efectuados a los parámetros existentes y contribuir a la mejora del anotador.

La revisión de los textos ha sido una de las tareas que más tiempo y esfuerzo ha consumido de entre las distintas fases de elaboración del corpus. No obstante, TEITOK proporciona un sistema cómodo y rápido de edición que permite modificar la información relativa a cada token sin necesidad de acceder al código interno del archivo XML-TEI cada vez que se detecta un fallo. Una de las principales utilidades de TEITOK es que el propio motor de búsquedas del corpus sirve para la revisión y corrección de errores (Janssen, Ausensi y Fontana, 2017, p. 4), por lo que conviene conocer el funcionamiento del buscador y las posibilidades de las expresiones regulares para realizar cambios en la edición del corpus (§ 4.8. *Búsquedas en el corpus basadas en lenguaje CQL*). Los distintos métodos para depurar errores aplicados a casos reales en el corpus se describen a continuación.

4.10.1. Modificar la información asociada a una palabra

Los errores puntuales detectados en el texto pueden ser corregidos fácilmente a través de una plantilla generada en la interfaz al clicar sobre la palabra correspondiente. En TEITOK, el posicionamiento del ratón sobre cualquiera de las palabras del texto genera una ventana emergente con la información relativa a cada token (Figura 67). Solo para los creadores del corpus está habilitada la opción de editar las palabras clicando sobre ellas. De esta manera, se abre una nueva ventana de edición en la que se podrán hacer modificaciones en cualquiera de los niveles de edición del token (Figura 68). Las correcciones realizadas por medio de esta plantilla se incorporarán al archivo fuente del documento, que se actualizará automáticamente una vez que se hayan guardado los cambios realizados.

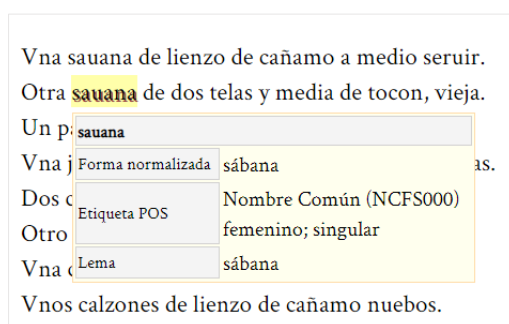


Figura 67: Selección del token en el texto

Token value (w-103): sauana	
pform Transcription (Inner XML)	sauana
form Transcribed form	
fform Expansion	
dipl Expanded form	
nform Normalized form	sábana
pos POS tag	NCFS000
lemma Lemma	sábana

Figura 68: Plantilla de edición de tokens en TEITOK

Esta plantilla permite realizar cambios en cualquiera de los niveles de edición (forma original, forma expandida, forma normalizada, etc.), pero para el caso de la anotación morfosintáctica presenta una ventaja más. Familiarizarse con el etiquetario utilizado conlleva tiempo y supondría aprender las etiquetas de memoria o bien ir consultando constantemente el etiquetario. Para evitarlo, la plantilla ofrece un constructor de etiquetas que, mediante un desplegable, permite elegir los distintos atributos asociados al contenido morfosintáctico del token, al igual que la búsqueda por etiqueta morfosintáctica (§ 4.9.2. *Búsqueda avanzada*). Además de la inserción de etiquetas por medio del constructor, la plantilla ofrece otra posibilidad a través de la función que en la jerga informática se conoce como *lookup*. Esta función busca si la palabra está presente más veces en el corpus y ofrece un resultado en forma de tabla del número total de

registros con los valores morfosintácticos que se le asocian. De entre las etiquetas encontradas se seleccionará la que coincida con el caso concreto que se quiere editar. En la Figura 69 se ofrece un *lookup* de la palabra *dicho*, la cual aparece en el corpus con cuatro anotaciones diferentes. Bastará seleccionar una de estas para elegir la etiqueta deseada.

Value Lookup		
Click to select pos for form = dicho		
Value	Display	Corpus count
AQ0MS0	Adjective; Qualitative; 0; masculine; singular; 0	3042
VMP0000	Verb; Main; participle; 0; 0; 0; 0	13
VMP00SM	Verb; Main; participle; 0; 0; singular; masculine	3
NCMS000	Noun; Common; masculine; singular; 0; 0; 0	3

Figura 69: Función *lookup* para la palabra *dicho*

Estas opciones agilizan la asignación de etiquetas morfosintácticas y facilitan el proceso de revisión a través de una interfaz intuitiva. Para facilitar la comprobación manual de etiquetas y lemas, la lectura de los textos se ha realizado seleccionando la visualización de las etiquetas morfosintácticas y las formas lematizadas para cada palabra en el documento (§ 4.7. *Triple visualización de los textos*).

Los cambios realizados a partir de este procedimiento de edición individual afectan únicamente al token seleccionado dentro del documento, independientemente de que la misma palabra se localice otras veces en el mismo texto o en otros en el corpus.

Este método de edición ha sido el habitual para la subsanación de errores considerados puntuales durante la comprobación de la anotación o cuando los fallos detectados no son recurrentes, es decir, cuando no se trata de palabras con elevada presencia en el corpus. Sin embargo, no es una práctica eficaz para aplicar cambios a mayor escala en palabras mal anotadas que se repiten en el corpus y donde se ven involucrados muchos documentos. Para estos casos, errores sistemáticos pueden corregirse de manera genérica utilizando otros procedimientos como los que se exponen a continuación.

4.10.2. Corregir errores en bloque mediante búsquedas en el corpus

TEITOK permite realizar cambios a gran escala utilizando el propio sistema de búsquedas del corpus, de modo que es posible corregir errores que afectan a un gran número de tokens realizando una sola acción. Este procedimiento permite aplicar un mismo cambio un número indeterminado de veces sin la necesidad de hacerlo documento a documento, lo que agiliza sobremanera las revisiones manuales. El sistema de edición individual de un solo token (descrito arriba) se ha sustituido por las comprobaciones en bloque a partir de búsquedas en el caso de errores que se sabe son recurrentes, especialmente en el nivel de la anotación morfosintáctica.

La palabra *cobre* ha sido reiteradamente mal anotada como forma del verbo *cobrar* en numerosas ocasiones. En un corpus de inventarios, es razonable pensar que, en la mayoría de los casos, el término hará referencia al metal del que están fabricados ciertos objetos y no a la acción de ‘cobrar’. Sin embargo, en los parámetros de NeoTag entrenados con el corpus *Post Scriptum* (§ 4.4 *Funcionamiento del anotador lingüístico en TEITOK*) era más habitual la forma verbal. Para corregir esta fallo en *ODE*, se ha utilizado la búsqueda normalizada [nform= "cobre"] para recuperar todos los casos del término. A partir de la concordancia obtenida se ha utilizado la función de edición múltiple de tokens para modificar todas las etiquetas simultáneamente (Figura 70). Esta función ofrece de nuevo la ventana de edición de la que ya se ha hablado más arriba. Los nuevos cambios se aplican a los casos seleccionados: todos los resultantes de la búsqueda o solo a aquellos en los que se desee aplicar el cambio (Figura 71).

contexto	Un cazo chiquito biejo de cobre en 2. Una espetera	1835 Almería
contexto	de cuero. <u>Hierro</u> , cobre y lata Un belon con	1845 Almería
contexto	pantallas. Un cazo de cobre . Dos planchas. Unas	1845 Almería
contexto	cristal; dos cazos de cobre y dos cuajaderas de id	1848 Almería
contexto	16. Un brasero de cobre y tarima de madera en	1882 Almería
contexto	67. Una cuajadera de cobre en una peseta. n	1882 Almería
contexto	n 70. Otra de cobre en cincuenta centimos. n	1882 Almería
contexto	. Una caldera grande de cobre para doce cantaros en cuarenta	1882 Almería
contexto	94. Una perola de cobre en una. n 95	1882 Almería
contexto	Dos cuajaderas, una de cobre [...]. Dos cazos de azofar	1896 Almería
contexto	. Una olla grande de cobre . Vidriado basto de cocina	1896 Almería
contexto	res. Vn brasero de cobre en seys res. Vna	1550 Almería
contexto	Casquer. Vna cantinplora de cobre . El qual dho ynbentario	1630 Almería
contexto	; mas vn caço de covre pequeño; vna cuçhara de	1726 Almería

Use this query for multi-token edit 

Figura 70: Búsquedas en el corpus de la forma *cobre* para la corrección de errores

form	Transcribed form	
fform	Expansion	
dipl	Expanded form	
nform	Normalized form	

pos	POS tag	NCMS000
lemma	Lemma	cobre

File ID	Sel.	Left context	Match	Right context
xmlfiles/ode/corpus/AL1714I0003.xml	☒	cántaros . Una cetra de cobre a medio servir . Un		
xmlfiles/ode/corpus/AL1721I0005.xml	☒	de moda . Peltre y cobre Ítem dos pedazos , digo		
xmlfiles/ode/corpus/AL1721I0005.xml	☒	89 Ítem dos cuajaderas de cobre con sus tapaderas . número		
xmlfiles/ode/corpus/AL1721I0005.xml	☒	90 Ítem dos cubiletes de cobre con sus tapaderas . número		
xmlfiles/ode/corpus/AL1721I0005.xml	☒	98 Ítem tres calderas de cobre , la una grande y		
xmlfiles/ode/corpus/AL1721I0005.xml	☒	100 Ítem un caldero de cobre mediano . número 101 Ítem		
xmlfiles/ode/corpus/AL1882I0044.xml	☒	16 . Un brasero de cobre y tarima de madera en		
xmlfiles/ode/corpus/AL1882I0044.xml	☒	67 . Una cuajadera de cobre en una peseta . número		
xmlfiles/ode/corpus/AL1882I0044.xml	☒	número 70 . Otra de cobre en cincuenta céntimos . número		
xmlfiles/ode/corpus/AL1882I0044.xml	☒	. Una caldera grande de cobre para doce cántaros en cuarenta		
xmlfiles/ode/corpus/AL1882I0044.xml	☒	94 . Una perola de cobre en una . número 95		
xmlfiles/ode/corpus/AL1896I0045.xml	☒	Dos cuajaderas , una de cobre . Dos cazos de azófar		
xmlfiles/ode/corpus/AL1896I0045.xml	☒	. Una olla grande de cobre . Vidriado basto de cocina		
xmlfiles/ode/corpus/AL1550I2675.xml	☒	reales . Un brasero de cobre en seis reales . Una		
xmlfiles/ode/corpus/AL1630I2662.xml	☒	Casquer . Una cantimplora de cobre . El cual dicho inventario		

Change selected select all • back to search 

Figura 71: Edición múltiple de tokens mediante selección de casos

En algunos casos, generar una concordancia que devuelva todas las ocurrencias del error que se pretende subsanar exige búsquedas más complejas que la que se ha mencionado para el caso de la palabra *cobre*. Para ello será especialmente útil el uso de expresiones regulares que afinen el resultado de las búsquedas. Puede incluso que para localizar en masa ciertos errores y modificarlos ya no baste con encontrar un solo token, sino también el que le sigue, o que sea necesario realizar búsquedas en las que se combinen varios niveles de edición del token. Para todos estos supuestos el sistema CQP cumple las necesidades mencionadas. Este tipo de búsquedas han sido una práctica muy común para corregir en bloque la anotación morfosintáctica de casos como el ejemplo que se menciona a continuación.

En el corpus, la palabra *traído* se emplea frecuentemente para indicar la calidad de la ropa⁵¹, más que como participio, por lo que se ha anotado como adjetivo al igual que el resto de su paradigma *traída/-as/-os*. No obstante, también es frecuente encontrar esta palabra en los inventarios en un contexto como el que sigue: “los bienes que ha traído al matrimonio”. Se observó que los casos como este último, en el que *traído* funciona como participio, eran mal anotados como adjetivos como consecuencia de los anteriores ejemplos de *traído* en esa función, mucho más numerosos en el corpus. Para

⁵¹ Traído/a: dicho principalmente de la ropa, usada, gastada, que se va haciendo vieja (*DLE*, s.v.)

corregir estos casos se realizó una búsqueda combinada en la que fuera posible especificar el contexto anterior, en este caso el auxiliar *haber*, para localizar exclusivamente aquellos casos objeto de modificación. La búsqueda [lemma = "haber"]@[nform = "traído"]⁵² devolvió todos los casos deseados. La búsqueda por lema engloba además cualquier forma del verbo auxiliar, de modo que es posible modificar todas las ocurrencias de *haber* + *traído* sin necesidad de probar varias combinaciones. Posteriormente y de manera inversa, es posible comprobar con una expresión regular [nform = "traíd.+"] que efectivamente están anotados como adjetivos aquellas formas que deben serlo. El mismo procedimiento se realizó con otras palabras que presentan también este tipo de ambivalencia: *andado*, *usado* o *servido*.

En definitiva, este método de revisión por medio de búsquedas CQP en los casos como los que se han mencionado ha sido el procedimiento habitual para agilizar hasta donde fuese posible la modificación de errores que afectan a un número considerable de tokens.

4.10.3. Verificar la anotación y corregir etiquetas mal construidas

Hasta ahora se han descrito distintos procedimientos con los que se ha revisado la anotación morfosintáctica cuando el error consistía en una etiqueta mal asignada. Aunque mucho menos frecuentes, también son usuales los errores debido a etiquetas mal construidas, ya sea por descuidos humanos o por fallos en el anotador causados por distintas razones. TEITOK permite identificar automáticamente etiquetas mal construidas buscando inconsistencias en el conjunto de todas ellas. Esta función está disponible en el menú de las funciones administrativas: *Check or update the NeoTag parameter set(s) > Check tagset consistency*.

El sistema compara la totalidad de etiquetas presentes en el corpus de entrenamiento con el conjunto de etiquetas posibles contempladas en el etiquetario del corpus. Esta comparación permitirá detectar etiquetas mal construidas debido a errores de posición o a la ausencia de valores esperables de acuerdo con la clasificación establecida.

⁵² En lenguaje CQP, el símbolo @ se sitúa delante del token que se va a modificar. Se indica así en una búsqueda cuál es el token que se desea corregir cuando en esta se pretenden recuperar dos o más tokens.

Tag	Count	Interpretation	Status
	10		(ok)
AO0FP0	17	Adjective;Ordinal;feminine;plural	(ok)
AO0FS0	277	Adjective;Ordinal;feminine;singular	(ok)
AO0MP0	4	Adjective;Ordinal;masculine;plural	(ok)
AO0MS0	83	Adjective;Ordinal;masculine;singular	(ok)
AQ0CN0	174	Adjective;Qualitative;common;invariable	(ok)
AQ0CP0	721	Adjective;Qualitative;common;plural	(ok)
AQ0CS0	1906	Adjective;Qualitative;common;singular	(ok)
AQ0CSP	15	Adjective;Qualitative;common;singular;participle	(ok)
AQ0FP0	1555	Adjective;Qualitative;feminine;plural	(ok)
AQ0FS0	4497	Adjective;Qualitative;feminine;singular	(ok)
AQ0FSD	8	Adjective;Qualitative;feminine;singular;?	Invalid D in position 5 for A; ←
AQ0MP0	1527	Adjective;Qualitative;masculine;plural	(ok)
AQ0MS0	7717	Adjective;Qualitative;masculine;singular	(ok)
AQAMS0	1	Adjective;Qualitative;augmentative;masculine;singular	(ok)
AQCCP0	38	Adjective;Qualitative;comparative;common;plural	(ok)

Figura 72: Comprobación de errores en las etiquetas morfosintácticas

En la imagen se pueden ver algunas de las combinaciones de etiquetas registradas en el corpus con el número total de ocurrencias y su descripción. En la última columna, para cada etiqueta en el corpus se asegura su correspondencia con el etiquetario o bien se advierte de un posible error. Tal y como se señala, hay 8 casos de palabras en el corpus que contienen una etiqueta errónea: una D en la última posición de la etiqueta no es un valor esperable para la categoría ‘adjetivo’. Para corregir el error se podrá recurrir de nuevo a una búsqueda CQP. En estos casos es necesario acceder a las ocurrencias mal anotadas desde el nivel de anotación morfosintáctica de los tokens. En el buscador la consulta será la siguiente: [pos= "AQ0FSD"]. La concordancia obtenida mostrará todos los casos de tokens con la etiqueta indicada, los cuales podrán modificarse de la manera ya descrita para ejemplos anteriores.

Tras corregir las etiquetas, la actualización del anotador es imprescindible para incorporar los nuevos cambios al corpus de entrenamiento y así evitar de nuevo errores asociados a la misma etiqueta mal construida.

4.10.4. Comprobar errores en la asignación de lemas

Tras la actuación automática del anotador, detectar errores en la lematización solo es posible a partir de la lectura atenta de cada documento. No existe en TEITOK un sistema que ayude a la comprobación de lemas mal asignados como, por ejemplo, el recién mencionado para el caso de las etiquetas morfosintácticas. Sin embargo, estos errores resultan más fáciles de corregir, ya que un solo lema puede englobar un gran número de formas, recuperables rápidamente a partir de una sola búsqueda.

Una vez que se ha anotado y revisado una cantidad notable de documentos, es posible hacer una estimación de qué palabras tienen mayor probabilidad de fallar en la

lematización automática a causa del propio sistema de funcionamiento del anotador (§ 4.4 *Funcionamiento del anotador lingüístico en TEITOK*). Ello ayuda a identificar cuáles son los errores más previsibles en la adjudicación de lema y localizarlos directamente a través de búsquedas amplias. Este ha sido uno de los métodos empleados en la revisión de lemas a gran escala.

Los errores de lematización suelen afectar a palabras de distintas categorías gramaticales que comparten terminaciones similares. Por ejemplo, el anotador identificó automáticamente la palabra *viudo*, así como sus variantes *viuda/-os/-as*, como formas verbales del lema **viur*. Por similitud con otras muchas palabras en el corpus anotadas como participios, el anotador lo consideró de igual manera en base a su terminación. Para corregir el lema en todos los tokens afectados, bastará una búsqueda por lema de la forma errónea **viur*, que devolverá una concordancia con todas las formas asociadas a él en la que se podrán seleccionar los tokens que se desean modificar. En la plantilla de edición habrá que indicar la forma correcta *viudo* en la casilla correspondiente al lema para aplicar el cambio en los casos seleccionados:

form	Transcribed form	
fform	Expansion	
dipl	Expanded form	
nform	Normalized form	
pos	POS tag	
lemma	Lemma	viudo

File ID	Sel.	Left context	Match	Right context
xmlfiles/ode/corpus/AL1704I0001.xml	☒	a María de Ponce ,	viuda	de dicho Pedro Rodríguez ;
xmlfiles/ode/corpus/AL1709I0002.xml	☒	por Felipa de Oña ,	viuda	por muerte de Antonio Fenoy
xmlfiles/ode/corpus/AL1754I0012.xml	☒	nupcias con Isabel López ,	viuda	de José Domene Ortiz y
xmlfiles/ode/corpus/AL1754I0048.xml	☒	lo referido Isabel López ,	viuda	de José Domene Ortiz ,
xmlfiles/ode/corpus/AL1754I0048.xml	☒	, pareció Isabel López ,	viuda	de José Domene Ortiz ,
xmlfiles/ode/corpus/AL1774I0014.xml	☒	declarado por Ramona Bornaz ,	viuda	de dicho difunto , y
xmlfiles/ode/corpus/AL1776I0017.xml	☒	doña Teresa Morales , su	viuda	, de la que recibíó
xmlfiles/ode/corpus/MD1703I0103.xml	☒	estando presente Catalina Sanz ,	viuda	y testamentaria de el dicho
xmlfiles/ode/corpus/MD1715I0106.xml	☒	Carvallido y Losada , su	viuda	, y de don Pedro
xmlfiles/ode/corpus/AL1715I0004.xml	☒	vecino de esta villa y	viudo	de doña Juana Quesada y
xmlfiles/ode/corpus/AL1754I0012.xml	☒	. Juan García Morales ,	viudo	de María Fernández Candela ,
xmlfiles/ode/corpus/AL1754I0012.xml	☒	vecino de esta villa ,	viudo	de María Fernández Candela ;
xmlfiles/ode/corpus/AL1754I0048.xml	☒	vecino de esta villa ,	viudo	de María Fernández Candela ,
xmlfiles/ode/corpus/MD1724I0107.xml	☒	presente en esta corte y	viudo	de doña Rosa Téllez Gudiel
xmlfiles/ode/corpus/MD1770I0130.xml	☒	y Ortega el mayor ,	viudo	de Teresa Sarrín difunta ,

Figura 73: Corrección del lema mediante búsqueda lematizada. Resultado para **[lemma = "viur"]*

Las palabras homógrafas han sido otro foco habitual de fallos en la lematización. La revisión manual del lema ha sido doblemente necesaria en estos casos en los que el anotador no puede elegir de acuerdo con criterios lingüísticos y lo hace en base a la

probabilidad de aparición de las etiquetas en el corpus de entrenamiento La revisión manual ha sido obligada en casos como los siguientes:

<i> fueron </i> tasadas	lema: <i>ser</i>
se <i> fueron </i> a la hacienda	lema: <i>ir</i>
una tenaja para tener <i> vino </i>	lema: <i>vino</i>
<i> vino </i> al dicho lugar	lema: <i>venir</i>
forrada por <i> fuera </i> en cuero	lema: <i>fuera</i>
como si <i> fuera </i> por sentencia	lema: <i>ser</i>
me instaron a que me <i> fuera </i>	lema: <i>ir</i>
garrafa de <i> cobre </i> mediana	lema: <i>cobre</i>
se le pague y se <i> cobre </i>	lema: <i>cobrar</i>

Las formas de los verbos *ir* y *ser* que coinciden en los pretéritos (*fui*, *fuera* o *fuese*) presentan problemas para el anotador a la hora de identificar el lema según el contexto. Estos casos se han corregido también siguiendo el método de edición por búsqueda lematizada en la que además se debe especificar la forma de la palabra correspondiente. Por ejemplo, mediante la búsqueda [nform = "fu.+ & lemma = ir"] obtendremos una lista con todas las formas en *fu-* que se han asociado automáticamente al lema *ir*. Solo será posible identificar los errores tras una lectura de todas ellas:

Sel.	Left context	Match	Right context
☐	en su caballo y se	fue	fuera de esta ciudad sin
☐	dicho Pedro Rodríguez , se	fueron	reconociendo los bienes muebles que
☐	alguacil mayor , y se	fueron	inventariando en la forma siguiente
☐	, por dichos tasadores se	fueron	viendo y reconociendo los referidos
☐	, inteligente nombrada , se	fueron	apreciando en esta forma :
☒	inventario ; a lo cual	fueron	presentes por testigos Andrés García
☐	hoy día de la fecha	fui	a las casas de la
☐	alguacil mayor de ella ,	fui	a las casas de la

Figura 74: Corrección de lemas mal asociados a la forma *ir* a partir de una búsqueda

El cambio en este caso no es global, sino que se aplicará únicamente a las formas seleccionadas cuyo lema se ha identificado erróneamente con el verbo *ir*. Una búsqueda equivalente deberá hacerse para aquellas palabras que lematizadas bajo la forma *ser*.

Por último, otro método empleado para revisar la asignación de lemas se ha basado en la realización de búsquedas aleatorias de posibles terminaciones de palabras en las que, por probabilidad, el anotador suele fallar. Sabemos que el anotador automático tiende a confundir los verbos de segunda y tercera conjugación cuyo lema debe ser la forma verbal en infinitivo. Interesa conocer si todas las formas verbales automatizadas con lema en *-er* son efectivamente verbos de la segunda conjugación o, por el contrario, si ha habido algún error en la identificación del lema. Para detectar

errores en los lemas de estos verbos se han realizado búsquedas en las que se especifica la terminación que nos interesa. La búsqueda [lemma = ".+er"] devolverá todas aquellas palabras que contengan cualquier carácter, una o más veces, cuya terminación sea *-er*, y lo mismo puede hacerse con aquellas palabras terminadas en *-ir*. Ambas búsquedas devolverán numerosos resultados para los que la forma lematizada cumple los requisitos mencionados. La concordancia surgida de esta búsqueda puede ordenarse en base a la forma lematizada, de manera que es posible conseguir una frecuencia por lema de las palabras obtenidas. El resultado será una larga lista de lemas con el recuento de los casos que en el corpus se asocian a una misma forma lematizada:

<u>añader</u>	1	<u>parecir</u>	1
ayer	2	omitir	2
atender	2	ocurrir	20
ascender	4	mártir	2
aparecer	5	morir	70
anteceder	1	mentir	5
alfiler	24	medir	27
Javier	3	<u>manifestir</u>	1

Figura 75: Extracto del resultado de la búsqueda de formas acabadas en *-er/-ir* ordenada según la frecuencia por lema

La obtención de resultados a partir de esta lista ha facilitado una revisión más afinada de los lemas que han fallado en la anotación automática y que han escapado a revisiones anteriores durante la lectura individualizada de cada texto. Una vez identificados los errores a partir de la lista estos se han localizado de nuevo a partir del buscador para su posterior corrección. La búsqueda por lema de los errores identificados en la Figura 75 (**añader*, **parecir*, **manifestir*) devolverá las palabras asociadas a dichos lemas, lo que permite modificar estos tokens de manera individual, dado que el fallo de cada lema afecta únicamente a una forma en el corpus.

La edición múltiple a partir del motor de búsquedas facilita la revisión y posterior corrección de lemas mal asignados con NeoTag. Asimismo, identificar errores en el lema ha servido para detectar, en muchos casos, también un fallo en la etiqueta morfosintáctica. En estos casos es de gran utilidad conocer el funcionamiento básico del anotador para predecir posibles fallos.

La búsqueda lematizada abarca un extenso número de resultados y es la más amplia entre los distintos niveles de marcación del texto en cuanto a que ofrece un mayor número de casos. Por ello, este método de edición ha resultado especialmente útil para corregir los lemas que de manera reiterada afectan a múltiples tokens en el corpus.

5. ESTADO DE LA CUESTIÓN

Las numerosas peculiaridades que caracterizan a los diminutivos en español han despertado el interés de los investigadores desde alguna de sus múltiples y muy variadas perspectivas lingüísticas. La exploración bibliográfica refleja que el estudio de los diminutivos se ha centrado sobre todo en los aspectos morfológicos y semánticos: las variantes formales que adoptan estos sufijos, en muchas ocasiones determinadas por las características de la base a la que se adhieren, o sus significados, una cuestión que ha sido ampliamente abordada y objeto de larga disputa entre los lingüistas debido a los numerosos valores que se le atribuyen. A estos aspectos hay que sumar las teorías sobre su origen o la enorme variación a la que están sometidos, no solo geográfica y cronológica, sino también social y pragmática. En suma, se trata de un tema que ha favorecido su análisis desde diferentes disciplinas como la gramática, la dialectología, la pragmática, la sociolingüística, la semántica o la historia de la lengua. Características de todas ellas se entrecruzan en el caso de los diminutivos, tratándose de un fenómeno complejo en el que resulta difícil discernir qué factores y en qué medida influyen en su uso.

En este capítulo se hará un repaso de los trabajos publicados sobre el tema con el fin de caracterizar el objeto de estudio. Los aspectos formales, así como las principales particularidades de los diminutivos y el lugar que ocupan en la gramática se tratarán en § 5.1. *Caracterización del diminutivo*. Aunque se distinguen hasta nueve clases distintas de sufijos diminutivos (NGLE § 9.1b, p. 627), nos centramos exclusivamente en *-ico*, *-ito* e *-illo*, por ser estos los más abundantes en la lengua española. En el uso de los diminutivos influyen en buena medida la variación diatópica y cronológica, consideradas como factores determinantes que condicionan su uso. A este aspecto va dedicado el apartado § 5.2. *Periodización y extensión geográfica de -ico, -ito e -illo*. El marco cronológico estudiado en esta tesis (ss. XVIII y XIX) se considera clave para la configuración del español contemporáneo, y de manera particular en lo que respecta al uso de los diminutivos, pues se trata de un episodio previo que sentará las bases para la distribución geográfica actual. Como resultado del periodo aquí analizado, el uso presente de los diminutivos, tanto en el español europeo como americano se analiza en § 5.3. *El diminutivo en la actualidad: frecuencia y distribución*.

Por último, en § 5.4. *Fuentes para el estudio de los diminutivos*, se hace una recapitulación de las principales fuentes que han servido para el estudio de la distribución geográfica de los diminutivos, prestando especial atención a los trabajos diacrónicos, línea en la que el análisis de esta tesis se inserta.

5.1. Caracterización del diminutivo

Los sufijos diminutivos forman parte, junto con los aumentativos y despectivos, de los morfemas apreciativos, denominados así por su capacidad para expresar distintas valoraciones de la base a la que se adjuntan. De entre las numerosas funciones semánticas que desempeñan, suelen ser mayoritarias en el español aquellas que

manifiestan valoraciones de tipo subjetivo, entre las que se cuentan nociones muy variadas, y en ocasiones difíciles de identificar, con las que mostrar afecto, ironía, atenuación, cortesía, emoción, cercanía o menosprecio. En otros casos, los diminutivos y aumentativos expresan también cualidades objetivas relacionadas con la pequeñez o la disminución de tamaño de lo designado, en el caso de los primeros, o el incremento o mayor tamaño, en el caso de los aumentativos (NGLE § 9.1a, p. 627).

La inclusión de los morfemas apreciativos dentro de las gramáticas ha sido tradicionalmente un aspecto problemático, pues constituyen un mecanismo particular para la formación de palabras con características afines tanto a la morfología derivativa como a la flexiva. Se considera un proceso derivativo en cuanto a que el sufijo aporta contenido semántico específico de la realidad nombrada, ya sea un significado dimensional relacionado con la magnitud o un sentido expresivo con el que mostrar afectividad. Sin embargo, comparten con la flexión otras características que los hacen diferenciarse claramente del resto de sufijos derivativos, entre las que destacan que los apreciativos no modifican la categoría gramatical de la base a la que se añaden, como en *mesa* → *mesita* (ambos sustantivos) o *pequeño* → *pequeñito* (ambos adjetivos). Asimismo, pueden seguir a otros sufijos derivativos, como en *parar* → *parada* → *paradita*, pero a ellos únicamente los pueden seguir los morfemas flexivos de plural (*paradita* → *paraditas*) (NGLE § 9.1g, p. 629).

La cuestión es mucho más compleja, especialmente en el caso de los diminutivos, que adoptan comportamientos particulares frente al resto de los apreciativos. Por ejemplo, a diferencia de los aumentativos y despectivos, los diminutivos heredan el género del lexema-base e imitan su terminación, como en *foto* → *foto* (pero *fotaza*) o *cura* → *curita* (Lázaro Mora, 1999, p. 4655 y ss.). Otro aspecto que ha llamado la atención de los lingüistas es su capacidad para la reduplicación: una misma palabra puede encadenar varios sufijos apreciativos con igual significado (*chiquito* → *chiqu-it-ito*, *chiqu-it-ico*), fenómeno conocido como recursividad (NGLE § 9.1g, p. 629), lo que hace que, en este sentido, se relacionen con determinados prefijos, a los que se asemejan en algunas funciones (Kornfeld, 2021, p. 277). Por otro lado, la manera en la que los diminutivos se añaden a la base manifiesta diferentes soluciones formales que ha hecho que algunos autores consideren que este tipo de elementos actúan no siempre como sufijos, sino como infijos o interfijos (Lázaro Mora, 1999, p. 4658; Fábregas, 2017).

El sufijo *-ito*, por ejemplo, manifiesta las variantes *-ito*, *-cito* y *-ecito*, cuya distribución difiere entre variedades del español, principalmente en lo que atañe a 1) las bases bisilábicas acabadas en vocal *-o/-a* y con diptongo en la primera sílaba (i.e. *viejo*) y 2) a las palabras monosílabas que acaban en consonante (i.e. *bar*). El español europeo prefiere las formas *viejecito* y *barecito*, mientras que las variedades americanas optan por *viejito* y *barcito*, respectivamente (Pato y Felú Arquiola, 2021, p. 77). Aun así, este tipo de variación podría regirse por otros factores como la clase social, el sexo o el nivel educativo y no solo por las diferencias exclusivamente dialectales (Eddington, 2017, p. 59).

Los sufijos apreciativos se sitúan a medio camino entre los procesos de flexión y derivación (Zacarías Ponce de León, 2008) y poseen propiedades morfológicas

singulares que difieren de los sufijos en general, por lo que se consideran una categoría independiente que ocupa en la gramática una sección aparte. Náñez Fernández analiza los cambios que ha experimentado el tratamiento del diminutivo a lo largo del tiempo en las principales gramáticas, en las que el concepto del diminutivo ha ido evolucionando desde su inicial definición como rasgo conceptual hasta su consideración axiológica (Náñez Fernández, 2006, pp. 34-93). Los aspectos formales de los diminutivos han sido extensamente tratados por Varela Ortega (1996, p. 87 y ss.) o Lázaro Mora (1999).

Otros trabajos se centran en la compatibilidad de determinadas bases léxicas con los sufijos diminutivos (Lázaro Mora, 1976), que difiere nuevamente entre las variedades del español (Callebaut, 2011; Kornfeld, 2021). Los sustantivos y adjetivos admiten con naturalidad la presencia de diminutivos, pero el diminutivo con valor afectivo puede darse también en otras palabras, como adverbios (*cerquita*), pronombres (*todito*), participios verbales (*perfumadito*) e interjecciones (*holita*), sobre todo en las variedades americanas, de manera especial en las que están en contacto con lenguas indígenas (Kornfeld, 2021, p. 276).

De entre todos los aspectos circundantes al diminutivo, el de su significado y las funciones semánticas que desempeña ha sido sin duda otro de los más estudiados y que más diversidad de opiniones ha generado⁵³. Dos son los valores fundamentales que de común acuerdo se le atribuyen: 1) el valor objetivo de disminución del referente designado y 2) el valor expresivo con el que se manifiesta la subjetividad del hablante. De estas funciones principales se derivan a su vez un gran número de matices muy diversos por lo común difíciles de discernir. La complejidad de establecer el valor semántico específico de una forma diminutiva concreta se ve acrecentada por el hecho de que todos sus valores posibles operan conjuntamente.

Los valores de carácter objetivo o denotativo se caracterizan por aludir al tamaño más pequeño o reducido de los referentes mencionados en la base. En estos diminutivos prevalece un sentido dimensional que se relaciona con la cualidad objetiva de la expresión de magnitud, por lo que son también llamados nocionales (Alonso, 1961) o conceptuales (Montes Giraldo, 1972). Frente a estos, las valoraciones de carácter subjetivo o connotativo implican matices de tipo expresivo sujetos a las apreciaciones particulares que se hace de las personas, animales o cosas designadas. Cabe destacar que estos valores referidos a la afectividad son de sobra los empleados con más frecuencia en el español, muy por encima de los que aluden esencialmente al tamaño de las cosas, a pesar de que el sentido nocional haya sido considerado de manera tradicional como el más básico y primario. Ambas funciones, las nocionales y afectivas, se pueden manifestar de muy diversas maneras y no son en absoluto excluyentes. En muchos casos, los sentidos primarios relacionados con la función disminuidora “no pueden considerarse con independencia de los matices afectivos” (NGLE § 9.6a, p. 651). El valor afectivo coexiste con el valor conceptual del sufijo en la mayoría de los sustantivos referidos a seres materiales, con mayor o menor predominio de una de estas

⁵³ Un amplio estado de la cuestión sobre la semántica de la sufijación valorativa en español puede verse en la tesis doctoral de Tirapu León (2014).

funciones según el caso, mientras que el sentido claramente afectivo sobresale en los sustantivos referidos a persona, los vocativos y los nombres propios (Zuluaga Ospina, 1970, p. 48). En cualquier caso, el sentido del diminutivo está fuertemente condicionado por el significado de la palabra base, así como por el contexto de aparición, razón por la que resulta difícil sistematizar el significado de estas formas y atribuirles una sola función determinada en cada caso.

A pesar del tiempo transcurrido, continúa siendo esencial el trabajo de A. Alonso (1961) en el que se subraya como básico el valor afectivo del diminutivo por encima del nocional. Para el autor, la interpretación originaria del diminutivo es la de ser “el signo de un afecto” (1961, p. 163). Siguiendo esta línea, otras ideas van más allá al considerar que el valor de aminoración opera siempre juntamente con el de afectividad y que, por tanto, “el diminutivo nocional no existe” (Lázaro Mora, 1999, p. 4651). El trabajo de Alonso dio lugar a numerosas reacciones, causando discrepancias entre los que se mostraban afines a sus postulados y aquellos que, contrarios a sus ideas, defendían como fundamental la prevalencia de lo nocional⁵⁴. Entre los de este pensamiento, destaca Coseriu, para quien “el valor básico del diminutivo es siempre el de «aminoración objetiva»” (Coseriu, 1965, p. 147). Monge (1965) integró ambos puntos de vista sosteniendo que, aunque con mayor peso de la valoración subjetiva, tanto el sentido afectivo como el nocional son inherentes al diminutivo y no es posible ni tiene sentido deslindar estas funciones. En realidad, gran parte del núcleo del debate se situó en la distinción teórica entre *lengua* y *habla* y en cuál de estos planos se asientan los distintos significados que el diminutivo es capaz de expresar: para Coseriu, el valor nocional más básico se sitúa en el plano del sistema de la lengua, mientras que las funciones afectivas dependen del contexto y conciernen al habla; para Monge (1965, p. 145), los valores nocionales y afectivos del diminutivo pertenecen ambos al plano de la lengua.

Los partidarios de la afectividad como valor prototípico del diminutivo justifican su postura en el hecho de que los usos expresivos son sobradamente los más empleados si se comparan con los de la función empequeñecedora. Ambas funciones se encuentran plenamente desarrolladas ya desde el latín, pero es la función afectiva con cualquiera de sus matices más diversos la que predomina también en nuestra lengua desde el castellano medieval. En la escritura, la presencia de los diminutivos estaría también determinada por el tipo de texto (González Ollé, 1962, p. 221). Esta cuestión se trata más detenidamente en § 6.4.2. *Valor semántico del diminutivo en el corpus*, donde se analiza de forma específica la presencia y el significado de los diminutivos en una tipología textual concreta: los inventarios de bienes.

La clasificación de corte estilístico sobre los valores del diminutivo de Alonso (1961) sentó precedentes e influyó notablemente en estudios posteriores sobre el estudio del diminutivo en español. Sobre sus interpretaciones semánticas han surgido complejas sistematizaciones en el intento de compilar los numerosos significados que se le

⁵⁴ Una revisión más completa de la polémica que suscitó el valor semántico del diminutivo en español puede verse en los trabajos de Monge (1965, 1988), Hummel (1997), Martín Zorraquino (2012), Nández Fernández (1997), González-Espresati (2015, pp. 35-43).

atribuyen (Fontanella de Weinberg, 1962; Zuluaga Ospina, 1970; Montes Giraldo, 1972; Soler Espiauba, 1996; Reynoso Noverón, 2005; *NGLE* § 9.6, pp. 651-656). Sus distintas interpretaciones semánticas se relacionan estrechamente con la intención comunicativa del hablante y el contexto de aparición, por lo que el diminutivo se ha estudiado también de manera más reciente en relación con la pragmática y parámetros estilísticos (Reynoso Noverón, 2002; Martín Zorraquino, 2012; Gasso-Villar, 2014; González-Espresati, 2015; Hummel, 2015).

Otro de los aspectos que no se han mencionado hasta ahora, pero relacionado en cierto modo también con los significados del diminutivo, es el de su capacidad para formar palabras nuevas en las que el referente ya no es el mismo de la base léxica. Entre las formaciones diminutivas, muchas de ellas han llegado a desarrollar sentidos específicos fruto de un proceso de lexicalización del sufijo (i.e. *bolsillo*, *tornillo*). Este procedimiento ya ocurría en el latín y se asocia de nuevo con la función expresiva del diminutivo, que también estaba presente en los sufijos latinos, y más particularmente, en el latín vulgar, el cual experimentó un aumento de los derivados con este valor (Lapesa, 2014, p. 80). De esta manera, el castellano heredaría muchas formas lexicalizadas que pasaron con valores semánticos especializados (i.e. *cuchillo*, *rodilla*, *zarcillo*). En nuestra lengua, al igual que sucedió en latín, el proceso de lexicalización afecta sobre todo al diminutivo *-illo*: su amplio uso hizo que se fuera desgastando como mecanismo para expresar afecto en favor de *-ito* (Hasselrot, 1957, p. 258; Monge, 1965, p. 140). Las causas de este proceso de cambio *-illo* > *-ito* se explican con más detalle en el siguiente apartado.

Las cuestiones hasta aquí mencionadas abordan el estudio de los diminutivos desde los ámbitos que más atención han recibido por parte de los lingüistas: el del aspecto formal y el de su significado. La unión de los sufijos a la raíz o la controversia en torno a sus valores semánticos han centrado muchos de los trabajos sobre el tema y dejan en evidencia que se trata de un elemento complejo dada su polisemia y las numerosas variantes morfofonológicas que posee. Al aspecto puramente formal hay que añadir además la variación diatópica a la que están sujeta los diminutivos. Es sabido que su uso está fuertemente condicionado por el origen geográfico, de ahí que algunos de ellos se consideren rasgos diferenciadores del español de determinadas regiones frente al resto. Si bien *-ito* es actualmente el diminutivo que predomina en el mundo hispánico, compite con *-ico* en algunas áreas geográficas concretas, donde actúa como caracterizador dialectal desde el periodo clásico. La variación diatópica está también determinada por el paso del tiempo, por lo que a la variable geográfica hay que añadir también la cronología del fenómeno. Sin embargo, no abundan, como más adelante se verá, los estudios de este tipo desde una perspectiva histórica (§ 5.4. *Fuentes para el estudio de los diminutivos*). En los apartados que siguen se hace una revisión de la variación cronológica y diatópica de los diminutivos en la bibliografía existente.

5.2. Periodización y extensión geográfica de *-ico*, *-ito* e *-illo*

El castellano mantuvo el empleo de los diminutivos heredados del latín y desarrolló también otros de etimología discutida (*-ico* e *-ito*)⁵⁵. Entre los de procedencia latina, *-illo* (< *-ELLUM*), *-uelo* (> *-OLLUM*) y *-ejo* (> *-ICULUM*), el más frecuente fue sin duda el primero, que hasta finales de la Edad Media se presentó bajo la variante en *-iello* (Clavería, 2019, p. 148). Este sería el diminutivo por excelencia para manifestar afectividad en el castellano antiguo, pero la preferencia por *-illo* comienza a experimentar un gran cambio desde finales del periodo medieval.

En el español clásico irrumpen en los textos literarios la presencia de dos nuevos sufijos, *-ico* e *-ito*, que disputarán la popularidad de *-illo*. No obstante, estos diminutivos podrían haber tenido mucho antes un uso más extendido del que comúnmente se cree, tal y como defiende González Ollé (1965), quien en este sentido da algunos apuntes sobre los primeros vestigios de estos dos diminutivos en el español. Las localizaciones de *-ico* se remontan a los siglos X-XI en escritos castellanos y siempre vinculadas a antropónimos. El primer caso es de un texto documental que data del año 959, mientras que su aparición en un texto literario se retrasa hasta 1280 y se aplica a un sustantivo común (*panicos*). Los antropónimos en *-ico* también se encuentran en documentos aragoneses del siglo XIV, sin embargo, no es hasta la segunda mitad del siglo XV cuando se puede hablar de un importante ascenso de este diminutivo, momento en el que comienza a incrementar gradualmente su frecuencia de aparición.

Por otro lado, algo parecido ocurrió con *-ito*. Su primera aparición en castellano se documenta en el año 1191 en un documento leonés, formando parte de un apodo. Los textos dejan de reflejar su uso en fechas posteriores, pero de nuevo en el siglo XV se registra un exponencial aumento de los casos en *-ito*, tanto en inventarios reales como en textos literarios. Para el autor, estos vacíos documentales de *-ico* e *-ito*, que van desde sus primeras apariciones durante el periodo medieval hasta finales del siglo XV y principios del XVI, cuando se constata su plena productividad, no deben interpretarse como síntomas de un fenómeno en desuso. Más bien al contrario, estos diminutivos debían gozar de vitalidad en el habla popular o en la esfera privada, pero no en la escritura de la época, menos aun en la lengua literaria o administrativa, que exigía un uso formal en el que no tenían cabida elementos considerados poco cultos o coloquiales. Por esta razón, *-ico* e *-ito* aparecen en la literatura medieval solo de manera tardía, una vez que en esta trasluce el auge de las clases populares (González Ollé, 1965, p. 551). Los textos literarios de principios del XVII seguían siendo el reflejo de los moldes tradicionales mientras que en otro tipo de textos trasluce el empleo de estos dos sufijos con mayor frecuencia (Ramírez Luengo, 2005).

Otra pista de la presencia de estos diminutivos en épocas tempranas la dan algunos topónimos. El sufijo *-ito* únicamente ha dejado huella en los nombres de algunas localidades de Huelva y Badajoz, arraigado en esta zona como consecuencia de la re-romanización durante el siglo XIII (Catalán, 1989, p. 252). Por el contrario, el atlas

⁵⁵ Sobre la etimología de estos sufijos véase Ariza (1998), Pharies (2002) y González Ollé (2006).

toponímico de Catalán pone en evidencia áreas del norte de Almería en las que predomina *-ico* sobre *-illo*, coincidiendo con la distribución actual de este diminutivo. En Zaragoza, donde también *-ico* forma parte del habla actual, solo se localizan tres topónimos, pero estos aumentan conforme se avanza hacia el sur, llegando incluso a sobrepasar a los de *-illo* en zonas de Murcia y Almería:

La densa área murciana de *-ico* en la toponimia parece sugerir que los colonos aragoneses de hacia 1300 preferían ya el diminutivo *-ico* al antiguo *-et(o)* del aragonés pirenaico, y que en los últimos siglos de la Edad Media su uso fue en aumento hasta competir ventajosamente con *-i(e)llo*: la zona fronteriza de Lorca y los territorios granadinos romanizados después de 1488 muestran la última etapa en esa progresión (Catalán, 1989, p. 251).

En el español clásico el uso de *-ico* se extiende de forma generalizada en amplias zonas del norte y centro de España (NGLE § 9.1k, p. 630). En el siglo XVI es la forma por excelencia que se emplea en la corte (Girón Alconchel, 2004, p. 861), pero en el siglo XVII, en cambio, pierde prestigio con respecto a *-ito* y su uso se considera un rasgo propio del habla popular o rústica, razón por la que Cervantes lo utiliza en boca de personajes de bajo rango social (Ariza, 1998, p. 353) y por lo que su presencia en la prosa técnica es casi inexistente (Ramírez Luengo, 2005, p. 33). Además, en este periodo *-ico* se caracteriza por ser exclusivo de algunas zonas determinadas. Se documenta en las regiones de León, Murcia, Navarra, Albacete, Teruel, Zamora, Almería, Huesca y Alicante –en ese orden de frecuencia–, que destacan frente al resto por un elevado número de casos en *-ico* (Calderón Campos, 2019c, p. 122). Su uso se va reduciendo de manera considerable a la par que adquiere un fuerte carácter dialectal.

Aunque en el XVI *-ito* e *-ico* experimentan un notable ascenso de casos, *-illo* será todavía la forma predominante hasta el siglo XVIII (Paredes García, 2023, p. 122). Si bien el uso de *-ico* cae drásticamente, el de *-ito* fue ganando prestigio y aumentando también su extensión de forma continuada con respecto a *-illo*, a quien finalmente logró alcanzar y superar.

Las gramáticas de la época vinculan los nuevos sufijos (*-ito* e *-ico*) con un carácter afectivo, mientras que *-illo* se reserva para el sentido objetivo de aminoración. Este hecho favoreció que un gran número de palabras con *-illo* adquiriera sentidos específicos: el sufijo se lexicaliza en estas formas y en consecuencia disminuye su función como pleno diminutivo (Monge, 1965, p. 140). Las primeras apariciones de *-ico* e *-ito* en apodos y antropónimos son una clara muestra de la función afectiva con la que se empleaban estos sufijos, probablemente heredada de *-illo*. Así, las causas del cambio *-illo* → *-ito* se deben a la pérdida de expresividad de aquel. Monge explica este proceso de la siguiente manera:

El que una formación diminutiva adquiera un sentido específico se apoya, evidentemente, en su atribución repetida al mismo objeto. La nueva denominación se convierte de frecuente en habitual y termina desplazando al primitivo – o haciéndose exclusiva si la realidad nombrada no tenía antes designación propia. Pero para ello es necesaria la desaparición (o simplemente la ausencia) de componentes afectivos de

cualquier tipo. En otras palabras, para que una formación diminutiva llegue a desarrollar un sentido independiente, se precisa la ausencia de componentes expresivos, hasta llegar a un contenido de significación estrictamente objetivo, nocional (Monge, 1965, p. 141).

Dejando a un lado las causas del cambio y volviendo a su periodización, se observa que las fechas concretas en las se produjo este cambio oscilan según la bibliografía consultada. Para Lapesa (2014, p. 334), quien se basa en trabajos sobre obras literarias, la ventaja de *-ito* sobre *-illo* no se produjo hasta el siglo XIX. Según Paredes García (2023, p. 122), el punto de inflexión se sitúa en la segunda mitad del XVIII, hecho que prueba a partir del análisis de textos documentales muy variados del corpus *CODEA+ 2015*. Por último, Calderón Campos (en prensa), según datos apoyados en cartas privadas del corpus *Post Scriptum*, adelanta este momento al siglo XVII, cuando constata un uso mayoritario de *-ito*, aunque no demasiado acentuado, circunstancia que sí ocurre claramente en el XVIII.

Lo que sí parece claro es que en el siglo XVIII ya se habría definido el orden general en el uso de los diminutivos: *-ito* > *-illo* > *-ico*. En el español dieciochesco, *-ico* es una forma residual únicamente productiva en áreas concretas que continuará en retroceso también en la centuria siguiente. Este orden se invierte en aquellas zonas de la franja oriental de la Península en las que *-ico* se constituye como un rasgo caracterizador de esa variedad lingüística. En Murcia, donde continúa siendo el sufijo más usado en el XVIII, se presenta además la variante palatalizada *-iquio*, exclusiva de esta región y marcada diastráticamente (Vivancos Mulero, 2012, p. 330). Sin embargo, en Málaga ya estaba en regresión (Torres Montes, 1990, p. 70) y a partir del siglo XVIII, también *-illo* irá reduciendo su uso (Náñez Fernández, 2006, p. 326; Pharies, 2002, p. 328). En el Madrid de los siglos XVIII y XIX *-ito* se difunde notablemente, sobre todo como rasgo del lenguaje popular (Sánchez-Prieto Borja et al., 2022, p. 835) y es el que mayor valores semánticos expresa (Seco, 1970).

El proceso de transformación que afecta a los diminutivos en el español peninsular se reproduce también en el español americano: a finales del XVIII y principios del XIX el triunfo de *-ito* sobre *-illo* es prácticamente general en el nuevo continente. Uritani y Berrueta (1985) destacan la importancia que las islas Canarias podrían haber tenido en todo este proceso como puente entre la Península y América. En las islas, la afluencia de población sevillana ya había hecho que se generalizara el uso de *-ito*, de modo que este fue el diminutivo preferente también en América (1985, p. 224).

En algunas zonas de América el cambio *-illo* > *-ito* fue abrupto y no se desarrolló de forma tan gradual como en el español peninsular. En el español de Buenos Aires, el porcentaje de *-illo* en los siglos XVI y XVII supera todavía al de *-ito*, mientras que un siglo después este sufijo representa ya el 86 % de los casos con diminutivo, muy por encima del 10 % correspondiente a *-illo* (Fontanella de Weinberg, 1987, p. 75). No obstante, la velocidad a la que *-ito* se impuso difiere según las zonas: en Venezuela se produce a finales del XVII, en la zona rioplatense tiene lugar en la primera mitad del XVIII, y en áreas más conservadoras, como Tucumán o Quito, a finales del XVIII e inicios de la siguiente centuria (Ramírez Luengo, 2006, p. 40).

Para muchos fenómenos morfosintácticos, el siglo XVIII supone en el español americano un punto de inflexión: a partir de este momento algunas transformaciones que se iniciaron en la otra orilla “no solo se han cumplido totalmente en América, sino que han avanzado más en muchas regiones que en el uso peninsular” (Sánchez Méndez, 2020, p. 176). Este hecho podría aplicarse de manera particular también al caso de los diminutivos, pues si bien se afianza la preferencia por *-ito*, homogénea en todo el mundo hispánico, en este siglo comienzan a desarrollarse características diferenciales con respecto a la variedad peninsular, tales como la recursividad o el aumento de los tipos de palabras que admiten sufijos diminutivos. Por ejemplo, en el español uruguayo, en los siglos XVIII-XIX el número de adjetivos diminutivos es menor que en periodos posteriores, al igual que los casos de diminutivos que se añaden a pronombres y adverbios (Bertolotti y Coll, 2014, p. 34).

En base a las diferencias en la periodización que se han ido mencionando son dos las reflexiones que se nos plantean: una de ellas relacionada con el tipo de fuentes textuales que se han empleado para el estudio de los diminutivos, y otra en torno a la procedencia geográfica de esas fuentes. La historia de los diminutivos es la historia de la documentación que ha llegado hasta nuestros días y que, como la de tantos otros fenómenos gramaticales, se caracteriza por la “excesiva atención al canon literario” (Eberenz, 2009, p. 191). Hasta el siglo XVI, cuando los textos comienzan a revelar una presencia abrupta de otros elementos además de *-illo*, es cuando cabe preguntarse sobre la posibilidad de que en la lengua de la época el espectro de variación fuera en realidad más amplio del que reflejan los textos estudiados hasta el momento. Es un hecho probado que la presencia de diminutivos es mayor en el habla popular, por lo que urge observar el fenómeno desde otros puntos de vista, en textos en los que afloren rasgos coloquiales. Por otra parte, el cambio de tendencia (de *-illo* a *-ito*) afectó de manera global al mundo hispánico, pero es lógico pensar que no se produjo a igual velocidad en todas las regiones. Partiendo de la información aquí recopilada, el análisis de esta tesis pretende ofrecer nuevos datos en esta dirección. Por un lado, la elección de inventarios como fuente documental va encaminada al estudio de textos con tendencia a mostrar un elevado número de diminutivos y de fenómenos cercanos a la oralidad, pues se trata de documentos que fueron elaborados sin una intencionalidad creativa. Por otro lado, la selección de inventarios de las provincias de Almería y Madrid se hace con el objetivo de detectar características divergentes tanto en los diminutivos empleados como en los procesos de cambio que afectan al uso de estos en dos modalidades lingüísticas actualmente muy diferenciadas.

En último lugar, el estudio de la cronología de los diminutivos se ha centrado sobre todo en el español clásico por ser este el periodo en el que ocurre el mayor reajuste. La nueva distribución de los elementos, si bien es cierto que es fruto de un largo proceso que abarca varias centurias, no se da por afianzada hasta el inicio del español moderno. Los siglos XVIII y XIX adquieren por tanto una importancia decisiva como periodo previo en el que se fija de manera definitiva la distribución de los diminutivos que pervive actualmente.

5.3. El diminutivo en la actualidad: frecuencia y distribución

El diminutivo *-ito* es actualmente el de mayor extensión en todo el mundo hispánico. En el español americano alterna con *-ico* en algunas zonas del Caribe (NGLE § 9.1j, p. 630), mientras que, en el español peninsular, conocido por el variado registro de sufijos diminutivos que posee (*-ín, -ino, -uco, -iño, -ete, -ejo, -uelo*, etc.), convive además con *-illo* e *-ico*, a los que nos ceñimos en esta tesis. El sufijo *-ico* se extiende por zonas de la franja oriental de España, desde el bajo Aragón hasta el sureste de Andalucía. Por su parte, aunque *-illo* mantiene su productividad de manera general en el español europeo, su uso destaca proporcionalmente en Andalucía (NGLE § 9.1l, p. 631).

Algunos trabajos que han abordado de manera general el uso de los diminutivos corroboran *grosso modo* esta distribución geográfica, aunque con algunos matices. Por ejemplo, a partir del análisis de los atlas lingüísticos, y en contra de lo que suele creerse, Uritani y Berrueta (1985) afirman que *-ico* no es el sufijo predominante en Aragón. Este solo aparece con frecuencia en la zona central (provincias de Zaragoza, Teruel y este de Navarra), ocupando la tercera posición (22.8 %) seguido de *-illo* (27.1 %) y de *-ete* (27.3 %). Este último se sitúa en la franja oriental (Huesca y este de Zaragoza y Teruel), con especial arraigo en las zonas próximas al dominio lingüístico catalán. En el mismo trabajo y con datos del ALEA, se constata el predominio de *-illo* casi en toda Andalucía (59.6 %), pero con mayor proporción en las provincias de la zona central y oriental, por oposición a las occidentales, donde predomina *-ito*. El uso de *-ico* en Andalucía oriental se documenta, de menos a más, en las provincias de Jaén, Granada y Almería (Uritani y Berrueta, 1985, p. 223).

Por otro lado, en el español americano, el uso de *-ico* en determinadas zonas del área caribeña podría también estar en retroceso. En el habla dominicana, su uso está fuertemente condicionado por razones de eufonía: se aplica a las bases léxicas que acaban en consonante dental /t/ (i.e. *zapatito, ratito, galletita*) y a los diminutivos dobles (*chiquitico*). Sin embargo, esta regla no se confirma en todos los casos, ya que en muchos otros en los que sería esperable su uso este coexiste con el del sufijo *-ito* (*tarjetita, cuartito, poquitito*). Este hecho señala el avance de *-ito* con respecto a *-ico* también en esta zona (Aleza Izquierdo, 2016, p. 43).

De forma general, una de las principales diferencias que se han señalado entre el español europeo y el español americano en cuanto al empleo de los diminutivos es la mayor frecuencia de formas diminutivas que suele caracterizar a las variedades americanas. Esta circunstancia se asocia al contacto del español con las lenguas indígenas, especialmente en los dialectos mestizos, donde se observa además una mayor variedad de valores semánticos asociados al diminutivo (Reynoso Noverón, 2001, p. 60).

A las diferencias geográficas que condicionan el empleo de determinados sufijos habría que añadir otras disimilitudes en el uso que se hace del diminutivo y que se relacionan con los contextos y la finalidad con la que los hablantes lo aplican. La comparación entre el español madrileño y el español mexicano revela que este último “se inclina por la disminución motivada por condiciones pragmáticas subjetivas”, mientras que el español madrileño utiliza indistintamente ambas categorías semánticas,

aunque con una ligera inclinación hacia los usos referenciales (Reynoso Noverón, 2002, p. 939). De esta manera, el comportamiento distintivo del diminutivo es también un fenómeno de caracterización dialectal en algunas variedades lingüísticas del español. La preferencia por los usos no referenciales se vería además apoyada en el hecho de que las variedades americanas utilizan el diminutivo sobre entidades que no admiten disminución, como los adverbios, en mayor proporción que el español peninsular. Los diferentes valores semánticos que subyacen a las diferencias cuantitativas en el uso de los diminutivos marcarían una isoglosa entre el español peninsular y los dialectos americanos (Company Company, 2002). Este hecho se confirma también en otras áreas de Hispanoamérica como Bogotá (Fontanella de Weinberg, 1962) o Caracas (Malaver, 2018), pero también se constata en el corpus del español madrileño analizado por Malaver y Paredes García (2020), por lo que, a falta de otros estudios, las diferencias en este aspecto podrían no ser tan relevantes entre el español de ambas orillas.

La distribución general de los diminutivos señalada arriba se complica a medida que se desciende a analizar una variedad concreta en profundidad, ya que varios sufijos coexisten en una misma región y en el uso de unos u otros intervienen a la vez factores muy diversos. Los estudios más recientes se han centrado en los factores de tipo pragmático y sociológico para detectar las diferencias que rigen el uso de los diminutivos.

En cuanto a las diferencias de género, el uso del diminutivo es más frecuente en las mujeres que en los hombres en un gran número de comunidades de habla (Fontanella de Weinberg, 1962, p. 570; Paredes García, 2015; Aleza Izquierdo, 2016; Malaver, 2018; León-Castro Gómez, 2020b). Supone una excepción el español andino, donde con respecto al español de Buenos Aires, Madrid o México, los hombres parecen hacer un mayor uso de este tipo de marcación (Reynoso Noverón, 2001, p. 120). Por otra parte, el diminutivo predomina especialmente en las clases más bajas o de menor nivel educativo (Fontanella de Weinberg, 1962, p. 571) y en los registros orales antes que en los escritos (Reynoso Noverón, 2001).

Conviene mencionar algunos trabajos que se han centrado en el estudio actual de los diminutivos en las variedades almeriense y madrileña, coincidiendo con la delimitación geográfica estudiada en esta tesis. Los datos de Almería y Madrid que aquí se presentan se pueden comparar, salvando las distancias, con los del español actual en estas dos provincias. En Almería predomina actualmente el uso de *-ito* (68 %), seguido de *-illo* (21 %) e *-ico* (11 %), aunque el uso de *-ico* aumenta en situaciones menos formales (López González, 2005, p. 242). Si se tiene en cuenta que estos datos se han extraído de un corpus de conversaciones en la radio local de Almería (medio de comunicación, por lo común, dado a la formalidad) las cifras podrían diferir en otros contextos de uso. En Granada, provincia colindante, el uso del diminutivo está muy marcado diastáticamente: el sufijo *-illo* (46.8 %) predomina sobre *-ito* (41.1 %), mientras que *-ico* (12.1 %) se encuentra en regresión y se adscribe a los niveles educativos más bajos (Manjón-Cabeza Cruz, 2012, p. 73). En el mismo trabajo se defiende el ascenso de *-ito*, en fase actual de difusión. De modo que el conocido uso de *-ico* en zonas de Andalucía oriental se debe matizar en el caso del español de Granada. El uso de *-ico* en estas zonas se entiende no como el diminutivo más frecuente, sino

como el que manifiesta un uso proporcionalmente más alto con respecto a otras zonas de la Península, ya que en las zonas donde se mantiene productivo actúa solo como diminutivo local (Manjón-Cabeza Cruz, 2016, p. 422).

En cuanto al uso del diminutivo en el español madrileño, se constata el predominio de *-ito* (62.8 %) con un mayor empleo de las funciones que implican subjetividad o evaluación de la realidad, seguido de lejos por *-illo* (19.1 %) (Paredes García, 2015, p. 90). Los factores sociales relativos a la edad y el sexo tienen un impacto considerable sobre el uso de los diminutivos en Madrid: las mujeres lo emplean con más frecuencia que los hombres y gana terreno entre los jóvenes (2015, p. 107). En cuanto al nivel de estudios, los hablantes que poseen únicamente estudios primarios lo emplean con mayor frecuencia que el resto (Malaver y Paredes García, 2020, p. 337).

5.4. Fuentes para el estudio de los diminutivos

El estudio de la distribución geográfica y cronológica de los diminutivos desde una perspectiva histórica se ha centrado tradicionalmente en el análisis de fuentes literarias. Al respecto, son ineludibles los trabajos sobre el español medieval de González Ollé (1962) y sobre el español clásico y moderno de Náñez Fernández (2006). En ellos se aborda de manera general el uso del diminutivo en estos periodos y siguen siendo hoy en día trabajos de referencia para el estudio diacrónico sobre la materia.

No faltan tampoco los que se centran en una obra concreta o en la bibliografía de un autor en particular, como los trabajos sobre el uso de los diminutivos en la lengua cervantina (Náñez Fernández, 1952, 1954 y 2010; Ramírez Luengo, 2005), en el *Libro de buen amor* del Arcipreste de Hita (Liria, 1974), en Galdós (García Pérez, 2009), en las obras del escritor navarro Germán Sánchez Espeso (Muela Ezquerra, 1979) o en los coloquios dramáticos del cura Gaspar Fernández y Avila para el estudio de los diminutivos en el habla malagueña del XVIII (Torres Montes, 1990). A estos se podrían añadir los trabajos que se han basado en un conjunto de obras cumbre de la literatura pertenecientes a un mismo siglo o periodo: para el siglo XVI (Pesqueira, 2006) o el XVII (Latorre, 1956).

En definitiva, los textos literarios han sido fundamentalmente el foco de aquellos estudios que han tratado de ofrecer un panorama sobre la evolución del diminutivo desde un punto de vista histórico. Por esta razón, también los manuales de historia de la lengua suelen explicar los cambios en la morfología derivativa del español basándose en estos trabajos (Pharies, 2002; Cano Aguilar, 2004; Lapesa, 2014; Ridruejo, 2019).

Continuando con los estudios de corte diacrónico, el campo de la toponimia ha sido otra de las vías utilizadas para el acercamiento al análisis de los diminutivos. Numerosos investigadores han visto en los corpus toponímicos una fuente especialmente útil y rica para abordar diferentes cuestiones lingüísticas en torno a la

formación de los nombres de lugares, siendo una de ellas la sufijación diminutiva como procedimiento para la creación de nuevas denominaciones⁵⁶.

La gran mayoría de los trabajos de la disciplina se centran en el estudio de zonas muy concretas, comarcas, provincias o comunidades autónomas, a lo sumo, lo que supone el análisis parcial de un espacio geográfico más bien reducido y como es de esperar, no todas las regiones han recibido la misma atención. Por lo general, las zonas de contacto lingüístico han sido objeto de mayor interés, de ahí que se encuentren más trabajos basados en zonas donde el castellano limita con otra de las lenguas vernáculas peninsulares. Algunos ejemplos son los estudios sobre la toponimia de León (Morala, 1986), de Huelva (Molina Díaz, 2008), de Extremadura (González Salgado, 2007) o de Cuenca (García Sánchez, 2015). Frente a estos trabajos más específicos, Diego Catalán (1989) presenta una visión integradora de la toponimia romance de la península ibérica y ofrece una distribución geográfica de los sufijos diminutivos a partir de mapas derivados del análisis que lleva a cabo. En cuanto a las provincias que más nos interesan para esta tesis –Madrid y Almería–, los estudios en este ámbito son más bien insuficientes. La escasez de trabajos en torno a la toponimia madrileña se extiende también de manera general a la zona centro peninsular. Para la región almeriense, solo en un trabajo más amplio como el recién nombrado de Diego Catalán se encuentran algunos datos.

Los trabajos mencionados se acercan a los intereses de esta investigación en tanto que pretenden delimitar la extensión geográfica de los diminutivos. No obstante, difieren considerablemente en el enfoque metodológico, así como en la disparidad de las fuentes utilizadas, por lo que, en términos comparativos, las conclusiones obtenidas en el ámbito de la toponimia no pueden aplicarse a las que se derivan del análisis realizado en esta tesis.

El estudio de la toponimia implica además varias dificultades añadidas con las que se debe lidiar a la hora de establecer una frecuencia de aparición de los diminutivos. Una de ellas es que la presencia de unos y otros sufijos se debe a la superposición de diferentes estratos lingüísticos sucedidos en el tiempo. Por lo tanto, es ineludible atender a los acontecimientos históricos de cada región, particularmente a la acción de la Reconquista, como hecho decisivo para la configuración de la toponimia: “las áreas toponímicas de los varios sufijos en los dos tercios meridionales de España dependen en su mayor parte, como la de tantos otros rasgos lingüísticos, de esa re-romanización medieval” (Catalán, 1989, p. 248). De modo que en muchas ocasiones la presencia o ausencia de determinados sufijos se explica por razones de historia externa, lo que lleva a considerar los datos extraídos de la toponimia como consecuencia del cúmulo de diferentes estratos cronológicos.

⁵⁶ Un riguroso estado de la cuestión de la disciplina puede verse en *Toponimia de España* (Gordón Peral, 2010). El volumen ofrece una visión de conjunto de los estudios lingüísticos realizados hasta el momento en las distintas comunidades del país, así como los trabajos en curso y las lagunas de investigación que aún restan por completar en aquellas regiones peninsulares que han sido, por lo general, más desatendidas.

Al mismo tiempo, cabe preguntarse si los diminutivos obtenidos en toponimia se corresponden con el uso que de estos se hace en el habla común de la época y si se pueden considerar rasgos característicos de la región. En este sentido, coincidimos con Trapero (1995, p. 24) en que el léxico de la toponimia, “frente al léxico de la lengua común o al propio de una especialidad, se manifiesta más conservativo”⁵⁷. Esto explicaría, por ejemplo, la ausencia en algunas zonas de sufijos que se consideran propios de determinadas regiones y también, al contrario, la presencia de sufijos que no se consideran autóctonos de la zona⁵⁸, razón por la que consideramos que las conclusiones derivadas de estos datos deben tomarse con cautela si quieren extrapolarse al habla viva de la época.

Con todo, los estudios en este campo han aportado nuevos datos sobre los diminutivos aun cuando la mayoría de los trabajos históricos han estado habitualmente orientados al análisis de fuentes literarias, pero al igual que estos, muestran serios inconvenientes para el estudio de la lengua hablada. En el caso de los corpus compuestos por textos literarios, la intención creativa oculta fenómenos cercanos a la oralidad y cuando se ofrecen muestras de la variación dialectal, estas suelen ser reproducciones estereotipadas. En los corpus toponímicos, como ya se ha indicado, debe tenerse en cuenta además la acumulación de estratos lingüísticos, fruto de las particularidades históricas de cada región.

No se puede negar el valioso aprovechamiento que han supuesto las fuentes literarias para el estudio del diminutivo en su vertiente diacrónica. Escasean en cambio los trabajos que abordan el tema a partir de documentación no literaria que permita contrastar, ya sea para corroborar o refutar, los resultados ya existentes con los que se pueda afinar y completar la visión del diminutivo en español a lo largo de la historia. La necesidad de estudiar el uso de los diminutivos en diversas fuentes textuales ya fue subrayada por Ramírez Luengo (2005). En este trabajo, la comparación entre una obra literaria como es el *Don Quijote de La Mancha* y otra técnica, como el *Libro del Arte de Cocina*, reveló diferencias significativas en cuanto al empleo de los sufijos en uno y otro texto, de lo que se concluye que el tipo de lengua influye directamente en la utilización de los diminutivos (2005, p. 33).

⁵⁷ A pesar de ello, para el autor no hay dudas de que el léxico de la toponimia debe considerarse como lengua funcional: “no es la toponimia un dominio en el que el léxico esté cerrado, inmovilizado y a expensas sólo de un grupo social o de una comunidad dialectal [...] al usarse por todos los hablantes, y en cualquier situación de habla, está expuesto a la evolución y al cambio al que está expuesto el léxico patrimonial de una lengua cualquiera” (Trapero, 1995, p. 24).

⁵⁸ Al contrario de lo que pudiera parecer, el considerado sufijo extremeño por excelencia (-ino) apenas presenta rendimiento en la formación de topónimos en esta comunidad. Para González Salgado (2007), ello se debe a que, por un lado, el sufijo alcanzó su máxima vitalidad en el habla en un momento en el que los topónimos ya se habían formado, con poca capacidad para crear nuevos nombres; por otro, los matices predominantemente afectivos que se le atribuyeron a -ino pudieron influir en que se descartara este sufijo en favor de aquellos en los que prevalece el sentido de referencia al tamaño, ya que no son los matices afectivos los que suelen intervenir en la creación de los topónimos (2007, p. 1978). Por otro lado, resulta llamativa en algunas zonas de Huelva la presencia de topónimos en -ín, -ino, mayor incluso que en Extremadura, e -iño, alejados de la frontera con Portugal. La presencia de -ín en Huelva se remonta a la repoblación medieval por su contacto con el dominio asturleonés, al igual que -iño, que más que por su cercanía con el portugués, se explicaría por el contacto con los repobladores gallegos también en esta época (Molina Díaz, 2008, p. 1432).

Otra circunstancia a la que ya aludimos anteriormente es que los sufijos *-ico*, *-ito* tienen orígenes muy distintos dependiendo de las fuentes consultadas. Las primeras apariciones de estos sufijos en textos documentales se remontan al siglo X (en el caso de *-ico*) y al siglo XI (en el caso de *-ito*). Sin embargo, “de atender únicamente al testimonio de la literatura, ambos sufijos resultarían modernos en castellano” (González Ollé, 1965, p. 551). El uso de *-ito* e *-ico* no se generaliza en la lengua literaria hasta fines del siglo XV, principios del XVI, por lo que el tipo de fuente utilizada resulta trascendental.

En esta misma dirección, otras investigaciones se han ocupado del análisis del diminutivo desde otras fuentes –no literarias– que permitan un mejor acercamiento al estudio de la oralidad: la correspondencia privada (Valeri, 2010; Ramírez Luengo, 2013), los inventarios de bienes (a los que nos referiremos un poco más adelante) o los atlas lingüísticos y los corpus orales, aunque estos últimos desde una perspectiva sincrónica y centrados en el español contemporáneo por razones obvias.

Los datos que proporcionan los atlas lingüísticos se prestan con facilidad a la cuantificación de los diminutivos, lo que permite un panorama general de la distribución geográfica de estos sufijos en distintas regiones de España. Basándose en aquellas que poseen atlas lingüísticos publicados, se pueden mencionar los trabajos de Enguita Utrilla (1984), Uritani y Berrueta de Uritani (1985), Hasselrot (1957), Mackowiack (2018) o Medina Domínguez (2022).

Por otro lado, hay que destacar una gran cantidad de publicaciones que se han dedicado total o parcialmente al estudio del diminutivo en alguna de las variedades del español actual: sobre el habla de Sevilla (León-Castro Gómez, 2020a, 2020b) de Granada (Manjón-Cabeza Cruz, 2012, 2016; Malaver, 2021), de Madrid (Paredes García, 2011, 2015) o de Canarias (Samper Hernández, 2019); y también del otro lado del Atlántico: en el español de Colombia (Fontanella de Weinberg, 1962), de República Dominicana (Aleza Izquierdo, 2016) o de México (Silva Almanza, 2012). Estos trabajos se acercan más al enfoque metodológico que se desarrolla en esta tesis en cuanto a que utilizan datos cuantificables con los que analizar la evolución de los diminutivos en términos estadísticos, sin embargo, lo hacen desde una orientación sociolingüística y sincrónica.

En el ámbito de la diacronía, pocos son los trabajos que tratan de analizar estadísticamente el fenómeno para el estudio de su cronología y distribución geográfica. Los inventarios de bienes constituyen una fuente documental que contrasta notoriamente con las ya mencionadas hasta ahora para los estudios históricos. El interés excepcional que este tipo de documentos tiene para el estudio de los diminutivos viene determinado por la proliferación de diminutivos que encontramos en estos textos, en los cuales se hace imprescindible aludir a las dimensiones de los objetos recurriendo para ello al uso de la sufijación. González Ollé (1962) llamó la atención sobre este mismo hecho en su principal trabajo sobre el diminutivo en el castellano medieval. Para el autor, el abundante uso de estos sufijos en los inventarios se justifica además por la desaparición del tono solemne que caracteriza a estos textos, ya que es de suponer que en ellos se describen las cosas de manera similar a como se haría en la lengua coloquial

(1962, p. 133). Por esta razón, los inventarios permiten el acercamiento a la oralidad en lo que respecta al uso de los diminutivos.

Trabajos más recientes han recurrido de nuevo a los inventarios de bienes, tan abundantes en todas partes de la geografía hispánica, con el objetivo de establecer una distribución diatópica y cronológica del uso del diminutivo. Pueden mencionarse al respecto los trabajos de Perdigüero (2015), Enguita Utrilla y Lagüéns Gracia (2019), Ortiz Cruz (2019) o Calderón Campos (2019c). Sin embargo, en ellos no se hace un análisis exhaustivo del uso de los diminutivos, sino que estos se suelen mencionar entre otros fenómenos de carácter fonético, morfológico o sintáctico también de interés presentes en los textos. Un análisis más completo sobre la distribución de los diminutivos es el que aporta Paredes García (2023), de especial importancia para esta tesis, ya que en él se toma como base el estudio de documentación notarial, entre la que se incluyen inventarios. Los principales resultados ya se mencionaron en § 5.2. *Periodización y extensión geográfica de -ico, -ito e -illo*. En este trabajo se emplea un corpus cuantificable para analizar cuantitativamente y en términos estadísticos la variación de los diminutivos. Esta técnica ha tenido una amplia acogida en el ámbito de la sociolingüística, donde son más comunes los análisis cuantitativos basados en modelos estadísticos (Mayoral Hernández, 2022), pero en cambio escasean en los trabajos sobre diacronía.

Tras haber hecho un recorrido por las principales fuentes que se han empleado para el estudio de la variación geográfica y cronológica del diminutivo a lo largo del tiempo, se observa una falta de trabajos centrados en ofrecer datos rigurosos y cuantificables de los que se puedan extraer conclusiones del uso de estos sufijos en la lengua oral. En este sentido, los inventarios de bienes constituyen una fuente fiable con la que poder estudiar de forma empírica fenómenos pertenecientes a la oralidad también en el plano morfológico de la lengua. Al mismo tiempo, aseguran la comparabilidad de los textos, ya que son documentos en los que se especifica de manera certera el lugar y la fecha en la que fueron creados. Por las características de los inventarios, los diminutivos proliferan en ellos, lo que permite disponer de numerosos datos que pueden ser contabilizados y contrastados estadísticamente. La idoneidad de esta tipología textual para el estudio de los diminutivos se vuelve a tratar en § 2.3.3. *Tipología textual* y § 6.4.2. *Valor semántico del diminutivo en el corpus*.

Esta tesis pretende contribuir al estudio de los diminutivos con un análisis exhaustivo de su evolución en los siglos XVIII y XIX aplicando un enfoque más próximo a la metodología de la lingüística de corpus: aquel que requiere el empleo de técnicas estadísticas refinadas con las que extraer conclusiones y probar nuevas hipótesis (Rojo, 2021, p. 49). Para ello es necesario contar con corpus que permitan agrupar los datos en base a diferentes características, según su fecha o procedencia, para poder contrastar los resultados obtenidos.

6. LOS DIMINUTIVOS -ICO, -ITO, -ILLO EN EL CORPUS

El presente capítulo se propone analizar las palabras con sufijación diminutiva en el corpus con el objetivo de diferenciar entre el valor transparente y lexicalizado del diminutivo. El planteamiento de la investigación y la delimitación del objeto de estudio se presentan más ampliamente en el apartado § 6.1. *Consideraciones generales*.

Como se sabe, la derivación diminutiva en español puede originar unidades léxicas independientes cuyo significado se aleja en mayor o menor medida del propio sentido del diminutivo. Para los efectos de esta tesis se distingue entre *diminutivo transparente* y *diminutivo lexicalizado* de acuerdo con los parámetros tradicionales que establece la NGLE. Se profundiza más sobre cada uno de estos conceptos en § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*.

Los criterios de selección que han llevado a la clasificación de los términos en estos dos grandes grupos se detallan en § 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*. El contexto ha determinado en la mayoría de los casos la clasificación entre unos y otros (§ 6.3.1. *El contexto como elemento desambiguador*), pero otros casos de compleja sistematización han exigido un estudio de mayor profundidad (§ 6.3.2. *Casos dudosos y problemas para la cuantificación*).

Los apartados § 6.4. *Diminutivos transparentes* y § 6.5. *Diminutivos lexicalizados* analizan las palabras en el corpus que se han considerado pertenecientes a uno y otro grupo. Entre los diminutivos lexicalizados se han detectado voces dialectales y otras poco documentadas (§ 6.5.2.1. *Regionalismos y voces poco documentadas*). Así como el caso concreto del término *bufetillo*, en el que la presencia del diminutivo se asocia con la evolución de un cambio semántico activo (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). Finalmente, todas las voces en las que el diminutivo se ha lexicalizado se incluyen en un glosario (§ 6.5.3. *Glosario de lexicalizaciones*).

6.1. Consideraciones generales

El análisis de las palabras con sufijación diminutiva que se plantea en esta tesis se centra en los sufijos -ico, -ito e -illo por ser los tres diminutivos más frecuentes en el español y aquellos que poseen mayor rendimiento. De los numerosos aspectos que han centrado su estudio (§ 5. *ESTADO DE LA CUESTIÓN*), esta investigación aborda en términos estadísticos la frecuencia de uso y la distribución diatópica y cronológica en Almería y Madrid en los siglos XVIII y XIX. Para alcanzar este objetivo es necesario seleccionar previamente las palabras del corpus en las que el sufijo sí se identifica con el auténtico valor de diminutivo, de manera que no todas las palabras entran en el recuento de formas diminutivas analizadas.

Para la clasificación de los diminutivos se tienen en cuenta algunas cuestiones relacionadas con su significado, ya que el valor semántico ha servido, como se expone más adelante, como criterio metodológico para la selección de las formas estudiadas. En cambio, no se analiza la presencia de los diversos sufijos en función de cuestiones

puramente semánticas o incluso fonéticas relacionadas con las bases con las que se combinan⁵⁹. El aspecto formal de los diminutivos ha sido extensamente tratado en muchos trabajos sobre el tema (§ 5.1. *Caracterización del diminutivo*), de modo que se prescinde aquí del análisis de cuestiones de tipo morfológico como las variantes que puede adoptar un mismo sufijo en la formación de derivados o las alteraciones que sufre la base.

Dadas las características textuales que conforman el corpus documental, los diminutivos se añaden principalmente a la base de sustantivos que denotan seres materiales relacionados con el ámbito doméstico, esto es, pertenencias propias de la vida cotidiana como utensilios, herramientas, muebles, prendas de vestir, joyas, alimentos, animales, etc. De entre todas estas formas cabe diferenciar las que representan verdaderos diminutivos –en los que predomina el sentido de alusión al tamaño–, de las que constituyen palabras ya lexicalizadas, entendidas como las que se han incluido en el diccionario como voces independientes pertenecientes al léxico de la lengua. La clasificación entre unas y otras ha estado marcada por una diferencia de carácter semántico de acuerdo con los mismos parámetros que establece la *NGLE* para la división de las palabras con sufijación apreciativa:

Las palabras que contienen sufijos apreciativos se dividen en dos grupos. Las llamadas de SIGNIFICADO TRANSPARENTE, o simplemente TRANSPARENTES, son aquellas cuyo significado se obtiene de la combinación de la base y el sufijo, como en *arbolillo*, *casita*, *cochazo*, *guapete*, *librote*, *vasico* [...] Las voces llamadas OPACAS, NO TRANSPARENTES o LEXICALIZADAS forman el otro grupo. Se trata de vocablos cuyo significado no se obtiene por la simple combinación de los dos componentes que la forman [...] Así, el diccionario contiene los sustantivos *centralita*, *cigarillo*, *cinturón*, *cuadernillo*, *estribillo*, *flequillo*, *hornilla*, *machote*, *manecilla*, *palacete*, *pañuelo*, *saquillo*, entre otros muchos de significado impredecible a partir del significado del diminutivo (*NGLE* § 9.3a-b, p. 635).

Partiendo de esta distinción, se trata de clasificar todas las palabras en el corpus de inventarios que contienen alguno de los diminutivos estudiados con el fin de realizar un análisis estadístico que permita obtener resultados concluyentes en torno a la distribución de los sufijos en las áreas y periodos estudiados. La intención fundamental es describir la frecuencia de uso de *-ico*, *-ito* e *-illo* en la variedad lingüística almeriense de los siglos XVIII y XIX por contraste con la representada en los inventarios de Madrid, los cuales constituyen el corpus de control (Capítulo 7).

No todas las formas en las que se encuentra aparentemente alguna de estas terminaciones que se identifican con un sufijo diminutivo se han considerado válidas para el análisis, por lo que ha sido necesario establecer unos criterios metodológicos claros que guíen la selección de estas formas. La distribución de los diminutivos se analiza teniendo en cuenta solo aquellas palabras en las que el significado se identifica

⁵⁹ En Lázaro Mora (1976) se aborda el problema de la compatibilidad entre lexemas nominales y sufijos diminutivos, que según este autor estaría relacionado con motivos semánticos y otros puramente fonéticos o fonológicos.

con un valor transparente y puede obtenerse combinatoriamente de la suma de la base y el sufijo. Quedan por tanto al margen del análisis estadístico posterior las numerosas palabras en las que el diminutivo ha desarrollado valores específicos, ya sea en el latín o en castellano, y no puede interpretarse independientemente del lexema base, como ocurre en *estopilla* (un tipo de tela) o en *almilla* (una prenda de vestir)⁶⁰. A efectos prácticos, únicamente se consideran verdaderos diminutivos las voces transparentes, por lo que cuando aludimos de manera general a los diminutivos para estudiar su frecuencia y distribución geográfica, se entiende que se está haciendo referencia solo a las voces en las que el diminutivo conserva tal carácter, a excepción de que se indique lo contrario.

La elección de palabras con diminutivos aptas para el análisis estadístico conlleva forzosamente la exclusión de otras muchas. Son las llamadas voces lexicalizadas en las que se ha desarrollado un proceso de cambio semántico. Determinar el mayor o menor grado de lexicalización del diminutivo e identificar sus posibles valores específicos ha sido especialmente difícil en algunos casos, sobre todo en las voces que parecen situarse en fases intermedias del proceso de lexicalización⁶¹. Como resultado de este estudio todos los diminutivos lexicalizados documentados en el corpus se han incluido en un glosario con sus respectivas acepciones y ejemplos (§ 6.5.3.4. *Glosario*).

A priori podría parecer que la consideración de las palabras con diminutivo como integrantes de uno u otro grupo es tajante, sin embargo, varias de las voces sufijadas en el corpus han planteado dificultades en torno a su clasificación. En los casos dudosos, la valoración como diminutivos transparentes o lexicalizados ha estado determinada por el mayor o menor grado de lexicalización de estas voces. Así, el valor semántico se ha establecido como el principal criterio diferenciador. En algunas palabras en las que se aprecia una cierta fosilización se ha prestado atención al grado de diferencia semántica que existe entre la forma con diminutivo y la forma original. Este planteamiento ha llevado a aplicar los siguientes criterios: 1) cuando en esta oposición únicamente predomina la diferencia de tamaño el término se ha considerado transparente; 2) cuando además del sentido dimensional la diferencia semántica afecta a la función o la utilidad del objeto designado, la voz se considera diminutivo lexicalizado. Así, entre *cuchara* y *cucharilla* el rasgo semántico diferenciador es únicamente el tamaño, razón por la que el derivado se estima como un diminutivo transparente⁶², mientras que entre *cortina* y *cortinilla* la forma derivada ha adquirido un sentido específico que va más allá del valor dimensional, considerada por tanto un

⁶⁰ Especie de jubón, con mangas o sin ellas, ajustado al cuerpo (*DLE* s.v. *ALMILLA*, acep. 1).

⁶¹ Sobre el concepto de *lexicalización* véase el apartado siguiente (§ 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*).

⁶² La consideración de *cucharilla* como término lexicalizado es actualmente polémica y los diccionarios difieren en su grado de lexicalización. Mientras que el *DLE* le atribuye solo un cambio de tamaño con respecto a *cuchara*, el *DEA* añade un cambio también en la función: “Utensilio de igual forma que la cuchara, pero de menor tamaño, utilizado para café u otras infusiones o para dulces” (*DEA* s.v. *CUCHARILLA*). Se retoma esta cuestión más detenidamente en los apartados § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo* y § 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*.

diminutivo lexicalizado con el sentido de ‘cortina pequeña que se coloca en la parte interior de los cristales’ (*DLE* s.v. CORTINILLA).

Los diminutivos transparentes son el principal objeto de estudio de esta tesis, ya que forman parte del análisis estadístico que nos permitirá estudiar la variación en el uso de los sufijos *-ico*, *-ito*, *-illo*. La cuantificación de resultados determinará si existen diferencias significativas entre los siglos analizados para las provincias de Almería y Madrid, por lo que la selección de los casos que formarán parte del cómputo numérico ha sido determinante para los resultados que aquí se muestran. Pero no por ello las voces lexicalizadas han sido objeto de menor atención en el devenir de la investigación.

Este capítulo se articula en base a dos ejes: por un lado, diferenciar entre los diminutivos transparentes y lexicalizados; por otro, analizar las formas derivadas en las que el diminutivo ha experimentado un proceso de lexicalización e incluirlas en un glosario. El análisis estadístico de la distribución diatópica y diacrónica de los diminutivos que se desarrolla en el Capítulo 7 parte asimismo de la distinción básica entre diminutivos transparentes y lexicalizados. Es necesario advertir que esta distinción se ha llevado a cabo con la totalidad de palabras del corpus, tanto en los textos almerienses como los pertenecientes a Madrid. La diferenciación entre corpus de estudio y corpus de control únicamente es pertinente a efectos de comparabilidad diatópica y diacrónica, por lo que en las páginas que siguen se alude al corpus analizado como un conjunto completo de textos para lo que no se considera relevante esta distinción.

El tratamiento de las voces con diminutivo en el corpus ha sido el paso previo y necesario para la consecución de los objetivos que aquí se han señalado. Para ello se ha realizado una clara distinción entre diminutivo transparente y diminutivo lexicalizado cuyos parámetros de estudio se van detallando en los apartados que siguen con ejemplos concretos de palabras en el corpus.

6.2. Significado transparente y lexicalizado de los sufijos *-ico*, *-ito*, *-illo*

El análisis planteado en esta tesis acerca de la distribución de los diminutivos implica seleccionar todas las palabras en las que la terminación en *-ico*, *-ito*, *-illo* se corresponde con la presencia de verdaderos diminutivos. En algunas palabras, la aparición del diminutivo se asocia con valores léxicos específicos ajenos a los propios de estos sufijos, por lo que distinguir estos casos de los que realmente interesan ha supuesto el principal punto de partida para el estudio de su distribución. Este planteamiento ha llevado a clasificar todas las palabras con sufijación diminutiva en el corpus en dos grandes grupos según la función que desempeña el sufijo en cada caso y que está determinada por una cuestión semántica. Por un lado, aquellas palabras en las que el diminutivo cumple con un significado transparente se han denominado *diminutivos transparentes*. Por otro, se han considerado *diminutivos lexicalizados* aquellas palabras que han experimentado un cambio de significado, más especializado o restringido que el original y más opaco. Se desarrollan a continuación las características de cada uno de estos grupos.

Se entiende por diminutivos transparentes aquellas palabras en las que el significado puede deducirse de la combinación de la base y el sufijo (NGLE § 9.3a, p. 365), como ocurre en *alfombrita*, *cajica*, *pajarito*, *espejillo*, *cerdicos*. El diccionario puede prescindir de ellas, ya que el diminutivo en estos casos no desarrolla ninguna especialización semántica, sino que depende del significado de la base a la que se une y a la cual modifica. En este sentido, se considera que son transparentes en cuanto a que todo hablante puede interpretar su significado de manera composicional sin acudir a un diccionario.

El valor semántico del diminutivo destaca por su capacidad para expresar numerosos y muy diferentes matices. Entre los diminutivos transparentes, es conocida la común división que se establece entre los dos valores principales que poseen, de los cuales se derivan otros muchos matices asociados a cada uno de ellos: 1) los valores que expresan matices de tipo afectivo (entre los que destacan aquellos que manifiestan atenuación, cercanía, afectividad, cortesía, ironía o desprecio), relacionados con la subjetividad del hablante; y 2) los que expresan una interpretación del tamaño, relacionados con un sentido objetivo de disminución. Se establece así una clara diferenciación semántica que estaría marcada por el mayor o menor grado de subjetividad que aportan estos sufijos.

Sin indagar más en el sentido específico de cada uno de los posibles valores de los diminutivos, pues no es este el objetivo principal, para la caracterización de las llamadas voces transparentes interesa fundamentalmente resaltar que existen dos significados principales que son propios de los sufijos diminutivos: los relacionados con la naturaleza afectiva o los meramente nocionales. Ambos, independientemente del sentido particular que predomine en cada caso, se alzan como valores transparentes del diminutivo, ya que en estas palabras el significado es fácilmente deducible por un procedimiento morfológico productivo. El significado particular de los diminutivos en el corpus documental de la tesis es una cuestión que se trata en detalle más adelante (§ 6.4.2. *Valor semántico del diminutivo en el corpus*). Por el momento, puede anticiparse que, debido a la naturaleza de los documentos, entre los diminutivos transparentes de los inventarios suele prevalecer el sentido objetivo de disminución, ya que la alusión al tamaño resulta natural y básica en los sustantivos que aluden a realidades materiales (NGLE § 9.6b, p. 651), tal es el caso de muchos de los nombres con diminutivos documentados en el corpus.

Los diminutivos opacos o no transparentes se oponen a los hasta ahora mencionados por constituir palabras en las que el significado no se obtiene por la combinación de los dos componentes que las forman, razón por la que deben incluirse en los diccionarios de la lengua y se consideran lexicalizados (NGLE § 9.3b, p. 635). En estas palabras el sufijo ha desarrollado un significado específico que se aleja de los valores prototípicos del diminutivo que indicamos arriba, como se observa en las palabras *abanico*, *almohadilla* o *lanilla* de los siguientes ejemplos:

- (1) Vn *abanico* de Ynglaterra, barillaje de marfil, en ziento y veynte r[eale]s (ODE – Madrid, 1845).

(2) Vna *almuadilla* de coser con sus tigeras (*ODE* – Madrid, San Lorenzo de El Escorial, 1789).

(3) Una colgadura de cama con zinco cortinas y zielo de *lanilla* muy antigua y dos cortinas con su zenefa de lo mismo (*ODE* – Madrid, 1701).

En cada una de estas palabras el diminutivo ya no cumple una función apreciativa, sino que ha formado una palabra distinta con la que se da nombre a una nueva realidad: en el caso de *abanico* ‘instrumento con el que se hace aire’, *almohadilla* se identifica con una ‘almohada pequeña para alfileres’ o ‘acerico’ y *lanilla* ha pasado a designar un ‘tejido de lana más fino y delgado’. Como puede observarse, en estos derivados con diminutivo la relación semántica que se establece con el referente denotado por la base es diferente, más o menos estrecha según el caso. Por ejemplo, la relación entre *abano* y *abanico* es mucho más opaca que la que se establece entre *lana* y *lanilla*. Tanto es así que muy probablemente algunos hablantes no reconozcan en la estructura morfológica de *abanico* la presencia del sufijo diminutivo *-ico* unido a la base *abano*, más aún porque con el tiempo esta forma ha sido desplazada por la diminutiva, aunque su significado se ha mantenido:

Cierto género de bastidór de madéra, mas ancho de abaxo que de arríba, el qual se suele poner en medio con unas tablas delgádas, ò con un lienzo, que le ocúpe todo, y colgádo del techo de la pieza con unas armellas por la parte que es mas angosta, y tirado con una cuerda, pasta por encima de la mesa haciendo áire, y quitando las moscas. Es uno antiguo, y oy se mantiene en muchas partes de la Mancha. Covarr. los llama Ventalles. (*Autoridades* s.v. *ABANO*).

En *lanilla*, en cambio, el diminutivo conserva el sentido de disminución de forma mucho más perceptible que en *abanico*. En este caso, *lanilla* designa un tipo de tejido hecho con lana, más fino y delgado que el habitual. Se retomarán los conceptos de opacidad y transparencia un poco más adelante.

Los diminutivos de las voces no transparentes no pueden interpretarse de manera independiente a las bases léxicas, se denominan por ello diminutivos lexicalizados en cuanto a que han pasado por un proceso de lexicalización en el que perdieron su significado original y han adquirido otro nuevo que ya no es deducible de la combinación de la base y el sufijo. En este punto, se constituyen como unidades léxicas de pleno derecho ya formadas que deben integrarse en el diccionario. Como es esperable, el cambio de significado ocurre mediante un proceso gradual en el que con el paso del tiempo se va perdiendo el sentido transparente del término. En lo que sigue, nos detendremos en la precisión del término *lexicalización* conforme a los parámetros con los que se entiende en esta tesis.

En un sentido amplio, el concepto de lexicalización mencionado hasta ahora es entendido como el cambio semántico por el que un elemento pasa a formar parte del léxico de la lengua y se incluye, por tanto, en los diccionarios. En un sentido más estricto, la lexicalización abarca el proceso mediante el cual una construcción gramatical se convierte en un elemento léxico. En este punto, se distinguen dos tipos de

lexicalización según el plano de la gramática del que parten (Blasco Mateo, 2002, p. 38). El primero de estos tipos se relaciona con los morfemas derivativos que se agregan a una base léxica para formar nuevas voces en las que ha habido una especialización semántica. El sufijo, que podía ser interpretado composicionalmente al inicio, pierde todo valor apreciativo y adquiere una nueva interpretación particular. En la forma, raíz y sufijo experimentan una integración total en la que el sufijo pasa a formar parte del lexema junto a la raíz de la base con la que se vincula ya de forma indivisible. Este proceso a nivel morfológico es el que se observa en numerosas palabras creadas por sufijación diminutiva en el corpus como *almohadilla*, *bolsillo* o *tornillo* en las que el diminutivo ha perdido su valor transparente.

Este tipo de lexicalización no debe confundirse con la que parte de la unión sintáctica de varios elementos para conformar nuevas unidades léxicas, la cual afecta a la creación de palabras compuestas⁶³. Así, el proceso de lexicalización puede partir entonces de diferentes secuencias entre las que se distingue una palabra derivada, un sintagma o una frase (Contreras y Suñer, 2004, p. 60). Hecha esta distinción, conviene aclarar que en este estudio interesan únicamente las lexicalizaciones creadas mediante el primero de estos procedimientos, aquellas desarrolladas en el plano morfológico. De manera que con el concepto de *lexicalización* entendemos el cambio lingüístico que se produce en algunas palabras derivadas y, más especialmente, en aquellas creadas por sufijación diminutiva, en las que el diminutivo pierde su valor prototípico y queda fusionado a la base con la que forma un nuevo lexema. Centrándonos ya exclusivamente en estos casos, trataremos de identificar el modo en que ocurre este proceso.

En español, *-illo* es el sufijo que ha dado un mayor número de lexicalizaciones comunes a todos los hispanohablantes, mientras que son mucho menos frecuentes los derivados en *-ito* o *-ico* (NGLE § 9.3d, p. 636)⁶⁴. Siendo *-illo* el diminutivo más frecuente en la lengua medieval, con el tiempo se verá desplazado por *-ito* e *-ico* en un proceso que Monge (1965) relaciona con la pérdida de expresividad que experimentó este sufijo. Para este autor, la adquisición de sentidos específicos de una palabra con diminutivo se apoya en dos hechos fundamentales: el primero de ellos se basa en el uso repetido de una misma forma con un significado determinado; el segundo tiene que ver

⁶³ En el corpus, este tipo de lexicalización sintáctica se observa en los compuestos *limpiadientes*, *sobrecama* o *guardapelo*, fusionados gráficamente, o los que aparecen sin fusionar –unidos o no mediante preposición–, como *mesa (de) camilla* u *hoja de lata*, cuyas apariciones se combinan también con la variante integrada *hojalata*. Todos ellos constituyen ejemplos de unidades léxicas creadas a partir de dos o más palabras, los cuales no son objeto de estudio para el análisis de esta tesis, a excepción de *camilla*, por contener el sufijo diminutivo en *-illa*.

⁶⁴ Deben hacerse algunas precisiones sobre el español americano. Sorprendentemente, el *Diccionario de americanismos* contiene un mayor número de voces lexicalizadas con el sufijo *-ito* (54 % del total de lemas contenidos con derivación diminutiva) que las producidas con *-illo* (42 %), mientras que las creadas con *-ico* representan solo el 4 % (Aleza Izquierdo, 2017). Cabe destacar que se incluyen en este recuento también los derivados de adverbios e interjecciones en *-ito* (hasta 109 lemas entre ambas clases de palabras), que son mucho más frecuentes en el español americano que en el europeo. No obstante, la preferencia por un tipo de sufijo u otro varía si se tiene en cuenta el campo semántico de las voces derivadas: mientras que las lexicalizaciones en *-illo* predominan en campos como el de la vegetación y la construcción, las formas en *-ito* sobresalen en otros muchos como el relativo a las personas, el mundo animal, la alimentación o los juegos y deportes, entre otros (Aleza Izquierdo, 2017, p. 743).

con la desaparición del carácter afectivo del diminutivo, pues para que la lexicalización se produzca “se precisa la ausencia de componentes expresivos, hasta llegar a un contenido de significación estrictamente objetivo, nocional” (Monge, 1965, p. 141). De este modo, la especialización semántica del derivado ocurriría solo cuando los valores afectivos se ven reducidos a la mínima expresión, esto es, a la pérdida o desaparición de sus funciones apreciativas, y ello solo cuando el empleo recurrente lo convierta en el sentido más habitual hasta el punto de desplazar a su significado primitivo.

Cabe destacar que toda esta transformación ocurre muy gradualmente durante un proceso en el que los cambios semánticos pueden confluir a su vez con cambios morfológicos, de manera que es posible detectar diferentes grados de lexicalización. Aleza Izquierdo (2017) explica el fenómeno de la lexicalización como el de un proceso conformado por varias fases o etapas. La primera de ellas correspondería con la *derivación libre*, en la que no hay ningún cambio apreciable ni en el referente ni en el significado. El derivado en esta fase únicamente se caracteriza por destacar alguna característica del lexema base. En una segunda etapa ocurre la *fosilización*, en la que tiene lugar un cambio en la realidad a la que se alude. No obstante, se sigue manteniendo una relación evidente entre el derivado y la forma primitiva. En la última fase del proceso se desarrolla la *lexicalización*, en la que los cambios tanto de referente como de significado provocan que la relación entre ambas formas no sea fácilmente perceptible (Aleza Izquierdo, 2017, p. 719). Un análisis más progresivo del proceso de lexicalización de los términos con diminutivo es el que hacen Gil y Torres (2011). Las autoras identifican en función del grado de lexicalización que presentan las formaciones lexicalizadas en el diccionario hasta cinco niveles diferentes. A las fases anteriores podrían añadirse aquellas en las que entre el derivado y su base se desarrollan relaciones de hiponimia y metaforización, hasta la desmotivación semántica total entre ambas formas (Gil y Torres, 2011, p. 264).

Tales fases pueden entonces disponerse en un *continuum* en el que ocupando los dos extremos podríamos situar las nociones de transparencia y opacidad. Por transparencia se alude al valor real del diminutivo en el que la voz derivada puede interpretarse por los elementos compositivos que la forman (diminutivos transparentes), mientras que con el concepto de opacidad se relacionan aquellas palabras de significado no composicional (diminutivos lexicalizados). En el extremo izquierdo de este proceso se encuentran las voces transparentes en las que el diminutivo conserva su auténtica función; el extremo opuesto lo ocupan las voces opacas, aquellas que han adoptado un significado específico y que puede estar más o menos vinculado con el de la forma primitiva. A su vez, entre los diminutivos lexicalizados puede haber términos más transparentes que otros, en los que es posible establecer una relación más o menos estrecha entre el significado original del diminutivo y el lexicalizado, más estrecha cuanto más transparente, pero la transparencia y opacidad de estas formas debe entenderse como nociones graduales sujetas a la conciencia lingüística y la cultura particular de cada hablante (NGLE § 9.3c, p. 635).

Si bien muchas de las palabras con diminutivos en el corpus no han planteado dudas a la hora de su clasificación como transparentes o lexicalizadas, en otras ha sido necesario un análisis más profundo que permita resolver esta disyuntiva. Los casos

dudosos que han supuesto un problema añadido para su clasificación se analizan en el apartado § 6.3.2. *Casos dudosos y problemas para la cuantificación*.

Algunas palabras se sitúan en estadios intermedios de lexicalización en los que las nociones de transparencia y opacidad son incluso más difíciles de acotar y objetivar, pues estando inmersas en un proceso evolutivo estos conceptos son necesariamente aproximativos (+ – transparente) / (+ – opaco). A la dificultad de valorar la diferencia entre el derivado y su primitivo en lo que al aspecto semántico se refiere, se añade en ocasiones el polimorfismo que presentan algunas voces ya lexicalizadas o en vías de lexicalizar. Este tipo de palabras han supuesto un problema metodológico a la hora de su clasificación, pues se ha debido analizar el grado de lexicalización para poder considerarlas como transparentes o lexicalizadas.

Durante el proceso de cambio semántico la voz derivada va ganando progresivamente autonomía, pero durante este transcurso puede sufrir alteraciones de diversa índole. Uno de estos cambios tiene que ver con el aspecto formal y es que es bastante habitual que en voces en vías de lexicalización se produzcan variaciones en el empleo del diminutivo. Así, algunas palabras en el corpus que manifiestan un evidente sentido específico presentan alternancia en el empleo del sufijo. Un ejemplo claro de ello es la palabra *bufetillo*, a la que se le ha dedicado una mayor atención dada su especial complejidad (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). Las ocurrencias de esta voz alternan con las de *bufetico*, y en menor medida con las de *bufetito*, a lo largo de todo el corpus sin que haya una clara distribución por zonas o épocas en el uso de un sufijo u otro. Este mismo hecho es señalado por Montero Curiel (2008) a raíz del proceso de lexicalización del diminutivo *ardilla*:

Antes de su lexicalización, el diminutivo *ardilla* pasó por sucesivos estadios de afianzamiento que dan como resultado el abundante polimorfismo del vocablo presente hasta nuestros días en los repertorios del léxico castellano [...] Otras variantes (como *hardilla*, *ardita*, *ardica*), pese a sus restricciones diacrónicas y diatópicas, son fundamentales para abarcar el proceso de fusión de la base con el sufijo en toda su complejidad (Montero Curiel, 2008, p. 119).

De forma análoga, Clavería Nadal (2003) relaciona la variación en el diminutivo de determinados nombres de plantas con un menor grado de fijación léxica en estos términos. Las voces que no registran alteraciones, como *manzanilla*, evidencian la lexicalización completa frente a otras en las que destaca la diversidad en el sufijo, que puede estar condicionado por la distribución geográfica (Clavería Nadal, 2003, p. 78). De esta manera, existen diferentes niveles de afianzamiento en los términos en vías de lexicalizarse. El proceso se considera culminado cuando existe la posibilidad de añadir al nuevo lexema otros morfemas derivativos. Una prueba fehaciente de ello en nuestro corpus son las palabras *mantillita* y *ovillo*, en las que el diminutivo ha quedado completamente integrado y los sufijos *-ita* e *-ico* se añaden a las voces *mantilla* y *ovillo* ya lexicalizadas para aportar de nuevo un sentido transparente de disminución.

Como se ha mencionado, la variación que experimenta el sufijo es el reflejo de fases previas a la fijación definitiva del término. Algunas palabras en las que se constata un cambio semántico evidente pueden no haber fosilizado en la forma, que todavía se encuentra sujeta a variación. El proceso de cambio semántico y morfológico, si bien se desarrollan a la par, podrían culminar en diferentes momentos. La alternancia en el sufijo disminuye progresivamente hasta que la consolidación del término se considera completa. Un claro ejemplo de esta circunstancia es el caso en nuestro corpus de la alternancia *puntilla/puntica* en el mismo documento:

- (4) Dos pares de calzoncillos ya mojados, los vnos con unas *puntillas* y los otros mas rejios [...] Vna corbata añadida y con *puntas* antiguas [...] Vn pañuelo con *punticas* antiguas, viejo (*ODE* – Madrid, 1700).

La forma *puntilla* como derivado de *punta* alcanzó la especialización semántica a partir del diminutivo como ‘encaje generalmente estrecho que forma ondas o picos en una de sus orillas y que se pone como adorno en el borde de pañuelos, toallas, vestidos, etc.’, definición con la que hoy en día se incluye el término en el diccionario (*DLE* s.v. PUNTILLA).

Entre los casos de *puntillas* en *ODE*, la ocurrencia de *punticas* constituye un vestigio de la variación sufijal que experimentó el término antes de que se generalizara la forma definitiva en *-illa*. Esta es la única variante que aparece registrada en las fuentes lexicográficas (*NTLLE*), sin embargo, a medida que se retrocede en el tiempo, surgen más ejemplos reveladores de las distintas formas con las que se designó este tipo de encaje. La consulta de *CorLexIn* (ejemplos de 5) y del resto de *ODE* (ejemplos de 6) demuestra que *puntilla* alternó no solo con *puntica*, sino también con *puntita* en contextos en los que se confirma ya un claro sentido lexicalizado de estos términos:

- (5) a. Otro paño de chuez entretallado con *puntica* alrededor en veinte y dos reales (*CorLexIn* – Albacete, Almansa, 1653).
 b. Vna toalla de ruan con sus redes y *punticas* en dos ducados. Vna toalla de ruan labrada de ylo amarillo con sus *puntas*, en diez y seis reales [...] Vna camissa de muxer de cuerpo de ruan y las faldas de lienço con unas *punticas*, en diez y seis reales (*CorLexIn* – Toledo, Escalona, 1644).
 c. Un armador de lama [sic] azul de Italia con *puntitas* de plata [...] Dos pares de calsones de tafetán, los unos colorados con pinos de çeda amarilla y *puntitas* de borzelana, y los otros de tafetán doble mescla, bordados con çeda azul y *puntas* de borcelana (*CorLexIn* – Bolivia, Potosí, 1677).
- (6) a. Una colcha de cotonia blanca con vnas *puntitas* a vn lado (*ODE*– Cáceres, 1675).
 b. Dos panos de mano de lienço, el vno con vnas *puntitas* y otro llano, en doze r[eale]s (*ODE* – Badajoz, 1664).

En todos los ejemplos de arriba el contexto garantiza el valor específico con el que se emplean indistintamente las voces *puntillas-punticas-puntitas*. Cabe destacar que los derivados con diminutivo conviven además con la forma primitiva *punta* o *puntas*

(usada mayoritariamente en plural con este sentido, tal y como indica también el *DLE* en su definición) y que abunda tanto en los inventarios de Almería y Madrid analizados como en los corpus complementarios. Sobre esta alternancia entre forma primitiva y derivada puede verse el mismo ejemplo mencionado en (4) o los extraídos del *CorLexIn* en (5b-c). A estos añadimos una reducida selección de la alternancia *puntas-puntillas* documentada en nuestro corpus de inventarios y en la que se puede apreciar que el significado con el que se emplean es el mismo:

- (7) a. Vna bara de tafetan carmesi con *puntilla* blanca (*ODE* – Madrid, 1750).
- b. Doce varas de *puntilla* para un juego de cama (*ODE* – Madrid, 1852).
- c. Otras dos [toallas] viejas con *puntillas* (*ODE* – Madrid, Algete, 1855).
- d. Un pañuelo de tul blanco bordado, ochenta r[eale]s. Uno id[em] de espumilla de olan, doscientos r[eale]s. Tres id[em] de tul con *puntilla*, cuarenta reales (*ODE* – Almería, 1844).
- e. Vna tohalla de medianillo con *puntas* (*ODE* – Almería, Abila, 1726).
- f. Una almilla de griseta encarnada con *punta* de plata (*ODE* – Almería, Batares, 1774).
- g. Un corte de *puntas* para mantos de encaxe finos y anchos (*ODE* – Madrid, 1750).

Tales muestras sirven para constatar la relación de sinonimia que se establece entre la forma primitiva (*puntas*) y sus derivadas (*puntillas-punticas-puntitas*), estando estrechamente relacionadas en su significado. En estos casos, en ese *continuum* evolutivo al que se aludía con anterioridad, el proceso de lexicalización se identifica más con la noción de transparencia. Este hecho queda reflejado también en la similitud con la que se definen ambos términos en el diccionario: la palabra base suele aparecer en la definición del derivado cuando el término se considera fosilizado, pero no lexicalizado (Aleza Izquierdo, 2017, p. 719)⁶⁵.

La alternancia de variantes en el corpus ha sido entonces entendida como prueba de la existencia de fases previas a la consolidación formal del término, las cuales son perfectamente identificables desde una perspectiva diacrónica. El estudio de los inventarios de Almería y Madrid permite observar estos cambios en la fosilización de los términos derivados de acuerdo con distintas fases o etapas. Así, la diversidad en el empleo del diminutivo se asocia con un proceso de lexicalización todavía incompleto, de manera que estos términos se han considerado diminutivos transparentes en aquellos casos en los que la diferencia semántica parece afectar únicamente al tamaño del referente. Algunos de estos casos se analizan de manera pormenorizada en § 6.3.2. *Casos dudosos y problemas para la cuantificación*. Sin embargo, en otras palabras, en las que se manifiesta del mismo modo la variación en el sufijo, la diferencia semántica trasciende el valor meramente nocional, por lo que estos casos sí se han tratado como diminutivos lexicalizados al considerarse en ellos que el valor del diminutivo ha dejado

⁶⁵ *Punta*, en el *DLE*, se define como ‘encaje que forma ondas o puntas en una de sus orillas’ (s.v. PUNTA, acep. 20), muy similar a la definición de *puntilla* que se mencionó más arriba: ‘encaje generalmente estrecho que forma ondas o picos en una de sus orillas y que se pone como adorno en el borde de pañuelos, toallas, vestidos, etc.’ (s.v. PUNTILLA).

de ser transparente, un ejemplo es el ya mencionado de *puntilla* y *bufetillo* (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*).

En muchos casos, la especialización semántica de la forma derivada no impide que esta deje de emplearse en su función claramente disminuidora. Lejos de esto, el diminutivo puede desempeñar diferentes interpretaciones, ya sea como significado transparente u opaco al mismo tiempo. Resolver esta ambivalencia es determinante para considerar el término como diminutivo transparente o lexicalizado de cara al análisis diatópico y diacrónico realizado en esta tesis, por ello, junto a los criterios que se detallan a continuación, el contexto de aparición que ofrecen los propios textos ha jugado un papel fundamental como elemento desambiguador, especialmente para la clasificación de diminutivos ambiguos.

6.3. Criterios para la desambiguación de diminutivos transparentes y lexicalizados

Los principios que se han tenido en cuenta para la clasificación entre diminutivos transparentes y lexicalizados han estado determinados por los criterios lexicográficos y semánticos que seguidamente se detallan. Así, se han considerado diminutivos transparentes los casos en los que se cumple alguno de los siguientes puntos:

a. Criterio lexicográfico:

- La palabra no se incluye en el diccionario puesto que su significado se deduce a partir de la combinación de la base y el sufijo (interpretación transparente del diminutivo), como por ejemplo sucede en *alfombrita*, *colchoncillo*, *canastica* o *librito*.
- La palabra se recoge en el diccionario, puesto que ha desarrollado sentidos lexicalizados, pero esto no impide que se pueda seguir empleando con un sentido transparente. El corpus documenta ambos sentidos de una misma palabra y es el propio contexto de aparición el que determina la desambiguación en palabras como *arillo*, *bolsillo* o *cortinilla*. Cada ocurrencia de estos términos ha sido etiquetada de manera individualizada en función del contexto.

b. Criterio semántico:

- La palabra se incluye en el diccionario con el significado que reflejan los textos, pero apenas existen diferencias, a excepción de las relacionadas con el tamaño, entre el derivado y su base, como por ejemplo en *cucharilla*, *platillo* o *bacinica*.

- La palabra se considera actualmente una voz lexicalizada, pero en la documentación de nuestro corpus se considera en una fase anterior a la lexicalización definitiva, como ocurre en *banquillo*, *mesilla* o *toallita*.

En lo que sigue, se explica de manera pormenorizada el proceso de aplicación de estos criterios, así como la motivación que ha llevado a adoptarlos, a partir del análisis de algunos casos que se juzgan interesantes.

En *alfombrita*, *cajica* o *embudillo* el valor semántico del diminutivo se asocia con un sentido dimensional: el que se refiere al tamaño menor del objeto expresado por las bases *alfombra*, *caja* y *embudo* respectivamente. Se trata de voces transparentes en su morfología, dado que su significado es fácilmente predecible a partir de la combinación de la base y el sufijo, razón por la que no se incluyen en los diccionarios (NGLE § 9.3a, p. 635).

Muchas de las palabras consideradas en el corpus como diminutivos transparentes han desarrollado a su vez uno o más significados opacos que sí quedan registrados en el diccionario, como por ejemplo ocurre en *artesilla*, *calderilla* o *romanilla*. La lexicalización de estos términos no impide que en otros casos se puedan seguir empleando con su sentido nocional más básico, pues, en principio, toda palabra que ha lexicalizado sigue estando disponible para la expresión apreciativa.

En aquellas palabras en las que el sufijo está sujeto a un doble valor semántico la consideración de diminutivo transparente o lexicalizado ha estado determinada por el contexto, el cual ha jugado un papel determinante en la clasificación de los diminutivos (§ 6.3.1. *El contexto como elemento desambiguador*). Así ha ocurrido, por nombrar solo algunos casos, con las palabras *cortinilla* o *arillo*, cuyo empleo en el corpus permite constatar tanto ejemplos de su uso transparente como lexicalizado⁶⁶. Los casos de (8) reflejan un empleo transparente del diminutivo en estos términos, al contrario que los señalados en (9), donde se constata un significado lexicalizado:

- (8) a. Dos *cortinillas* de tafettan con una zenefa uiejas (ODE – Madrid, 1701).
 b. Un *arillo* de yerro con quarenta y ocho escarpas para trabaxar con su peso en medio (ODE – Madrid, 1711).
- (9) a. Dos *cortinillas* de ventana en cinco r[eale]s (ODE – Madrid, San Lorenzo de El Escorial, 1858).
 b. Mas vnos pendientes de aljofar con sus *arillos* de oro, todo en nobenta reales (ODE – Madrid, 1724).

Otras veces, el significado opaco se ha desarrollado con posterioridad al lapso que representan los textos analizados, por lo que no se ha considerado que tales formas aparezcan en los inventarios con un sentido fosilizado. Es el caso de *mesilla* (analizado con más detalle en § *Mesilla*) o *maquinilla*, cuyo significado actual como ‘máquina de

⁶⁶ El sentido lexicalizado de *arillo* en el corpus se identifica con el de *arete* ‘tipo de pendiente’ (DLE s.v. ARILLO, acep. 2). Para la desambiguación de los valores transparentes y lexicalizados de *cortinilla* véase más adelante § *Cortinilla*.

afeitar’ es posterior al del sentido transparente que se refleja en los ejemplos “una *maquinilla* de hacer filetes”, “una *maquinilla* de volver hojas de canelón” (ODE – Madrid, Aranjuez, 1870) y “una *maquinilla* para hacer café” (ODE – Madrid, 1877), donde la voz no desempeña un significado específico, sino más bien el genérico de ‘artefacto de pequeñas dimensiones movido por algún mecanismo’.

Entre las voces con diminutivo que han pasado a formar parte del léxico de la lengua existe una diferencia entre aquellas cuyo significado guarda una estrecha relación con el término del que proceden y aquellas en las que la especialización léxica ya apenas mantiene relación con el significado del término original. Por ejemplo, entre *aro* y *arillo*, junto a la diferenciación en el tamaño (un *arillo* es siempre más pequeño que un *aro*), se da también un cambio significativo en la función (un *arillo* es un tipo de pendiente) y en su constitución (un *arillo* puede estar hecho de metales preciosos), características que lo diferencian claramente de la forma *aro* de la que deriva. Sin embargo, a pesar de la interpretación específica que adquiere *arillo* como elemento lexicalizado, sigue siendo perceptible el nexo entre ambos términos. Esa relación, en cambio, ya no se reconoce entre *cinto* y *cintillo*, pues sus significados no se asemejan ni por la forma ni por la función de los objetos a los que se refieren, aunque la diferencia de tamaño sigue estando presente de alguna manera: se asume que un *cintillo* debe ser algo pequeño y, efectivamente, un *cinto* es una faja y un *cintillo* es una sortija pequeña (DLE s.v. CINTO, CINTILLO). En estas formas derivadas (*arillo* y *cintillo*) se sigue manteniendo la referencia al tamaño pequeño de los objetos, pero no se aprecian en ellas de igual manera otros cambios semánticos relacionados con el uso, material, aspecto, etc. que las hacen adoptar un sentido muy específico y alejado del original.

Existen términos en los que consideramos que la diferencia de significado entre derivado y forma originaria está marcada sobre todo por la desigualdad de tamaño, de modo que el diminutivo desempeña su función prototípica sin que tenga sentido hablar de lexicalización. Algunos ejemplos en el corpus son *cucharilla*, *cacillo* o *bacinilla*. Las definiciones del DLE para estos términos no son diferentes a las de ‘cuchara pequeña’, ‘cazo pequeño’ y ‘bacina pequeña’, según el uso que de tales formas se hace en el corpus. Aun estando incluidos en el diccionario, no muestran diferencias con el significado del término del que derivan, por lo que se han considerado como diminutivos transparentes. A *bacinilla* y *cucharilla* se alude de forma particular en los apartados § *Bacinilla* y § *Cucharilla*.

La variación en el sufijo de una misma palabra ha tenido consecuencias para la selección de los diminutivos transparentes. En el corpus, la alternancia *azadilla-azadica*, *cucharilla-cucharita*, *cacillo-cacito* o *platillo-platito-platico* ha favorecido la consideración de diminutivos transparentes en estos casos, puesto que la variación en el sufijo no se suele manifestar en otros términos claramente lexicalizados en los que la relación semántica entre derivado y forma original no es tan evidente⁶⁷. De este hecho podemos inferir que cuanto más estrecha es esa relación, mayor grado de libertad

⁶⁷ Solo las lexicalizaciones de *puntilla* y *bufetillo* reflejan en el corpus la alternancia en el sufijo atestiguada en las formas *puntica* y *bufetico-bufetito* (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*).

demuestra el sufijo de la forma derivada⁶⁸. En los casos mencionados arriba, la alternancia en el sufijo constata una menor fijeza léxica de los términos derivados con diminutivo, razón por la que se han considerado plenos diminutivos.

En algunas de las palabras que se han ido mencionando, la variación del diminutivo no establece un dominio claro de un sufijo sobre otro y la alternancia en estos términos sería el reflejo de un proceso a medio camino hacia la lexicalización definitiva. Estos términos, cuya forma no se ha asentado aún en el léxico de la lengua, favorecen la alternancia en el uso de los sufijos en contraposición con los que ya se han fijado completamente. La alternancia en el diminutivo de formas sufijadas incluidas en el diccionario se ha visto como un indicador de la inestabilidad léxica del término todavía en la época de los textos, razón por la que se ha valorado su consideración como diminutivos transparentes si el significado única o principalmente parece asociarse con el cambio de tamaño.

Algunos términos, como *banquillo* y *platillo*, aunque evidencian en algunos ejemplos del corpus un proceso avanzado de lexicalización, se han considerado finalmente como diminutivos transparentes a la vista de la alternancia que presentan con sus formas base (*banco*, *plato*) en las que se identifican los mismos valores semánticos. Esta decisión es polémica si se tiene en cuenta que *banquillo*, por ejemplo, es considerada una lexicalización para algunos autores (Perdigüero, 2015; Laguna Álvarez, 2017). Estas palabras forman parte de aquellos casos considerados dudosos en los que se asume lo discutible que puede llegar a ser hablar de diminutivos transparentes o lexicalizados según qué caso. Los criterios que han llevado a considerarlos como diminutivos transparentes se exponen detenidamente en § *Banquillo* y § *Platillo*.

Resumiendo todo lo anterior, el principal criterio por el que algunas de las palabras han sido tenidas en cuenta como diminutivos transparentes se ha visto justificado por el hecho de que solo se puede constatar una diferencia de tamaño entre el valor semántico del diminutivo y el término del que procede. Este criterio ha prevalecido frente al de la inclusión de los derivados en el diccionario, pues la información que ofrece el contexto proporcionado por los inventarios y los datos lexicográficos no han sido suficientes para determinar un cambio significativo en la función de los objetos. Así, se ha visto como entre *cucharilla* y *cuchara*, *azadilla* y *azada*, *mesilla* y *mesa*, *bacinica* y *bacina* o *banquillo* y *banco* no existen diferencias más allá de las referidas al tamaño que nos hagan pensar en un sentido específico del diminutivo alejado del representado por la base.

En algunos de estos derivados se constata además la variación en el diminutivo, frente al término lexicalizado en *-illo/-illa* se encuentran otros con *-ito/-ita* o *-ico/-ica*.

⁶⁸ Algunos de los aspectos que centran este capítulo han sido publicados de forma preliminar en Arrabal Rodríguez (en prensa) de forma mucho más reducida. La publicación, derivada de este análisis, se centra en la dificultad para diferenciar el valor transparente y lexicalizado del diminutivo, acentuada en aquellos casos en los que el contexto no permite la desambiguación. La alternancia en el sufijo (*-ito/-ico/-illo*) ha sido en algunos casos el criterio que ha servido para probar el sentido transparente del diminutivo, favorecida además por la relación semántica estrecha que se establece entre algunas formas con diminutivos y las bases léxicas de las que derivan. Este planteamiento se desarrolla ampliamente a lo largo del capítulo a raíz de casos concretos.

Esta alternancia en el sufijo ha sido también un criterio determinante a la hora de considerar el carácter transparente del diminutivo, pues, como se ha explicado, la variación sufijal parece darse con más frecuencia en casos en los que los dos términos se relacionan estrechamente en su significado, esto es, cuando el término no está plenamente fijado en el léxico. El criterio semántico se ha impuesto al de la inclusión de las voces en el diccionario y ha sido el que ha llevado a considerar como plenos diminutivos algunos casos de palabras derivadas que son consideradas como formas lexicalizadas por algunos autores o fuentes lexicográficas. Esto es así porque en algunas ocasiones los propios diccionarios no son consecuentes con la técnica lexicográfica cuando incluyen definiciones del tipo “cuchara pequeña” o “cazo pequeño” como definiciones de *cucharilla* o *cacillo* (*DLE* s.v. CUCCHARILLA, acep. 1 y CACILLO, respectivamente).

Para terminar, se condensa el proceso que se ha llevado a cabo para determinar el valor transparente del diminutivo. En un primer lugar, la consulta del diccionario ha sido decisiva para considerar el carácter de diminutivo del término. La ausencia de la forma en los repertorios lexicográficos ha facilitado su inclusión en este grupo. El análisis del contexto ha sido clave para valorar aquellos casos en los que el término sí se recoge en el diccionario, pero cuando la definición no coincide con el sentido con el que se emplea en los inventarios, se ha tratado igualmente de plenos diminutivos. Por otro lado, el criterio de la similitud semántica entre derivados y formas originales ha predominado en las palabras que se incluyen en el diccionario ya como lexicalizaciones y también en aquellas cuya alternancia de género no implica cambios de significado. En algunos de estos casos, la variación observada en el empleo del diminutivo se ha considerado como un signo de las fases intermedias que atraviesan los términos en su camino hacia la lexicalización. El proceso definitivo no estaría culminado del todo en esta época, por lo que se ha estimado más acertada su valoración como diminutivos transparentes todavía en este tiempo. De este modo, cuando el criterio lexicográfico no ha bastado para la clasificación de los términos con diminutivo, el criterio semántico ha prevalecido en la toma de decisiones.

La consideración como diminutivos transparentes o lexicalizados de las palabras en el corpus ha estado determinada en buena parte por el análisis del contexto (§ 6.3.1. *El contexto como elemento desambiguador*), pero en muchas otras palabras el criterio adoptado puede ser discutible en algunos casos. Por esta razón los términos de compleja sistematización se discuten de forma aislada, donde se analizan de manera pormenorizada las circunstancias que nos han llevado a considerar como plenos diminutivos palabras que podrían ser juzgadas como lexicalizaciones si no fuesen considerados como válidos los criterios que aquí se han descrito (§ 6.3.2. *Casos dudosos y problemas para la cuantificación*).

6.3.1. El contexto como elemento desambiguador

Numerosos diminutivos en los inventarios presentan la particularidad de ser ambiguos, puesto que pueden tener un valor apreciativo o estar plenamente lexicalizados. Esta circunstancia afecta sobre todo a los derivados en *-illo* debido al alto número de lexicalizados con este sufijo. Así, la palabra *arillo* puede poseer un valor apreciativo cuando se corresponde con el significado de ‘aro pequeño’ o un sentido específico adquirido tras un proceso de lexicalización cuando se identifica con el de *arete* ‘pendiente’ (*DLE* s.v. ARILLO, acep. 2). En el primer caso, se trata de un auténtico diminutivo en el que el sufijo mantiene su valor nocional más básico referido al tamaño, mientras que en el segundo estamos ante una forma lexicalizada cuyo significado se recoge en el diccionario. Distinguir entre el valor transparente del diminutivo y el opaco de la forma lexicalizada es una cuestión determinante de cara al análisis estadístico llevado a cabo en esta tesis, pues de ello dependerá la inclusión o no del término en el recuento de los diminutivos transparentes.

Además de la ambigüedad léxica que conllevan los términos con diminutivo, hay que sumar los no pocos casos en los que una voz lexicalizada ha desarrollado a su vez diferentes sentidos específicos. No es raro que en los inventarios un mismo término ofrezca ejemplos para varias de sus diferentes acepciones recogidas en el diccionario. *Camilla*, por ejemplo, se localiza en nuestra documentación con hasta tres significados diferentes: el de ‘mesa camilla’, el de ‘cama portátil’ y el de ‘enjugador’⁶⁹ (§ 6.5.3.4. *Glosario* s.v. CAMILLA). La desambiguación de estos casos no influye en el análisis de los diminutivos, pero sí ha contribuido a dar cuenta de las diferentes dimensiones, formas, funciones, etc. que podían tener los objetos de la época y que en ocasiones no se encuentran reflejadas en las fuentes lexicográficas. Casos como estos han servido para ejemplificar los varios significados de una misma palabra y enriquecer así las definiciones que aportan los diccionarios a partir de los ejemplos del corpus.

De esta manera, la clasificación de los términos con diminutivo ha supuesto una doble dificultad fruto de las distintas ambigüedades que ocasionalmente pueden presentar: 1) la que tiene lugar entre el valor transparente del diminutivo y el opaco de la forma lexicalizada; 2) la que ocurre entre los distintos significados del derivado lexicalizado. Para resolver las posibles ambigüedades ha sido necesario recurrir a diferentes estrategias favorecidas por el contexto para una correcta interpretación semántica.

Identificar los objetos de manera inequívoca era imprescindible de cara a la tasación, por lo que, en muchos de estos casos, es el propio escribano quien se sirve de diferentes fórmulas para evitar la ambigüedad léxica de muchos de los términos empleados (Morala, 2019). Uno de estos procedimientos está relacionado con la precisión con la que se describen los objetos presentes en la tasación. Las alusiones a la forma, el tamaño o la función son más exactas cuando se refieren a objetos que pueden confundirse en alguna de sus características con otros más comunes y cuya frecuencia

⁶⁹ ‘Especie de camilla con un enrejado de cordel en la parte superior, que servía para secar y calentar la ropa’ (*DLE* s.v. ENJUGADOR).

de aparición es mayor. Las matizaciones de este tipo se incluyen a fin de evitar ambigüedades cuando el contexto lo requiere. Partiremos de un ejemplo para aclarar este hecho.

El término *brasero* está ampliamente representado en el corpus con el sentido de ‘recipiente en el que se echan las brasas para calentarse’, siendo esta su acepción más general. Los objetos que acompañan al brasero (tarima, caja de madera, badil, copa, etc.), así como las referencias al material (cobre, azófar, hierro), dejan claro que este era el sentido habitual del término, por lo que no es necesario señalar la función que desempeñaban ni añadir especificaciones de cualquier otro tipo. Sin embargo, es lógico que cuando se mencione un brasero con un uso distinto al habitual, se incluyan referencias más precisas que lo diferencien del brasero común. Es lo que sucede en el ejemplo “1 *braserillo* p[ar]a fumar de plata con mangos de madera” (*ODE* – Madrid, 1843), donde se añaden matices importantes relacionados con su función –“para fumar” – y el material –“de plata”–, características bastante inusuales para un brasero y que estarían refiriéndose a un objeto diferente, probablemente, a un utensilio para encender o quemar el tabaco. De especial importancia han sido aquellos casos en los que, similar a este, las precisiones del escribano han permitido determinar el significado del término, y si el término actúa como pleno diminutivo o como voz lexicalizada. Se mencionan a continuación otros ejemplos concretos.

El *pañó*, entendido como ‘trozo de tela’, es un objeto de uso general y presencia recurrente en los inventarios. Los múltiples usos a los que se destinaba hacen que sea necesario especificar sus características para poder identificarlo y tasarlo adecuadamente. En consecuencia, es bastante frecuente encontrar la distinción habitual entre “pañó de manos” (también “toalla de manos”) y “pañó de cama”. En el corpus se localizan las voces con diminutivo *pañico/pañito* y *toallitas*. Dado que ambas han desarrollado un significado lexicalizado que recoge el diccionario, sería difícil establecer el valor del diminutivo si no se hicieran mayores precisiones en los textos. En el caso de *toallitas*, la intuición lleva a pensar que se trata de un verdadero diminutivo, pues el único valor lexicalizado como ‘toallita húmeda’ es bastante moderno. Sin embargo, *pañito* se encuentra en el diccionario como voz lexicalizada con el significado de “trozo de tela adornado o labor hecha de encaje, ganchillo, etc., que se usa para cubrir o embellecer bandejas, sillones, mesas y otros objetos” (*DLE* s.v. PAÑITO), por lo que se precisa de mayor contexto para deshacer la ambigüedad. Junto a los paños ya mencionados, también se localizan los “paños de mesa”, los paños empleados en el oficio religioso identificados como “pañó de holán para bautizar”, “pañó de bautismos”, “pañó de altar” o “pañó de servicio”, u otros destinados al aseo personal: “pañó de peinar”, “pañó de afeitarse”. Entre los ejemplos de este último, *ODE* ha servido para ampliar los casos de inventarios de otras provincias en los que se evidencia su uso para limpiar la navaja con la que afeitarse. El paño o toalla empleado con este fin era de un tamaño menor, por lo que en ocasiones se emplea el diminutivo con un claro sentido de disminución:

- (10) a. Un peinador, tohalla y *pañico* de navajas (*ODE* – Almería, 1736).
- b. Dos toallas y *pañitos* para afeitarse y peinarse (*ODE* – Sevilla, 1795).

c. Tres *thoallitas* p[ar]a limpiar la navaja (ODE – Almería, Cuevas de Almanzora, 1766).

El valor nocional del diminutivo queda fuera de toda duda si se compara con estos otros ejemplos: “dos *paños* de afeitar” (ODE – Madrid, Estremera, 1836), “un *pañó* de afeitar” (ODE – Almería, Escúllar, 1837), “vn *pañó* para limpiar la nauaja” (ODE – Murcia, Lorca, 1761), los cuales corroboran el valor denotativo del sufijo en los ejemplos de arriba. Puntualizaciones como las señaladas hacen que no sea necesaria la alusión al tamaño por medio del diminutivo o de adjetivos relacionados con las dimensiones del objeto, el cual ya quedaba perfectamente identificado en los ejemplos en los que se emplea el genérico *pañó*.

El mismo procedimiento se ha empleado para otros tantos ejemplos que definen y en los que el diminutivo puede cumplir una doble función: “una *calderilla* de ylar seda” (ODE – Almería, Abia, 1726), “una *paletilla* de yerro p[ar]a raer zera” (ODE – Madrid, 1711). Estas precisiones sirven para distinguir claramente el sentido dimensional del diminutivo sin que se le atribuya alguna otra especialización léxica. Los casos de *calderas* y *paletas* que no cuentan con mayores puntualizaciones se vinculan con la acepción más general de estas palabras como objetos utilizados en la cocina. Su aparición en los inventarios junto a otros utensilios afines en su uso como asadores, parrillas, trébedes o sartenes los diferencia claramente de la *calderilla* y las *paletillas* de los ejemplos, las cuales deben ser descritas con mayor precisión y evitar que se confundan con las de uso más común.

El diminutivo se asocia con un significado ya fosilizado en “vna *almuadilla* de coser con sus tigeras” (ODE – Madrid, San Lorenzo de El Escorial, 1789), donde el término se relaciona estrechamente con las labores de costura, tratándose, por tanto, de una voz ya lexicalizada empleada con el significado de ‘acerico’⁷⁰ (DLE s.v. ALMOHADILLA, acep. 5).

En los casos mencionados hasta ahora las precisiones realizadas por el escribano han permitido deshacer la ambigüedad existente entre el valor léxico y el valor apreciativo del diminutivo. En otros casos, estas mismas precisiones han servido para distinguir entre los significados de aquellas lexicalizaciones que han adoptado acepciones diferentes.

Del lexicalizado *cepillo* se registran distintos significados gracias a las especificaciones “cepillo de carpintero” (ODE – Almería, Alboloduy, 1819), “cepillo de limpiar ropa” (ODE – Madrid, San Lorenzo de El Escorial, 1858) o “zepillo p[ar]a limpiar poluo” (ODE – Almería, Cuevas de Almanzora, 1766), con las cuales es posible detectar distintos usos de este objeto como instrumento de carpintería o como utensilio para la limpieza, ya sea del polvo o la ropa. Para los objetos con amplia presencia en el corpus o con funciones tan diversas como las que tiene un cepillo, los casos como los mencionados en los que se describe de manera inequívoca el objeto son los que facilitan luego poder identificarlos con mayor claridad cuando no se alude a ellos con tanta precisión o simplemente se mencionan sin mayor detalle. Por ejemplo, en “una azuela,

⁷⁰ ‘Almohadilla que sirve para clavar en ella alfileres o agujas’ (DLE s.v. ACERICO, acep. 1).

un zepillo, un escoplo y otras erramientillas” (*ODE* – Almería, Abrucena, 1784) o en “tres serruchos, un cepillo y una azuela” (*ODE* – Madrid, Aranjuez, 1870) el contexto y los objetos circundantes, entre ellos *escoplo* y *azuela* –relacionados indiscutiblemente con el ámbito de la carpintería–, posibilitan poder reconocer de manera más precisa el tipo de instrumento al que se hace referencia, así como identificar la acepción del término *cepillo*, en ambos casos, con la labor de carpintería. Detectar las diferencias de significado es imprescindible para vincular los ejemplos con las distintas acepciones de la voz en el glosario de diminutivos lexicalizados (§ 6.5.3.4. *Glosario*).

Hasta aquí, se ha expuesto cómo las puntualizaciones del escribano para determinados objetos han sido determinantes para la desambiguación entre los distintos valores del diminutivo. Sin embargo, para otros casos confusos en el corpus y en los que el escribano no ha descrito con precisión el objeto al que se refiere, se han debido buscar otras estrategias que faciliten la interpretación. Para ello ha sido necesaria una lectura más amplia del contexto con el fin identificar el sentido con el que se emplea la palabra.

Los inventarios se caracterizan por contener extensas listas de objetos, pero en raras ocasiones estas enumeraciones se presentan de manera arbitraria. En ellas predomina un orden interno en el que los objetos se agrupan en función de su significado, el material del que están hechos o las estancias de la casa en la que se localizan (Morala, 2019, p. 388), de tal manera que los objetos se clasifican en los textos por tipologías onomasiológicas. En algunas ocasiones, esta clasificación se manifiesta explícitamente en los inventarios cuando es el propio escribano el que con anotaciones al margen hace una división por el género de ropa (ropa blanca, vestidos, ropa de cama, etc.), por la materia (plata, madera, cobre, hierro, etc.), por los aposentos donde se encuentran los bienes inventariados (cocina, dormitorio, sala principal, alcoba, etc.) o incluso por la persona a la que pertenecen los bienes (ropa del difunto, ropa de caballero, regalos del novio)⁷¹. Otras veces, tales indicaciones pueden aparecer también como encabezado de distintas listas dentro del mismo inventario⁷². Asimismo, en otros muchos inventarios simplemente se nombran todos los bienes de corrido y no hay divisiones como las mencionadas, pero incluso en estos casos subyace también un cierto orden lógico en la presentación de los objetos, los cuales aparecen agrupados por categorías.

El ejemplo “dos *paletillas* en tres r[eale]s” (*ODE* – Madrid, Navalafuente, 1809) no aporta un contexto suficiente que permita aclarar si estamos ante un diminutivo con el valor apreciativo de ‘paleta pequeña’ o ante alguno de los significados lexicalizados que se recogen en el diccionario y que podrían encajar también con el que se usa en los inventarios: ‘en algunas reses, como el cordero, cuarto delantero’ o ‘candelero bajo con asa’ (*DLE* s.v. PALETILLA, acep. 2, 4). Este es un claro ejemplo de cómo la lectura del

⁷¹ Puede verse un ejemplo en el inventario «Bienes que Pedro Fernández lleva a su segundo matrimonio tras la muerte de Lucía Fernández, su primera mujer» (*ODE* – Madrid, 1747). Enlace al texto: <http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/MD1747I0110.xml>

⁷² Un ejemplo es el documento «Bienes que Bernardo Fernández recibe como dote por su casamiento con su futura esposa Mónica Jenara García» (*ODE* – Madrid, 1852). Enlace al texto: <http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/MD1852I0124.xml>

contexto ayuda a la clasificación de los términos con diminutivo en el corpus, pues solo una lectura más amplia de los objetos circundantes ayudará a una interpretación adecuada del significado. Las *paletillas* del ejemplo se nombran entre chocolateras, cazos, cucharas y sartenes, todos ellos utensilios de cocina. La relación semántica que los une es lo que permite la desambiguación léxica de *paletillas* como un diminutivo transparente de *paleta* en el sentido de ‘instrumento de hierro utilizado en la cocina compuesto por un mango largo y base agujereada’ y no como un término lexicalizado.

El sentido lexicalizado se aprecia en el caso de *estufilla*, en el que es el contexto el que favorece la diferenciación entre acepciones del término ya lexicalizado⁷³. El ejemplo de esta voz por sí solo en “vna *estufilla* ordinaria bien tratada en diez y ocho” (*ODE* – Madrid, 1724) no sirve de ayuda para aclarar con cuál de los dos sentidos que da *Autoridades* se identifica el uso de la voz en el ejemplo señalado:

1. s. f. Manguito pequeño hecho de pieles finas: como Martas, para traher abrigadas las manos en tiempo de Invierno (*Autoridades* s.v. ESTUFILLA, acep. 1).
2. Se llama tambien un género de braserillo, a manera de un caxoncito con su tapa agujereada, hecha de hoja de lata, de que usan las mugeres para calentarse, el qual se pone debaxo de las faldas (*Autoridades* s.v. ESTUFILLA, acep. 2).

Es necesario acudir al inventario para comprobar qué otras cosas se mencionan junto a la *estufilla* para obtener alguna pista. La voz se incluye precisamente en uno de los inventarios clasificado por tipologías formando parte de los ‘vestidos’, entre los que también se incluyen, además de vestidos, casacas o almillas, otras prendas y accesorios de mujer como medias, abanicos y guantes:

- (11) [*En el margen: Bestidos*] Vn bestido de teleton, casaca y basquiña, color de flor de bioleta, nuevo, en quatrocientos y veinte r[eale]s.
 Vn guardapies de raso escarolado casi nuevo en duzos y quarenta r[eale]s.
 Vna basquiña de lamparilla monperada, casi nueva, en nobenta r[eale]s.
 [...]
 Mas quatro pares de medias de seda de color de fuego, vsadas, en sesenta r[eale]s.
 Vn abanico de talco nuevo en treinta r[eale]s.
 Otro abanico de lamina con barillas de concha en treinta y siete r[eale]s.
 Otro abanico ordinario biexo en seis.
 Vna *estufilla* ordinaria bien tratada en diez y ocho r[eale]s.
 Vn par de guantes nuevos, bordados de plata, en siete r[eale]s (*ODE* – Madrid, 1724).

⁷³ Debido a la antigüedad de la documentación, no se contempla en este caso la posibilidad de que *estufilla* actúe como pleno diminutivo con el sentido de ‘estufa pequeña’. El concepto actual de *estufa* como ‘aparato con el que calentar un recinto por medio de electricidad o combustión’ (*DLE* s.v. ESTUFA) es de invención más reciente. Así lo prueba la ausencia de la palabra en inventarios de *ODE* anteriores al siglo XIX (y, como es obvio, también en *CorLexIn* en el XVII). Las primeras apariciones del término se localizan precisamente en dos de los inventarios madrileños de *ODE* a finales del XIX: “seis *estufas* nuevas de cobre” (Madrid, Aranjuez, 1870), “vna *estufa* chubesqui para carbon” (Madrid, Alcalá de Henares, 1890).

Por consiguiente, gracias al contexto podemos asegurar con total certeza que la *estufilla* del ejemplo aparece con el sentido de ‘manguito’ y no con el de ‘brasero’, tratándose de la primera acepción que señala *Autoridades* para este término. Este caso, junto con el arriba mencionado para *paletillas*, es una prueba clara de que a veces es también necesario atender a los objetos circundantes para la correcta interpretación de la palabra.

Los dos procedimientos de desambiguación mencionados hasta ahora –la mayor precisión con la que se describen algunos objetos y los facilitados por el contexto mediante la agrupación de las pertenencias–, se combinan en el caso particular de *tenacillas*. Este ejemplo es un caso peculiar, ya que no todas sus ocurrencias se han vinculado con un mismo valor semántico. Además de la distinción entre valor apreciativo y valor léxico del diminutivo, *tenacillas* es otra de esas muchas palabras ambiguas en el corpus que han desarrollado diferentes significados. De las seis apariciones de la forma en los textos, en una de ellas el uso del diminutivo se ha considerado como pleno diminutivo (12), mientras que en las cinco restantes se asocia con significados ya lexicalizados (13):

(12) Unas *tenacillas* en trescientas mil[esima]s (*ODE* – Madrid, Aranjuez, 1870).

- (13) a. 1 *tenacillas* de plata p[ar]a tomar azucar para el café (*ODE* – Madrid, 1843).
b. Vnas *tenacillas* para azucar, al parecer de plata [...] Vnas *tenacillas* de oro para fumar (*ODE* – Madrid, Alcalá, 1890).
c. Cuatro pares de *tenacillas* para rizar ropa (*ODE* – Madrid, 1877).
d. Vnas *tenacillas* de plata y un limpia oídos de plata (*ODE* – Madrid, Alcalá, 1890).

Veamos más detenidamente los aspectos que han llevado a tomar tal decisión. Los ejemplos de arriba en los que se señala la función –“para azúcar”, “para fumar”, “para rizar ropa”–, se corresponden con esos casos en los que el escribano, consciente de la existencia de diferentes tipos de tenazas que pueden variar tanto en las dimensiones como en el uso, debe evitar la ambigüedad y acotar el sentido de la palabra precisando su función. La palabra, usada siempre en plural, consta de cinco acepciones diferentes recogidas en el *DLE*, de las cuales se aprecian en los ejemplos los sentidos lexicalizados cuando funciona como ‘tenaza para coger terrones de azúcar’ y como ‘instrumento con el que coger el cigarrillo mientras se fuma’. Las “*tenacillas* para rizar ropa” del ejemplo aparecen inventariadas entre los utensilios de cocina –especificado en el margen del folio– y se empleaban para hacer dobleces o formar figuras en las telas (*Autoridades* s.v. RIZAR). Relacionado también con esta tarea de rizar la ropa se nombran en el mismo inventario “cinco planchas de yerro” y “un pie de yerro para colocar la plancha”. Entre los significados lexicalizados de *tenacillas* no se incluye en el diccionario ninguno que lo vincule con la función descrita, pero sí con el “instrumento, a manera de tenaza pequeña, que sirve para rizar el pelo” (*DLE* s.v. TENACILLAS, acep. 2), por lo que el diminutivo en este caso se ha considerado también lexicalizado como en el resto de los casos contenidos en (13).

Por otra parte, en (12) y (13) es evidente que los ejemplos así aislados sin mayor contexto no permiten identificar la realidad a la que se alude. La ausencia de especialización por parte del escribano apunta inicialmente al significado general del término *tenaza* que se emplearía en esta ocasión con la forma sufijada (*tenacillas*). Es necesario ampliar el contexto y considerar el resto de los objetos con los que se nombra para identificar el sentido. De esta manera puede comprobarse que el valor de diminutivo transparente se cumple para (12), ya que las *tenacillas* de este ejemplo se nombran entre las herramientas –especificado así en el encabezado de la lista–, donde se incluyen punzones, cortafríos, alicates, serruchos y también tenazas (sin el sufijo).

Por el contrario, en (13) no es tan fácil identificar el significado. Si bien en algunos casos el empleo de la palabra con diminutivo constituye un claro ejemplo de lexicalización, en otros la falta de descripción dificulta la identificación del objeto al que se alude. En los casos mencionados arriba, el contexto contribuye fácilmente al tratamiento de los términos en el corpus como lexicalizaciones o como verdaderos diminutivos, sin embargo, otros muchos casos inducen a duda, bien por la falta de detalle del escribano al inventariar los objetos, o bien porque el investigador actual desconoce por completo la realidad de los objetos que se mencionan. Para este último supuesto, a veces ni las fuentes lexicográficas ni la escasez de estudios o documentación sobre el término sirven de gran ayuda.

El inventario en el que se localiza el ejemplo de (13) no incluye una división de los bienes por categorías que facilite la desambiguación. Las *tenacillas* se nombran entre objetos muy variados y de valor considerable como un alfiler de oro, una sortija con un diente o un abanico de plata con nácar. Entre las pertenencias se encuentran objetos predominantemente masculinos, los cuales fueron propiedad de Policarpo Cortecedo, como corbatas, alfileres, gemelos o “anillos de caballero” y también una “fosforera de plata” mencionada poco antes a las *tenacillas* del ejemplo⁷⁴. La falta de otros detalles no nos permite asegurar el sentido exacto, el cual podría interpretarse con alguno de los sentidos lexicalizados ya mencionados. En este caso, aunque el contexto sirve para descartar algunos de los significados del término, así como el valor transparente del diminutivo, no es suficiente para asegurar su sentido real. A pesar de lo ambiguo del ejemplo, nos inclinamos por el significado de ‘instrumento para fumar’ dado que se nombra poco antes junto a una fosforera y en este mismo inventario ya se citan otras tenacillas con el mismo fin. De esta manera, el ejemplo se relaciona con esta acepción en su correspondiente entrada en el glosario de lexicalizaciones (§ 6.5.3.4. *Glosario s.v. TENACILLAS*).

La imprecisión léxica lleva a no poder concretar el sentido real del diminutivo como ocurre en este otro ejemplo: “un cantaro y una *cantarilla*” (*ODE* – Madrid, Navas del Rey, 1832), donde es prácticamente imposible conocer si el término con sufijo diminutivo *cantarilla* hace referencia a una cántara pequeña o en realidad se trata del significado lexicalizado que se recoge en el *DLE* (“vasija de barro, sin baño, del tamaño

⁷⁴ «Inventario de los bienes que quedaron por la fin y muerte de Policarpo Cortecedo y Ayuso» (*ODE* – Madrid, Alcalá de Henares, 1890): <http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/MD1890I0143.xml>

y forma de una jarra ordinaria, y de boca redonda”, s.v. CANTARILLA). La proximidad semántica entre ambos términos –el derivado con diminutivo y la forma lexicalizada– impide tomar una decisión fiable. La solución quedará sin descifrar, pues únicamente podrían resolverlo los presentes en el momento mismo de la tasación, quienes conocían de primera mano los objetos incluidos en el inventario. En este caso, solo el reconocimiento visual serviría para precisar el objeto al que se refiere el texto. A pesar de ello, es necesario establecer un criterio, por lo que en los casos como este en los que no es posible determinar con total seguridad si la palabra actúa como diminutivo o como lexicalización se han considerado como lexicalizaciones, siempre y cuando la voz se contempla en el diccionario con un uso específico distinto al del tamaño. Casos similares a los mencionados se suceden con cierta frecuencia en el corpus y son un ejemplo de la dificultad existente a la hora de diferenciar los diminutivos auténticos de los lexicalizados. En el siguiente apartado se analizan algunos de ellos.

En cualquier caso, el carácter ambiguo de una voz lo es únicamente para el lector actual que se acerca a las realidades reflejadas en los inventarios, pues no existía tal confusión para los escribanos o notarios de la época que, cuando nombraban uno u otro objeto, lo hacían sabiendo que sería fácilmente identificable por sus coetáneos. Resulta natural pensar que serían superfluas las precisiones del escribano, –tan necesarias hoy en día al investigador actual para su correcta interpretación–, si estos documentos estaban dirigidos a quienes compartían con él la familiaridad de los objetos que los rodeaban, pertenecientes a su vida cotidiana y por ello de sobra conocidos e identificables.

Es evidente que no todas las ambigüedades léxicas presentes en los textos serán de fácil interpretación, sin embargo, en este apartado se ha querido demostrar cómo para la gran mayoría de los casos el contexto que nos ofrecen los propios inventarios mediante alguno de los procedimientos ya explicados ha bastado como elemento desambiguador. Para el análisis llevado a cabo en esta tesis ha sido determinante el poder identificar la ambivalencia que se produce entre el valor transparente del diminutivo y el opaco de la forma lexicalizada, aunque fruto de este análisis también se ha procedido a identificar los distintos significados que de una misma voz se documentan en el corpus y que se recogen en el glosario (§ 6.5.3.4. *Glosario*).

6.3.2. Casos dudosos y problemas para la cuantificación

La selección de los diminutivos transparentes implica consecuencias determinantes de cara al análisis estadístico que se realiza en esta tesis sobre la frecuencia y distribución de los sufijos *-ico*, *-ito* e *-illo*. De la clasificación minuciosa de los derivados depende el recuento total de las formas que se consideran parte del análisis y que influirá directamente en el resultado de los datos. Sin embargo, esta clasificación conlleva establecer unos criterios metodológicos que no son de fácil aplicación en determinados casos. Algunos de ellos ya se han mencionado en el apartado anterior, donde explicábamos cómo la falta de contexto dificulta poder averiguar el sentido exacto de la palabra con diminutivo, como ocurre con *tenacillas* o *cantarilla*. Otros casos merecen

una consideración aparte dada su mayor complejidad. Por esta razón se ha estimado oportuno dedicar lo que sigue a la mención de algunos de ellos, ya que para su análisis se han tenido en cuenta diferentes factores que van más allá de la inclusión o no de los términos en el diccionario y de la aplicación de los criterios que ya se han señalado (§ 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*).

Por otro lado, conviene mencionar que para algunos casos la consideración de diminutivo transparente no se ha podido determinar con fiabilidad y, a fin de no alterar los datos analizados estadísticamente, se han eliminado del recuento. Como consecuencia, estos casos se han incluido siempre como lexicalizaciones ante la falta de evidencias que permitan corroborar su valor como pleno diminutivo. Al respecto se puede mencionar el caso de *cantarilla*, señalado en el apartado anterior o el de *bufetillo-bufetico-bufetito* (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*).

En las palabras que en este apartado se detallan se cumplen una serie de características que han sido determinantes para su consideración como diminutivos transparentes o lexicalizados, pero que podrían analizarse de manera diferente bajo la mirada y los criterios de otro investigador. Se detallan, por orden alfabético, los casos de *azadilla*, *bacinilla*, *banquillo*, *braserilla*, *cencerrica*, *cortinilla*, *cucharilla*, *mesilla*, *platillo* y *puntilla*.

La cuantificación de los diminutivos de uno y otro grupo se recoge en § 6.4.1. *Recuento de formas diminutivas en el corpus* y § 6.5.1. *Recuento de diminutivos lexicalizados en el corpus*. Por otro lado, el listado completo de los diminutivos transparentes en el corpus puede consultarse en § Anexo IV. *Relación de diminutivos transparentes incluidos en el análisis*, mientras que el total de formas lexicalizadas se incluye en el glosario (§ 6.5.3.4. *Glosario*).

Azadilla

La forma *azadilla* se ha considerado pleno diminutivo pese a que esta forma se encuentra registrada en el *DLE* como sinónimo del arabismo *almocafre*: “instrumento que sirve para escardar y limpiar la tierra de malas hierbas, y para trasplantar plantas pequeñas” (*DLE* s.v. ALMOCAFRE). La escasez de ejemplos en *ODE* ha hecho plantearse algunas dudas en torno a la consideración del valor transparente u opaco del diminutivo en estos casos:

- (14) a. Una açhica pequeña p[ar]a cortar leña. Una *azadica* pequeña. Otra mas pequeña. Una hoz para segar trigo (*ODE* – Almería, Cuevas de Almanzora, 1714).
b. Dos azadones llanos, viejos, en veinte r[eale]s. Vn destrál y otros dos picos viejos en veinte r[eale]s. Vna *azadilla* para destallar olivos en ocho r[eale]s (*ODE* – Madrid, Ambite, 1845).

En el primero de ellos, la poca especificidad en la descripción impide conocer con certeza el tipo de azada al que se hace referencia. No obstante, parece poco probable que si el término se empleara en este caso con un sentido plenamente lexicalizado se utilizara la forma *azadica* y no *azadilla* como la ya fijada en la lengua. En (14) la finalidad “para destallar olivos” con la que se usa es muy distinta del valor específico que adquiere *azadilla* como elemento lexicalizado según el *DLE*, por lo que todo indica que podría tratarse de un diminutivo con significado transparente.

El derivado *azadilla* aparece más en *ODE* en el ejemplo “dos *azadillas* garvanzeras” (*ODE* – Jaén, Villacarrillo, 1796), en el que también se señala su función, en este caso, como instrumento empleado para la cosecha del garbanzo. Este uso se aleja de nuevo de la definición que nos ofrece el diccionario, pero la clave para confirmar el valor transparente del diminutivo la obtenemos en otro ejemplo jienense: “otras dos *azadas* pequeñas para escauar garbanzos” (*ODE* – Jaén, Villacarrillo, 1724), en el que el mismo objeto aparece esta vez como “azada pequeña” y no *azadilla*, lo que indica que el valor del diminutivo en los casos de *azadilla* sería básicamente nocional.

En definitiva, en las palabras *azadica* y *azadilla* de nuestro corpus se ha considerado que el diminutivo representa un valor apreciativo motivado principalmente por estas dos causas: la variación del sufijo en el primero de los casos y la función con la que se emplea la *azadilla* en el segundo ejemplo, distinta a la descrita en el diccionario. Los ejemplos del término en *ODE* han servido para la desambiguación entre diminutivo transparente y opaco de los casos localizados en los inventarios de Almería y Madrid.

Bacinilla

En el *DLE* la acepción principal de *bacinilla* (“vasija pequeña para diversos usos”) no presenta diferencias de significado con respecto a su primitivo *bacina* o *bacín*. Ambos términos se definen de igual manera en el diccionario a pesar de que el derivado con diminutivo estaría considerado una forma lexicalizada. A este significado se añade también el de ‘orinal’, de nuevo vinculado tanto a *bacina* como a *bacinilla*.

El derivado ha estado sujeto a la variación del diminutivo desde las primeras apariciones lexicográficas: el diccionario académico señala la alternancia entre *bacinica-bacinilla* desde *Autoridades*. La alternancia en el uso del sufijo que se refleja en los diccionarios se constata también en los ejemplos del término con diminutivo que se recogen en *ODE* (9 casos de *bacinica*, frente a 5 de *bacinilla*):

- (15) a. Dos *bazinicas* de metal uixas y rrotas (*ODE* – Cáceres, 1639).
- b. Una *bacinica* de aljófar (*ODE* – Badajoz, 1670).
- c. Tres peroles grandes y uno mas mediano. Una *basinilla* de asofar (*ODE* – Sevilla, 1703).
- d. Dos saluillas, dos *uasinillas* y una palangana, todo de peltre (*ODE* – Cádiz, 1738).
- e. Una *bazinica* de cobre (*ODE* – Madrid, 1793).

La falta de mayor contexto en los ejemplos no permite identificar con cuál de los dos significados descritos arriba se utiliza el término en cada caso. Sin embargo, estos casos, y más particularmente el documentado en Madrid en nuestros inventarios (15), se consideran diminutivos transparentes dada la similitud semántica que asocia la forma original *bacina* con sus derivados *bacinilla-bacinica*, entre los que se establecería una relación de transparencia. Las definiciones lexicográficas para cada uno de los términos no señalan diferencias en el uso o la función de los objetos, por lo que ambos términos podrían ser equivalentes. La diferencia de significado entre una *bacina* y una *bacinilla* quedaría sobre todo vinculada al cambio de tamaño, siendo este el principal valor del diminutivo, a pesar de estar considerada como voz lexicalizada en las fuentes lexicográficas.

Banquillo

En *ODE*, la palabra *banquillo* se asocia con dos significados principales: 1) ‘banco pequeño’ y 2) ‘soporte que sirve como pie de un mueble’.

El primero de estos significados se corresponde con el diminutivo de *banco* en su acepción principal como ‘asiento, con respaldo o sin él’. En este sentido, se entiende que en los siguientes ejemplos de *banquillo* el sufijo cumple con el valor transparente del diminutivo:

- (16) a. Dos messas pequeñas de pino en diez y ocho r[eale]s. Vn *banquillo* de pino en quatro reales. Dos tauretes [sic] de pino de rrespaldo en treze rr[eale]s (*ODE* – Madrid, Colmenar de Oreja, 1722).
- b. Una banqueta y dos *banquillos* (*ODE* – Madrid, Navas del Rey, 1832).

Otras fuentes consideran *banquillo* como un término ya lexicalizado con el sentido de ‘asiento alargado sin respaldo’ (Perdiguero, 2015, p. 1568). En el diccionario de Terreros y Pando (1786) *banquillo* se iguala a *escabel*⁷⁵ y se define como “un banco pequeño sin respaldo alguno”. El diccionario académico, sin embargo, y creemos que, de manera acertada, nunca ha recogido esta acepción de *banquillo*, pues consideramos que se trata de un diminutivo de *banco*. En nuestros textos, el concepto de ‘escabel’ aparece equiparado al de *banco*, sin diminutivo, en este ejemplo: “dos bancos o escaueletes, madera de pino” (*ODE* – Almería, Vera, 1791), por lo que *banco* y *banquillo* serían dos muebles de características muy parecidas que se diferenciaban principalmente en sus dimensiones.

⁷⁵ La sinonimia entre *banco/banquillo* y *escabel* o *escabelo* se constata en estos otros ejemplos extraídos del *CDH* (1535-1557, Gonzalo Fernández de Oviedo, *Historia general y natural de las Indias*):

- Iban los españoles a ver a Montezuma, e mandábales dar duhos, que son unos *banquillos* o escabelos, en que se sentasen (muy lindamente labrados e de gentil madera) e decíanles qué querían, que lo pidiesen e dárselo hían.
- E al cacique e hombres principales poníanles unos *banquillos* de palo, muy bien labrados, de lindas maderas, e con muchas labores de relieve e concavadas, entalladas y esculpidas en ellos, a los cuales *bancos* o escabelo llaman duho.

Son más numerosos en los textos de Madrid los casos que se identifican con la segunda acepción del término. Los *banquillos* entendidos como ‘soporte’ podían servir de apoyo para diferentes objetos. Con este significado se entienden los numerosos casos de *banquillos* que aparecen vinculados a la cama y que servían como soporte de esta:

- (17) a. Vna cama de tablas con sus *banquillos* (ODE – Madrid, Brunete, 1751).
- b. Una cama de cinco tablas y *banquillos* de pino dada de verde [...] Una cama de quatro tablas dada de verde con los *banquillos* torneados (ODE – Madrid, Pinto, 1784).
- c. Una cama de cinco tablas con sus *banquillos* de pino, buena (ODE – Madrid, Brunete, 1800).
- d. Un tablado de cama con sus *banquillos* y respaldo (ODE – Madrid, 1801).
- e. Un tablado de cama y dos *banquillos* (ODE – Madrid, Algete, 1834).
- f. Un capero de pino y un tablado de cama con sus *banquillos* (ODE – Madrid, Aranjuez, 1870).

El tablado, de diferentes dimensiones según el número de tablas que lo componen, se apoyaba sobre los *banquillos*, que servían como soporte o patas de aquel. El conjunto completo de la cama estaba formado entonces por el tablado y sus *banquillos*, de modo que se inventariaban y se tasaban juntos. Esta relación queda manifiesta por medio del posesivo “sus” en la mayoría de los ejemplos de arriba, los cuales dan cuenta de que la armadura de una cama de la época estaba constituida por el tablado y los *banquillos* de forma inseparable. Estos podían tener algún tipo de decoración, como se evidencia en el ejemplo de (17), en el que se especifica que son torneados. Este ejemplo en particular ratifica el significado de *banquillo* como el de soporte de la cama, pues esta característica en el corpus suele ir asociada de manera general con las patas o pies de diferentes muebles como mesas, sillas o camas: “vna cama de quatro tablas anchas, pies torneados, dada de varniz” (ODE – Madrid, 1745), “vna mesita de nogal con sus pies torneados” (ODE – Madrid, 1715).

Relacionados también con el significado de ‘soporte’ se identifican más ejemplos en los que los *banquillos* sirven como apoyo de otros muebles, no solo de la cama:

- (18) Un baul forrado de piel con sus *banquillos* tasado en tres pesetas [...] Un baul pequeño forrado de piel con sus *banquillos*, en una peseta (ODE – Madrid, 1877).
- (19) a. Un tablero de diez pies largo con tres *banquillos* en un escudo. Otro tablero grande con sus *banquillos* en dos escudos [...] Dos tableros pequeños viejos con sus *banquillos* (ODE – Madrid, Aranjuez, 1870).
- b. Un *banquillo* de labar (ODE – Madrid, San Lorenzo de El Escorial, 1858).

Los baúles de (18) se elevaban sobre el suelo apoyados en los *banquillos*, que se colocaban perpendicularmente a la base como travesaños en los que se asentaban. El mueble al completo estaba formado por el baúl y sus *banquillos*, muy probablemente

dos, uno a cada extremo. De nuevo la asociación semántica entre *banquillo* como ‘pies o patas de un mueble’ se evidencia si se considera este otro ejemplo en el mismo inventario: “un baul grande largo de pino, forrado de piel, con sus pies de madera” (ODE – Madrid, 1877), a partir del cual *pies* y *banquillos* quedarían vinculados por sinonimia, ambos con el mismo significado.

Como se constata en los ejemplos de (19), los *banquillos* podían ser también las patas de tableros que cumplen con las funciones de una mesa que podía usarse para diversas tareas o labores. En este sentido, *banquillo* se asocia muy estrechamente con el primitivo *banco* en su concepto de ‘mesa de trabajo’, del mismo modo que actualmente se utiliza el término *banco de carpintero* como la mesa especialmente destinada a las labores de este oficio⁷⁶. Por extensión, el término *banco-banquillo* se utilizaría también para otras mesas dedicadas a otros trabajos y así se explicaría el caso de “un *banquillo* de lavar” que se recoge en (19), el cual haría alusión a un tipo de mesa utilizada especialmente para este fin. También en nuestros textos, aunque no con el diminutivo, se localiza otro banco de trabajo vinculado esta vez con las tareas de prensado: “un *banco* y otro con usillos⁷⁷” (ODE – Almería, Escúllar, 1834). Como justificación de esta hipótesis se encuentran más ejemplos en otras provincias de ODE y CorLexIn en los que el término con o sin diminutivo se utiliza para designar el banco de trabajo de diferentes oficios:

- (20) a. Vn *banco* de platero con toda la herramienta pertenecientes al dho off[ici]o (ODE – Sevilla, 1572).
- b. Vn *banco* de aparg[ater]o (ODE – Granada, Baza, 1590).
- c. Un *banco* de coser (ODE – Granada, Baza, 1591).
- d. Un *banco* grande de carpintero, con sus tornos y puerta (CorLexIn – Huelva, 1617).
- e. Una mesa de gonzes con sus *banquillos* sin cadena (CorLexIn – Murcia, 1637).
- f. Dos *bancos* grandes de ofiçio, con un hierro grande con que se aprieta la madera (CorLexIn – Badajoz, Alburquerque, 1645).
- g. Un *banco* de matar marranos (CorLexIn – Granada, Montefrío, 1661).
- h. Un *banco* de herrador con una bigornia (CorLexIn – Badajoz, 1683).
- i. Un *banco* de cordonero de coser suelas (ODE – Granada, Monachil, 1723).
- j. Un *banquillo* con su bigornia en ochenta y cinco reales (ODE – Granada, 1796).

La acepción de *banquillo* que venimos comentando no encuentra apoyo lexicográfico, sin embargo, los ejemplos que avalan el significado de ‘soporte’ son bastante numerosos y claros. De aquellos que aparecen vinculados a la cama se han señalado aquí únicamente los que aparecen en Almería y Madrid, pero son también abundantes

⁷⁶ Este significado se recoge en el DLE como una de las acepciones de *banco*: “Madero grueso escuadrado que se coloca horizontalmente sobre cuatro pies y sirve de mesa para labores de carpinteros y otros artesanos” (s.v. BANCO, acep. 2).

⁷⁷ El *husillo* (“tornillo de hierro o madera que se usa para el movimiento de las prensas y otras máquinas” DLE s.v.) aparece en otro inventario almeriense relacionado con el prensado de la uva: “un usillo de estruxar uvas” (§ 6.5.3. Glosario de lexicalizaciones s.v. HUSILLO).

los que se localizan en el resto de *ODE* y *CorLexIn*, y lo mismo se puede decir de los *banquillos* que funcionan como patas de tableros, baúles, arcas, artesas, etc. En todos los casos *banquillo* significa una cosa diferente a ‘banco pequeño’ con una forma y una función muy distinta. Este cambio semántico evidencia ya en esta época una lexicalización avanzada del sufijo que, por otra parte, aparece en todos los casos bajo la forma en *-illo* (recuérdese la importancia que otorgamos a la variación en el sufijo *-ausente* en este caso— para determinar la lexicalización como proceso inacabado)⁷⁸. Sin embargo, todavía se localizan algunos ejemplos en los que *banquillo* alterna con *banco* en contextos idénticos. En *ODE*, el primer registro del término con diminutivo data de 1552: “vna cama de madera de tres *vanq[ui]llos* viejos y siete tablas” (Granada, Baza). Los “bancos de cama” son más numerosos en el siglo XVI, pero aparecen también hasta finales del siglo XVIII:

- (21) a. Vnos *bancos* de cama e cañiso (*ODE* – Sevilla, 1500).
- b. Çinco tablas de cama y dos *bancos* (*ODE* – Granada, Baza, 1552).
- c. Otra armadura de cama e *bancos* y tablas (*ODE* – Badajoz, 1624).
- d. Una manta rayda, *bancos* y tablas de cama (*ODE* – Badajoz, 1628).
- e. Una cama de tablas y *bancos* nueva (*ODE* – Huelva, 1704).
- f. Una cama de tablas y *bancos* (*ODE* – Cádiz, 1739).
- g. Una cama de *bancos* y tablas con el tertero maqueado, encarnada (*ODE* – Cádiz, Puerto Real, 1776).
- h. Un tablado de cama con cinco tablas y dos *bancos* (*ODE* – Murcia, Lorca, 1792).

Los ejemplos de arriba de *banco* vendrían a constatar la convivencia en el XVIII con *banquillo* (véanse los ejemplos de (17) en Madrid), por lo que el término primitivo con este significado aún no ha desaparecido a pesar del avanzado estado de lexicalización de su equivalente con diminutivo. Del mismo modo alternan derivado y primitivo cuando los *bancos* o *banquillos* funcionan como patas de otros muebles. Es por esta razón por la que *banquillo* se ha considerado a efectos del análisis de esta tesis como un diminutivo transparente, ya que la variación entre *banco-banquillo* no comporta diferencia alguna de significado, aunque el empleo del diminutivo irá aumentando en detrimento de *banco*. Asimismo, *banquillo* podría interpretarse como un diminutivo transparente de *banco* en su acepción como soporte.

En esta misma dirección apuntan también los ejemplos de *banquillo* con el significado de ‘asiento’. En nuestros inventarios se localiza la alternancia *banco-banquillo* en contextos prácticamente idénticos: “un *banco* y dos banquetas de pino” (*ODE* – Madrid, Brunete, 1800), “una banqueta y dos *banquillos* (*ODE* – Madrid, Navas del Rey, 1832). De modo que la distinción podría limitarse a la diferencia de tamaño entre ambos objetos. Por otra parte, en algunas ocasiones se añade la precisión “con respaldo”: “un *banco* de pino con respaldo” (*ODE* – Madrid, Estremera, 1836), “un *banco* largo de respaldo” (*ODE* – Madrid, San Lorenzo de El Escorial, 1858). De

⁷⁸ Véase al respecto § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*.

modo que, si la diferencia entre un *banco* y *banquillo* fuera la ausencia de respaldo en este último (como indican algunas fuentes), serían extrañas las referencias al respaldo en los casos de *banco*. Siendo tan ocasionales las precisiones de este tipo, parece razonable pensar que el banco común podía caracterizarse también por no tener respaldo, pero cuando sí constaba de él era imprescindible mencionarlo en el inventario para su correcta identificación. En vista de los numerosos ejemplos que ofrecen los corpus (en los que muy ocasionalmente *banquillo* se emplea con otro sufijo diminutivo), es posible que el término estuviera en un proceso a medio camino de la lexicalización en el que en ocasiones alternaba con el de *banco*. En cualquier caso, lo que parece claro es que el *banco* y su derivado con diminutivo apenas se diferenciaban en la forma (ambos podían carecer de respaldo) y la función (los dos se utilizaban como asiento), mientras que sí parecían distinguirse notablemente en el tamaño, siendo el *banquillo* más bajo y pequeño que el banco convencional. Esta estrecha relación semántica entre ambos términos es la que favorece la alternancia entre *banco* y *banquillo* en ejemplos como los que se han ido señalando. No parece claro que un *banquillo* sea muy diferente de un ‘banco pequeño’ del mismo modo que *cucharilla* no difiere de una ‘cuchara pequeña’, razón por la que el significado del diminutivo en ambos términos se ha considerado transparente.

En definitiva, las ocurrencias de *banquillo* en cualquiera de las acepciones mencionadas se han considerado diminutivos transparentes en consonancia con el criterio del valor semántico similar al de la base sin el sufijo. En la primera de ellas como ‘asiento’, esta decisión se justifica conforme al criterio semántico, según el cual consideramos que la diferencia entre ambos referentes es principalmente dimensional. Para el significado de *banquillo* como ‘soporte, si bien se aprecia un proceso de lexicalización avanzado, este no se considera todavía acabado, a la vista de la convivencia con *banco* en la misma época y sin cambio de significado. Por otro lado, ya que *banco* también se atestigua como ‘soporte’, los casos de *banquillo* no serían más que diminutivos transparentes de *banco* en una acepción que no ha sido documentada lexicográficamente, pero sí abundantemente en los corpus.

Braserilla

La palabra *braserilla* se documenta en nuestros inventarios en una sola ocasión y sin mayor contexto que nos señale la forma, el material o la función con la que se utilizaba. Aparece inventariada entre los objetos pertenecientes al menaje de cocina:

- (22) Dos candeleros y una palmatoria. Una [sic] almirez con su mano. Una *braserilla*. Dos candeleros labrados. Un guisqué (ODE – Madrid, 1840).

Este uso aislado de la voz en femenino contrasta con el resto de las ocurrencias en el corpus del término derivado en masculino. Los casos de *braserillo* y *braserito* se han asociado fácilmente con el significado transparente de ‘brasero pequeño’, por lo que se

han considerado auténticos diminutivos de *brasero*, documentado este término abundantemente en los inventarios de *ODE*.

De la forma en femenino, en cambio, el ejemplo señalado es el único testimonio localizable. No se han encontrado otras ocurrencias en los corpus complementarios consultados y tampoco arroja algún dato la búsqueda de *braserilla* o la de su primitivo *braser* en las fuentes lexicográficas (*NTLLE*). Esta última se ha localizado únicamente en *CorLexIn* en un inventario de Cuenca de 1690: “La *braser* del número ziento y ochenta y dos se remató en Gerónimo Rodríguez en tres reales y medio”.

Ante la ausencia de mayor contexto, no parece que *braserilla* haya adquirido un sentido diferente al de su equivalente en masculino como ‘brasero pequeño’, por lo que estaríamos más bien ante un uso aislado sin alcance reseñable en el que no se constata un cambio de significado respecto de la forma estandarizada en masculino. *Braserilla*, por tanto, es un hápax que localizamos una sola vez en nuestra documentación. La falta de más datos, además de los pocos mencionados, nos ha hecho considerarla como un diminutivo transparente.

Cencerrilla

Este caso ha despertado dudas en cuanto a la consideración de diminutivo transparente, ya que el uso del masculino *cencerro* sí está ampliamente documentado y se recoge en el diccionario con el significado de ‘campana pequeña y cilíndrica, tosca por lo común, hecha con chapa de hierro o de cobre, que se usa para el ganado y suele atarse al pescuezo de las reses’ (*DLE* s.v. CENCERRO).

El uso del femenino *cencerra* remite al término en masculino desde *Autoridades* y así se mantiene en la edición actual del *DLE*, por lo que la variación de género no conlleva en este caso un cambio de significado: ambas variantes actúan como sinónimos en los diccionarios académicos. El mismo criterio adopta el *DEA*, donde además el término en femenino se marca como *raro*, en señal de su uso anticuado.

En *ODE* predomina el uso de la variante en masculino (12 casos) frente a las solo cuatro ocurrencias del femenino en un documento sevillano y en los inventarios de Almería y Madrid, considerando también las formas con diminutivos:

- (23) a. Una *zenzerrica*. Dos cadenas para mulas. Un collar de hierro para una mula (*ODE* – Almería, Cuevas de Almanzora, 1714).
b. Vna *senserra* arriera vale 20 (*ODE* – Almería, Fiñana, 1832).
c. Un collar de *senserrillas* (*ODE* – Sevilla, Osuna, 1752).
- (24) Veinte y dos *zençerros* medianos de yerro. Otro *senzerro* grande de yerro. Treinta y nueve *senzerrillas* de yerro (*ODE* – Madrid, 1711).

El empleo de los diminutivos en (23) parece apuntar a un sentido transparente en el que el sufijo tendría un valor real de disminución. Este planteamiento se vería justificado por la alternancia en el empleo del diminutivo según la provincia y por el uso de la

forma simple en femenino *cencerra* (23). Más dudoso, en cambio, parece el ejemplo de *cencerrillas* en (24), dado que esta forma se combina con el uso del masculino *cencerro*, lo cual podría indicar una variación del significado entre ambas variantes. El femenino plural está considerado en el *DLE* una voz dialectal propia de Álava con el significado de “cencerros para las caballerías”. No obstante, en los casos registrados en *ODE* nos inclinamos a pensar por diminutivos transparentes en los que el uso del femenino responde a un rasgo dialectal con significado equivalente al del término más habitual en masculino. Dada la ausencia de más casos en *ODE* y en *CorLexIn*, no se puede determinar con precisión la extensión geográfica de la voz en femenino.

Cortinilla

Los casos que se señalan en (25) se han valorado como plenos diminutivos por contraposición a aquellos en los que claramente se distingue el significado lexicalizado de ‘cortina pequeña que se coloca en la parte interior de los cristales’ (*DLE* s.v. CORTINILLA) y que se refleja en los ejemplos de (26):

- (25) a. Dos *cortinillas* de tafettan con una zenefa uiejas (*ODE* – Madrid, 1701).
 b. Una *cortinita* y otro paño viejo en vn real de vellon (*ODE* – Madrid, 1728).
 c. Una cortina con su fleco en ocho. Siete *cortinillas* de muselina andadas (*ODE* – Madrid, Algete, 1855).
- (26) a. Una cortina de percal guarnecida de cinta usada. Seis *cortinillas* de bidriera de muselina estampada. Dos id[em] mayores. Tres cortinas de lienzo ordinario para ventana (*ODE* – Madrid, 1840).
 b. Dos *cortinillas* de ventana en cinco r[eale]s (*ODE* – Madrid, San Lorenzo de El Escorial, 1858).
 c. Un par de *cortinillas* del balcon en una peseta. Unas *cortinillas* blancas para el balcon tasadas en dos pesetas (*ODE* – Madrid, 1877).

La primera edición del diccionario académico en reflejar el sentido lexicalizado de *cortinilla* fue la de 1884 (*NTLLE*). De igual modo, los ejemplos del corpus, el más antiguo de 1840, restringen la aparición de este sentido en torno a la segunda mitad del siglo XIX. El término en su sentido lexicalizado alterna en nuestra documentación en este otro caso con la forma simple *cortina*: “cuatro *cortinas* de balcon y alcoba con sus correspond[ien]tes varillas” (*ODE* – 1870, Madrid, Aranjuez), lo que probaría que el sentido lexicalizado de *cortinilla* estaba aún por afianzarse. Siendo este el sentido habitual del término hoy en día, equivalente al de *visillo*, no debió serlo en la época de los inventarios, pues de otra manera, no serían necesarias las especificaciones “de ventana”, “de vidriera”, “para el balcón” o similares que aparecen en los textos y con las cuales se diferenciaban estas cortinas de las otras más comunes. A pesar de lo dudoso de los ejemplos, la ausencia de este tipo de precisiones en los casos agrupados en (25) ha favorecido su consideración como plenos diminutivos. A este planteamiento se añade en (25) la presencia de la variante *cortinita* en lugar de *cortinilla* que probaría el valor

apreciativo del diminutivo (la alternancia en el sufijo se entiende como muestra de que el término lexicalizado no se ha fijado del todo, véase al respecto § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*).

Cucharilla

La palabra *cucharilla* es otro de esos casos polémicos en los que resulta difícil determinar el grado de lexicalización del diminutivo y las fuentes lexicográficas difieren en su definición. *Cucharilla* se incluye en el *DLE* como un diminutivo lexicalizado con la principal acepción de “cuchara pequeña” (*DLE* s.v. CUCHARILLA, acep. 1). Sin embargo, a efectos del análisis que aquí se lleva a cabo, todas las formas del lema en el corpus han sido tratadas como diminutivos transparentes, ya que en ellas se considera que el sentido lexicalizado no es distinto del valor semántico que ya aporta el diminutivo por sí mismo. Entre *cuchara* y *cucharilla* el principal cambio de significado está relacionado exclusivamente con la desigualdad de tamaño, sin que la palabra con diminutivo evoque otras diferencias con respecto a la forma o el material del que están hechos estos objetos. Su inclusión en el diccionario con este significado se corresponde con una mala definición, pues no puede haber lexicalización en ‘cuchara pequeña’. En *cucharilla*, el diminutivo no conlleva implicaciones semánticas de otro tipo más que las referidas al tamaño, de modo que esta forma apenas difiere de su base salvo en lo que respecta a las dimensiones del objeto designado. Tomando como referencia este criterio semántico, no se estima oportuno que estos casos en los que el sufijo cumple con un valor transparente deban considerarse como diminutivos lexicalizados.

El *DEA* añade a esta definición un cambio en la función del objeto, señalando que es “utilizado para café u otras infusiones o para dulces” (*DEA* s.v. CUCHARILLA). Sin embargo, tampoco consideramos que este sea un significado lexicalizado, a la vista de que también puede utilizarse una *cuchara* para consumir postres o mover líquidos, como por ejemplo ocurre en este caso: “Otras seis cucharas de platta pequeñas para almibar, todas en sesenta r[eale]s” (*ODE* – Granada, 1781). El empleo de una *cucharilla* o *cuchara* quedaría de nuevo supeditado al tamaño del objeto, más grande o pequeño según la necesidad. *ODE* recoge multitud de casos de cucharas y cucharillas grandes, medianas o pequeñas, por lo que en el uso del diminutivo predomina sobre todo el tamaño.

Junto al criterio semántico, la variación del sufijo en este caso también se ha tenido en cuenta para la consideración de diminutivo transparente. En el corpus se atestigua la alternancia *cucharilla-cucharita* con una distribución muy similar de ambas variantes: de los once casos registrados en los inventarios de Madrid y Almería, seis aparecen con el diminutivo *-illo* y los otros cinco se emplean con *-ito*, lo que demuestra el uso de los diminutivos indistintamente.

Mesilla

El diminutivo *mesilla* constituye hoy en día un término claramente lexicalizado con el significado de ‘mesa de noche’. Esta es la principal acepción del término, que alterna con *mesita* en buena parte del mundo hispánico. Sin embargo, no es el valor semántico que se registra en los inventarios de *ODE*, cuyas ocurrencias bajo cualquiera de las variantes registradas (*mesilla-mesita-mesica*) se ha relacionado con un sentido transparente del diminutivo.

El significado de esta voz como ‘mesa de noche’ es de aparición tardía, tal y como reflejan las fuentes lexicográficas. En el diccionario académico, esta no es la principal acepción hasta la edición de 1992, aunque las formas *mesilla* o *mesa de noche* se registran unos años antes por primera vez en la publicación de 1970 (*NTLLE*). La moderna aparición del concepto ‘mesa de noche’ en los diccionarios parece coincidir con los ejemplos que del compuesto *mesa de noche* se localizan en nuestra documentación. La concreción “de noche” aparece únicamente en cuatro ocasiones en dos inventarios madrileños de finales del siglo XIX y junto al núcleo *mesa* sin el sufijo diminutivo:

- (27) a. Una luna pequeña de espejo sin marco, un velador sin pintar, cuatro cuadros de la historia de Zulema, una tarima redonda con su brasero y una *mesa de noche* en doce escudos y seiscientos mil[esima]s (*ODE* – Madrid, Aranjuez, 1870).
b. Una *mesa de noche*, madera de pino, chapeada de caoba con piedra de mármol, tasada en cinco pesetas. Otra id[em] de pino barnizada, tasada en tres pesetas [...] Una *mesa de noche*, madera de pino barnizada, tasada en cuatro pesetas [...] Una *mesa de noche*, madera de pino barnizada, usada, tasada en cuatro pesetas (*ODE* – Madrid, 1877).

Estos ejemplos constituyen los primeros indicios en el corpus de la aparición de un nuevo concepto, cuyo término en el marco cronológico que abarcan los inventarios aún no había adoptado el sufijo diminutivo. Dada la datación de estos casos, parece razonable considerar el resto de ocurrencias anteriores como verdaderos diminutivos del sustantivo *mesa*:

- (28) a. Vna *mesilla* bieja con su cajon, tasada en quatro reales. Otra *mesilla* bieja en tres reales (*ODE* – Madrid, 1700).
b. Una *mesica*, madera de pino, pequeña (*ODE* – Almería, Cuevas de Almanzora, 1714).
c. Vna *mesita* de nogal con sus pies torneados cubiertos de badana (*ODE* – Madrid, 1715).
d. Otra *mesica* pequeña de serbal, pequeña (*ODE* – Almería, Fiñana, 1745).
e. Una papelera con su *mesita* de nogal (*ODE* – Madrid, 1765).
f. Otra *mesita* mas pequeña (*ODE* – Almería, Vera, 1791).
g. Dos *mesitas* con juego de damas, muy estropeadas (*ODE* – Madrid, Estremera, 1836).
h. Una *mesita* pequeña de tocador (*ODE* – Madrid, 1843).

El contexto de aparición justifica el sentido apreciativo con el que se emplea el sufijo en estos casos. Como ya se ha comentado, a ello se añade la aparición posterior a estos ejemplos del significado lexicalizado del diminutivo. Por otra parte, la variación sufijal avala la poca fijeza léxica del término, el cual estaría sujeto a la procedencia geográfica, siendo más probable el valor eminentemente nocional del diminutivo.

Platillo

El término *platillo* se considera una lexicalización cuando se refiere a una ‘pieza pequeña de forma semejante al plato, cualquiera que sea su uso y la materia de que esté formada’ (*DLE* s.v. PLATILLO). Los usos lexicalizados incluirían aquellos casos en los que la forma del objeto se asemeja a un plato, pero su uso difiere de este en que no se emplea para servir alimentos o comer en él. Independientemente del tamaño, este sería el principal aspecto en el que se diferencia un *platillo* de un *plato* en su sentido lexicalizado, según la definición que de ambos se ofrece en el *DLE*.

Los *platillos* de los ejemplos siguientes acompañan a tazas de café, jícaras, salseras, se incluyen con la romana (balanza) o son parte de elementos decorativos:

- (29) a. Una caxa de plata lisa cuadrada, prolongada con *platillo* en la tapa (*ODE* – Madrid, 1750).
- b. Una jicara de china y *platillo* en quinze r[eale]s. Otra jicara y *platillo* de lo mismo en 15 (*ODE* – Madrid, Pinto, 1784).
- c. Cuatro candeleros de plaque. Un *platillo* y unas espavilad[era]s de lo mismo (*ODE* – Madrid, 1840).
- d. 1 marcerina gemela, *platillo* figura de barco con dos tapas de plata [...] Nueve jicaras grandes con sus *platillos* de china [...] Una romana con *platillo* [...] Dos salseras con sus *platillos* de china (*ODE* – Madrid, 1843).
- e. Un servicio de china para café con adornos dorados compuesto de doce tazas con sus *platillos* (*ODE* – Madrid, 1877).

En estos ejemplos parece más evidente el sentido lexicalizado si se tiene en cuenta que el contexto en el que aparecen sugiere que la función que desempeñan es otra diferente a la habitual y esos *platillos* no se refieren a los platos comunes empleados para comer. Sin embargo, al mismo tiempo podría defenderse el valor transparente del diminutivo como referencia al tamaño si se compara con otros ejemplos de *ODE* donde los mismos contextos recogidos en (29) aparecen también con la forma sin diminutivo:

- (30) a. Quattro *platos* pa jicaras, barro de Genoba (*ODE* – Almería, Cuevas de Almanzora, 1766).
- b. Dose *platos* y tasas blancas (*ODE* – Almería, Escúllar, 1775).
- c. Cuatro *platos* de postre, doce *platos* para café, cinco tazas finas de Talavera (*ODE* – Madrid, Alcalá de Henares, 1890).
- d. Unas espabiladeras con su *plato* de p[la]ta (*ODE* – Cádiz, Puerto Real, 1784).
- e. Un belon de açofar con su *plato* de peltre debaxo (*ODE* – Cáceres, 1664).

Los ejemplos de *platillo* podrían entenderse entonces como lexicalizaciones, pero también como plenos diminutivos, a la vista de los casos de *plato* empleado con los mismos valores. La consideración como diminutivo transparente o lexicalizado plantea dudas en muchos casos, y sobre todo en los siguientes ejemplos, en los que, a diferencia de los anteriores, no es posible por el contexto averiguar ante qué tipo de plato nos encontramos:

- (31) a. Tres *platicos* pequeños finos azules (*ODE* – Almería, Cuevas de Almanzora, 1714).
 b. Doze *platillos* trincheros de plata (*ODE* – Madrid, 1715).
 c. Doze *platillos* de china blancos con flores (*ODE* – Almería, Cuevas de Almanzora, 1766).
 d. Tres *platitos* en dos reales y m[edi]o (*ODE* – Madrid, Navalafuente, 1809).
 e. Cuatro *platillos* de Balencia (*ODE* – Almería, Abla, 1833).
 f. Diez gicaras para chocolate, cuatro *platitos* encarnados (*ODE* – Madrid, Alcalá de Henares, 1890).

En estos casos, ya dudosos por la falta de contexto, se constata la alternancia documentada en el empleo del diminutivo *platillo-platito-platico*. Especialmente en los casos de *-ito* e *-ico*, se podría defender un valor transparente del diminutivo sujeto a la procedencia geográfica: los casos de *-ico* se documentan en Almería, mientras que los de *-ito* en Madrid, coincidiendo con la distribución general del uso de los diminutivos en estas regiones y en los periodos reflejados (§ 7.4.3.3. *Distribución geográfica*). La variación en el diminutivo sería entonces una prueba del empleo transparente del diminutivo.

La alternancia del sufijo, unido al uso similar que parecen revelar algunos ejemplos de *plato* y *platillo*, nos ha llevado a considerar todas las formas de *platillo-platico-platito* en el corpus como diminutivos transparentes, sin embargo, podría defenderse también lo contrario, dado que el valor lexicalizado parece más evidente en algunos de estos casos, como se ha mencionado. Por tanto, la separación entre diminutivo transparente y lexicalizado es polémica en algunos casos y los ejemplos no ofrecen indicios suficientes para la desambiguación. Siendo conscientes de este hecho, para este estudio se han considerado diminutivos transparentes. Por un lado, la alternancia que se establece entre el diminutivo (*platillo*) y la forma base (*plato*) para los mismos valores vendría a significar que en los términos sufijados prima la diferencia de tamaño con respecto al resto. Este mismo criterio es el que se ha valorado también en otras palabras, como por ejemplo a la hora de considerar *banquillo* como un diminutivo de *banco* (véase más arriba § *Banquillo*). Por otro lado, la alternancia del sufijo nos hace pensar en un empleo transparente del diminutivo y no tanto lexicalizado de algunos casos considerados más bien dudosos.

Puntilla

El término *puntilla* se ha considerado diminutivo transparente en los tres casos siguientes localizados en un mismo inventario de Almería. Se añade el contexto ampliado en el que puede verse cómo el término se relaciona con joyas o alhajas:

- (32) Dos pares de aretes de oro con tres adarnes, sesenta r[eale]s. Un par de id[em] con *puntillas*, cuarenta r[eale]s. Una cruz con diez diamantes rosas, cien r[eale]s. Una sortija antigua con esmeraldas, sesenta r[eale]s. Una id[em] con una esmeralda, veinte r[eale]s. Tres id[em] con *puntillas* de diamantes, ciento veinte r[eale]s. Una id[em] esmaltada, treinta r[eale]s. Una id[em] con un topacio, diez r[eale]s. Dos id[em] de oro lisas, veinte r[eale]s. Un alfiler con un topacio, cincuenta r[eale]s. Otro id[em] con *puntillas* de diamante, sesenta r[eale]s (*ODE* – Almería, 1844).

Puntilla se interpreta en estos casos como diminutivo de *punta* en su sentido de ‘extremo’ del diamante. El término puede entenderse, o bien como referencia únicamente a la punta aguda del diamante, o bien como la presencia de una pequeña cantidad de esta piedra preciosa en las joyas que se mencionan. En cualquier caso, ambas interpretaciones estarían muy relacionadas semánticamente y serían compatibles con el sentido real de disminución del diminutivo, razón por la que se ha optado por incluir estos tres casos en el recuento de formas diminutivas en el corpus.

Este uso particular de *puntillas* difiere de los casos mucho más numerosos que en el corpus aparecen ligados a telas y prendas de vestir. En ellos el término se emplea en su sentido lexicalizado de ‘encaje’ y sí aparece registrado en los diccionarios, por lo que con este significado se ha considerado como un diminutivo lexicalizado (sobre la lexicalización de esta voz véase § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*).

En su sentido lexicalizado, la voz *puntilla* alude a un tipo de adorno realizado en encaje que se cose en el borde de algunos textiles como vestidos, pañuelos o toallas. La variación en el uso del diminutivo que se observa en los diferentes corpus consultados respondería a una fase previa a la lexicalización definitiva en *-illa*. El corpus *ODE* refleja la variación *puntilla/puntica/puntita*, aunque únicamente la variante en *-illa* es la que aparece registrada en las fuentes lexicográficas. En efecto, la voz *puntilla* es la que se generalizó como término lexicalizado y la que aparece en la mayoría de los ejemplos de corpus, mientras que *puntica/puntita* tiene un carácter mucho más residual. La variante *puntica* aparece solo en una ocasión en el corpus de Madrid y es también residual en el *CorLexIn*, donde se registra en algunos puntos como Zaragoza, Albacete y Toledo; y más tarde en el ejemplo al que nos referimos de nuestro corpus:

- (33) a. Vna toalla de ruan con sus redes y *punticas* en dos ducados [...] Vna camissa de muxer de cuerpo de ruan y las faldas de lienço con unas *punticas* (*CorLexIn* – Toledo, Escalona, 1644).
b. Calçones, ropilla y jubón de ormesí negro con *punticas* al canto, número 225, muy viejo (*CorLexIn* – Zaragoza, 1647).

- c. Otro paño de chuez entretallado con *puntica* alrededor en veinte y dos reales (*CorLexIn* – Albacete, Almansa 1653).
- d. Vn panuelo con *punticas* antiguas, viejo (*ODE* – Madrid, 1700).

Aunque el uso de *puntica* es eventual, la variación se registra en su correspondiente entrada en el glosario (6.5.3.4. *Glosario* s.v. PUNTILLA, -CA).

6.4. Diminutivos transparentes

Se incluyen en este apartado las palabras en las que el sufijo diminutivo manifiesta un valor apreciativo que se relaciona, principalmente, con el significado objetivo de disminución. Quedan por tanto fuera de este grupo las voces que han adquirido sentidos específicos fruto de un proceso de lexicalización de las que se hablará más adelante.

En este apartado se habla de diminutivos transparentes conforme a los criterios ya descritos con anterioridad en § 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*. El cómputo numérico como resultado de la aplicación de estos criterios se ofrece junto a otros datos de interés en § 6.4.1. *Recuento de formas diminutivas en el corpus*. Una mención más detallada sobre el significado de los diminutivos transparentes en el corpus se acomete en § 6.4.2. *Valor semántico del diminutivo en el corpus*.

Conforman este apartado los diminutivos que se incluirán en el análisis estadístico sobre la distribución geográfica y cronológica de los diminutivos en las provincias de Almería y Madrid en los siglos XVIII y XIX, cuyo listado completo de formas junto a los ejemplos se puede consultar en § *Anexo IV. Relación de diminutivos transparentes incluidos en el análisis*.

6.4.1. Recuento de formas diminutivas en el corpus

Para este recuento ha sido imprescindible la anotación morfosintáctica de palabras con sufijos diminutivos en el corpus de inventarios de Almería y Madrid, de manera que todas ellas se han recuperado posteriormente a partir de una sola búsqueda. Recordemos que la derivación apreciativa como rasgo morfológico se indica con la letra D en la última posición de la etiqueta para el caso de los sustantivos, y en la segunda posición para el caso de los adjetivos (§ 4.5.1. *Etiquetario*). Por otro lado, en el proceso de asignación de lema, las palabras con sufijos diminutivos se han lematizado bajo la forma sin el diminutivo, de tal manera que una búsqueda lematizada devolverá todas las variantes formales de la base, incluidas aquellas con diminutivo. Véanse algunos ejemplos:

<i>cofrecito</i>	NCMS00 D	lema: cofre
<i>arquitas</i>	NCFP00 D	lema: arca
<i>chiquito</i>	AQ D MS0	lema: chico
<i>pequeñita</i>	AQ D FS0	lema: pequeño

Fruto de una búsqueda por etiqueta morfosintáctica en los inventarios de Almería y Madrid, obtenemos el siguiente recuento de palabras con diminutivo:

		Diminutivos transparentes
Sufijo	-illo	173
	-ito	147
	-ico	52
	Total	372

Tabla 12: Recuento de formas diminutivas en el corpus documental

El desglose de estas cifras por siglos y por provincia se analizará más adelante en § 7.4. *Resultados*. Por el momento, conviene precisar que en este cómputo entran en consideración todos los casos en los que se localiza una palabra con diminutivo, independientemente de que esta pueda repetirse varias veces en el corpus. Es por tanto necesario distinguir entre formas, tipos y lemas (Brezina, 2018, p. 39).

El resultado que se muestra en la tabla hace referencia al total de casos observados de alguno de los sufijos en el corpus. 372 es el número de veces que aparece una forma con diminutivo, sin importar si esa misma forma ya existía antes en el corpus. Cada ocurrencia de una palabra con diminutivo equivale a un caso. Esta noción de forma como ocurrencia única es la que se identifica comúnmente con el concepto de *token* en lingüística de corpus y tiene que ver con el total de casos de un fenómeno. Según nuestro análisis, en el corpus figuran 372 formas con diminutivo o, si se prefiere, 372 tokens.

Los tipos, en cambio, se corresponden con las formas diferentes que puede adoptar una palabra –*niño*, *niños*, *niñitos* son tipos de la palabra ‘niño’–. Es evidente que en el corpus habrá muchos menos tipos que formas, ya que los tipos de palabras engloban los casos de formas repetidas. Para esta investigación no se ha contabilizado el número de tipos diferentes con diminutivo. Por último, el concepto de *lema* sería el más amplio de los mencionados hasta ahora, pues bajo él quedarían incluidas todas las variantes flexivas de una misma palabra, incluidas sus formas diminutivas.

Se recopila todo lo anterior partiendo de un ejemplo. Supongamos que en el corpus se localiza la palabra *cajita* en cuatro ocasiones y *cajica* en dos. Cada una de estas seis ocurrencias constituye una sola forma de la palabra con diminutivo, que se incluye directamente en el recuento total de diminutivos, sea cual sea el sufijo empleado. Si se piensa además que encontramos *cajitas* en una ocasión, el cálculo ascendería ya a siete formas con diminutivo, aunque palabras diferentes o tipos de palabras solo tengamos tres (*cajita*, *cajica* y *cajitas*). Todas ellas se englobarán bajo el lema *caja*, pues son distintas variantes formales de una misma palabra. En definitiva, tenemos siete formas o tokens con diminutivo –*cajita* (4 casos), *cajica* (2 casos) y *cajitas* (1 caso)–, tres tipos o palabras diferentes –*cajita*, *cajica*, *cajitas*–, y un solo lema –*caja*–.

En las tablas siguientes se recogen los lemas (141 en total) que en el corpus aparecen con alguno de los tres diminutivos estudiados. En algunos casos el diminutivo se añade sobre la base de aumentativos ya lexicalizados, como ocurre con *arcón*, *azadón*, *cajón*, *jubón*, *lechona*, *mangote*, *medallón*, *pipote*, *rejón* o de otros diminutivos, como en *cazuela*, *mantilla*, *muleta*, *paleta*, *pañuelo* o *pileta*.

-illo/-illa		-ito/-ita		-ico/-ica
arcón	jubón	alfombra	fino	bacina
aro	lámpara	almendra	gordo	bancal
artesa	mangote	asa	jarra	baúl
azadón	máquina	bandeja	libro	bote
banco	medallón	blusa	marco	canasta
bigornio	muleta	bocal	medalla	cántaro
bolso	paleta	broche	melindre	engaste
botón	pipote	cacerola	nudo	grano
cabecero	puerta	cajón	pájaro	hacha
cachorro	rambla	caldero	pantalón	jícara
campana	rejón	camisa	pañuelo	lechona
cedazo	saco	capa	peana	manta
cerco/a	serón	cazuela	pedazo	mantilla
colchón	sierra	collar	perla	olla
corneta	tabla	concha	pomo	orón
chispa	taburete	contador	queso	ovillo
daga	tarima	corona	sartén	pañó
espejo	telar	chapa	tartera	perro
fleco	tenaza	chico/a	tinaja	pieza
frasco	toalla	dosel	urna	pileta
haba	tubo	estera	zafa	retal
harnero	valona			silla
herramienta	vieja			
hierro	zaranda			
jamugas				

Tabla 13: Bases en el corpus a las que se añade un sufijo diminutivo

Entre las palabras que presentan variación en el uso del diminutivo se encuentran además las voces recogidas en la Tabla 14. Algunas de ellas se caracterizan por presentar una triple alternancia en el sufijo, mientras que predominan las voces en las que *-illo* alterna con cualquiera de los otros sufijos, *-ito* o *-ico*. Solo en cinco palabras la variación se produce entre estos últimos.

-ito/-ico/-illo	-ito/-illo	-ico/-illo	-ito/-ico
arca	brasero	azada	barril
cerdo	cadena	cofre	caja

mesa	caldera	embudo	cuadro
orza	cazo	pequeño/a	escritorio
pila	copa	sábana	lámina
plato	cortina	taza	
romana	cuchara		
	perol		
	rosa		
	vaso		

Tabla 14: Palabras con variación en el uso del diminutivo

Los sufijos diminutivos en el corpus están ligados sobre todo a la base de sustantivos y, en menor medida, a la de adjetivos calificativos que acompañan a estos nombres y con los que se señala alguna de sus cualidades. Entre los adjetivos predominan los diminutivos de *pequeño* y *chico* y solo en tres ocasiones el diminutivo se añade sobre otros adjetivos calificativos: *gordo*, *fin* y *vieja*, dando lugar a *gorditos*, *finita* y *viejecilla* en el corpus. Con respecto a la distribución por género, no existen diferencias notables que nos hagan reparar en un empleo distinto del diminutivo según el género de la palabra (Tabla 15). El desequilibrio numérico es evidente entre los derivados sustantivos y los adjetivos, aunque, como se verá más adelante, la relación entre el diminutivo y el procedimiento de la adjetivación es estrecha (§ 6.4.2. *Valor semántico del diminutivo en el corpus*). No se encuentran ocurrencias de palabras con sufijos diminutivos aplicados a otras categorías gramaticales tales como los adverbios.

	Sustantivos	Adjetivos
Género masculino	161	13
Género femenino	182	16
Total	343	29

Tabla 15: Distribución de diminutivos según la categoría gramatical

6.4.2. Valor semántico del diminutivo en el corpus

En este apartado se aborda el valor semántico que cumple el diminutivo particularmente en el corpus documental de esta tesis, conformado por los inventarios de Almería y Madrid de los siglos XVIII y XIX.

Es innegable que la capacidad para expresar tanto cualidades objetivas como matices afectivos es consustancial a la semántica de los diminutivos. Cabe destacar que en la mayoría de los casos ambas funciones están presentes a la vez y no parece tan claro que los valores meramente nocionales, en particular aquellos que aluden al tamaño, puedan manifestarse independientemente de los valores afectivos (*NGLE* § 9.6a, p. 651). Según este planteamiento, un diminutivo podría desarrollar un valor afectivo o bien afectivo y conceptual, pero nunca este último en exclusividad. Planteada

así la cuestión, se podría determinar entonces si se manifiesta en una forma diminutiva cada uno de sus valores posibles, ya sean nocionales o expresivos. En lo que al corpus de esta tesis se refiere, lejos de negar toda función afectiva en los diminutivos documentados, la función nocional prevalece por encima de ninguna otra en la mayoría de los ejemplos estudiados. En lo que sigue trataremos de probar esta afirmación.

Se ha demostrado que el valor con el que se emplea el diminutivo parece estar fuertemente ligado al contexto de la situación comunicativa. Los diminutivos registran una mayor frecuencia de uso en la lengua oral que en la escritura, debido a que el carácter predominantemente afectivo encuentra mayor arraigo en la lengua popular y familiar (Martín Zorraquino, 2012, p. 557). Entre las fuentes escritas, ciertos tipos de textos favorecen la presencia de los diminutivos más que otros y, del mismo modo, el predominio de una de sus funciones sobre la otra. Así, los diminutivos con valor afectivo abundan en el género literario, periodístico o publicitario y, en general, en aquellos en los que es permisible que afloren los matices subjetivos. Por el contrario, se prescinde de su uso en los textos de carácter científico, didáctico, jurídico o administrativo en los que predomina un tono solemne (NGLE § 9.6p, p. 656)⁷⁹.

Los inventarios de bienes se aproximan más a este segundo grupo de géneros discursivos en cuanto a que se trata de documentos notariales en los que predomina una escritura diplomática más o menos regulada. En ellos desaparece el tono solemne y excesivamente formulario y se caracterizan por su practicidad en los que la enumeración y descripción de cosas cotidianas favorece la presencia de diminutivos. En los inventarios confluye el carácter objetivo con el que se describen las pertenencias de la manera más fiel posible con la designación de objetos familiares que son nombrados muy frecuentemente. Las referencias inequívocas al tamaño, el uso, la forma o la composición de estos objetos para su correcta identificación y tasación incitan a pensar en una función nocional del diminutivo. Esta función es, por otra parte, la que resulta más natural en los nombres contables que denotan seres materiales (NGLE § 9.6b, p. 651) y la que parece prevalecer en casos como los siguientes:

- (34) a. Una criba, un *arnerillo* y un arnero grande (ODE – Madrid, 1834).
b. Un cazo, una sarten, *cacillo* de azofar y otro de yerro (ODE – Madrid, 1801).
c. Ytt[em] otra caja negra p[ar]a tabaco de concha con gonzes de platta, vieja, en quattro r[eale]s. Ytt[em] una *cagita* de madera pintada en dorado p[o]r fuera en diez y siete m[a]r[avedí]s (ODE – Almería, Cuevas de Almanzora, 1766).

En los ejemplos de arriba la forma derivada con diminutivo concurre con la forma primitiva sin él. La diferencia entre un *harnero* y un *harnerillo*, un *cazo* y un *cacillo* o una *caja* y una *cajita* parece centrarse en el tamaño más que en una valoración afectiva

⁷⁹ Así lo demuestra González Ollé para los textos medievales en los que analiza la presencia y ausencia de los diminutivos en base al género textual de las obras analizadas. Matiza el autor esta afirmación con la idea de que la presencia del diminutivo “guarda una estrecha relación con la expresividad, pero en cuanto que ésta se manifieste en forma espontánea, es decir, no sometida a elaboraciones intelectuales o retóricas” (1962, p. 250), hecho que explica la presencia del diminutivo en determinadas circunstancias en géneros en principio no afines a él como la épica, la prosa didáctica o la histórica.

de estos objetos. La alusión al tamaño por medio del diminutivo en estos casos opera conjuntamente al resto de características descriptivas que sirven para detallar las particularidades de cada uno de ellos. En los inventarios predomina sobre todo la intención descriptiva, por lo que resulta menos probable que su empleo se deba a la subjetividad del que escribe más que a una interpretación objetiva del tamaño. En contra de esta afirmación podrían aducirse los otros muchos ejemplos en el corpus en los que el diminutivo se presenta junto a una forma adjetiva. Sin embargo, lo que aquí se sostiene es que el carácter práctico de los inventarios haría predominar el valor nocional sobre el afectivo, aunque este también pueda estar presente en las formas registradas.

En efecto, los diminutivos sintéticos (*sartencita*) concurren en el corpus con los analíticos (*sartén pequeña*) para la expresión del tamaño. Tampoco faltan los ejemplos en los que se combinan ambos procedimientos y el diminutivo se añade sobre la base del sustantivo (*sartencita pequeña*) o la del adjetivo (*sartén pequeñica*). Mediante el recurso de la adjetivación el uso del diminutivo se combina con adjetivos que denotan pequeñez y que actuarían como reforzadores de la función nocional del diminutivo sin que se perciba por ello redundancia (NGLE § 9.6b, p. 651).

Para algunos autores los diminutivos analíticos vendrían a ratificar la función objetiva y solo en estos casos puede hablarse de un verdadero sentido dimensional. Insiste Alonso en esta idea afirmando que “cuando el sentido central es realmente el de disminución, se suele insistir en la idea de pequeñez con otros recursos (*una cajita pequeña, una cosita de nada*, etc.)” (Alonso, 1961, p. 163). En la misma dirección apunta González Ollé (1962, p. 256) cuando constata una mayor presencia de los diminutivos analíticos sobre los sintéticos en los textos medievales que analiza y en los que se busca un tono neutro carente de subjetividad. Justifica el autor esta circunstancia en el hecho de que el predominio de manera general de la función afectiva sobre la minorativa en el uso de los diminutivos ha llevado a prescindir de estos en contextos en los que se desea eliminar todo rasgo de afectividad para evitar suscitar valores equivocados. Según esta afirmación, en la conciencia del hablante un diminutivo es sobre todo afectivo, por lo que cuando no se pretende que se interprete así se tiende a evitar su uso. Este planteamiento es lo que ha hecho que en los textos de carácter científico y similares se prescinda de estas formas en favor de otros procedimientos como el de la adjetivación.

En esta tesis interesan solo los derivados con diminutivo, por lo que en las formas analíticas se han valorado aquellos casos en los que el diminutivo se añade a alguno de los elementos del sintagma. Los ejemplos de este tipo son numerosos en el corpus; se expone aquí una muy sucinta selección, dada la similitud de todos ellos:

- (35) a. Una alfombra vieja de moqueta para los pies en seis pesetas. Una *alfombrita* de moqueta vieja, *pequeña*, tasada en dos pesetas (ODE – Madrid, 1877).
- b. 1 sortija con cinco brillantes. 1 sortija con quince *brillantes pequeñitos* (ODE – Madrid, 1843).
- c. Un cazo *chiquito* biejo de cobre en 2 (ODE – Almería, Fiñana, 1835).
- d. Vna *arquita pequeña* de albercoquero con zerrad[ur]a y llaue (ODE – Almería, Cuevas de Almanzora, 1766).

- e. Seis *quadritos pequeños* en quatro r[eale]s (ODE – Madrid, Brunete, 1751).
f. Vna mesa de vna tabla de nog[a]l mediana y viexa. Ottra *mesica pequena* de serbal, pequeña (ODE – Almería, Fiñana, 1745).

En los casos en los que la sufijación se combina con el uso de adjetivos trasluce de manera inequívoca el sentido nocional del tamaño. De nuevo, en algunos de los ejemplos de arriba puede verse que los diminutivos analíticos aparecen junto a las formas simples (*alfombra* y *alfombrita pequeña*, *brillantes* y *brillantes pequeñitos*, *mesa* y *mesica pequeña*), de las cuales se diferencian por el tamaño como objetos de la misma especie. La sufijación diminutiva en los adjetivos, –especialmente en los de pequeñez como *pequeño* y *chico*– se ha asociado con un valor superlativo en el que se debe suponer una intensificación de la cualidad expresada por el adjetivo (González Ollé, 1962, p. 231) o una interpretación cuantificativa (NGLE § 9.6g, p. 653).

Como se ha visto, el significado real de disminución es evidente cuando se combina con otros procedimientos como el de la adjetivación, pero sería extraño que en los casos en los que el diminutivo no aparece junto a un adjetivo fuera otra la función principal del diminutivo. No obstante, no se puede negar con rotundidad la existencia de los matices afectivos en el uso de los diminutivos, los cuales podrían justificarse en el hecho de que estos se aplican a objetos domésticos usados de manera recurrente en la vida cotidiana y en el lenguaje familiar⁸⁰. Aunque la función expresiva sea la más frecuente en el español, no es la única ni tampoco la más importante en otros casos como los que se han ido mostrando a lo largo de este apartado a partir de los ejemplos del corpus. En los inventarios, pese a tratarse de una prosa administrativa o judicial desprovista, *a priori*, de intencionalidad afectiva, abundan los diminutivos, lo que se justificaría por el sentido nocional con el que se emplean estas formas. La función minorativa se expresa en muchos casos por medio de diminutivos analíticos en los que el sufijo confluye con adjetivos que denotan pequeñez, pero también en las formas sintéticas en las que no cabe sospechar que el sentido prototípico del diminutivo sea distinto.

En definitiva, se considera que entre los diminutivos transparentes de nuestro corpus predomina sobre todo el sentido nocional. La función disminuidora, aunque minoritaria en el español, se localiza perfectamente y sin equívocos en casos determinados, siendo los inventarios de bienes un claro ejemplo de ello dada la naturaleza que los caracteriza. En este apartado, lejos de negar la presencia de la función afectiva en los diminutivos del corpus, se ha intentado probar la existencia del valor objetivo del diminutivo, menos frecuente y más difícil de documentar en las fuentes escritas de que disponemos.

⁸⁰ En consonancia con este planteamiento cabe destacar lo siguiente: “el hecho de que diminutivos y aumentativos se vinculen preferentemente a palabras empleadas con elevada frecuencia habla en favor de la pertenencia de estos al lenguaje familiar. Con diminutivos se designan sin problemas los animales domésticos y las bebidas preferidas, pero no los animales raros y las bebidas insólitas; también se designan objetos familiares sólo si se utilizan palabras familiares. Se trata de referentes con los que el hablante tiene una especial relación personal, que él expresa en el diálogo familiar” (Hummel, 1997, p. 196).

6.5. Diminutivos lexicalizados

Forman parte de este grupo todas las palabras del corpus en las que el diminutivo ha desarrollado valores específicos que se alejan de su significado compositivo original. En ellas el significado del sufijo ha dejado de ser transparente tras haber experimentado un proceso de lexicalización⁸¹, razón por la que se las denomina también voces opacas o lexicalizadas. Bajo esta denominación únicamente se han tenido en cuenta las voces que han experimentado un cambio lingüístico en castellano, por lo que no se incluyen aquí las que proceden de voces latinas que ya contenían un sufijo diminutivo o aquellas de procedencia extranjera que se importan ya con el diminutivo.

El recuento numérico de las voces que integran este grupo se incluye en el apartado § 6.5.1. *Recuento de diminutivos lexicalizados en el corpus*. La mayoría de estas palabras se asocian a campos semánticos muy concretos relacionados con los objetos de la vida cotidiana y suelen compartir entre ellas algunas características generales que se mencionan en § 6.5.2. *Características de los diminutivos lexicalizados*.

Algunos de los diminutivos lexicalizados son especialmente interesantes por constituir casos de voces con escasa o nula documentación lexicográfica. El corpus ha servido para desvelar su uso, así como el significado y la extensión geográfica de algunas de ellas, apoyándonos sobre todo en la consulta de fuentes lexicográficas y en la del corpus *ODE* en su conjunto y de *CorLexIn*. Asimismo, la localización concreta que desvelan los corpus para algunas de estas palabras ha permitido considerarlas voces regionales; se analizan en § 6.5.2.1. *Regionalismos y voces poco documentadas*. Por otro lado, la variación en el diminutivo de una palabra ya lexicalizada, aunque poco habitual, se percibe en el caso de algunas voces. De entre estas, se analiza detenidamente la alternancia entre *bufetillo/bufetico/bufetito*, la cual pone de manifiesto la dificultad que supone en algunos casos distinguir entre el uso transparente y lexicalizado del diminutivo (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*).

Finalmente, el total de los diminutivos lexicalizados en la documentación analizada se incluye a modo de glosario al final de este apartado. En él se define y ejemplifica con las citas del corpus cada una de las palabras documentadas (§ 6.5.3.4. *Glosario*).

6.5.1. Recuento de diminutivos lexicalizados en el corpus

Numerosos diminutivos en los inventarios de Almería y Madrid forman parte de palabras lexicalizadas. Bajo esta denominación se consideran todos los términos que deben incluirse en el diccionario, puesto que en ellos el diminutivo ha perdido su valor original y ya no es posible obtener el significado de la palabra o sus matices expresivos a partir de la combinación de la base y el sufijo (*NGLE* § 9.3c, p. 636). Mientras que en

⁸¹ Sobre los términos con los que se entiende el concepto de *lexicalización* en esta tesis véase § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*.

los diminutivos transparentes el significado se interpreta morfológicamente, en los diminutivos lexicalizados la palabra conforma ya una unidad léxica por sí misma con significado independiente, quedando incluso diluida en algunas ocasiones el significado original del sufijo diminutivo.

En el total de textos que conforman el corpus documental de esta tesis, el número de ocurrencias de diminutivos lexicalizados dobla al de los diminutivos transparentes de acuerdo con la siguiente distribución:

		Diminutivos transparentes	Diminutivos lexicalizados
Sufijo	-illo	173	709
	-ito	147	11
	-ico	52	49
	Total	372	769

Tabla 16: Recuento de diminutivos transparentes y lexicalizados en el corpus

De entre el total de lexicalizaciones, *-illo* es el que reúne un mayor número de casos, seguido de *-ico* e *-ito*, en último lugar. Con el diminutivo *-ico* se localizan 49 formas ortográficas que se identifican con los lemas *abanico*, *vainica*, *bufetico* y *puntica*. En lo que respecta a *-ito*, sus doce formas corresponden a los lemas *manguito*, *bufetito* y *francesita*. Cabe destacar que algunos de estos términos alternan en el corpus con sus respectivas formas en *-illo*: *puntilla-puntica*, *francesilla-francesita*⁸² y *bufetillo-bufetico-bufetito*, por lo que la lexicalización en *-ito* o *-ico* no es exclusiva.

El diminutivo *-illo* es el que, con sobrada diferencia, registra el mayor número de lexicalizaciones en el corpus con un total de 709 ocurrencias entre las que se identifican hasta 100 lemas diferentes con sus respectivos significados y ejemplos recogidos en el glosario (§ 6.5.3.4. *Glosario*). Al igual que con los diminutivos transparentes, se han contabilizado todos los casos de palabras con diminutivo independientemente de que una misma forma pueda repetirse varias veces en el corpus (véase la diferencia entre formas, tipos y lemas explicada con anterioridad a propósito de los diminutivos transparentes en § 6.4.1. *Recuento de formas diminutivas en el corpus*).

En este recuento se han desestimado todas aquellas formas en las que el sufijo estaba ya lexicalizado en latín. Estos términos han pasado al español tras un proceso de derivación por sufijación que tuvo lugar previamente en la lengua latina. Estas voces ya gozaban de significado pleno en latín, donde el proceso de lexicalización ya se había completado antes de pasar a nuestra lengua. Forman parte de este grupo las voces

⁸² ‘Prenda de vestir femenina’. El *DHTT* considera la forma *francesilla* una tela de algodón a cuadros, a menudo rayada (s.v. *FRANCESILLA*). La única ocurrencia del término en nuestro corpus (“vna *francesita* de duroy nueva [...] otra usada de estameña”, *ODE* – Madrid, San Lorenzo de El Escorial, 1789) nos induce a pensar más en una prenda de vestir que en una tela. Algunos aspectos relacionados con esta voz se examinan en § *Francesita*.

*angarilla*⁸³ (ANGARIELLAE), *anillo* (ANELLUS), *amarillo* (AMARELLUS), *borrico* (BURRICUS), *cuchilla* y *cuchillo* (CUTELLUS), *hebilla* (*FIBELLA), *martillo* (MARTELLUS), *novillo* (NOVELLUS), *ovillo* (GLOVELLUS), *rastillo* o *rastrillo* (RASTELLUS), *rodilla* (ROTELLA), *sencillo* (*SINGELLUS), *traílla* (*TRAGELLA), *vajilla* (VASCELLA) y *zarcillo* (CIRCELLUS).

6.5.2. Características de los diminutivos lexicalizados

El estudio de los inventarios de bienes desvela un inmenso caudal léxico perteneciente al ámbito de la vida cotidiana. Tejidos, muebles o utensilios nuevos iban surgiendo para adaptarse a los cambios que experimentaba la sociedad del momento y a los que era necesario dar un nombre. Fruto del incesante intercambio entre países europeos como consecuencia de un comercio cada vez más desarrollado, muchos de los productos importados en España toman el nombre de la lengua de origen del lugar del que proceden. Proliferan en los inventarios los préstamos léxicos, sobre todo galicismos (*marmita*, *batista*, *plaqué*, *franela*, *lamparilla*, *droguete*...) e italianismos (*balcón*, *brocado*, *casaca*, *festón*, *gaveta*...). Pero también es destacable la aparición de palabras creadas a partir de la sufijación y, más en particular, de la sufijación apreciativa.

En lo que respecta al léxico de la vida cotidiana, las lexicalizaciones a partir de sufijos diminutivos han resultado ser muy productivas en campos como el de las telas o el de la indumentaria. Así queda evidenciado en el corpus de inventarios, donde estos dos campos acumulan un mayor número de términos lexicalizados frente a otros. No es de extrañar, dado que la península ibérica evidencia un claro desarrollo de la industria textil ya desde la Edad Media (Cabrera Lafuente, 2005) y es en el siglo XVIII cuando asiste a una auténtica revolución textil caracterizada “por una fuerte influencia de la situación política, económica y artística francesa” (Amaro Martos, 2018, p. 698). Uno de los principales motivos de esta transformación estará marcada por la desaparición paulatina de los gremios de artesanos y la aparición de los mercaderes, lo que supone el origen de las tiendas y del negocio estable en detrimento del nómada (Amaro Martos, 2018, p. 688).

En lo relativo a las telas, el diminutivo *-illo* confiere por lo general el significado de ‘cierto tipo de’. En estos casos, el diminutivo unido a la base, más que una referencia al tamaño implica una relación de similitud con la realidad nombrada. Esta relación se evidencia en un gran número de tejidos en el corpus como *algodoncillo*, *beatilla*, *bocadillo*, *hiladillo*, *holandilla*, *espumilla*, *estopilla*, *jerguilla*, *lanilla*, *picotillo*. En estas palabras, junto al significado de ‘tipo de’, el sufijo podía encerrar además otras connotaciones relacionadas con la calidad o el valor de la tela, normalmente siempre de

⁸³ ‘Armazón de cuatro palos clavados en cuadro, de los cuales penden unas bolsas grandes en las que se transporta a mano o en caballería cualquier tipo de carga’ (DLE s.v. ANGARILLA, acep. 6). En el corpus, usado en singular y plural en los ejemplos “una *angarilla* de llevar cristales” (ODE – Madrid, Aranjuez, 1870) y “unos palos en forma de *angarillas* para sacar mies” (ODE – Almería, Cuevas de Almanzora, 1704).

calidad inferior, e incluso también aludir a su consistencia, más ligera o más gruesa, según el caso. Así, la *jerguilla* es una tela que se parece a la jerga, la *lanilla* es un tejido más fino que la lana o el *picotillo* es un picote de calidad inferior.

A estos tejidos se pueden sumar otros no documentados lexicográficamente, pero de igual presencia en nuestro corpus como *albornocillo*, *batistilla*, *borloncillo*, *escamilla*, *apretadillo* y *medianillo*⁸⁴ –en estos dos últimos la derivación es de tipo adjetival y toma como base los términos *apretado* y *mediano* para conformar dos nuevas lexías aplicadas al lienzo–. Sus ocurrencias en los textos dejan ver la misma lógica que subyace a los derivados con *-illo* en nombres de telas como los anteriores. La lexicalización de cada uno de ellos como un tipo de tela del que deriva se evidencia en la coexistencia en el corpus del término base y su derivado. De esta manera, es posible localizar ejemplos de *albornoz* (“otro debantal de *albornoz*” ODE – Madrid, 1724), pero también de *albornocillo* (“cinco baras de *albornocillo* negro” ODE – Madrid, 1711), donde se aprecia ya un claro proceso de lexicalización.

El sufijo diminutivo *-illo* es el más productivo entre los sufijos derivativos para la formación de lexicalizaciones. De la consulta de los corpus complementarios y estudios sobre el tema se desprende la existencia de otros derivados como *caserillo*, *damasquillo*, *durancillo*, *telilla*, *rayadillo* (en *CorLexIn* y en otras provincias de ODE), *escarlatilla*, *cameloncillo*, *castorcillo*, *velillo* (Stala, 2014), *tramadillo* o *telilla* (Perdiguero, 2015).

Los casos señalados ponen en evidencia que este mecanismo de derivación a partir del diminutivo *-illo* estaba ampliamente extendido para poner nombre a nuevas telas que se diferenciaban de sus equivalentes en la calidad o el grosor, lo que les confería un nuevo estatus y las convertía en un tipo diferente de tejido. El uso que de estas palabras se hace en los inventarios nos hace considerarlas ya en la época como unidades léxicas de pleno derecho en las que, unido a la consideración de un tipo de tela derivada de su análoga, se añadía también en ocasiones el de inferior calidad. En algunas zonas, estos nuevos términos prosperaron hasta el punto de dialectalizarse, sin llegar a calar en la lengua general como ocurre con *medianillo*, *apretadillo* y *batistilla* (§ 6.5.2.1. *Regionalismos y voces poco documentadas*).

Estrechamente ligados con los tejidos, son también numerosos en los inventarios de esta época los términos que aluden a las diferentes técnicas de urdimbre y decoración de las telas. Es a partir de la segunda mitad del XVIII cuando se desarrolla el talento creativo de sastres y modistos (Amaro Martos, 2018) y en los inventarios comienzan a aparecer con mayor frecuencia prendas y vestidos decorados y telas más nobles. Con respecto al siglo anterior, el carácter utilitario de las prendas deja de ser el principal aspecto tenido en cuenta, dejando paso a la ornamentación y la originalidad de los diseños. En este sentido, resulta interesante la evolución que experimenta el uso de la *mantilla*, de accesorio de abrigo a prenda que se decora y se usa como adorno y realce (sobre este cambio semántico véase más adelante § 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*).

⁸⁴ El estudio de estas palabras no contempladas en los diccionarios (NTLLE) se aborda con detalle en § 6.5.2.1. *Regionalismos y voces poco documentadas*.

Se documentan en el corpus términos creados a partir del mismo proceso de derivación con *-illo* que aluden a los tipos de labores o adornos que podían aplicarse a almohadas, colchas, manteles, toallas o prendas de vestir: *bolillo*, *canutillo*, *confitillo*, *cordoncillo*, *felpilla*, *granillo*, *gusanillo*, *puntilla*, *vainica*. En estas palabras, el diminutivo sí implica una referencia al tamaño como indicador de una labor menuda, aunque no necesariamente, como por ejemplo ocurre en *felpilla*, donde el diminutivo comporta una relación de similitud y adquiere de nuevo el significado de ‘tipo de’. En este caso, por *felpilla* se entiende cierta especie de hilo, similar a la felpa.

En el campo semántico de la vestimenta también se localizan numerosas lexicalizaciones a partir de *-illo* en prendas como *almilla*, *bolsillo*, *calzoncillo(s)*, *capotillo*, *chaquetilla*, *coletillo*, *cotilla*, *golilla*, *justillo*, *manguito*, *mantilla*, *petillo*, *ropilla* o *tontillo*. En casi todas ellas impera el significado ‘tipo de’ o ‘especie de’; en otras, a la prenda se le confiere además un uso distinto del que ya tenía, como ocurre con *calzoncillos* (destinado a prenda interior, al contrario que los calzones), *cotilla* (usado por las mujeres, a diferencia de la cota, que era para los hombres), *mantilla* (como prenda para cubrir la cabeza y los hombros, a diferencia de la manta) o *bolsillo* (como bolso cosido a prendas de vestir), en el que además se mantiene el significado de ‘pequeño’ referido al tamaño.

También se documentan términos lexicalizados en el ámbito de las joyas, como en *almendrilla*, *arillo*, *broquelillo*, *cintillo*, *gargantilla*, *horquilla*, *manilla* o *mostacilla*, donde el diminutivo conserva su alusión al tamaño, pero también un uso específico distinto al que se destina el objeto, en este caso, como accesorio ornamental generalmente usado por las mujeres.

Además de los ámbitos mencionados, los inventarios están salpicados de otras lexicalizaciones correspondientes a una gran variedad de objetos domésticos relacionados con diferentes estancias de la casa y destinados a muy diversos usos. Entre ellos se encuentran algunos útiles de labranza (*escardillo*, *esportillo*, *horquillo*), herramientas de trabajo variadas (*cepillo*, *husillo*, *palomilla*, *tornillo*), instrumentos para tejer o relacionados con la costura (*almohadilla*, *bastillo*, *bolillo*, *palillo*), utensilios de cocina (*botillo*, *cocinilla*, *cuartilla*, *cuartillo*, *molinillo*, *morillo*, *paletilla*, *parrilla*, *tenacillas*), mobiliario del hogar (*banquillo*, *bufetillo*, *camilla*, *pajecillo*) o materiales para su construcción (como camas de *granadillo* y sillas con asientos de *soguilla*).

La gran mayoría de todas estas palabras en el corpus evidencian el proceso de lexicalización ya completo. Esto no solo se pone de manifiesto cuando el significado del término con diminutivo deja de ser transparente y se asocia con otro concepto, también cuando al término le sigue algún adjetivo referido al tamaño y de manera especial, cuando el adjetivo es grande, prueba de que el morfema apreciativo ha perdido su significado originario de aminoración. Sirven como ejemplos “campanilla pequeña” (ODE – Madrid, Estremera, 1836), “parrillas chicas” (ODE – Almería, Zurgena, 1825), “tontillo grande” (ODE – Madrid, 1750), “bufetico pequeño” (ODE – Almería, 1721).

Los casos más interesantes de diminutivos lexicalizados han estado motivados por su escasa documentación en los diccionarios o por el carácter dialectal que revela el empleo del término en los corpus. A estos casos se añaden otros que presentan alternancia en el uso del diminutivo, como *puntilla* y *bufetillo* (§ 6.5.2.2. *La variación*

del diminutivo en voces opacas: el caso particular de «bufetillo»). Se les ha dedicado una sección especial, y más concretamente al estudio de *bufetillo*, por ser un ejemplo que refleja a la perfección el proceso de evolución semántica del término, así como la dificultad de diferenciar entre usos transparentes y lexicalizados del diminutivo, esencial para el desarrollo de la investigación de esta tesis doctoral.

6.5.2.1. Regionalismos y voces poco documentadas

En este apartado se aborda el análisis de los diminutivos lexicalizados que se documentan en los inventarios de *ODE* de Madrid y Almería y que han recibido escasa o nula atención por parte de las fuentes lexicográficas. En estos casos, la documentación analizada ha servido para sacar a la luz voces que han permanecido ocultas hasta el momento o de las que apenas se posee información. Los ejemplos de *albornocillo*, *algodoncillo*, *borloncillo*, *escamilla*, *raspilla* y *apretadillo* constituyen las únicas huellas que hemos podido encontrar de la existencia de estas palabras y para las que *ODE* ha sido la única fuente de la que poder obtener una definición aproximada.

En algunos casos, el corpus atestigua variantes que difieren de las recogidas en los diccionarios, como el uso de las formas en masculino *bastillo*, *granillo*, *horquillo*, por contraposición a las femeninas (*bastilla*, *granilla*, *horquilla*), documentadas lexicográficamente y en las que se asemejan o no en su significado, como se verá más adelante. En otros casos, estas variantes refieren el empleo de un diminutivo distinto, como *confitillo*, del que solo se señala en los diccionarios la forma *confitico*.

Asimismo, el corpus documenta el uso de nuevas acepciones no descritas con anterioridad, como por ejemplo para los términos *francesita* o *manilla*, y aporta nuevos datos que contribuyen a ampliar la información que se conoce de algunas voces, como ocurre en *batistilla*.

En la mayoría de los casos la falta de documentación de los términos coincide con la reducida extensión geográfica de estos, algunos de los cuales han sido considerados voces regionales por su localización limitada exclusivamente a una zona geográfica concreta. De esta manera, se analiza el carácter dialectal de las voces *bastillo*, *apretadillo*, *horquillo*, *medianillo*, y *pajecillo*, y en una de las acepciones de *manilla*, según los datos que ofrecen los corpus consultados. Estas palabras se han marcado en el glosario con la etiqueta *reg.* (§ 6.5.3. *Glosario de lexicalizaciones*) a imitación del criterio que se establece en el *DEA* para palabras que no son de uso general. Esta marca señala que la palabra se emplea en un ámbito geográfico limitado, de diversa extensión, sin más especificación de cualquier otro tipo (*DEA*, *Guía del lector*, p. xv).

A continuación, se analizan detalladamente cada uno de estos términos en orden alfabético, algunos de los cuales se analizan conjuntamente por su afinidad.

Algodoncillo, albornocillo y borloncillo

Los términos *algodoncillo*, *albornocillo* y *borloncillo* son voces de las que no se ha encontrado correspondencia en otras fuentes históricas y de las que únicamente nos consta el único caso que de estas tres voces se documentan en *ODE*. De todos ellos únicamente se puede decir que se trata de telas semejantes al albornoz, al algodón y al borlón respectivamente, sin que los ejemplos del corpus ayuden a una definición más precisa:

- (36) a. Mas dos colchas de *borloncillo* traydas (*ODE* – Cáceres, 1666).
 b. Una ttoalla de *algodoncillo* (*ODE* – Madrid, 1701).
 c. Una mantilla p[ar]a ombros de baieta negra. Cinco baras de *albornoncillo* negro. Una senefa de olandilla encarnada en dos pedazos (*ODE* – Madrid, 1711).
 d. Dos colchas blancas de *borlonzillo*, nuevas (*ODE* – Almería, 1721).

Bastillo

Del término *bastillo* documentado una sola vez en un inventario almeriense de *ODE*, las fuentes lexicográficas únicamente recogen la forma en femenino (*bastilla*), la cual se define en el *DLE* como “doblez que se hace y se asegura con puntadas, a manera de hilván menudo, a los extremos de la tela para que esta no se deshilache” (s.v. BASTILLA). Esta definición se mantiene desde *Autoridades*, donde se añade que la palabra “es término de sastres y costureras”. El ejemplo en masculino que se recoge en *ODE* también se relaciona con las labores de costura, pero no parece coincidir con el significado que dan los diccionarios para la variante en femenino:

- (37) Un [sic] canttarera, un belador y tres candiles y unas debanaderas y un *bastillo* de lino (*ODE* – Almería, Serón, 1715).

Por el contexto, inventariado junto a unas devanaderas⁸⁵, se ajusta más a la definición que da el *TLHA* como “rastrillo del telar”. La voz en masculino, por tanto, alude al instrumento utilizado en la preparación del lino para su posterior confección y se circunscribe a la región de Almería según los dos casos documentados en el *ALEA*⁸⁶ y el inventario de 1715 de nuestra documentación. Con esta motivación se ha señalado como un regionalismo en el glosario y se toma el criterio del *TLHA* para su definición.

Batistilla

Del término *batistilla* no se ha encontrado ningún resultado en las obras lexicográficas que comprende el *NTLLE*. En nuestra documentación la voz aparece dos ocasiones en

⁸⁵ ‘Armazón de cañas o de listones de madera cruzados, que gira alrededor de un eje vertical y fijo en un pie, para que, colocadas en aquel las madejas del hilado, puedan devanarse con facilidad’ (*DLE* s.v. DEVANADERA).

⁸⁶ El *ALEA* lo documenta en Lúcar, un municipio de la zona cercano a Serón (Tomo IV, lámina 987).

un mismo inventario de Almería y los escasos ejemplos adicionales que se han encontrado de ella en el resto de *ODE* y en *CorLexIn* se localizan en la región colindante de Murcia:

- (38) a. Vna thoalla de *vatistilla* biexa [...] vna camisa de crea p[ar]a muger, mangas de *batistilla*, poco vsada (*ODE* – Almería, Cuevas de Almanzora, 1766).
b. Dos paños de narizes de *batistilla* (*CorLexIn* – Murcia, Cieza, 1661).
c. Vn sobrepelliz de *batistilla* viejo [...] vn paño de afeittear de *battistilla* nuevo con su encage [...] vna toalla de *batistilla* sin estrenar [...] vn alba de *battistilla* fina con encages antiguos vsada. Otra alva de *battistilla* con encage vsada (*ODE* – Murcia, Lorca, 1761).

La falta de más ejemplos sugiere un uso reducido del término y contrasta con la ampliamente documentada *batista*, la forma de la que deriva y de la que es fácil localizar ejemplos en los corpus, así como información en otras fuentes. La *batista* se identifica con un ‘lienzo fino muy delgado’ que comenzó a elaborarse en Francia. Su invención se le atribuye al fabricante Baptiste, de quien tomó el nombre (*DLE* s.v. BATISTA). Para su fabricación se empleaba un hilo de calidad superior y más fino que para el lienzo, dando lugar a un tejido más delgado y blando, por lo que se empleaba para la confección de pañuelos, blusas y prendas delicadas de lencería, las cuales podían bordarse para mayor adorno, según se desprende del *Diccionario histórico de telas y tejidos* (*DHTT* s.v. BATISTA). En efecto, los casos de la voz en nuestros corpus corroboran este uso:

- (39) a. Mas vna camisa de mujer de true guarnizada de batista y encajes en los puños (*ODE* – Madrid, 1745).
b. Otro [pañuelo] de batista con cenefa blanca (*ODE* – Madrid, 1818).
c. Otro pañuelo de seda de la Yndia, nuebo. Otro de batista bordado, nuevo (*ODE* – Madrid, 1834).
d. Tres camisolos de batista con pecheras y vueltas de encage (*ODE* – Madrid, 1843).
e. Un pañuelo de batista bordado (*ODE* – Madrid, 1852).
f. Cuatro pañuelitos blancos de la mano en 16. Otros dos de batista (*ODE* – Madrid, Algete, 1855).

Conviene precisar las características de la *batista* para intentar esclarecer lo que se conocía por *batistilla*. El extendido uso de la *batista* provocó la aparición de otros muchos tejidos derivados de este que diferían en calidad o finura. Probablemente, la *batistilla* fuera uno de estos tejidos derivados de la *batista* del que se puede colegir su calidad inferior por medio del uso del diminutivo. Recuérdese que el diminutivo *-illo* confiere en muchos de los términos lexicalizados del ámbito de las telas el valor semántico de ‘cierto tipo de’, al que se une en ocasiones la consideración de ‘calidad inferior’ (§ 6.5.2. *Características de los diminutivos lexicalizados*).

En consonancia con esta hipótesis, el *DHTT* recoge el término *batistilla* como “tela parecida posiblemente a la *batista*, o de calidad inferior a esta, que se empleaba

para confeccionar pañuelos, y que se vendía en las tiendas vallisoletanas de telas en 1860”. A falta de poder contrastar los lugares donde se comercializaba este tejido, lo que sí podemos confirmar es que la *batistilla* se empleaba también para la confección de toallas y prendas de vestir, y no solo para la elaboración de pañuelos, como se desprende de los ejemplos del término en los corpus (38). Entre las prendas de vestir, destacan además las que eran usadas por los eclesiásticos, concretamente la sobrepelliz y las albas⁸⁷ que se documentan en O.

Por otro lado, no puede delimitarse con certeza la extensión geográfica del término, la cual se vincula a la zona sur oriental de la Península (Almería y Murcia), de acuerdo con los pocos casos que nos proporcionan los corpus. Si bien de *batista* puede confirmarse un uso general, no parece que *batistilla* también lo fuera. El hecho de que el término no haya sido recogido por los diccionarios, —excepto por aquellos más específicos (*DHTT*)—, y tampoco se encuentren numerosos ejemplos en los corpus de inventarios, parece probar que su uso era bastante limitado. Por esta razón se ha marcado como regional.

Confitillo

A pesar de que es una voz general y bien documentada en los corpus, se ha decidido dedicar una mención especial al término *confitillo*, ya que los corpus no reflejan la alternancia de variantes que se señala en los diccionarios académicos. El término se define en la actual edición del *DLE* como ‘labor que presentan algunas colchas’ y se recoge bajo el lema *confitico*. El término no se incluye en *Autoridades* y el primer registro del término en el diccionario académico (concretamente en la edición de 1780) refleja la variación *-ico/-illo*, que se mantuvo hasta la edición de 1803, incluida. Posteriormente, todas las ediciones desde 1817 hasta la de 1992, llegaron a recoger en el lema la variación de los tres diminutivos (*confitico*, *confitillo*, *confitito*). El resto de los diccionarios copiaron este mismo patrón y únicamente el diccionario de Gaspar y Roig (1853) recogió el término como *confitillo*, sin hacer alusión a las otras variantes posibles (*NTLLE*). La actual edición del *DLE* elimina la variación en el término y lo lematiza como *confitico*. Sin embargo, a pesar de que los diccionarios lo catalogan como un término lexicalizado sujeto a cambios en el sufijo, esta alternancia no se refleja en ninguna de las ocurrencias del término en *ODE*, tampoco en *CorLexIn* ni en el *CDH*.

- (40) a. Otra colcha de *confitillos* con rapazejos mui bieja (*ODE* – Madrid, 1703).
- b. Una colcha blanca de *confitillo* nueva (*ODE* – Almería, 1736).
- c. Otra [colcha] de *confitillo* (*ODE* – Madrid, San Lorenzo de El Escorial, 1779).

⁸⁷ *Sobrepelliz*: ‘Vestidura blanca de lienzo fino, con mangas perdidas o muy anchas, que llevan sobre la sotana los eclesiásticos, y aun los legos que sirven en las funciones de iglesia, y que llega desde el hombro hasta la cintura poco más o menos’ (*DLE* s.v. SOBREPPELLIZ). *Alba*: ‘Vestidura o túnica de lienzo blanco que los sacerdotes, diáconos y subdiáconos se ponen sobre el hábito y el amito para celebrar los oficios’ (*DLE* s.v. ALBO, -A, acep. 4).

- d. Otra colcha, labor de *confitillos*, en treinta reales (*ODE* – Madrid, Pinto, 1784).

De esta búsqueda en los corpus mencionados se desprende, por un lado, que el término se registra mucho antes de que fuera incluido por primera vez en un diccionario (1783): el primer ejemplo localizado data del año 1625 (*CDH*). Por otro lado, todos los casos del término reflejan únicamente la variante en *-illo*. Es por esta razón por la que en *ODE* se ha decidido lematizar la palabra como *confitillo* e incluirla también bajo esta forma en el glosario, a diferencia de lo que ocurre en la mayoría de las fuentes lexicográficas. Se ha eliminado, por tanto, cualquier rastro de la alternancia entre variantes, la cual no se constata en los corpus consultados.

Escamilla y raspilla

Escamilla y raspilla son otro tipo de telas que no ha sido posible definir con precisión a falta de más datos que los propios ejemplos localizados en *ODE*. Se incluyen a continuación los dos únicos casos documentados en Madrid de *escamilla* y otros de *raspilla* localizados en Almería, Sevilla y Jaén:

- (41) a. Un jubon de seda azul de *escamilla* nuevo (*ODE* – Madrid, 1806).
b. Un jubon de *escamilla* de color (*ODE* – Madrid, 1815).
- (42) a. Un armador de espolin [...] Otro de *raspilla* (*ODE* – Almería, Tabernas, 1822).
b. 1 guardapie *raspilla* color de pimienta (*ODE* – Sevilla, 1795).
c. Un jugon [sic] de *raspilla* morado (*ODE* – Jaén, Villacarrillo, 1797).
d. Vna almilla de felpa negra estrenada [...] Otra de *raspilla* encarnada en cinquenta. (*ODE* – Jaén, Pozo Alcón, 1797).
e. Huna almilla de cartulina blanca [...] Otra de *raspilla* azul con tafettan encarnado (*ODE* – Jaén, Pozo Alcón, 1797).
f. Vna almilla de terciopelo negro. Otra de *raspilla* negra (*ODE* – Jaén, Pozo Alcón, 1797).

Las ocurrencias de ambos términos en el corpus revelan que eran textiles con los que se confeccionaban, por lo general, los armadores o jubones⁸⁸. Estos tejidos podrían aludir a un tipo de tela fina y delgada, si se considera que es este uno de los significados que habitualmente adquiere el diminutivo *-illo* en la lexicalización de los términos referidos a las telas (§ 6.5.2. *Características de los diminutivos lexicalizados*).

En el ejemplo (41) la *escamilla* se aplica a un jubón de seda, por lo que también podría indicar que se trata de una forma característica que se le daba a la tela, la seda en este caso, quizás urdiendo los hilos formando ligamentos parecidos a escamas, y de ahí

⁸⁸ *Armador* se considera un sinónimo de *jubón* según el *DLE*, este último se define como ‘vestidura que cubría desde los hombros hasta la cintura, ceñida y ajustada al cuerpo’ (s.v. JUBÓN). Asimismo, la *almilla* es también una especie de jubón.

su nombre. Esta idea parece más convincente para el caso de *escamilla*, pero el hecho de que no exista una tela denominada *raspa*, y tampoco *escama*, nos incita a pensar que en estos casos el diminutivo no se refiere a un tipo distinto de tela dentro de las de su clase. También para *raspilla* adquiere más fuerza la idea de que el término se refiera a un tipo de técnica con el que se tejía la tela. En nuestra opinión, ambos términos podrían actuar incluso como sinónimos entre sí y aludir al mismo tipo de técnica, dado que la motivación de sus nombres parece ser la misma y aparecen ligados al mismo tipo de prendas.

La teoría que se ha mencionado es la que nos parece más acertada para la definición de ambos términos, sin embargo, no encontramos información lexicográfica que apoye esta hipótesis en ninguna de las fuentes consultadas que han servido para la elaboración del glosario con los términos lexicalizados incluidos en el corpus (§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*).

Francesita

Tampoco ha sido posible proporcionar una definición exacta del término *francesita* localizado en un inventario madrileño de *ODE*. La forma *francesilla* se encuentra en el mismo corpus, también en Cádiz:

- (43) Vna *francesita* de duroy nueva [...] otra usada de estameña (*ODE* – Madrid, San Lorenzo de El Escorial, 1789).
- (44) Un bestido de tela celeste y plata guarnecida la *francesilla* de encaje de plata. Un traje de cotonia fina. Una *francesilla* de terciopelo negro. Una *francesilla* de raso liso negro (*ODE* – Cádiz, 1775).

Las dos francesitas de (43) aparecen inventariadas entre los objetos que Gaspar Ballesteros entrega a su sobrina como dote por su reciente casamiento y se localizan junto a varios zagalejos, un manteo, manteletas, jubones y una cotilla, todas ellas prendas, la mayoría de abrigo, usadas por las mujeres⁸⁹. Deducimos que la *francesita* es también una prenda de vestir femenina y, más aún, que era una de las piezas del vestido, según se deduce del ejemplo en (44). En este ejemplo se hace un apunte sobre la francesilla, adornada de manera diferente al resto del vestido, del que sería una de sus partes, probablemente perteneciente a la parte inferior de este. A pesar de ello, el término *francesilla* se recoge en el *DHTT* como un tipo de tela. Concretamente se define como ‘tela de algodón a cuadros, a menudo rayada’. No obstante, el significado de

⁸⁹ *Zagalejo*: ‘Refajo que usan las lugareñas’ (*DLE* s.v. ZAGALEJO). *Manteo*: ‘Ropa de bayeta o paño que llevaban las mujeres, de la cintura abajo, ajustada y solapada por delante’ (*DLE* s.v. MANTEO). *Manteleta*: ‘Especie de esclavina grande, generalmente con puntas largas por delante, que usan las mujeres, a manera de chal, para abrigo o como adorno’ (*DLE* s.v. MANTELETA). Enlace al documento en *ODE*: <http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/MD1789I0133.xml>

‘prenda de vestir’ parece coincidir más con el sentido con el que se emplea el término en los ejemplos de corpus señalados, por lo que se ha optado por considerarla como tal. La falta de ulteriores detalles en los inventarios y la ausencia de más información sobre el término en otras fuentes no han facilitado una definición más precisa de la voz en su correspondiente entrada del glosario (§ 6.5.3.4. *Glosario* s.v. FRANCESITA).

(Lienzo) apretadillo y (lienzo) medianillo

Frente al resto de voces analizadas hasta ahora, los términos *apretadillo* y *medianillo* se caracterizan por ser adjetivos lexicalizados a los que se les ha añadido un sufijo diminutivo. Se analizan conjuntamente por esta razón, y también porque forman parte de grupos nominales en los que actúan como modificadores del sustantivo *lienzo*, el cual puede o no aparecer de forma explícita. Se relacionan además semánticamente al tratarse de un tipo específico de lienzo.

En los inventarios analizados, ambas formas están ampliamente documentadas y se localizan exclusivamente en la provincia de Almería:

- (45) a. Vna cortina de lienzo *apretadillo* con varetas y manillas de yerro [...] Vnas naguas blancas de *apretadillo* viejas (ODE – Almería, Fiñana, 1745).
 b. Siete cuartas de *apretaillo* [...] Un cenador de *apretaillo* (ODE – Almería, Abia, 1833).
 c. Dos cavezeras de *apretadillo* nuevas, tramadas de lana y pobladas en 70. Un mullidor nuebo de *apretadillo* tramado en lana con listas pagizas y blancas en 12. Otro mullidor biejo de *apretadillo* [...] Dos cabeceras biejas bacidas de *apretadillo* blancas en 10 (ODE – Almería, Fiñana, 1835).
 d. Dos toallas *apretadillo* en una peseta (ODE – Almería, Fiñana, 1882).
- (46) a. Ytt[em] doze tablas de manteles grandes nuevas de *medio anillo*. Ytt[em] ciento y catorze servilletas de *medio anillo* nuevas de vara de largo [...] Ytt[em] seis paños de mano del mismo *medio anillo* nuevos con rapazejo [...] Ytt[em] seis savanas de *medio anillo* nuevas (ODE – Almería, 1721).
 b. Mas vna tohalla de *medianillo* con puntas (ODE – Almería, Abia, 1726).
 c. Ocho almohadas de lienzo *medianillo* [...] quatro sauanas de *medianillo* [...] una tabla de manteles de *medianillo* [...] dos dosenas de seruilletas de *medianillo* (ODE – Almería, 1736).
 d. Dos almoadas de *medianillo* con botones en las vocas [...] una camisa p[ar]a hombre de *medianillo* y mangas delgadas a medio serbir [...] unos calzoncillos de *medianillo* biexos en un r[ea]l [...] vna camisa de muger de *medianillo* sin coser (ODE – Almería, Cuevas de Almanzora, 1766).

Los ejemplos se refieren a un tipo de lienzo con el que se elabora principalmente la ropa de hogar (cortinas, toallas, manteles, servilletas o sábanas), pero también algunas prendas de vestir como enaguas, camisas o calzoncillos. En el caso del (*lienzo*) *apretadillo*, se alude a un tipo de tela labrada con las hebras más juntas o apretadas,

dando lugar a un lienzo más grueso. El (*lienzo*) *medianillo*, por el contrario, hace referencia a un lienzo de consistencia más fina (Calderón Campos, 2018b, p. 325).

En estos casos las menciones a cualquiera de estas dos formas deben entenderse como parte de un grupo nominal en el que se produce la elisión del sustantivo *lienzo*: “una toalla de Ø apretadillo”. El núcleo se omite la amplia mayoría de las veces, pero no en otras donde se menciona de forma expresa el compuesto sintagmático *lienzo apretadillo* / *lienzo medianillo* (45) y (46). Algunos de los casos en los que se omite el núcleo, este queda sobrentendido por medio de una relación anafórica con antecedente explícito –véanse de nuevo los ejemplos de (45) y (46)–. En la mayoría de los ejemplos, sin embargo, no hay mención del sustantivo *lienzo*, pero es el contexto el que favorece la interpretación del núcleo tácito, a diferencia de los adjetivos sustantivados, que “no requieren del contexto lingüístico o de la situación para ser interpretados” (NGLE § 13.6d). La aparición de estas formas en inventarios de bienes y el propio contexto en el que aparecen favorecen la interpretación de estas formas como ‘lienzo’. Fuera de contexto, la simple mención de “un medianillo” o “un apretadillo” no haría recuperable la forma *lienzo*. Por esta razón no se considera que los ejemplos mencionados sean casos de un proceso de sustantivación en el que *apretadillo* y *medianillo* alteran su categoría gramatical como adjetivos, sino que más bien funcionan como modificadores de un nombre elíptico (NGLE § 13.6d, p. 938). En todos los casos se recupera el núcleo *lienzo*, tratándose de construcciones en las que *apretadillo* y *medianillo* continúan siendo adjetivos y Ø representa un sustantivo tácito que se corresponde con *lienzo*. De esta forma, se trata de lexicalizaciones del adjetivo que se insertan en construcciones nominales en las que hay que suponer un sustantivo que no se expresa, pero se sobrentiende.

En la forma, ya se ha dicho, los términos pueden aparecer acompañados o no del sustantivo al que modifican: (*lienzo*) *medianillo* / (*lienzo*) *apretadillo*. En algunas construcciones el adjetivo puede aparecer introducido por preposición (*lienzo de medianillo* / *lienzo de apretadillo*). El único ejemplo de este caso se localiza en el *CorLexIn*: “Otras dos almohadas llenas, de *lienzo de medianillo*, en once ducados” (Almería, Canjáyar, 1646). En el caso particular de *medianillo*, se localiza la forma *medio anillo* en uno de los inventarios almerienses de nuestro corpus (46), surgida probablemente por etimología popular.

En lo que sigue analizaremos el uso de estos términos en cuanto a su extensión geográfica. Con respecto a *medianillo*, nuestra documentación únicamente restringe su uso a la provincia de Almería y al siglo XVIII: se documentan 31 casos en un total de cuatro inventarios. Su presencia en los corpus *ODE* y *CorLexIn* confirma el carácter dialectal del término a partir de su alta frecuencia en inventarios granadinos y en otros de Córdoba, Málaga, Almería y Jaén (Calderón Campos, 2018b).

Ante la falta de información procedente de fuentes lexicográficas, los ejemplos de corpus nos sirven para considerar esta voz como un regionalismo circunscrito al extremo oriental andaluz. Los casos de *medianillo* en nuestros inventarios de Almería constatan su uso en el extremo de Andalucía oriental y atestiguan su presencia, al menos, hasta 1766. A partir de este momento no se registran más casos en *ODE*. Este hecho podría coincidir con la decadencia que en el siglo XVIII experimenta la industria

de los telares en la zona debido a la importación de telas procedentes de otras partes de la Península e incluso del extranjero (Calderón Campos, 2018b, p. 323).

Con respecto a *apretadillo*, su presencia es mucho más limitada y únicamente se documentan los casos señalados en Almería (once ocurrencias en total). Se trata de una voz sin representación en los corpus y tampoco en las obras lexicográficas. Precisamente su ausencia en corpus complementarios sugiere un uso dialectal. Al igual que *medianillo*, se ha marcado como regional en el glosario y se ha descrito conforme a la información que proporcionan los propios ejemplos de corpus (§ 6.5.3.4. *Glosario s.v. APRETADILLO*).

Granillo

Granillo es también una palabra de escasa documentación lexicográfica que contrasta con la presencia que sí posee en los diccionarios su equivalente en femenino (*granilla*).

Con *granilla* se alude al ‘grano que por el revés tiene el paño’ (*DLE* s.v. GRANILLA) y vendría a ser un tipo de técnica con la que se elabora cierto tipo de tejidos (*DHTT*). No obstante, en el corpus de esta tesis únicamente se registra el término en masculino en un inventario de Madrid:

- (47) Una colcha con guarnicion blanca, bien tratada, en 160. Otra blanca de *granillo* en 120 (*ODE* – Madrid, 1815).

Aunque el contexto de aparición de este ejemplo no sirve de gran ayuda para su definición, lo que sí parece claro es que se trata de un tipo de urdimbre con la que se decora ciertos tejidos. Una posible definición de *granillo* la encontramos en el siguiente fragmento extraído de *Nona*, la novela póstuma del escritor murciano José Selgas:

Las servilletas resplandecían de puro blancas, *olían a limpio* [cursiva en el original] y eran grandes como manteles. Estaban hechas de lino cogido en los bancales del Juncar, hilado en la casa y curado al sol y al sereno, tejido a conciencia, formando esa labor menuda que llaman allí *granillo de trigo*; dos listas azules marcaban en los extremos la anchura de la urdimbre, y en los ángulos contrapuestos había pequeñas iniciales bordadas al traspaso con hilo encarnado (*BVMC*. José Selgas y Carrasco, 1883).

El autor se refiere a *granillo* como “labor menuda” y vendría a coincidir con el sentido con el que se emplea en *ODE* (47), similar al de *granilla*. El uso del masculino no se registra en las fuentes lexicográficas (*NTLLE*), donde la mayoría de las ocasiones se define simplemente como ‘diminutivo de grano’. El *DEA*, en cambio, sí recoge con un significado similar la variante *granito*, la cual adscribe al léxico específico del textil y define como “tejido cuya superficie presenta pequeños puntos de relieve” (s.v. GRANITO). Este sentido es el que mejor se relaciona con los numerosos ejemplos que se han localizado en *CorLexIn* y en otras provincias de *ODE* en los que se muestra además la alternancia *granito* (48) / *granillo* (49)(48). Se recoge aquí una selección:

- (48) Media pieza de mantelería de *granitto* [...] dos manteles de *granitto* [...] ocho servilletas de *granitto* [...] dos docenas de servilletas de *granito* (ODE – Cádiz, 1765).
- (49) a. Quatro manteles de los criados de *granillo*, d estopa y estopilla, de bara y terciá de largo y vara de ancho, más otros ocho manteles para el aparador de *granillo*, de zerro y estopilla, de dos varas de largo y bara y quarta de ancho (CorLexIn – Vitoria, Álava, 1638).
- b. Quatro paños de manos de *granillo*, con sus puntillas, nuevos [...] Yten, quatro baras de lienço de *granillo* ancho de estopa para manteles. (CorLexIn – Guipúzcoa, Oñate, 1617).
- c. Cinco servilletas de *granillo de ceuada*, viejas [...] Otros manteles de *granillo de cebada*, traídos (CorLexIn – Soria, Burgo de Osma, 1612).
- d. Dos piezas de manteles nuevas, digo de servilletas de *granillo de cebada* (CorLexIn – Guadalajara, Atienza, 1641).
- e. Tres tablas de manteles d estopa pequeñas, labor *granillo de trigo* [...] tres servilletas de lienço de *granillo de trigo* [...] Yten, otra pieza de manteles de *granillo de trigo* de diez baras [...] Vna tabla de manteles de *granillo de trigo* nueva [...] Yten, vna colcha de lienço de *granillo de trigo*, con su red en medio y sus puntas al rededor, nueva [...] Yten, siete servilletas de *granillo de trigo* [...] (CorLexIn – Madrid, Torrelaguna, 1658).

Las ocurrencias de *granillo/granito* en los corpus lingüísticos incitan a pensar que el término en masculino hacía alusión a un tipo de labor aplicada a la manufactura de textiles. Una referencia explícita es la que se documenta en el inventario madrileño de CorLexIn con la mención “labor granillo de trigo” (49). El término *granillo* se vincula por tanto al de labor o técnica con la que se entrelazan los hilos. Muy probablemente, por la forma en que se tramaba la tela, se formaban pequeños granos y el resultado era similar visualmente al de una espiga de trigo, razón por la que es frecuente en CorLexIn encontrar la expresión *granillo de trigo*. Este corpus la documenta hasta en once ocasiones en ese mismo inventario de Madrid del que se han mostrado arriba algunos ejemplos. Esta misma relación de semejanza es la que motiva también la aparición de la forma *granillo de cebada* documentada en otros inventarios de Guadalajara y Soria (49). El parecido entre sí de ambos cereales con la labor de costura origina estas formas compuestas aplicadas a la elaboración de las telas, comúnmente del lienzo, según se desprende de las varias ocasiones en las que se menciona en los ejemplos. Sin embargo, esta labor podía también aplicarse a otro tipo de tejidos, como la estopa o la estopilla (49).

Los casos de *granillo* reproducen las mismas estructuras sintácticas que otros términos en el corpus que también significan un tipo de labor, como *confitillo* o *gusanillo*. De ellos hay ejemplos en los que se menciona expresamente su significado: “Una colcha grande, *labor de gusanillo*. Otra colcha, *labor de confitillos*” (ODE – Madrid, Pinto, 1784), aunque son mucho más numerosos los casos en los que los términos se emplean de forma independiente sin que sea necesaria la precisión “labor

de”, la cual queda sobrentendida en el propio significado de la forma lexicalizada con diminutivo.

Con respecto a la extensión diatópica, los ejemplos de corpus no permiten fijar su uso en una zona concreta y la escasa documentación del término no se relaciona, como en otros casos, con el carácter dialectal del término.

Granillo es una voz escasamente documentada en los corpus lingüísticos y sin presencia en los diccionarios en su acepción de ‘labor menuda’. Los datos que de ella se han mencionado aquí nos llevan a confirmar algunas cuestiones semánticas y formales. En cuanto a su significado, los ejemplos de corpus atestiguan su sentido como ‘labor menuda’, lo que ha llevado a elaborar una definición propia de acuerdo con la información que se ha ido recogiendo de las distintas fuentes consultadas. Con respecto a su forma, por un lado, los corpus constatan la alternancia en el uso del diminutivo *granito/granillo*. Por otro lado, su uso se da en construcciones sintácticas como integrante de un grupo nominal en el que se deduce el sustantivo *labor*.

Horquillo

El término *horquillo* no se recoge en las obras lexicográficas académicas (*NTLLE*). La *horquilla*, en femenino, sí se recoge desde *Autoridades* con un significado similar al que se registra en nuestro corpus como instrumento de labranza y de forma parecida a la horca, de la que deriva por sufijación: “horca pequeña, que ordinariamente sirve para colgar y descolgar las cosas que están altas, o para afianzar y asegurar alguna cosa en el suelo” (*Autoridades* s.v. HORQUILLA). Los únicos ejemplos de corpus que se han localizado del término en masculino aparecen en dos ocasiones en *ODE* (50) y una en el *CorLexIn* (51), relacionado en todos los casos con los útiles de labranza e inventariado entre palas, cribas, harneros o rastros:

- (50) a. Tres trillas inservibles en 18. Un allegador en 4. Seis *orquillos* y tres palas en 18. Cuatro cribas anchas y dos angostas en 18. Tres arneros viejos en 9 (*ODE* – Madrid, Estremera, 1836).
b. Vn rastro para llegar las parbas. Diez *orquillos* andados. Seis crivas angosteras. Cuatro arneros. Seis palas. Cuatro rodillos o rastros pa remeter la parba (*ODE* – Madrid, Ambite, 1845).
- (51) Vna çesta vieja de dos asas. Vna pala vieja. Quatro orcas. Vn *orquillo*. Quatro açes de esparto. Quatro hoçes mediadas de segar. Vnas debanaderas viejas (*CorLexIn* – Cuenca, Villamayor, 1635).

Los ejemplos no dejan duda de que se trata de un instrumento para labrar. En los inventarios madrileños de nuestro corpus el *horquillo* se nombra entre los “muebles de

madera”, en el caso de (50), y entre los “aperos de labor y otros trastos” en (50)⁹⁰. Estos casos concuerdan más con la definición que ofrece el *DEA* como “horca agrícola de madera con dos púas” (s.v. HORQUILLO), en el que además se señala como voz de uso regional.

La voz en femenino no aparece en nuestro corpus con el significado de *Autoridades* que mencionábamos arriba. No obstante, sí se registra un caso de *horquilla* con el significado de ‘pieza que utilizan las mujeres para sujetarse el pelo’. Este sentido no se incluyó en el diccionario académico hasta su edición de 1884. La aparición más tardía de esta acepción coincide con el ejemplo madrileño de finales del XIX registrado en *ODE*: “dos *horquillas* de oro, una con perlas y otra idem” (*ODE* – Madrid, Alcalá de Henares, 1890). De esta manera, la alternancia de género (*horquilla*, *horquillo*) tiene en el corpus implicaciones semánticas y, aunque ambos proceden del mismo étimo, aluden a realidades diferentes. Por esta razón se ha optado por incluirlos en el glosario en entradas diferentes: la variante en femenino alude al accesorio para el pelo, cuyo significado se toma del *DLE*; la forma en masculino se registra con el sentido de ‘instrumento de labranza’ y se ha marcado como regionalismo conforme a los ejemplos de corpus y las fuentes consultadas (§ 6.5.3.4. *Glosario* s.v. HORQUILLA, HORQUILLO).

Manilla

Una mención destacada merece el caso de *manilla*, el cual revela en el corpus su uso con dos sentidos distintos, el segundo de ellos, diferente al que se le atribuye lexicográficamente de manera tradicional:

- (52) a. Unas *manillas* de aljofar y granates (*ODE* – Madrid, 1701).
b. Dos bueltas de *manillas* de aljofar menuda y granates (*ODE* – Madrid, 1703).
c. Vnas *manillas* de perlas de boemia [...] dos collares de lo mismo de las *manillas* (*ODE* – Madrid, Torrelaguna, 1717).
d. Dos bueltas de *manillas* de alxofar de genero de medio rostrillo compuestas de mil treszientos y nouenta granos que yualmente se expresa en dha hixuela, pesan veinte y cinco adarmes y terzia y salen de ochozientos y setenta y siete granos y tres cuartos en onza y vale segun su calidad quinientos y sesenta y quatro rs de vn que es su valor (*ODE* – Madrid, 1750).
e. Vnas *manillas* de oro pesan una onza y un adarme (*ODE* – Madrid, 1834).
- (53) Vna cortina de lienzo aprettadillo con varetas y *manillas* de yerro (*ODE* – Almería, Fiñana, 1745).
b. Zinq[uen]ta *manillas* de metal y hierro para cortinas [...] dos atados con treintta y seis *manillas* de hierro y bronze pa cortina [...] una cortina de lienzo delgado con vareta de hierro y *manillas* pa la sala del valcon [...] otra corttina de almarieta con vareta de hierro y *manillas* pa la puerta de la sala q esta a la

⁹⁰ Enlace al documento en *ODE*:

<http://corpora.ugr.es/ode/index.php?action=file&cid=xmlfiles/ode/corpus/MD1845I0139.xml>

d[e]r[ech]a como se entra a la del valcon, a medio servir [...] otra cortina con vareta de hierro y *manillas* de almarieta, a mas de medio serbir pa la puertta de la sala de la solana [...] vna corttina de lienzo delgado con su vareta de hierro, *manillas* y alcayattas [...] otra corttina de almarieta con vareta de hierro, *manillas* y alcayatas remendada [...] una corttina, lienzo delgado, con vareta de hierro, alcayatas y *manillas* [...] otra corttina de almarietta con vareta, alcayatas y *manillas* de hierro (*ODE* – Almería, Cuevas de Almanzora, 1766).

La primera acepción de *manilla* se refiere al abalorio similar a una pulsera. Este significado del término vinculado al ámbito de las joyas femeninas está ampliamente documentado (Herradón Figueroa, 2019; Egido, 2016; Arbeteta, 1998) y aparece descrito en los diccionarios académicos desde *Autoridades*: “el adorno que trahen las mugeres en las muñecas, compuesto de unas sartas que dan varias vueltas, de perlas, corales, granates o otras cuentas. Llámase assí por traherse en las manos” (s.v. MANILLA). En *ODE* se localiza en numerosas ocasiones en Madrid (52) y también en otros puntos repartidos por la geografía en el resto de *ODE* y *CorLexIn* que no se muestran por ser similares a los madrileños.

Los inventarios almerienses de *ODE* recogen el término *manilla* con un significado muy distinto al de ‘joya’ (53): se emplea como sinónimo de ‘anilla’, esto es, cada uno de los aros de donde se cuelgan las cortinas. La analogía entre *manilla* como abalorio y *manilla* para las cortinas surge probablemente por el parecido en la forma de ambos objetos, pero este último sentido no se recoge en ninguno de los diccionarios consultados (*NTLLE*). La búsqueda en el resto de *ODE* y *CorLexIn* arroja dos nuevos casos de esta acepción:

- (54) a. Vna cortina de tafetán carmesí con sus *manillas* (*CorLexIn* – Granada, Montefrío, 1662).
- b. Quatro cortinas de veatillas blancas con sus *manillas* y varas (*ODE* – Granada, Baza, 1703).

Estos dos nuevos casos, junto con los ya señalados en dos inventarios almerienses de Fiñana y Cuevas de Almanzora (53), circunscriben esta acepción a las provincias de Granada y Almería, por lo que constatan un uso regional de la voz en estas zonas. Ante la falta de más ejemplos, resulta complicado delimitar la extensión de la voz y corroborar su uso en otras regiones, pero los casos ya presentados permiten confirmar el valor dialectal del término empleado en, al menos, esta parte de Andalucía oriental. Por tanto, el corpus no solo nos permite corroborar la ambivalencia semántica de *manilla*, sino sacar a la luz una acepción que hasta ahora no había sido descrita y que parece confirmar su uso claramente dialectal.

Pajecillo

Pajecillo se identifica con un marcado carácter dialectal en Andalucía. El término se relaciona con *bufetillo/bufetico* en su acepción de ‘mesa auxiliar donde se coloca la luz’

(§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). *Pajecillo* es claramente una voz lexicalizada, pero la preferencia por el diminutivo *-illo* para formar la lexicalización, no pareció estar tan clara hasta un periodo más bien tardío (siglo XIX), según se infiere de las fuentes consultadas.

De acuerdo con los casos que se registran de esta voz en el *CorLexIn*, la variación *-ito*, *-illo* se documenta en el XVII en documentos andaluces: *pajecillo* en Cádiz y Huelva, y *pajecito* en Sevilla (Morala, 2015, p. 158). A comienzos del XVIII, el término tampoco parece estar del todo fijado: los diminutivos se emplean indistintamente del lugar de origen. En *ODE* se localizan ejemplos con *-illo*, *-ito* y también *-uelo*, de nuevo en Sevilla y Cádiz:

- (55) a. Otro bufete chico y un *pagesillo* (*ODE* – Sevilla, 1703).
b. Otro bufetico q[ue] llaman *pajezuelo* forrado de felpa encarnada guarnezido con galon de oro en ziento y zinquentta r[eale]s (*ODE* – Sevilla, Écija, 1706).
c. Un *pajesito* de el belon en quatro p[eso]s (*ODE* – Cádiz, Conil, 1710).
d. Un vufete grande, un *pajesillo* y un escriptorio (*ODE* – Cádiz, Villamartín, 1748).

Las fuentes lexicográficas contemplan *pajecillo* como un término plenamente lexicalizado, pero dejan constancia de la alternancia del sufijo en el propio lema. Así, todavía en algunos diccionarios del XIX se sigue reflejando la variación *-ico*, *-illo*, *-ito* en las entradas del término (RAE, 1832-1869; Salvá, 1846; Domínguez, 1853), según se desprende de la consulta del *NTLLE*.

La ausencia de otras ocurrencias de la voz en el corpus impide determinar con mayor exactitud el alcance cronológico de esta variación. El siguiente ejemplo del término es el que aporta el subcorpus de Almería, el cual se documenta a comienzos del siglo XIX con el diminutivo *-illo*, variante que acabará por afianzarse:

- (56) It[em] un *pagecillo* en treinta (*ODE* – Almería, 1806).

En lo semántico, *pajecillo* pertenece al conjunto de objetos que en los inventarios de bienes se suele incluir entre el mobiliario relacionado con el dormitorio o junto a mesas y sillas. En el *DLE*, el término se describe en su primera acepción como ‘mueble en que se pone la palangana’. Así, *pajecillo* vendría a significar lo mismo que *palanganero*, siendo ambos sinónimos. De hecho, en numerosas obras lexicográficas, incluidas las ediciones del diccionario académico hasta la de 1970, el término siempre ha remitido a *palanganero* como recurso para su definición (*NTLLE*).

Adicionalmente, *pajecillo* cuenta con una segunda acepción marcada en el diccionario como un andalucismo léxico: ‘bufete pequeño en que se ponen los velones y candelabros’ (*DLE*). Este significado coincide con el empleo de *pajezuelo* en el ejemplo sevillano mostrado más arriba (55) y especialmente en (55). En ellos se asocia de manera explícita la voz *bufete* con *pajecillo*, pudiendo ser este un tipo específico de bufete. Más aún, *pajecillo* se encuentra estrechamente relacionado con el ya también lexicalizado *bufetillo/bufetico*, por lo que, en lo que a esta acepción se refiere,

podríamos encontrarnos ante dos términos que en Andalucía actuaban como sinónimos: *pajecillo* y *bufetillo*. Así, el *pajecillo* en esta región se asocia con el *bufetillo* o *bufetico* de luces, como demuestran los ejemplos recogidos con anterioridad y que han servido para confirmar el carácter dialectal de esta voz.

No obstante, este uso dialectal del término no impide que en Andalucía también se utilice *pajecillo* con el valor de *palanganero*. Con este sentido se registra una ocasión en Málaga, según los datos del ALEA⁹¹.

En el ejemplo almeriense citado arriba (56) la ausencia mayores precisiones dificulta vincular el uso de *pajecillo* a una u otra acepción. El objeto aparece enumerado entre una cama de tablas y un velador, por lo que nada permite asegurar que podría tratarse de un *palanganero* y no del mueble donde se coloca la luz, o viceversa. Sin embargo, lo relevante de este ejemplo es su localización geográfica en una región oriental como es Almería, pues hasta el momento el término *pajecillo* únicamente se había localizado en documentos andaluces occidentales (Morala, 2015). Adicionalmente, la consulta del término en ODE ha servido para constatar la fuerte vacilación en el uso del diminutivo incluso en el siglo XVIII.

6.5.2.2. La variación del diminutivo en voces opacas: el caso particular de «*bufetillo*»

No son frecuentes en el corpus las voces que presentan variación en el uso de los diminutivos aun cuando estas se asocian con significados ya lexicalizados. En algunas palabras, las distintas fuentes consultadas dan muestra de cómo se ha producido la variación del diminutivo en algún momento de la historia de estas palabras, aunque esta alternancia no se constata propiamente en nuestro corpus. Esta circunstancia se mencionaba para los casos de *confitillo*, *francesilla*, *granillo* o *pajecillo*, analizados en el apartado anterior. En cambio, el corpus sí recoge la variación de las voces lexicalizadas *puntilla/puntica* (§ *Puntilla*) y *bufetillo/bufetico/bufetito*.

La variación en el uso del diminutivo se ha relacionado con un menor grado de fijación léxica, motivo por el que esta no suele ser habitual en la gran mayoría de las voces completamente lexicalizadas atestiguadas en el corpus. Por el contrario, en las palabras que sí reflejan esta alternancia en el sufijo se descubren fases previas a la fijación definitiva del término, de modo que se pueden establecer diferentes niveles de afianzamiento en los términos que se consideran lexicalizados, prueba de un proceso vivo de cambio lingüístico⁹².

En el apartado § 6.3.2. *Casos dudosos y problemas para la cuantificación*, quedó demostrado que determinar el grado de lexicalización de algunas de las ocurrencias con diminutivo es una tarea compleja que implica examinar diversos aspectos con mayor profundidad. En este sentido, el estudio que seguidamente se lleva a

⁹¹ ALEA II, mapa 698 “palanganero”.

⁹² Esta cuestión se aborda con detalle en § 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo*, donde se explicó cómo la variación del sufijo forma parte de las diferentes fases del proceso de lexicalización.

cabo de *bufetillo/bufetico/bufetito*, y al que se le ha querido otorgar cierto protagonismo, pone de manifiesto la enorme dificultad que entraña diferenciar en algunos casos el valor apreciativo del valor léxico del diminutivo. El caso concreto de esta voz ha supuesto un especial interés, pues esta clasificación supuso un problema teórico difícil de resolver para el que fue necesario aplicar un mayor nivel de análisis. Con este fin se ha querido desarrollar un estudio meticuloso que implica la ardua tarea de tener que distinguir entre usos diminutivos transparentes de aquellos que no lo son, pero que al mismo tiempo es trascendental de cara al análisis de los diminutivos en esta tesis. De este modo, se ha intentado trazar, no solo la evolución semántica de la palabra, sino también el desarrollo de las distintas designaciones que ha recibido a lo largo de la historia este objeto, entre las que se encuentran los derivados con diminutivo. Su estudio nos sirve para evidenciar la inestabilidad del diminutivo en usos en los que no ha sido fácil asegurar el valor léxico. Ello nos ha llevado a analizar pormenorizadamente cada caso y descubrir una evolución en el uso y las formas del objeto, así como en su nombre: de *bufete* a *bufetillo/bufetico/bufetito*.

Con *bufetillo* se alude a un objeto perteneciente al mobiliario del hogar que gozó de una amplia presencia en todo el territorio hispánico hasta aproximadamente principios del siglo XIX, desde su incorporación como galicismo en el 1557, fecha en la que Corominas fija su primera documentación (*DCECH*, s.v. BUFETE). Por lo general, cumple las funciones de una mesa, aunque se destinó a múltiples y variados usos a lo largo de la historia. Su localización en diferentes estancias de la casa y sus tan distintas formas y tamaños nos hace pensar en un mueble caracterizado por su gran versatilidad.

Llegado a cierto punto, sus usos eran tan diversos que surgió la necesidad de diferenciarlos. Aparecen así las designaciones “bufete de estrado”, “mesa bufete”, “bufete tocador”, “bufetillo”, entre otras. Si en un primer momento un *bufetillo* es propiamente un bufete pequeño, con el tiempo la voz alcanza la especialización y pasará a designar un tipo concreto de bufete. En este momento podemos hablar de *bufetillo* como un término lexicalizado y diferenciado semánticamente del que deriva. En lo que sigue, tratará de darse justificación a esta hipótesis, apoyada en los ejemplos de los corpus que se han analizado. Se emplean preferentemente aquellos procedentes de *ODE* de las provincias de Almería y Madrid (pertenecientes al corpus documental analizado en esta tesis), aunque a lo largo de estas páginas se combinan con los de los corpus complementarios (el resto de *ODE*, *CorLexIn* o *CORDIAM*), especificados en cada ejemplo.

Previo al análisis de *bufetillo*, conviene detenerse en la forma de la que deriva, *bufete*, para entender mejor por qué surgió la necesidad de especialización.

El término *bufete* es una palabra recurrente en los inventarios de bienes de todo el ámbito hispánico. En los textos, es posible localizar bufetes en numerosas partes de la vivienda como integrante del mobiliario de la cocina, el dormitorio, el despacho, la sala de estar o el estrado⁹³. Se trata de un objeto de indiscutida presencia en los hogares de la época, desde los más humildes hasta los más pudientes. Tanto es así que Terreros y

⁹³ ‘Lugar o sala de estar alfombrada donde se sentaban las mujeres y recibían las visitas’ (*DLE* s.v. ESTRADO, acep. 6).

Pando (1786) se refiere a él como “mesa común de cualquier casa”. En los diccionarios, la definición más completa se encuentra en *Autoridades*:

Mesa grande, ò à lo menos mediana y portatil, que regularmente se hace de madéra, ò piedra, mas ò menos preciosa, y consta de una tabla, ù dos juntas, que se sostienen en piés de la misma, ù otra materia. Sirve para estudiar, para escribir, para comer, y para otros muchos y diversos usos. Es tomado del francés *bufet*, que vale lo mismo (*Autoridades*, 1726, s.v. BUFETE).

En cuanto al material, solían ser de madera, mayoritariamente de pino y nogal, aunque también se localizan de cedro, castaño, caoba o roble, así como bufetes de hierro o piedra. En algunas ocasiones, el bufete pasó a ser considerado un artículo de lujo cuando el material del que se fabricaba era de más valor. Se encuentran inventariados bufetes de marfil, jaspe, concha y, mayoritariamente, de plata. En estos casos era un objeto precioso, señal de riqueza y ostentación, por lo que se vio supeditado a los reales decretos que regulaban el lujo (Varela Merino, 2009).

Se encuentran bufetes de todos los tamaños (grandes, pequeños, medianos) a los que a veces también los acompañan sillas o bancos. El bufete común solía constar de uno o varios cajones, gavetas o navetas, en ocasiones, también con cerradura y llave. Estas características se mencionan en muchos de los ejemplos de corpus:

- (57) a. Un *bufete* de plata y caxuela pequeña (*CorLexIn* – México, Guaxocingo, 1622).
- b. El *bufete* pequeño de madera con su cajón (*ODE* – Badajoz, 1643).
- c. Un *bufete* de madera mediano con su cajon y dos bancos pequeños (*ODE* – Badajoz, 1645).
- d. Dos *bufetes* grandes de nogal herrado (*ODE* – Almería, 1721).
- e. Vn *bufete* con los pies torneados (*ODE* – Huelva, 1725).
- f. Ytt[em] el *bufette* de pino con gavetas (*ODE* – Almería, Dalías, 1750).
- Dos *bufetes*, vno grande y otro pequeño, con gavetas, en ocho r[eale]s (*ODE* – Almería, Serón, 1754).
- g. Un *bufete* de nogal de una tabla (*ODE* – Madrid, Pinto, 1784).
- h. Otra mesa de igual madera que la anterior, de vara y quarta, con *vufete*, zerradura y llave, en treinta r[eale]s (*ODE* – Almería, Zurgena, 1825).
- i. Un *bufete* con su gaveta y cerradura, diez r[eale]s (*ODE* – Almería, Zurgena, 1829).
- j. Una mesa *bufete* con pie de abuja (*ODE* – Almería, Escúllar, 1834).

Podemos deducir que por *bufete* se entendía una mesa de calidad y nivel estético algo mayor que una mesa convencional. Precisamente, el bufete como un tipo de mesa se evidencia en las formas “mesa bufete”, “mesa con bufete” y “mesa de bufete” que se localizan en varios inventarios del corpus de Almería. Solía estar hecha de madera u otros materiales más exclusivos y podía tener patas ornamentales (recuérdese la apreciación de *Autoridades* como mesa “más o menos preciosa”). Además, de la consulta de los corpus se desprende que el tener cajones era una característica típica del

bufete, los cuales podían cerrarse con llave. Así, la alusión expresa a los cajones del bufete es algo recurrente en los numerosos ejemplos.

En cuanto a los usos a los que podía destinarse un bufete, las fuentes consultadas revelan una gran diversidad de funcionalidades entre las que destacan su uso para escribir, estudiar y comer. El carácter de mesa portátil se señala en numerosos diccionarios y así queda reflejado en aquellos ejemplos que aluden al bufete como una mesa que se ponía y se quitaba atendiendo a la necesidad. Entre estos usos también se recogen ejemplos de bufetes en los dormitorios que servían para guardar la ropa o en otras estancias como mesas pequeñas de juego (Varela Merino, 2009, p. 742).

Por otro lado, los ejemplos extraídos de los corpus revelan que el bufete, además de mesa auxiliar, también cumplía las funciones de soporte sobre el que se apoyaban otros muebles:

- (58) a. Más, dos *bufetillos* de nogal sobre que se ponen los scritorios de arriba. Más, otro escritorillo pequeño aforrado en baqueta, guarneçido con p[...] verdes que sirbe papeles a la mano, con vn *bufetillo* de nogal sobre que se pone (*CorLexIn* – Madrid, 1649).
b. Mas un *bufete* pie de d[ic]ha papelera, forrado en vadana, los pies de nogal y la tabla de pino (*ODE* – Madrid, 1700).
c. It[em] un escriptorio grande de nogal con zerradura y llave, y por pie un *bufete* de la misma madera, herrado (*ODE* – Almería, 1721).
d. Yt[e]n una mesa con *bufete* en cuar[en]ta y seis (*ODE* – Almería, Tabernas, 1822).
e. Una mesa con *bufete* sin llabe (*ODE* – Almería, María, 1834).

Como puede verse, el bufete podía ser el pie sobre el que se colocan otros objetos como escritorios, papeleras⁹⁴ o contadores; objetos destinados por lo común al almacenaje de papeles y material de oficina. El mueble en su conjunto estaría conformado por el bufete, que actúa como mesa-soporte, y otra parte superior destinada al almacenaje. En todos los ejemplos de (58) el bufete se asocia con la mesa que sirve de soporte del escritorio superior.

Además del uso en el despacho, donde los bufetes y escritorios se dedican al almacenaje de papeles y a las labores de escritura, los corpus atestiguan otros usos y funciones, sobre todo cuando el bufete tiene un tamaño más pequeño. Así, se encuentran bufetes pequeños o bufetillos que ocupan las alcobas y estancias personales, más concretamente, aquellas pertenecientes a las mujeres de la casa. El siguiente ejemplo muestra este cambio en la funcionalidad de los bufetes, ahora usados fuera del ámbito de la oficina para guardar objetos femeninos:

- (59) Vn escritorio de madera de ciprés con sus cajones, viejo, y se alló en vna de las gabetas dél dos pares de medias, vnas de lana, plateadas, de Yngalatierra, y otras de seda berde, traídas. En otra gabeta se alló vna gargantilla pequeña de aljófar y

⁹⁴ Entendida con el sentido, actualmente en desuso, de ‘escritorio, mueble para guardar papeles’ (*DLE* s.v. PAPELERA, acep. 8).

panecilos de açabache. Y en otra gabeta se alló vna estufilla de conejillo, y en los demás no hauía cosa de consideración y se bolbió a cerrar, el qual está sobre vn *bufetillo pequeño* de nogal con su caxón, en el qual no se alló nada (*CorLexIn* – Madrid, 1649).

Tal y como se puede leer, el bufete se destinó también a guardar enseres femeninos como joyas, prendas de vestir, peines, etc. Por consiguiente, figurará en numerosos testamentos y dotes entre los bienes de la mujer. Este nuevo concepto de bufete como mueble de uso femenino coincide con el de tocador y obligará a formular especificaciones para diferenciar los usos de la mesa destinada a la escritura de los que no lo son. De hecho, el concepto de ‘tocador’ se reitera en estos otros ejemplos de nuestro corpus: “una *mesa de bufete* con tocados” (*ODE* – Almería, Tabernas, 1868), y “una *mesita pequeña de tocador*” (*ODE* – Madrid, 1843). Estos casos reflejan que las mesas actúan como un sinónimo del bufete tocador y en ellas se podían guardar los accesorios del peinado y adornos de la cabeza que utilizaban las mujeres.

Volviendo al ejemplo de arriba (59), el *bufetillo pequeño* se asocia nuevamente a la mesa con cajón sobre la que se coloca el escritorio, aunque en esta ocasión se guardan objetos diferentes que nada tienen que ver con las labores de oficina. La variación en el fin al que se destina el bufete comportará también un cambio en el uso del término, pues comienza a aparecer de manera reiterada el empleo del diminutivo *bufetillo* o *bufetico* para aludir a este tipo de bufete en el que podían guardarse objetos cotidianos femeninos. El bufete adquiere entonces un nuevo uso cumpliendo las funciones de tocador, para lo que se empleará el término con diminutivo, considerado ya una lexicalización en los ejemplos de abajo:

- (60) a. Vn *bufetico* de nogal con su cajón, *que es tocador*, y se abrió y se alló lo siguiente: vn espejo mediano de dos puertezillas, vna toalla con sus puntas de gusanillo, vnos paños de labar la cara y vn aril del espejo y otros trastos de mujer de poca consideración (*CorLexIn* – Madrid, 1649).
- b. Vn *bufetillo* pequeño de tаража con su caxoncito, el qual yço abrir y en él se alló vn açerico de prender alfileres echo de avxa de seda de colores; [...] y las piezas de vn juego de damas y vn abanico de tafetán berde y negro; y vn apretador de bidrio con su rascamoño y flor de piedras blancas y berdes; y vnas cintas de lana berdes y paxiças (*CorLexIn* – León, Valderas, 1647).
- c. Vn tocador con su espejo de media vara, su moldura de nogal y un *bufetito* p[o]r pie (*ODE* – Cádiz, 1732)⁹⁵.

Se enfatiza especialmente el ejemplo de (60), que constituye la prueba de lo que venimos argumentando. El cambio en la funcionalidad del bufete ya fue señalado por Corominas: “en esta fecha es ‘tocador’ [se refiere a 1587] y en 1617 significa ‘mesa de comer’, pero a mediados del siglo se aplica a un escritorio” (*DCECH* s.v. BUFETE). Sin

⁹⁵ Este ejemplo constata a la perfección el uso del bufete o bufetito como el de soporte de cajoneras, contadores, etc. que se ha comentado más arriba y que ahora también actúa como mesa-soporte de tocadores y espejos.

embargo, no podemos determinar que estos cambios en el uso estuvieran concatenados y bien delimitados en el tiempo, pues los ejemplos de los corpus señalan que en la misma época convivieron diferentes usos para el mismo objeto, si bien este fue adquiriendo diferentes formas.

Las dataciones exactas que proporciona Corominas, en nuestra opinión, deben tomarse con cautela, pues los casos estudiados no permiten establecer una cronología precisa que ayude a asegurar con certeza los usos que fueron originarios y cuáles surgieron posteriormente. Lo que sí podemos afirmar es que el bufete llegó a significar todas estas cosas (mesa de escritorio, mesa tocador, mesa de comer, etc.). En cuanto a las fuentes lexicográficas incluidas en el *NTLLE*, a partir del último tercio del XVIII, los diccionarios comienzan a definir el bufete con un sentido similar al actual, esto es, como una “mesa que se destina a estudiar, escribir u otros usos semejantes”, de lo que se deduce que el bufete común está enfocado en esta época, más concretamente, a los asuntos de escritorio.

La consideración de *bufetillo* como forma lexicalizada y no como diminutivo es el resultado de una extensa labor de cotejo de diversas fuentes, así como fruto del análisis de los muchos ejemplos que ofrecen los corpus estudiados. En numerosas ocasiones, la falta de un contexto mayor en los inventarios que defina de manera más extensa el objeto ha obstaculizado poder determinar con exactitud si estamos ante un verdadero diminutivo o bien ante una voz lexicalizada. Además de esto, esa falta de detalle en los manuscritos dificulta a veces poder precisar la forma y la finalidad a la que se destinaba en la época un bufetillo frente a un bufete. Aun con todo, sí podemos asegurar que el bufete fue un mueble de uso genérico y multifuncional en el ámbito doméstico, circunstancia que facilitó la aparición de numerosas designaciones que sirvieron para diferenciar unos objetos de otros, entre las que se encuentra *bufetillo*.

Las múltiples funciones a las que podía destinarse un bufete forzaron a precisar el tipo de mueble de que se trataba. Prueba de ello son algunas de las especificaciones que encontramos en los inventarios: “mesa (de) bufete”, “bufetes de cocina”, “bufete pie”, “bufete tocador”, “bufetillo de estrado”, “bufetillo / bufetico”. En lo que sigue nos centraremos en los usos con diminutivo.

El bufete pasó a integrar el mobiliario del estrado, un espacio que ocupaban las mujeres y donde podían recibir visitas y pasar el tiempo. La estancia solía estar alfombrada, con almohadas y cojines para acomodarse en sillones o incluso en el suelo. El estrado podía ocupar un lugar elevado, colocado en una parte de la sala principal o sala de estar de la casa, situándose en estos casos sobre una tarima de madera. Los inventarios de bienes reflejan fielmente el mobiliario de que se componían. De esta manera, entre los elementos textiles, se refieren alfombras y estereras para el suelo, colgaduras para la decoración de las paredes y almohadas o cojines para la comodidad de sus ocupantes, de mayor o menor suntuosidad según el estatus de la familia. Así, encontramos alfombras turquescas, estereras de junco, almohadas de raso y damasco o cojines de terciopelo carmesí, por mencionar algunos ejemplos.

Entre los bienes muebles, el estrado solía estar ocupado por braseros para caldear la estancia y láminas, cuadros o imágenes religiosas como elementos

decorativos. El mobiliario lo componían sofás, bancos, taburetes o sillas, por lo común pequeñas:

- (61) a. Un estrado compuesto de sofá, dos butacas y doce sillas de nogal forradas de damasco de seda encarnado, en buen uso, con fundas blancas; tasado en cuatrocientas pesetas (*ODE* – Madrid, 1877).
- b. Una dosena de laminas de estrado con sus vidrieras (*ODE* – Cádiz, 1702).
- c. Doze halmoadas de estrado con caras de damasco encarnado y asiento de terciopelo del mismo color, nuevas (*ODE* – Almería, 1721).
- d. Tres sillas pequeñas de estrado con asiento de anea, biexas (*ODE* – Almería, Cuevas de Almanzora, 1766).
- e. Una alfombra de estrado vieja y remendada (*ODE* – Cádiz, 1769).

Junto al mobiliario destinado al asiento y los elementos que servían como decoración, era común encontrar otros muebles para guardar cosas. Uno de ellos sería el bufete pequeño, de tamaño menor que el bufete corriente, que pasó a denominarse bajo la forma diminutiva *bufetico* o *bufetillo*:

- (62) a. Un *bufettico* de estrado y un brasero con su copa (*ODE* – Málaga, 1700).
- b. Dos *bufeticos* de estrado forrados en terziopelo encarnado guarnecidos con galon de oro (*ODE* – Sevilla, Écija, 1706).
- c. Vn escriptorito de caova de una bara de largo y poco mas de media de ancho, con zinco nauetas y un *bufetico* de estrado por pie de evano y perfilado de marfil viejo (*ODE* – Madrid, 1715).
- d. It[em] tres *bufeticos* de nogal herrados para estrado (*ODE* – Almería, 1721).
- e. Un bufete grande de nogal con sus hierros. Otro pequeño de lo mismo para estrado (*ODE* – Almería, 1736).

El uso del diminutivo para nombrar al bufete que de manera particular formaba parte del estrado es ya recurrente en el siglo XVII, según se desprende de la consulta del *CorLexIn*, aunque siguen siendo todavía numerosas las designaciones “bufete de estrado”, sin diminutivo. Con este fin comenzará a especializarse hasta alcanzar la lexicalización. El siguiente ejemplo refleja esto mismo y resulta especialmente valioso, pues apuntaría a que la voz con diminutivo se encuentra ya plenamente lexicalizada en la primera mitad del siglo XVIII para designar precisamente a este tipo de bufetes destinados al estrado:

- (63) Un *bufetico* pequeño de los que llaman de estrado forrado en cañamazo, colores blanco y pajizo (*ODE* – Almería, 1721).

Al igual que el bufete, el bufetillo también gozó de diferentes usos. Además de los generalizados bufetillos tocador y bufetillos de estrado señalados arriba, los inventarios reflejan que el bufetillo era considerado en algunos casos un objeto precioso por los materiales de mayor estimación con los que estaba construido, como plata, marfil, palo

santo, ébano o concha, aspecto que lo diferenciaba de otras mesas más ordinarias, realizadas por lo general con madera de pino:

- (64) a. Dos *bufetes* de palo santo y marfil, de estrado (*CorLexIn* – Madrid, 1649).
b. Vn *bufetillo* de nogal para el recado del oratorio (*CorLexIn* – Madrid, 1649).
c. Un *bufetillo* de estrado y dos Niños Jesus de bultto (*ODE* – Málaga, 1700).
d. Vn escaparate de bara y media de alto con quince bidrios xptales; y dentro de el una imagen de N[uest]ra Señora de la Conzepcion de talla, con su corona de plata y su *bufetito* de evano, concha y perfiles de bronze correspondiente a la vrna. Otro *bufetico* de lo mismo que el antezedente y enzima de el un santissimo xpto cruzificado (*ODE* – Madrid, 1715)⁹⁶.

Los ejemplos de (64) desvelan que el *bufetillo* fue también una pieza exclusiva relacionada con el rezo donde se podían guardar objetos vinculados con la oración o sobre los que se colocaban crucifijos o imágenes religiosas.

Otro uso que de estos muebles se señala en los inventarios es el relacionado con la mesa auxiliar en la que se colocaba la luz:

- (65) a. Otro *befetillo* [sic] pequeño de caoba en que se ponen las luzes (*CorLexIn* – Madrid, 1649).
b. Le mandó a esta testigo la d[ic]ha D[oñ]a Ana María q[u]e le alumbrase con una lus que tenía ensima de un *bufete* (*CORDIAM* – Argentina, Jujuy, 1686).
c. Un *bufetico* de luzes cubierto de palo s[ant]o (*ODE* – Madrid, 1711).
d. Vn *bufetico* de luzes de concha perfilado de bronze (*ODE* – Madrid, 1715).
e. Otro *bufetico* q[ue] llaman *pajezuelo* forrado de felpa encarnada guarnezido con galon de oro en ziento y zinquentta r[eale]s (*ODE* – Sevilla, Écija, 1706).

En los anteriores ejemplos la voz *bufetillo* o *bufetico* se considera ya lexicalizada y vendría a designar una mesita auxiliar donde se colocaban objetos de iluminación, tales como luces, velones o candelabros. En este sentido, el concepto de ‘bufetillo’ se relaciona con el de *pajecillo*, al que ya se ha aludido con anterioridad a raíz de su consideración como regionalismo (§ *Pajecillo*). El término era usado en Andalucía para nombrar al bufete cuando su finalidad era hacer de soporte de la luz. El ejemplo de (65) resulta muy interesante por aludir explícitamente a esta particularidad.

Por tanto, de *bufete*, mueble que por lo común cumplía las funciones de escritura en el despacho, se derivó la voz con diminutivo, *bufetillo*, para designar un mueble más pequeño y que cumplía con alguna de las siguientes funciones que se han ido desarrollando:

- a. Como tocador o joyero donde las mujeres guardaban sus pertenencias. En estos casos, el *bufetillo* podía localizarse en el dormitorio femenino.

⁹⁶ Nótese en este ejemplo que en la primera aparición de *bufetito* el diminutivo parece tener valor de aminoración, mientras que en la segunda aparición (*bufetico*) actúa como elemento ya lexicalizado. La variación en el uso del diminutivo se analiza más adelante.

- b. Como parte del mobiliario del estrado, bien como objeto decorativo, bien como elemento utilitario. En esta acepción de soporte podríamos incluir también el uso para colocar imágenes religiosas.
- c. Como mesa auxiliar donde poner la luz.

Como conclusión de toda la cuestión semántica que aquí se ha planteado, podemos afirmar que el término *bufetillo* se especializa cuando designa al bufete pequeño que cumple con alguna de las tres funciones arriba mencionadas, las cuales se diferencian del resto de usos que se adjudican al bufete. Este significado de *bufetillo* coincide con la definición que da *Autoridades* (1726): “dimin. de bufete. En lo literal significa el que es pequeño; pero de ordinario se suele tomar por el que sirve para el tocador de las mujeres, ò para adorno en los estrados”, y supone la primera aparición de *bufetillo* en las fuentes lexicográficas como voz plenamente lexicalizada (*NTLLE*). Sin embargo, en muchos diccionarios posteriores, *bufetillo*, aunque con entrada independiente, seguirá mencionándose simplemente como el diminutivo de bufete (*Autoridades*, 1770; Terreros y Pando, 1786; Núñez de Taboada, 1825; Salvá, 1846; Domínguez, 1853). Tampoco el diccionario académico llegó a incorporar esta voz en ninguna de sus ediciones con alguno de los significados que venimos exponiendo. Solo Toro y Gómez incluye en su *Nuevo diccionario enciclopédico ilustrado de la lengua castellana* (1901) el término *bufetillo* con el significado de ‘arquilla’ (*NTLLE*), quizás como referencia a los casos en los que el *bufetillo* sirve como mueble pequeño en el que se guardan cosas.

Pasando ya a la cuestión puramente lingüística sobre el uso de los diminutivos, a lo largo de estas páginas se ha visto cómo en los ejemplos el uso de los términos con diminutivo alterna indistintamente entre *bufetillo-bufetico* y, en menor medida, también con *bufetito*. Del mismo modo, las formas diminutivas conviven con las de *bufete* sin sufijo. El uso del término ya lexicalizado no evitará que se sigan empleando al mismo tiempo fórmulas especificativas compuestas para aludir al objeto: “bufetico de luces”, “bufetico de estrado”. Este último resulta uno de los más recurrentes en *ODE* y aparece siempre con diminutivo excepto en dos ocasiones. Resulta extraño pensar que el diminutivo no se encuentra lexicalizado en todos los casos de abajo:

, sesenta reales. Dos	bufetes de estrado	chafados, vno grande y	1703 Granada
de la Piedad. Dos	bufetes de estrado	con sus trauas de yerro	1705 Murcia
con zinco nauetas y un	bufetico de estrado	por pie de evano y	1715 Madrid
pta. Ytt, un	bufetico de estrado	de serca de tres quartas	1712 Cádiz
tercia en cuadro. Un	bufetico de luzes	cubierto de palo so.	1711 Madrid
quartas de largo. Vn	bufetico de luzes	de concha perfilado de bronze	1715 Madrid
y doszos reales. Dos	bufeticos de estrado	forrados en terziopelo encarnado guarnezidos	1706 Sevilla
haze veintinuee rrs. Tres	bufeticos de estrado	enbutidos en carai y marfil	1707 Granada
y vn rrs. Dos	bufeticos de estrado	con pies torneados chapados de	1753 Jaén
y un rrs. Dos	bufeticos de estrado	con pies torneados chapados de	1753 Jaén
ocho rs. Ytt vn	bufetillo de estrado	madera de albercoquero nuevo en	1766 Almería
un rl. Ytt un	bufetillo de estrado	, viejo, en qtro	1766 Almería
y veinte reales. Vn	bufetillo de estrado	chapeado de carey, noventa	1720 Sevilla
y vte rs. Un	bufetillo de estrado	nobenta y cinco rs.	1670 Badajoz
pos. Ytt, dos	bufetillos de estrado	embutidos en ocho pessos.	1703 Cádiz
bufette de nogal. Un	bufetico de estrado	y un brasero con su	1700 Málaga
merendero de brocattel. Un	bufetillo de estrado	y dos Niños Jesus de	1700 Málaga

Figura 76: Resultados de la búsqueda ‘bufete/bufetillo de estrado/de luces’ en ODE

En otros casos, el uso del término con el sufijo diminutivo se acompaña además de calificativos para indicar tamaño, lo que supone un indicio claro de lexicalización, especialmente cuando el adjetivo es *grande*. Así, el empleo del diminutivo en los siguientes ejemplos no se relaciona con un valor objetivo de aminoración sobre las dimensiones del bufete, ya que el uso de los adjetivos es un claro indicador de que el diminutivo ha dejado de actuar como tal:

- (66) a. Y un *bufetillo* grande y otro pequeño (ODE – Badajoz, 1644).
b. Y dos *bufetillos* de madera, uno mediano y otro pequeño (ODE – Badajoz, 1646).
c. Yten sinco bufetes grandes, tres pequeños y dos medianos, y otros dos *bufetillos* (CorLexIn – Bolivia, Chuquisaca, 1700).
d. Un *bufettillo* de pino pequeño. Ottro *bufettillo* pequeño de nogal. Ottro bufette de nogal pequeño con caxon, zerradura y llaue. Ottro bufette grande de nogal grande [sic] con attrabesaños de yerro (ODE – Madrid, 1701).
f. Dos *bufetillos* pequeños (ODE – Huelva, 1704).

Adicionalmente, algunos de los ejemplos de arriba, especialmente (66), reflejan la aparición conjunta de bufetes y bufetillos de manera continuada en el mismo inventario. La ausencia de cualquier otra especificación referida al uso podría indicar que la sola distinción entre *bufete* y *bufetillo* tenía implicaciones léxicas y era suficiente para identificar sin equívocos ambos muebles. Por lo tanto, el diminutivo estaría lexicalizado en estos términos. Cabe destacar que esta consideración de los diminutivos analíticos

como lexicalizados puede resultar problemática a la vista de que para algunos autores es precisamente el uso de la adjetivación lo que refleja de manera evidente el valor objetivo del diminutivo (Alonso, 1961; Lázaro Mora, 1999). Esta cuestión se ha tratado más a fondo en § 6.4.2. *Valor semántico del diminutivo en el corpus*.

En el análisis morfológico llevado a cabo en esta tesis, todos los casos del término con diminutivo se han considerado lexicalizaciones. Esta decisión se fundamenta en el hecho de que la mayoría de ellos reflejan un significado distinto al de *bufete* que va más allá de las alusiones al tamaño. No plantean dudas los casos en los que junto al término *bufetillo* se añaden las precisiones como mueble de estrado, tocador, elemento decorativo religioso o mesa para sostener la luz, tal y como se ha ido mencionando en los ejemplos seleccionados hasta ahora. Este tipo de precisiones evidencian que el objeto se destina a lugares y usos que no se corresponden con los habituales del bufete corriente. Asimismo, son inequívocos los casos en los que el término *bufetillo* se identifica con alguna otra denominación del objeto. Al respecto, destacábamos los fragmentos señalados en cursiva, que han servido para interpretar el significado del término en otros muchos casos no tan evidentes.

La falta de mayores detalles en la descripción de algunas de las ocurrencias del término ha hecho imposible asegurar ante qué tipo de bufetillo nos encontramos y si efectivamente se corresponde con un uso o no lexicalizado del diminutivo. Se han considerado por ello lexicalizaciones con el fin de no alterar el recuento de diminutivos transparentes. Esta decisión implica que algunos de los términos en los que el diminutivo simplemente cumple con una función apreciativa referida al tamaño no serían contabilizados. Sin embargo, asumir esta posibilidad no es tan desacertado teniendo en cuenta que muchos de los ejemplos reflejados hasta ahora evidencian, cada vez más, usos lexicalizados del término. Los casos de diminutivos transparentes serían, en nuestra opinión, los menos, además de poco probables, dado el alto grado de especialización léxica que va adquiriendo *bufetillo* y que se demuestra en los ejemplos de corpus. Se añaden a continuación algunos de estos casos sujetos a una doble interpretación en los que el contexto no permite confirmar ni desmentir que se trate de verdaderos diminutivos; a pesar de ello se han considerado lexicalizaciones:

- (67) Más, çinco caretas con ojuela que están en vn *bufetico* y vn escritorio pequeño (*CorLexIn* – Sevilla, 1650).
- b. Vn escritoritto con un *bufettito* (*ODE* – Huelva, 1709).
- c. Vn *bufetito* ordinario de cañamazo de dos terzias de largo (*ODE* – Madrid, 1715).
- d. Mas un bufete pie de d[ic]ha papelera, forrado en vadana, los pies de nogal y la tabla de pino. Mas otro *bufetillo* de nogal pequeno de vara menos, sesma de largo y dos terzias de ancho. Mas otro *bufetillo* de nogal, menor q el de arriba con su cajon (*ODE* – Madrid, 1700).
- e. Un arca pequeña vieja. Vn *bufetico*. Vna mesica. Vna artesa. Cinco sillicas (*ODE* – Almería, Roquetas de Mar, 1748).

Con respecto a la variación en el sufijo, se ha analizado la frecuencia de aparición de las distintas formas del término en los corpus. Para el siglo XVII se han contabilizado los

casos presentes en el *CorLexIn* (Tabla 17). Los datos del total de *ODE* se han utilizado para el siglo XVIII (Tabla 18), mientras que para el XIX este corpus no registra ningún caso. En el momento de redacción de estas líneas, los casos de *bufete*, así como los de las formas con diminutivo, se recogen en los corpus conforme a la siguiente distribución⁹⁷:

*FO: frecuencia observada

S. XVII	Sin diminutivo	Con diminutivo		
	<i>bufete</i>	<i>bufetillo</i>	<i>bufetico</i>	<i>bufetito</i>
N.º casos (FO)	331	45	21	2
Total (FO)	331	68		

Tabla 17: Frecuencia de bufete y sus derivados con diminutivo en el siglo XVII (*CorLexIn*)

S. XVIII	Sin diminutivo	Con diminutivo		
	<i>bufete</i>	<i>bufetillo</i>	<i>bufetico</i>	<i>bufetito</i>
N.º casos (FO)	182	22	23	7
Total (FO)	182	52		

Tabla 18: Frecuencia de bufete y sus derivados con diminutivo en el siglo XVIII (*ODE*)

Estos recuentos constatan que el bufete sigue estando muy presente en ambos siglos y que las formas en *-illo* son las más empleadas, sobre todo en el XVII. Sin embargo, alternan con las de *bufetico*, muy especialmente en el siglo XVIII, casi en igualdad de condiciones. Por otro lado, la aparición de *bufetito* es prácticamente anecdótica.

Si se tiene en cuenta la variable geográfica, estos mismos datos reflejan que en el siglo XVIII se encuentran *bufeticos* en áreas donde predominaba el uso del sufijo *-illo*⁹⁸. Centrándonos en las provincias de interés para nuestro estudio, las once ocurrencias del término en Madrid se distribuyen entre *bufetico* (5 casos), *bufetillo* (4 casos) y *bufetito* (2 casos). No sorprende en cambio la presencia esperada de *-ico* en Almería, donde, por otro lado, también se localizan *bufetillos* casi en la misma proporción. Las seis ocurrencias del término en esta provincia se distribuyen de manera muy igualada entre *bufetico* (3 casos) y *bufetillo* (2 casos). Por tanto, la presencia de *bufeticos* en Madrid confirmaría la lexicalización del término en esta región. Esta hipótesis coincide con el planteamiento de Perdiguero (2015), para quien el uso del diminutivo *-ico* en algunas palabras ha perdido su valor de aminoración incluso en zonas donde el empleo de este

⁹⁷ Debe tenerse en cuenta la diferencia de tamaño de los corpus contrastados: 800 000 palabras (*CorLexIn*) y 314 700 palabras (*ODE*). Cifras aproximadas.

⁹⁸ Especialmente en Madrid. El análisis en esta tesis de los diminutivos transparentes (§ 7.4.3.3. *Distribución geográfica*) refleja el mayoritario uso de *-illo* en el siglo XVIII frente al uso de *-ico* en Almería durante el mismo periodo.

sufijo no es un rasgo característico. Y señala precisamente el caso de *bufetico* en un inventario segoviano como prueba de lexicalización (Perdiguero, 2015, p. 1571).

El análisis de los casos demostraría que el uso de un sufijo u otro no atiende a patrones dialectales concretos cuando se trata de una lexicalización, a diferencia de lo que ocurre cuando estos mismos sufijos actúan como verdaderos diminutivos (Véase la distribución de los diminutivos transparentes en las provincias de Almería y Madrid analizadas en esta tesis § 7.4. *Resultados*).

Bufetillo y *bufetico* alternan en *ODE* en Sevilla, Cádiz, Huelva o Málaga. En *CorLexIn* se encuentran *bufeticos* en zonas no caracterizadas por el uso de este sufijo, como Madrid, Toledo, Guipúzcoa, y de nuevo en las provincias andaluzas de Huelva, Sevilla, Córdoba y Málaga; todas estas, regiones en las que también se encuentran *bufetillos*. Estos casos podrían indicar que la variable geográfica no es determinante para la elección del sufijo en el caso de los términos lexicalizados.

El sufijo *-ito* juega un papel mínimo en los procesos de lexicalización dada la escasa frecuencia de aparición de *bufetito* en los corpus. Si nos centramos únicamente en los casos de *-ico* e *-illo*, la alternancia entre *bufetillo/bufetico* incita a pensar que nos encontramos en un episodio previo a la fijación definitiva del término, la cual en el siglo XVIII todavía no se había producido. Y no llegaría a producirse nunca, pues la falta de datos en el XIX, posiblemente por la desaparición en este siglo del objeto, nos impide asegurar el triunfo de un sufijo sobre otro⁹⁹.

Como se ha expuesto, la variación en el empleo de los sufijos diminutivos en el caso de *bufetillo*, *bufetico* y *bufetito* juega un papel importante como parte del proceso de lexicalización. No cabe duda de que el empleo del diminutivo comenzó haciendo referencia al tamaño, pues, en efecto, se aludía a un tipo de mueble más pequeño que el bufete común. Pero posteriormente este mueble de dimensiones menores adoptó diferentes usos y su significado llegó a especializarse pasando a designar objetos con distintas funciones y formas a las originarias.

Este estudio respecto del término *bufetillo* y sus variantes pone asimismo de manifiesto la dificultad de determinar en algunos casos la diferencia entre el valor transparente del diminutivo y su uso lexicalizado, trascendental para el análisis morfológico planteado en esta tesis. A pesar de la alternancia en el diminutivo, la cual puede ser motivo de duda en algunos de los ejemplos recopilados, todos los casos de la voz con diminutivo se han considerado lexicalizaciones, lo que ha supuesto la exclusión de estas voces del cómputo numérico analizado estadísticamente (§ 7. *ANÁLISIS ESTADÍSTICO DE LOS DIMINUTIVOS -ICO, -ITO, -ILLO*).

Para terminar, del análisis llevado a cabo en torno a esta voz se recopilan las siguientes conclusiones:

⁹⁹ Resulta llamativa la desaparición en el XIX de *bufetillos* o *bufeticos*, un mueble que había tenido un uso bastante generalizado en periodos anteriores, según los datos de los corpus. En el XIX también desciende drásticamente el uso de bufetes: 39 ocurrencias en el XVIII, frente a las 14 en el XIX (casos solo en nuestro corpus de Almería y Madrid).

1. Los corpus comienzan a desvelar el uso de un tipo concreto de bufete, más pequeño que el bufete común, y circunscrito al ámbito de la mujer. En la práctica, cumple con las funciones de tocador en el dormitorio o como elemento decorativo en el estrado, diferenciándose del bufete ligado al despacho y a las tareas de escritura o como mesa para comer.
2. Este tipo de bufete comienza a especializarse por medio de un sufijo diminutivo (*bufetico* o *bufetillo*), el cual se lexicaliza. Para los usos lexicalizados, en los siglos XVII y XVIII, el sufijo *-illo* es más frecuente que *-ico*, mientras que la presencia de *-ito* es insignificante en estos casos.
3. La variación existente en el uso de los sufijos *-ico* e *-illo* como parte de la lexicalización pone de manifiesto dos fenómenos relevantes:
 - a. Por un lado, esta vacilación en el uso del sufijo revela que no se había producido todavía en el XVIII una lexicalización definitiva del término que confirmase un claro triunfo de *-illo* sobre *-ico*.
 - b. Por otra parte, la variación no parece responder a patrones dialectales concretos, de manera que las ocurrencias de *bufetico* no se adscriben a zonas geográficas donde predomina este sufijo, y lo mismo ocurre con *bufetillo*. Este hecho nos lleva a considerar la última de las conclusiones.
4. La distribución dialectal no parece ser un elemento decisivo en el uso de los sufijos (*-ico*, *-ito*, *-illo*) cuando lexicalizan, los cuales no se comportan igual en el proceso de lexicalización que cuando actúan como plenos diminutivos. Esto explicaría la presencia de *bufetico* en áreas como Madrid, donde *-ico* es prácticamente ausente en el siglo XVIII (§ 7.4.2.2. *Relación estadística entre diminutivo y siglo en el corpus de Madrid*).

6.5.3. Glosario de lexicalizaciones

Todos los diminutivos lexicalizados con presencia en el corpus que se han ido abordando a lo largo de estas páginas se incluyen en el glosario (§ 6.5.3.4. *Glosario*). En total, 100 entradas en las que se recogen las palabras derivadas con sufijos diminutivos (*-ico*, *-ito*, *-illo*) en las que ha tenido lugar un proceso de lexicalización que ha ocasionado la formación de una nueva unidad léxica. De todas ellas se aborda su significado con sus respectivos ejemplos, las diferentes variantes ortográficas con presencia en el corpus y la extensión diatópica de aquellas que no son de uso general.

El glosario da cabida únicamente a los significados con los que se emplean las palabras en el corpus, de manera que una palabra puede tener más sentidos, pero solo se recogen los que se ven reflejados en los ejemplos del corpus. En algunas ocasiones, el

uso que se hace de una misma palabra en los textos ha posibilitado documentar distintas acepciones. En estos casos, las entradas del glosario reflejan los diferentes significados, apoyados documentalmente con las citas del corpus.

El corpus ha permitido trazar la evolución semántica de algunas de las palabras contenidas en él y contribuir a su estudio histórico añadiendo nuevas precisiones que completan o afinan la información bibliográfica que poseemos de determinadas voces. En este sentido, numerosas definiciones se han reelaborado y enriquecido con aportaciones propias en las que se añaden nuevos detalles sobre el material, el uso o la forma de los nombres que se describen. Ello contribuye a un mayor acercamiento a la vida cotidiana que reflejan los inventarios de bienes y al mejor conocimiento y más preciso de los objetos empleados en los siglos XVIII y XIX.

6.5.3.1. Estructura de las entradas

El contenido de cada entrada del glosario se ha estructurado en dos partes: encabezamiento y cuerpo. El encabezamiento se inicia con el lema de la palabra, seguido de sus variantes formales y el número total de casos que se documentan en el corpus. En un tamaño de letra menor, el cuerpo de la entrada del glosario está constituido por la definición del término y los ejemplos del corpus que ilustran el uso de la palabra descrita.

La microestructura de cada entrada del glosario contiene los siguientes elementos por el mismo orden de aparición:

a. Lema

Cada entrada del glosario está encabezada por el lema de la palabra, el cual aparece en negrita y en un tamaño de letra ligeramente superior al resto del texto del artículo. El lema muestra la forma normalizada que registra el *DLE* en su edición actual, sin embargo, se han tomado decisiones en la lematización que difieren de las que se reflejan en los diccionarios, fruto del análisis que se ha realizado sobre algunas palabras en apartados precedentes. Por ejemplo, de algunas palabras el lema recoge la alternancia de formas que existía en la época representada en los textos (*puntilla/puntica*, *bufetillo/bufetico/bufetito*), para otras, la variación de la que informan los diccionarios ya no se refleja en los textos (*vainilla/vainica*, *confitillo/confitico*). Así, el término *confitico* aparece lematizado como tal en el *DLE*, pero todos los casos de corpus reflejan la variante *confitillo*, razón por la que se ha lematizado en el glosario bajo esta última forma. Por su parte, *bufetillo* aparece en *Autoridades*, pero no las numerosas formas en *-ico* documentadas en el corpus y también algunas en *-ito*, que reflejan un significado lexicalizado (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). Esta alternancia se ha reflejado en el lema.

b. Formas ortográficas

Debajo de cada lema se recogen en cursiva las diferentes variantes ortográficas con que la palabra se documenta en el corpus. Estas incluyen todas las formas flexivas de la palabra, ya sean en singular o plural.

c. Número de casos

De cada lema se indica el número de casos en los que la voz aparece en el corpus. Este recuento incluye el total de veces que la palabra se incluye en los textos independientemente de su variante ortográfica o de que alguna de sus formas se repita varias veces. Este dato aparece en línea aparte y bajo la abreviatura *n.º casos*.

d. Remisiones internas

Parte de la información contenida en las definiciones es consecuencia del análisis detallado que se ha realizado de algunas palabras a lo largo de todo este capítulo. La definición, por tanto, es tan solo el resultado de un estudio más profundo en el que se desarrollan algunos aspectos de cada término. No se ha querido dejar de hacer referencia en el glosario a estos apartados en los que se analizaron en detalle cuestiones que han resultado ser relevantes para la definición de las palabras. El lector que lo desee puede ampliar la información acudiendo a esos apartados para consultar, por ejemplo, cuáles son los factores que nos han llevado a marcar una palabra como regional o a modificar la definición que establece determinado diccionario.

Las remisiones internas se incluyen entre paréntesis y bajo el símbolo §, seguido del título de la sección en esta tesis donde se ha abordado con detenimiento el estudio de la palabra correspondiente.

e. Definición

Está encabezada por la referencia al género de la palabra mediante las abreviaturas *m.* o *f.* y al número, en caso de las palabras empleadas generalmente en plural (*pl.*). Para la definición de los términos se tienen en cuenta los significados que han proporcionado las fuentes lexicográficas, sobre todo académicas, a lo largo de la historia, pero también la información que aporta el contexto de los casos en el corpus y que en ocasiones es determinante para precisar las características de los elementos que se describen.

Las definiciones del glosario aglutinan la información que proporcionan las distintas fuentes consultadas y se apoyan principalmente en las ediciones del diccionario académico: desde *Autoridades*, coetáneo al marco temporal contemplado en el corpus, hasta la edición actual del *Diccionario de la lengua española*, el *DLE*. La referencia a las fuentes consultadas se muestra en las propias entradas del glosario de forma abreviada y entre paréntesis. Las fuentes que han servido como base, así como el procedimiento que se ha seguido para la redacción de las definiciones, se especifican con mayor detalle en el apartado siguiente (§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*).

f. Citas o ejemplos del corpus

Cada definición va acompañada en línea aparte de las citas extraídas del corpus de inventarios de Almería y Madrid y ordenadas cronológicamente de la más antigua a la más reciente. Se incluyen los ejemplos debidamente datados de la palabra en su contexto. Encabeza cada cita el año del documento en negrita, seguido de la provincia (indicada con los códigos AL y MD), y separado por guion el municipio de origen. A continuación, se proporciona la transcripción paleográfica de las citas, en las que no se modifica ningún aspecto gráfico, lo que ha llevado a indicar con [sic] la copia literal de formas que pueden ser consideradas como erratas. Esta marca se añade siempre justo detrás de la forma en cuestión.

La extensión de las citas puede variar en cada caso según haya sido necesario ampliar el contexto para facilitar la comprensión. Algunos fragmentos han sido suprimidos cuando no se consideran relevantes para ilustrar el término correspondiente, lo que se indica con el empleo de [...].

No todos los ejemplos en el corpus para una misma voz aparecen en el glosario. Cuando son muchos o muy repetitivos se incluye solo una selección de los más representativos, los cuales permitan constatar mejor el sentido reflejado en la definición y corroborar el significado expresado en esta. Con la misma intención no se han recogido ejemplos ambiguos en los que no ha sido posible relacionar el sentido con una acepción concreta del término. Un ejemplo de ello es este de la palabra *canutillo*: “Sesenta id[em] [ojas de lata] de *canutillo* en ocho escudos” (ODE – Madrid, Aranjuez, 1870). El significado no parece ajustarse al de ‘hilo de oro o de plata rizado para bordar’ con el que se relacionan otras ocurrencias de la misma voz, por lo que se ha dejado fuera de los ejemplos que ilustran esta acepción.

Para facilitar la lectura de las citas, la palabra que interesa en cada caso se marca en cursiva y los ejemplos correspondientes se delimitan mediante el uso de la figura ▫.

g. Acepciones

En el glosario únicamente se recogen las acepciones que se ven reflejadas en los ejemplos del corpus. En el caso de que la palabra cuente con más de una acepción, tal es el caso de *almohadilla*, *camilla*, *canutillo*, *cepillo*, *tenacillas* o *varilla*, estas aparecen numeradas desde las más general a la más particular, y en párrafos independientes encabezados con el número en negrita. La acepción consta de dos partes: la definición y las citas del corpus. Las citas o ejemplos que se relacionan con cada uno de los significados señalados se incluyen por separado para cada acepción.

6.5.3.2. Fuentes utilizadas para la elaboración del glosario

En este apartado se alude a las fuentes que han servido para elaborar el glosario en cuanto a lo que a sus definiciones se refiere. La totalidad de los ejemplos que se

incluyen en el glosario para ilustrar las diferentes acepciones proceden de nuestro propio corpus documental, por lo que no se incluyen ejemplos tomados de otras fuentes.

En primera instancia, el *Diccionario de Autoridades* (RAE, 1726-1739) se ha considerado como fuente principal por su contemporaneidad con el corpus, siempre y cuando este ha permitido definir con exactitud los objetos que se encuentran inventariados en los textos. En el glosario se incluye la adaptación de las definiciones de *Autoridades*, normalizadas ortográficamente, y con ligeras modificaciones que facilitan la comprensión en algunos casos. En segundo lugar, se ha optado por la consulta de la actual edición del *DLE* (2022) cuando el significado de *Autoridades* no equivale al sentido que revelan los ejemplos del corpus. En estos casos, el glosario refleja la información literal contenida en este diccionario. La preferencia por las definiciones del *DLE* se da en los siguientes supuestos:

- a. Cuando, similares a las de *Autoridades*, las definiciones del *DLE* son más completas y contribuyen a una mejor comprensión de los términos conforme a los contextos en que las voces aparecen en el corpus. Así ocurre en las entradas de *barquillo*, *bolillo*, *bolsillo*, *boquilla*, *casquillo*, *cuartillo*, *cuartilla*, *esterilla*, *tornillo*, *trabilla*, *trencilla*, *vainica*.
- b. Cuando el significado recogido en el corpus no aparece en *Autoridades*, pero sí en el *DLE*: *calzoncillo*, *chaquetilla*, *clavillo*, *coletillo*, *confitillo*, *escardillo*, *jeringuilla*, *marquilla*, *mostacilla*, *picotillo*, *rejilla*, *soguilla*, *sombrilla*, *tenacillas*. Esto ha ocurrido también en otros casos solo para alguna de las acepciones de la palabra, las cuales no se documentan en *Autoridades*, como en *camilla*, *horquilla* o *varilla*.
- c. Cuando *Autoridades* únicamente define el término como ‘diminutivo de’ tal y como ocurre en *barandilla*, *campillo*, *cocinilla*, *toquilla*, *trampilla* y *redecilla*.

A continuación, nos detenemos en la mención de un par de ejemplos que creemos conveniente señalar para ilustrar mejor el procedimiento llevado a cabo en la elaboración de las acepciones de cada entrada en el glosario.

Se ha escogido la definición de una u otra fuente en base al significado que se refleja en el corpus. En algunos casos, se aprecia cómo el significado que muestran las obras lexicográficas se ve modificado de acuerdo con la evolución de los objetos y los usos a que estos se destinan. Así, el término *manguito* refleja en el corpus unas características más acordes con las señaladas en *Autoridades* que con las que actualmente se distingue esta prenda hoy en día. La definición del *DLE* como ‘media manga de punto que usaban las mujeres ajustada desde el codo a la muñeca’, limita el material del que podía confeccionarse esta prenda al punto, mientras que en el corpus es posible localizar además manguitos de piel, de droguete o de seda, lo que le confería mayor calidad a la prenda y la dotaban de un carácter más ornamental. En este caso, la definición de *Autoridades* (“cierto género de manga abierta por ambos lados, hecha de martas o otras pieles adobadas [...] oy se hacen tambien de

pluma, seda y otras cosas, para mayor adorno”) concuerda más con los casos reflejados en el corpus que incluyen manguitos de piel, droguete o los decorados con técnicas como el entorchado o el bordado de cartulina, por lo que para el glosario se toma la definición de *Autoridades* y no del *DLE* (§ 6.5.3.4. *Glosario* s.v. MANGUITO).

Por el contrario, otras veces la definición del *DLE* se ajusta más a los ejemplos de *ODE* que la de *Autoridades*, como ocurre para una de las acepciones de *camilla*. La mayoría de los ejemplos de esta voz, se relacionan con el significado de ‘mesa camilla’, esto es, la mesa que por lo común se coloca en el comedor o salón y lleva una tarima en los pies donde se coloca el brasero:

- (68) a. Vna mesa de pino comedor, grande, en cuar[en]ta. Otra *camilla* para comedor en veinte r[eale]s (*ODE* – Madrid, Ambite, 1845).
- b. Una mesa larga en diez r[eale]s. Yd[em] otra *de camilla* en catorce (*ODE* – Madrid, Algete, 1855).
- c. Una *camilla* de pino con tapete de bayeta verde (*ODE* – Madrid, 1877).
- d. Vna *mesa camilla* con tapete y tablero para brasero, de pino y bayeta verde (*ODE* – Madrid, Alcalá de Henares, 1890).

Estos casos suelen especificar el material del que está fabricada la mesa (por lo general de pino) y solo en uno de ellos se menciona el compuesto *mesa camilla* (68). En los demás casos aparece el término *camilla* sin la designación de ‘mesa’, sin embargo, este significado no aparece en *Autoridades*. En efecto, los ejemplos del corpus, el más temprano de 1845, corroboran la aparición más tardía de esta acepción, la cual no se recoge en los diccionarios académicos hasta la edición de 1992 (cfr. *NTLLE*). Por ello, en el glosario se ha incluido la definición del *DLE* para plasmar este uso.

Otros ejemplos en el corpus se asocian con los significados anteriores que sí incluye *Autoridades*: el de ‘cama portátil’ (69) y el de ‘enjugador para secar la ropa’ (70), este último denominado *camilla* por su parecido con una cama de cordeles:

- (69) Una cama de pino de cinco tablas con sus pies, vieja. Una *camilla* de ruedas tanvien de pino (*ODE* – Madrid, 1703).
- (70) Una *camilla* de pino para enjugar ropa (*ODE* – Madrid, 1832).

Para las dos acepciones señaladas en los ejemplos de arriba se ha preferido la definición que ofrece *Autoridades*. Por lo tanto, para una misma voz, la consulta de un diccionario u otro depende del origen de las acepciones: las más tempranas se toman de *Autoridades*, mientras que para las de origen posterior se recogen las definiciones del *DLE*. Esta misma circunstancia se repite en otras palabras como *bolsillo*, *calzoncillo(s)* u *horquilla*, donde los significados contemplados en el corpus no se recogen aún en *Autoridades*.

En otras ocasiones, las definiciones de *Autoridades*, y en menor medida las del *DLE*, han sido reelaboradas y completadas con la información que nos proporciona el propio corpus o la consulta de otras fuentes. Al respecto ha servido de gran utilidad el *Diccionario histórico de telas y tejidos* (*DHTT*) para la definición de los numerosos

diminutivos lexicalizados pertenecientes al campo de las telas o el *Glosario de términos de la indumentaria regia y cortesana en España: siglos XVII y XVIII (GTIRCE)* para otras tantas palabras referidas a la vestimenta o a los accesorios que formaban parte de esta en la época documentada en el corpus. Ocasionalmente se han añadido también referencias al *Diccionario castellano con las voces de ciencias y artes* de Terreros y Pando (1786) cuando esta fuente aporta algún detalle relevante que completa la definición que se incluye en las fuentes ya citadas, como por ejemplo ocurre para *almohadilla* y *batistilla*. La información contenida en el *Diccionario del Español Actual (DEA)* ha servido también en el caso de otras palabras (*camilla*, *horquilla*, *husillo*). Así, una misma palabra puede ser descrita a partir de la consulta de una o más fuentes, de manera que en el glosario se incluye una definición que integra toda la información de las obras consultadas. Algunos ejemplos de definiciones combinadas son *beatilla*, *bocadillo*, *tontillo*, *picotillo* o *puntilla*, entre otras muchas. En el caso de las voces consideradas regionales, se ha consultado el *Tesoro léxico de las hablas andaluzas (TLHA)* y el *Vocabulario andaluz* (Alcalá Venceslada, 1999) para comprobar su inclusión en estas obras. Solo de algunas palabras se ha podido corroborar la extensión y definición que nos ofrece el corpus (*bastillo*, *manilla* y *pajecillo*).

El propio corpus de inventarios constituye también una fuente valiosa para mejorar o ampliar las definiciones lexicográficas. Por ejemplo, la palabra *botillo* se define en *Autoridades* como “pellejo para echar vino u otro liquor” y en el *DLE* simplemente para llevar vino. Nuestro corpus arroja el ejemplo de un “*botillo* con dos f[anega]s de sal” (*ODE* – Almería, Dalías, 1750), lo que revela que este objeto podía contener otras sustancias además de vino o cualquier otro líquido. La definición en el glosario se ha adaptado, por tanto, para reflejar este uso. En otros casos el corpus aporta mayores detalles en la elaboración o confección de algunos objetos, como en *esportillo*, ‘capacho de esparto o palma’ según el *DLE*, pero que también podía hacerse de paja, según se observa en este ejemplo de un inventario almeriense: “un *esportillo* de paja de centeno” (*ODE* – Almería, Fiñana, 1835).

Otro caso similar es el de *bocadillo*; a partir de los ejemplos de corpus se constató que su empleo era más amplio de lo que recogían las fuentes lexicográficas. Se trata de una especie de lienzo delgado bastante común en la época empleado para la elaboración de cortinas, según se refiere en *Autoridades* y en el *DHTT*. La actual edición del *DLE* también lo relaciona con las telas utilizadas para el cortinaje: “tela delgada colgada entre el visillo y la cortina” (s.v. BOCADILLO, acep. 5). Sin embargo, a partir del corpus podemos saber que también se empleaba para la elaboración de otros textiles del hogar e incluso en la confección de prendas de vestir:

- (71) a. Vnas contramangas de *bocadillo* (*ODE* – Madrid, 1700).
- b. Un pañuelo usado y de *bocadillo* (*ODE* – Madrid, 1701).
- c. Una toalla nueva sin mojar de *bocadillo* guarnezida de encajes de niebe. Dos almuadas de *bocadillo* vien tratadas (*ODE* – Madrid, 1703).
- d. Unas enaguas de *bocadillo* con deshilados (*ODE* – Madrid, 1711).
- e. Tres cortinas de lienzo de *bocadillo* a medio traer (*ODE* – Madrid, 1715).

- f. Dos baras menos quarta de *bocadillo* en ocho reales [...] Vn pañuelo de *bocadillo* hecho del agua en tres reales (*ODE* – Madrid, 1728).
- g. Dos corbatas de *bocadillo* (*ODE* – Almería, Vera, 1730).
- h. Quatro almoadas de *vocadillo* sin moxar con sus encaxes hordinarios [...] Otro deuantal de *bocadillo* sin moxar [...] Tres pares de enaguas de *bocadillo* sin moxar con sus fluecos [...] Una casaca de tafetan negra, forrada en *bocadillo* y las faldillas en tafetan de porzelana (*ODE* – Madrid, 1750).

Se documentan en el corpus pañuelos, almohadas, toallas, delantales y también camisas, casacas, corbatas, contramangas y enaguas, todo fabricado de bocadillo. En *CorLexIn* también se localizan paños, colchas, coetillos y peinadores en documentos en su mayoría de Andalucía (Pérez Toral, 2017). En nuestro corpus, las numerosas ocurrencias en el corpus se circunscriben prácticamente a Madrid (20 casos), con la excepción de un caso en Almería. También en *ODE* la voz se localiza en inventarios de Extremadura, Murcia y distintos puntos de Andalucía, por lo que era un término de uso general.

En otros casos, el corpus ha permitido perfilar las indicaciones que los diccionarios incluyen en cuanto al uso del término. Al respecto se puede mencionar el caso de *coetillo*. Este término no se incluye en *Autoridades*, pero el *DLE* lo define como “corpiño sin mangas, usado por las serranas de Castilla”. A pesar de esta apreciación, el término se documenta en seis ocasiones en nuestro corpus, cinco de ellas en Almería, por lo que su empleo era más generalizado del que se señala en el diccionario¹⁰⁰. De *capotillo*, ‘ropa corta hasta la cintura’, en *Autoridades* se señala que “se usa ya poco o nada”. Sin embargo, goza de cierta vitalidad en el corpus, registrado en doce ocasiones, casi todas en inventarios de Madrid, a excepción de uno.

Nos detenemos a continuación en el caso de *mantilla*, ya que refleja a la perfección cómo el corpus sirve también para constatar el cambio de significado y uso de algunos objetos a medida que pasa el tiempo. El primer significado de esta prenda de vestir utilizada por las mujeres se asocia al de prenda de abrigo. Así se define el término en *Autoridades*: “la cobertura de bayeta, grana o otra tela, con que las mugeres se cubren y abrigan: la qual descende desde la cabeza hasta más abaxo de la cintura” (s.v. MANTILLA, acep. 1). Los ejemplos más tempranos de Madrid y Almería evidencian mantillas fabricadas en su gran mayoría de bayeta, pero también de paño, telas gruesas y abrigadas:

- (72) a. Una *mantilla* de paño encarnado con un encaje de porzelana por guarnizion, forrada en tafetan anteadado (*ODE* – Madrid, 1703).
- b. Una *mantilla* p[ar]a ombros de baieta negra (*ODE* – Madrid, 1711).
- c. Vna *mantilla* blanca de baieta (*ODE* – Madrid, Colmenar de Oreja, 1722).
- d. Vna *mantilla* de baieta guarnezida con liston de color de oro (*ODE* – Madrid, 1724).

¹⁰⁰ El término se documenta asimismo en otras zonas de la geografía peninsular como constata la consulta de *ODE* (en Huelva, Sevilla, Cádiz, Granada y Extremadura) y *CorLexIn* (en Alicante, Murcia y Navarra, entre otras provincias andaluzas).

- e. Vna *mantilla* de baieta blanca con riute negro [...] Vna *mantilla* de damasco azul zelete antiguo con encajes forrada en olandilla (ODE – Madrid, 1728)
- f. Vna *manttilla* de raso verde forrada en olandilla (ODE – Almería, Fiñana, 1745).
- g. Vna *manttilla* de bayeta de Ynglaterra poco vsada (ODE – Almería, Cuevas de Almanzora, 1766).
- h. Una *mantilla* de cristal (ODE – Almería, Fiñana, 1776).
- i. Una *mantilla* de estopilla andada en veinte (ODE – Madrid, 1800).
- j. Una *mantilla* de sarga de seda lisa [...] una *mantilla* de franela con cinta (ODE – Madrid, 1815).
- k. Una *mantilla* de mosolina mediada (ODE – Almería, Fiñana, 1822).
- l. Una *mantilla* bordada en sesenta [...] una *mantilla* fina de bayeta en veinte r[eale]s (ODE – Almería, Tabernas, 1822).
- m. Una *mantilla*, tafetan de Francia con guarnicion de blonda en doscientos r[eale]s (ODE – Almería, Fiñana, 1824).
- n. Vna *mantilla*, lana color negro en diez r[eale]s (ODE – Almería, Zurgena, 1825).
- ñ. Una *mantilla* de clarin en 7. Otra id[em] mediada. Una *mantilla* de seda biexa (ODE – Almería, Abia, 1833).
- o. Una *mantilla* de franela con cinta de terciopelo (ODE – Madrid, Algete, 1834).
- p. Vna *mantilla* de encage de blonda negra nueva (ODE – Madrid, 1834).
- q. Dos *mantillas*, una de blondas con casco rayado y otra lisa con velo (ODE – Madrid, San Lorenzo de El Escorial, 1858).

En las fuentes lexicográficas, la alusión de *mantilla* como prenda destinada al abrigo de las mujeres desaparece en la edición de 1803 y así se mantiene en los diccionarios académicos posteriores. En la edición de 1869 ya se añade a la definición de *mantilla* la posibilidad de llevar guarnición de tul o encaje: “ropa suelta con guarnición de tul o encaje, o sin ella, con que las mujeres se cubren la cabeza y parte del cuerpo” (cfr. *NTLLE*), lo que la convierte en una prenda de adorno y realce para las mujeres y elaborada con tejidos de mayor lustre. En efecto, son mayoritarios los ejemplos datados de finales del XVIII en adelante que aluden a mantillas de seda, blonda, tul, raso, muselina o clarín, todas ellas telas muy finas destinadas más al adorno que como abrigo, y no tanto de bayeta, lo que revelan un uso más ornamental y menos utilitario que comenzaría a generalizarse más tarde, aunque siguieran existiendo aquellas destinadas al abrigo (72). Este uso como prenda decorativa acabará por derivar en el significado actual de la mantilla utilizada en fiestas o actos solemnes, distinción que no se incluye en el diccionario académico hasta la edición de 2001. De esta manera, y dado que la definición de *Autoridades* referida solo a la prenda como abrigo no sería acertada del todo, en el glosario se aplica a esta voz una definición combinada que aglutina ambos usos de acuerdo con los ejemplos del corpus.

En casos como los mencionados, las definiciones que proporcionan los diccionarios se han adaptado con el fin de ajustarse al uso que revelan los ejemplos en el corpus y se señalan con la especificación *adaptada de*. Este tipo de definiciones normalmente constituyen una redacción propia basada únicamente en la información que nos aporta el propio corpus, lo que se indica con la marca *def. propia*.

Se ofrece también una definición propia cuando el término no se recoge en las fuentes lexicográficas o la definición de las fuentes no se ajusta al uso contemplado en el corpus. Así, las definiciones de algunos términos (*algodoncillo*, *albornocillo*, *bocadillo*, *borloncillo*, *cencerrillas*, *escamilla*, *francesita*, *granillo*, *manecilla*, *manilla* –acep. 2–, *medianillo* y *raspilla*) son elaboraciones propias a partir de la información contextual que aportan los inventarios de Almería y Madrid y para las que también ha servido la consulta de los corpus complementarios en algunos casos (resto de *ODE* y *CorLexIn*). Se trata, sobre todo, como se ha visto anteriormente, de voces escasamente documentadas, lo que ha dificultado en ocasiones aportar una definición precisa, la cual no hemos sido capaces de abordar con detalle en los casos de *algodoncillo*, *albornocillo*, *borloncillo*, *escamilla*, *francesita* y *raspilla* (§ 6.5.2.1. *Regionalismos y voces poco documentadas*).

Los casos que se han ido señalando constituyen ejemplos de cómo la documentación de corpus ayuda a perfilar la definición que ofrecen los diccionarios generales de los términos, mejorando así la información lexicográfica de que disponemos. Adicionalmente, el estudio de los casos ha servido para documentar la variación de género o en el sufijo que se produce en algunos términos, información muchas veces no incluida en los diccionarios y que se ha añadido en las entradas correspondientes de este glosario. Este análisis ha servido también para la detección de regionalismos y de voces escasamente documentadas que se marcan en el glosario con la abreviatura *reg.*

6.5.3.3. Abreviaturas y signos empleados

Se recogen en las siguientes tablas las abreviaciones gráficas y marcas que se han empleado en la elaboración de las entradas del glosario:

Abreviaturas y signos	
<i>m.</i>	Sustantivo de género masculino
<i>f.</i>	Sustantivo de género femenino
<i>adj.</i>	Adjetivo
<i>reg.</i>	Voz regional
n.º casos	Número de casos en el corpus
AL	Ejemplo procedente de Almería
MD	Ejemplo procedente de Madrid
§	Remite al apartado que se indica dentro de esta tesis
~	Representa el lema de la entrada
[...]	Supresión de palabras en el ejemplo
def. propia	Definición propia

Tabla 19: Abreviaturas y signos empleados en el glosario

Claves bibliográficas	
<i>Autoridades</i>	<i>Diccionario de Autoridades</i> (1726)
<i>DEA</i>	<i>Diccionario del español actual</i> (2008)
<i>DHTT</i>	<i>Diccionario histórico de telas y tejidos</i> (2004)
<i>DLE</i>	<i>Diccionario de la lengua española</i> (2021)
<i>GTIRCE</i>	<i>Glosario de términos de la indumentaria regia y cortesana en España: siglos XVII y XVIII</i> (2006)
Terreros y Pando	<i>Diccionario castellano con las voces de ciencias y artes</i> , E. de Terreros y Pando (1786)
<i>TLHA</i>	<i>Tesoro léxico de las hablas andaluzas</i> (2000)
Venceslada	<i>Vocabulario andaluz</i> , A. Alcalá Venceslada (1999)

Tabla 20: Claves bibliográficas empleadas en las definiciones del glosario

6.5.3.4. Glosario

A

abanico

abanico, abanicos

n.º casos: 39

m. Instrumento que regularmente se hace de cabritilla o de papel, pegado a unas varillas muy delgadas de madera, concha o marfil, las cuales se unen por el extremo con un clavillo remachado por ambas partes y hecho un pliegue del papel o cabritilla en cada varilla, se abre y cierra fácilmente, y sirve para hacerse aire. También se suele hacer de concha, talco y de otras materias (*Autoridades* s.v. ABANICO).

1724, MD-Madrid: Vn *abanico* de talco nuevo en treinta reales. Otro *abanico* de lamina con barillas de cocha en treinta y siete reales. Otro *abanico* ordinario biexo en seis ▪ **1745**, MD-Madrid: vn *abanico* de Ynglaterra, barillaje de marfil, en ziento y veynte r[eale]s. Mas vn *abanico* franzes maltratado en quarenta r[eale]s ▪ **1750**, MD-Madrid: Un *abanico* de barillas blancas enrexadas [...] Un *abanico* de cabritilla con su varillaxe de hueso en treinta reales. Otro *abanico* de una lamina con su varillaxe

acharolado, bien tratado ▪ **1766**, AL-Cuevas de Almanzora: un *abanico* antiguo con vareta y pie de hueso ▪ **1818**, MD-Madrid: Un *abanico* con guias de azero ▪ **1843**, MD-Madrid: Seis *abanicos* para chimenea ▪ **1844**, AL-Almería: Un *abanico* con pie de nacar. Uno grande de seda. Un *abanico* de cabritilla ▪ **1855**, MD-Algete: un *abanico* de concha moderno ▪ **1890**, MD-Alcalá: Vn *abanico* de plata con nacar con moldura dorada.

albornocillo

albornocillo

n.º casos: 1

(§ *Algodoncillo, albornocillo y borloncillo*)

m. Tela de lana probablemente similar al albornoz, pero de grosor más fino y elaborada con las hebras muy torcidas (def. propia).

1711, MD-Madrid: Cinco baras de *albornocillo* negro.

algodoncillo

algodonzillo

n.º casos: 1

(§ *Algodoncillo, albornocillo y borloncillo*)

m. Tela similar al algodón, probablemente de consistencia más fina (def. propia).

1701, MD-Madrid: Una ttoalla de *algodonzillo*.

almilla

almilla, armilla, almillas, armillas

n.º casos: 31

f. Prenda interior masculina y femenina similar al jubón y ajustado al cuerpo usado por lo común en invierno para reparo y defensa del frío (*Autoridades* s.v. *ALMILLA*) y con mangas o sin ellas (*DLE*). Tiene su origen en la indumentaria militar, usada por los soldados debajo de la armadura y sobre la camisa para evitar roces (*GTIRCE*).

1700, MD-Madrid: Dos *almillas* de cotonia traídas con sus mangas ▪ **1724**, MD-Madrid: Vna *almilla* de sempiterna encarnada, vsada ▪ **1745**, AL-Fiñana: vna *almilla* de raso encarnado viexa ▪ **1745**, MD-Madrid: dos *almillas* de cotonia de mujer de true, guarnezida la vna con gasa ▪ **1774**, AL-Bacares: Una *almilla* de griseta encarnada con punta de plata ▪ **1786**, AL-Cuevas de Almanzora: unos rettales de las *armillas*. Una *armilla* blanca con un cordon morado [...] tres *armillas*, una de espolin de colores, otra de glase y la otra negra de terciopelo ▪ **1833**, AL-Abla: Una *armilla* de sarga biexa. Una *armilla* de felpa biexa ▪ **1840**, MD-Madrid: Dos *almillas* de lana nuevas.

almohadilla

almoadilla, almuadilla

n.º casos: 5

f. Almohada pequeña para ciertos usos, pero más especialmente la que es empleada generalmente para la costura, bien para la labor blanca de las mujeres y costureras, que prenden sobre ella el lienzo de la ropa que cosen y labran (*Autoridades* s.v. *ALMOHADILLA*, acep. 1). Suele tener su caja para meter las cosas necesarias para la labor (Terreros y Pando). También empleada a modo de acerico como cojín pequeño donde

se clavan los alfileres o agujas de costura (*DLE*).

1724, MD-Madrid: Vna *almoadilla* de cañamazo en quinze reales ▪ **1779**, MD-Escorial: Una *almoadilla* de coton en tres r[eale]s [...] Un dedal de plata con su caxa en ocho r[eale]s. Una *almuadilla* de mue en veinte r[eale]s ▪ **1789**, MD-Escorial: vna *almuadilla* de coser con sus tigas ▪ **1825**, MD-Madrid: Una *almoadilla* regalada por d[om]ña Carmen Nestosa en cuarenta r[eale]s.

apretadillo

apretadillo, apretaillo, aprettadillo

n.º casos: 11

(§ (*Lienzo*) *apretadillo* y (*lienzo*) *medianillo*)

adj. reg. También LIENZO ~. Tipo de lienzo en el que la trama de los hilos se realiza de manera que queden más compactos o apretados. Es voz regional, documentada en Almería (def. propia).

1745, AL-Fiñana: Vna cortina de lienzo *aprettadillo* con varetas y manillas de yerro [...] Vnas nagueas blancas de *aprettadillo* viexas ▪ **1833**, AL-Abla: Siete cuartas de *apretaillo* [...] Un cenador de *apretaillo* ▪ **1835**, AL-Fiñana: Dos cavezeras de *apretadillo* nuevas, tramadas de lana y pobladas en 70. Un mullidor nuevo de *apretadillo* tramado en lana con listas pagizas y blancas en 12. Otro mullidor biejo de *apretadillo* [...] Dos cabeceras biejas bacidas de *apretadillo* blancas en 10 ▪ **1882**, AL-Fiñana: Dos toallas *apretadillo* en una peseta.

arillo

arillos

n.º casos: 2

m. Aro pequeño usado como adorno en las orejas (*Autoridades* s.v. *ARILLO*). El término también podía referirse al anillo (*GTIRCE*).

1701, MD-Madrid: Dos pares de pendientes de aljofar con sus *arillos* de

platta sobredorada ▪ **1724**, MD-Madrid: Mas vnos pendientes de aljofar con sus arillos de oro.

arrimadillo

arrimadillo

n.º casos: 2

m. Estera o friso que se pone en las piezas y estrados arrimada y clavada en la pared a fin de evitar las manchas de polvo o yeso (*Autoridades* s.v. ARRIMADILLO).

1766, AL-Cuevas de Almanzora: vn *arrimadillo* de yndiana vsado [...] otra estera q[ue] fue *arrimadillo*, biexa.

B

barandilla

barandilla, varandilla

n.º casos: 3

f. Antepecho de columnas de madera, hierro, bronce u otra materia que se colocaba en la cama para evitar la caída a modo de protección (def. propia).

1721, AL-Almería: una cama de pilarotes y *barandilla* cuya madera no se distingue cual sea ▪ **1766**, AL-Cuevas de Almanzora: vna cama de campo de marca maior, madera de nogal torneada, de *varandilla*, en doszientos r[eale]s [...] una *varandilla* de pino torneada para cama en seis r[eale]s.

barquillo

barquillo, varquillos

n.º casos: 3

m. Especie de canuto pequeño, especialmente el que se hace con harina sin levadura, azúcar o miel y, por lo común, canela (def. propia, adaptada del *DLE* s.v. BARQUILLO).

1711, MD-Madrid: Un *barquillo* p[ar]a colar zera ▪ **1720**, MD-Pinto: Vnos yerros de hazer *varquillos* ▪ **1782**, MD-Madrid: Otro vestido de lo mismo, color de *barquillo*.

barrilla

barrilla, varrilla

n.º casos: 3

f. Hierba delgada que apenas sale de la tierra, cuyas hojitas relucen como plata, las cuales se cogen, y quemadas, de su ceniza se forma un género de piedra, que sirve para el vidrio, y también para el jabón. Se encuentra con abundancia en el Reino de Murcia (*Autoridades* s.v. BARRILLA) y se empleaba antiguamente para blanquear la ropa (*DLE* s.v. BARRILLA, acep. 2).

1748, AL-Roquetas de Mar: Quarenta quintales de *barrilla* ▪ **1750**, AL-Dalías: el oron y simiente de *varrilla* ▪ **1786**, AL-Cuevas de Almanzora: un corcico [sic] pequeño con un poco *varrilla*.

bastillo

bastillo

n.º casos: 1

(§ *Bastillo*)

m. reg. Rastrillo del telar. Es voz regional documentada en Almería (*TLHA* s.v. BASTILLO).

1715, AL-Serón: tres candiles y unas debanaderas y un *bastillo* de lino.

batistilla

batistilla, vatistilla

n.º casos: 2

(§ *Batistilla*)

f. reg. Tela similar a la batista (“lienzo fino muy delgado”, *DLE* s.v. BATISTA), pero más fina y probablemente de menor calidad que esta. Es voz regional, documentada en la zona sur oriental de la Península (def. propia). Se empleaba para confeccionar pañuelos y se vendía en las tiendas vallisoletanas de telas en 1860 (*DHTT*). Se distinguían diferentes tipos de batista según lo tupido de la urdimbre (Terreros y Pando), pudiendo ser uno de ellos el que en esta región se nombraba como *batistilla*.

1766, AL-Cuevas de Almanzora: vna thoalla de *vatistilla* biexa [...] vna camisa

de crea p[ar]a muger, mangas de *batistilla*, poco vsada.

beatilla

beatilla, beatillas

n.º casos: 3

f. Cierta tela de lana, lino o algodón muy delgada y casi transparente (*GTIRCE*). Las beatas y monjas recoletas empleaban este tejido para la confección de tocas, de ahí su nombre, aunque también porque eran empleadas por las mujeres para taparse la cabeza en los actos religiosos (*Autoridades, GTIRCE, DHTT*). Podía tener varias medidas, documentadas según *Autoridades* de vara, de dos tercias y de media vara de ancho (*Autoridades* s.v. BEATILLA).

1722, MD-Colmenar: Dos pares de senaguas de *beatillas* con encajes en treinta y ocho reales ▪ **1724**, MD-Madrid: Vna toalla nueva de *beatilla* de Biscaia en quinze reales ▪ **1791**, AL-Vera: dos cortinas de lienzo de *beatilla*.

bocadillo

bocadillo, vocadillo

n.º casos: 21

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

m. Especie de lienzo delgado que constituye la clase más ínfima entre los lienzos finos y con el que se solían hacer cortinas y visillos para los balcones (*Autoridades*), pero también todo tipo de género de ropa del hogar, como manteles, toallas y pañuelos, así como prendas de vestir (def. propia). Con este término también se denomina hasta finales del XVII a la tela o forro que sobresalía de algunas prendas con el fin de adornarlas (*GTIRCE*).

1700, MD-Madrid: Vnas contramangas de *bocadillo* ▪ **1701**, MD-Madrid: Un pañuelo usado y de *bocadillo* ▪ **1703**, MD-Madrid: Una toalla nueva sin mojar de *bocadillo* guarnizada de encajes de niebe. Dos almuadas de *bocadillo* vien tratadas ▪ **1711**, MD-Madrid: Unas enaguas de *bocadillo*

con desilados ▪ **1715**, MD-Madrid: Tres cortinas de lienzo de *bocadillo* a medio traer ▪ **1728**, MD-Madrid: Dos bars menos quarta de *bocadillo* en ocho reales [...] Vn pañuelo de *bocadillo* hecho del agua en tres reales ▪ **1730**, AL-Vera: Dos corbatas de *bocadillo* ▪ **1750**, MD-Madrid: Quatro almoadas de *vocadillo* sin mojar con sus encaxes hordinarios [...] Otro deuantal de *bocadillo* sin mojar [...] Tres pares de enaguas de *bocadillo* sin mojar con sus fluecos [...] Una casaca de tafetan negra, forrada en *bocadillo* y las faldillas en tafetan de porzelana.

bolillo

bolillos, volillos

n.º casos: 2

1. *m.* Palo pequeño y torneado que sirve para hacer encajes y pasamanería. El hilo se arrolla o devana en la mitad superior, que es más delgada, y queda tirante por el peso de la otra mitad, que es más gruesa (*DLE* s.v. BOLILLO, acep. 1).

1714, AL-Cuevas de Almanzora: Quatro *volillos*. M[edi]a libra de hilo blanco.

2. *m.* Vuelo de gasa o de encaje (*DLE* s.v. BOLILLO, acep. 4).

1703, MD-Madrid: Una cama de nogal de dos cabezeras, torneada, q[ue] la faltan algunos *bolillos* y el alto de colgar.

bolsillo

bolsillo

n.º casos: 3

m. Bolsa pequeña, con una abertura fija, que se cose a una prenda de vestir o se añade a otros objetos (*DLE* s.v. BOLSILLO, acep. 2).

1843, MD-Madrid: Doce pañuelos para el *bolsillo* ▪ **1870**, MD-Aranjuez: un reloj de *bolsillo* en veinte y siete escudos, cien milesimas ▪ **1890**, MD-Alcalá: Cuarenta y seis pañuelos para el *bolsillo*, blancos y de color.

boquilla

voquillas

n.º casos: 1

f. Pieza pequeña que se adapta al tubo de ciertos objetos y que se introduce en la boca para beber, fumar o soplar (adaptado del *DLE* s.v. BOQUILLA).

1766, AL-Cuevas de Almanzora: otros dos cocos con pies y *voquillas* de plata en sesenta r[eale]s.

borloncillo

borlonzillo

n.º casos: 1

(§ *Algodoncillo, albornocillo y borloncillo*)

m. Tela semejante al borlón (“tela de lino y algodón sembrada de borlitas”, *DLE* s.v. BORLÓN), aunque probablemente más fina (def. propia).

1721, AL-Almería: Ytt dos colchas blancas de *borlonzillo* nuevas.

botillo

botillo

n.º casos: 1

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

m. Pellejo pequeño empleado generalmente para llevar vino, pero que servía también para contener otros líquidos o sustancias como sal u especias y de capacidad de una arroba o inferior (def. propia adaptada de *Autoridades* s.v. BOTILLO).

1750, AL-Dalías: la votija de vidrio vale zinco. Ytt[em] el *botillo* con dos f[anega]s de sal del acopio valen zinquentta rr[eale]s.

broquelillo

broquelillos

n.º casos: 6

m. Botón pequeño con el que se adornaban las mujeres las orejas y del que podían colgar piedras, almendras o arracadas (“arete con adorno colgante” *DLE*) (*Autoridades*, *DLE* s.v. BROQUELILLO).

1750, MD-Madrid: Dos arracadas de oro, el reuerso tallado y picado, compuestas de dos *broquelillos*, dos copetes de oxas y seis colgantes [...] Dos *broquelillos* de oro, el reberso liso guarnezidos con diez y siete diamantes rosas y delgados de barios tamaños sin uno que le falta [...] Dos pendientes de alxofar puestos en d[ic]hos *broquelillos* ▪ **1779**, MD-Escorial: Un par de *broquelillos* de piedra de Francia ▪ **1784**, MD-Pinto: Unos *broquelillos* de oro quar[en]ta r[eale]s ▪ **1843**, MD-Madrid: Once pares de *broquelillos* o sean pendientes de oro.

bufetillo, -ico, -ito

bufetico, bufetillo, bufetillo, bufetito

n.º casos: 16

(§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*)

1. *m.* Mueble perteneciente al estrado que podía servir como elemento utilitario en el que guardar cosas o como objeto decorativo como soporte de imágenes y figuras religiosas.

1715, MD-Madrid: Vn escaparate de bara y media de alto con quinze bidrios xptales; y dentro de el una ymagen de N[uest]ra Señora de la Conzepcion de talla, con su corona de plata y su *bufetito* de evano, concha y perfiles de bronze correspondiente a la vrna. Otro *bufetico* de lo mismo que el antezedente y enzima de el un santissimo xpto cruzificado con su cruz de evano y guarnezida la cruz de bronze, feligraneado con unas piedras falsas y unas figuras de marfil en la peana. Vn *bufetito* ordinario de cañamazo de dos terzias de largo ▪ **1721**, AL-Almería: tres *bufeticos* de nogal herrados para el estrado [...] un *bufetico* pequeño de los que llaman de estrado forrado en cañamazo, colores blanco y pajizo ▪ **1766**, AL-Cuevas de Almanzora: vn *bufetillo* de estrado, madera de albercoquero, nuevo en quinze r[eale]s [...]

un *bufetillo* de estrado, viejo, en q[ua]tro r[eale]s.

2. *m.* Mesa auxiliar donde se colocaba la luz, candelabros o velones.

1711, MD-Madrid: Un *bufetico* de luzes cubierto de palo s[ant]o ▪ 1715, MD-Madrid: Vn *bufetico* de luzes de concha perfilado de bronce. Otros dos *bufeticos* de pino pequeños con sus cuviertas de baqueta.

3. Para otras acepciones no documentadas en el corpus documental de esta tesis vid. § 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo».*

C

cabritilla

cabritilla

n.º casos: 5

f. La piel de animales pequeños como cabrito o cordero, la cual se adoba, adereza y colorea. Habitualmente se utilizan las pieles de los cabritos y de ahí su nombre (*Autoridades* s.v. CABRITILLA).

1750, MD-Madrid: Un abanico de *cabritilla* con su barillaxe de hueso ▪ 1834, MD-Madrid: Un abanico de nacar en *cabritilla* ▪ 1844, AL-Almería: Un abanico de *cabritilla* [...] Doce pares de guantes de seda y *cabritilla*.

calzoncillo

calzoncillos, calzocillos, calzonzillos

n.º casos: 38

m. Prenda interior masculina que cubre desde la cintura hasta el muslo, de longitud variable. Usado más en plural (def. propia adaptada del *DLE* s.v. CALZONCILLO).

1700, MD-Madrid: Vna camisa y *calzoncillos* nuevos de lienzo gallego crudo [...] Vnos *calzonzillos* de lienzo gordo [...] Dos pares de *calzonzillos* ya mojados, los vnos con unas puntillas y los otros mas rejios ▪ 1714, AL-Cuevas de Almanzora: Un

par de *calzonzillos* para hombre de lienzo de lino tramado estopa, a m[edi]o servir. Otro par de *calzonzillos* de mismo jenero. Otro par de *calzonzillos* de lienzo de lino, nuevos

▪ 1829, AL-Zurgena: ocho pares de *calzonzillos* blancos en treinta y dos r[eale]s

▪ 1896, AL-Laujar: Cuatro pares de *calzoncillos* blancos. Cuatro de *calzoncillos* elasticos y cuatro camisetas.

camilla

camilla

n.º casos: 6

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

1. *f.* También MESA ~ o MESA DE ~. Mesa, generalmente redonda, bajo la cual suele haber una tarima para colocar el brasero (*DEA, DLE* s.v. CAMILLA, acep. 1).

1845, MD-Ambite: Vna mesa de pino comedor, grande, en cuar[en]ta. Otra *camilla* para comedor en veinte r[eale]s ▪ 1855, MD-Algete: una mesa larga en diez r[eale]s. Yd[em] otra de *camilla* en catorce ▪ 1877, MD-Madrid: Una *camilla* de pino con tapete de bayeta verde ▪ 1890, MD-Alcalá: Vna *mesa camilla* con tapete y tablero para brasero, de pino y bayeta verde.

2. *f.* Cama pequeña y estrecha que se lleva sobre varas a mano o sobre ruedas, para transportar enfermos, heridos o cadáveres (*Autoridades* s.v. CAMILLA, acep. 1).

1703, MD-Madrid: Una cama de pino de cinco tablas con sus pies, vieja. Una *camilla* de ruedas tanvien de pino.

3. *f.* Por semejanza de la cama de cordeles, se llama una rejuela de madera y alambres o cuerdas por encima, donde se pone a enjugar la ropa, que sirve lo mismo que el enjugador (*Autoridades* s.v. CAMILLA, acep. 2).

1832, MD-Madrid: Una *camilla* de pino para enjugar ropa.

campanilla

campanilla

n.º casos: 2

f. La campana pequeña, y regularmente la que se hace para tocarla, tomándola en la mano, y la que ponen por dije a los niños. (*Autoridades* s.v. CAMPANILLA, acep. 1).

1836, MD-Estremera: una *campanilla* pequeña ▪ **1843**, MD-Madrid: 1 escribanía de plata con tres piezas y *campanilla*.

campillo

campillo

n.º casos: 1

m. Lo mismo que *ejido*, campo que no se labra y donde suelen reunirse los ganados o establecerse las eras (*DLE* s.v. EJIDO).

1784, AL-Abrucena: una aza en el *campillo*.

canutillo

canutillo, cañutillo

n.º casos: 2

m. Hilo de oro o de plata rizado que se utiliza para bordar, vistoso y de valor. El bordado que se hace con él recibe también el mismo nombre (*Autoridades* s.v. CAÑUTILLO, acep. 2, *DLE* s.v., acep. 4).

1750, MD-Madrid: Un aderezo de *cañutillo* en treinta reales.

capotillo

capotillo

n.º casos: 12

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

m. Ropa corta a manera de capa y utilizada como abrigo por encima del vestido que llegaba hasta la cintura. De uso general en el siglo XVIII.

1745, MD-Madrid: vn *capotillo* de griseta azul guarnecido de plata forrado en felpa larga. Mas vn *capotillo* guarnecido de raso liso azul ▪ **1750**, MD-Madrid: Un *capotillo* de media grana bordado de blanco ▪ **1765**, MD-Madrid: una manteleta y *capotillo* de bayeta ▪ **1770**, MD-Estremera: un *capotillo* color de zereza de terciopelo con galon de

oro al cantto ▪ **1779**, MD-Escorial: Un *capotillo* encarnado de perzelina [...] Otro *capotillo* de bayeta fina ▪ **1784**, MD-Pinto: Un *capotillo* de mediana encarnado.

casquillo

casquillo

n.º casos: 1

m. Anillo o abrazadera de metal, que sirve para reforzar la extremidad de una pieza de madera (*DLE* s.v. CASQUILLO, acep. 1).

1724, MD-Madrid: Vn palillo de ebano de hazer media con su *casquillo* y cadenilla de plata.

cepillo

cepillo, zepillo, cepillos

n.º casos: 6

1. *m.* Instrumento de carpintería formado por un prisma cuadrangular de madera dura, que lleva embutido, en una abertura transversal y sujeto por una cuña, un hierro acerado con filo, el cual sobresale un poco de la cara que ha de frotar con la madera cuya superficie se quiere alisar (*Autoridades* s.v. CEPILLO, acep. 1).

1784, AL-Abrucena: [...] una azuela, un *zepillo*, un escoplo y otras erramientillas ▪ **1819**, AL-Alboloduy: Un *cepillo* de carpintero ▪ **1870**, MD-Aranjuez: Tres serruchos, un *cepillo* y una azuela en un escudo.

2. *m.* Limpiadera con que se quita el polvo, y limpian los vestidos, hecha de manojitos de cerdas, metidas y apretadas con cola en unos agujeros, ejecutados a proporción en una tabla o zoquetillo, de calidad que queden iguales las cerdas. Por el revés se suele forrar en guadamecil, o otra cosa, que sirve para distintos usos de limpieza (*Autoridades* s.v. CEPILLO, acep. 2).

1766, AL-Cuevas de Almanzora: un *zepillo* en un r[ea]l p[ar]a limpiar poluo ▪ **1855**, MD-Algete: dos *cepillos* p[ar]a la ropa en doce.

cerilla

zerilla

n.º casos: 5

f. Vela delgada y larga para diversos usos, entre ellos para alumbrarse y para que no se deshilen los cortes de las telas de seda, al pasar por ellos la cerilla encendida ligeramente (*Autoridades* s.v. CERILLA, acep. 1).

1711, MD-Madrid: Dos pailas de cobre grandes con asas y la una de ellas con dos ganchos en ella para azer *zerilla* [...] Un yerro de azer *zerilla* [...] Dos tornos con sus pies de pino p[ar]a azer *zerilla*. Un tornillo para azer *zerilla* con dos zigüañas de yerro. Una fuente de pino para azer *zerilla*.

chaquetilla

chaquetilla

n.º casos: 1

f. Chaqueta corta y ajustada, generalmente con adornos (*DLE* s.v. CHAQUETILLA).

1840, MD-Madrid: Una *chaquetilla* de color de la lana y un pantalon verde todo de paño.

cintillo

cintillo, sintillo, zintillo, zinttillo, cintillos

n.º casos: 11

m. Sortija pequeña guarnecida de diamantes u otras piedras con que se adornan los dedos (*Autoridades* s.v. CINTILLO, acep. 2).

1703, MD-Madrid: Una sortija, *zintillo* de oro pulido [sic], guarnecida con un diamante delgado, dos esmeraldas ▪ **1721**, AL-Almería: un *cintillo* de oro y esmeraldas de tres piedras [...] otro *cintillo* de oro con tres puntas de diamante ▪ **1736**, AL-Almería: un *sintillo* de oro con tres diamantes ▪ **766**, AL-Cuevas de Almanzora: un *zintillo* de oro con una esmeralda y una chispilla de diamante en sesenta. Otro *zinttillo* con vna esmeralda puntiaguda ▪

1845, AL-Cuevas de Almanzora: Dos *cintillos* de oro.

clavillo

clavillo

n.º casos: 1

m. Capullo seco de la flor del clavel. Tiene la forma de un clavo pequeño, con una cabeza redonda formada por los pétalos y rodeada de cuatro puntas, que son las divisiones del cáliz, de color pardo oscuro, de olor muy aromático y agradable, y sabor acre y picante. Es medicinal y se usa como especia en diferentes condimentos (*DLE* s.v. CLAVO, acep. 4).

1786, AL-Cuevas de Almanzora: un botecico de oja de latta con zafrán y *clavillo*.

cocinilla

cocinillas

n.º casos: 1

f. Fogón pequeño y portátil (adaptado del *DLE* s.v. COCINILLA, acep. 1).

1870, MD-Aranjuez: Seis *cocinillas* economicas en siete escudos.

coletillo

coletillo, colettillo, colettillos

n.º casos: 6

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

m. Corpiño sin mangas que cubría el torso. Procede del italianismo *coletto* y hacía referencia a las prendas militares italianas de cuero que llegaron a España para sustituir las corazas metálicas de la armadura (*GTIRCE*). Del *coletto*, generalmente hecho de ante, pieles o cuero, se derivó el *coletillo*, confeccionado de telas más finas y empleado también por las mujeres. Lo vestían las serranas de Castilla, pero su uso era bastante más generalizado (*DLE* s.v. COLETILLO).

1728, MD-Madrid: Vn *coletillo* viejo de ante en diez reales ▪ **1745**, AL-Fiñana: Vn

colettillo viexo de raso de conor [sic] de gayumba [sic] [...] Vn *colettillo* de damasco verde mediado ▪ **1750**, AL-Dalías: los *colettillos* de nobleza rrosada, digo almilla ▪ **1819**, AL-Alboloduy: Un *coletillo* de seda en veinte y cinco r[eale]s ▪ **1835**, AL-Fiñana: Un *coletillo* de seda dorado con ramos en 6.

confitillo

confitillo, confitillos

n.º casos: 4

(§ *Confitillo*)

m. Labor menuda que tienen algunas colchas, parecida a los confites (‘golosina pequeña’) (*DLE* s.v. CONFITICO).

1703, MD-Madrid: Otra colcha de *confitillos* con rapazejos mui bieja ▪ **1736**, AL-Almería: una colcha blanca de *confitillo* nueva ▪ **1779**, MD-Escorial: Otra [colcha] de *confitillo* en quarenta y cinco r[eale]s ▪ **1784**, MD-Pinto: Otra colcha, labor de *confitillos*, en treinta reales.

cordoncillo

cordoncillo, cordonzillo

n.º casos: 22

1. *m.* Labor que en algunos tejidos hace la trama, sin formar dibujo alguno y que solo muestra alguna aspereza (*Autoridades* s.v. CORDONCILLO, acep. 2).

1714, AL-Cuevas de Almanzora: Un jubon de *cordonzillo* de tela de Lorca [...] zinco varas de lienzo de cañamo de *cordonzillo* para manteles ▪ **1825**, AL-Zurgena: Vna servilleta de *cordonzillo* á medio vsar en dos r[eale]s. Dos manteles de almarieta de *cordonzillo* á medio vsar en treze r[eale]s [...] Vnos manteles de *cordonzillo*, lienzo de cañamo, en quatro r[eale]s. Tres manteles, lienzo de lino, dos de *cordonzillo* y el otro de labor ▪ **1834**, MD-Algete: tres costales, dos de *cordoncillo* y uno de gerga ▪ **1844**, AL-Cuevas de Almanzora: Nueve servilletas viejas con *cordoncillo* ▪ **1855**, MD-Algete: un vestido de gró de *cordoncillo* en muy buen estado.

2. *m.* Labor realizada en el canto de la moneda para que no puedan ser falsificadas o cercenadas fácilmente para modificar su valor (*DLE* s.v. CORDONCILLO, acep. 2).

1834, MD-Madrid: En dinero contante y sonante en monedas de *cordoncillo* de oro, 25 000.

cortinilla

cortinillas

n.º casos: 5

(§ *Cortinilla*)

f. Cortina pequeña que se coloca en la parte interior de los cristales de balcones, ventanas, puertas vidrieras, portezuelas de coches, etc., para resguardarse del sol o impedir la vista desde fuera (*DLE* s.v. CORTINILLA, acep. 1).

1840, MD-Madrid: Seis *cortinillas* de bidriera de muselina estampada ▪ **1858**, MD-Escorial: Dos *cortinillas* de ventana en cinco r[eale]s ▪ **1870**, MD-Madrid: Un par de *cortinillas* de balcon en una peseta [...] Unas *cortinillas* blancas para el balcon.

cotilla

cotilla, cotillas

n.º casos: 11

f. Jubón sin mangas hecho de dos telas, embutido con barba de ballena, y pespuntado, sobre el cual se visten las mujeres el jubón o casaca, y se ajusta al cuerpo (*Autoridades* s.v. COTILLA).

1724, MD-Madrid: Vna *cotilla* nueva de olanda cruda en sesenta reales ▪ **1745**, MD-Madrid: vna *cotilla* de tafetan blanco en ziento y veynte r[eale]s. Mas otra *cotilla* de tafetan encarnada andada en treinta r[eale]s ▪ **1747**, MD-Madrid: vna *cotilla* de damasco carmesi bien tratada ▪ **1750**, MD-Madrid: Una *cotilla* nueva azul en setenta y cinco reales ▪ **1784**, MD-Pinto: Una *cotilla* con forro de medio tapiz ▪ **1765**, MD-Madrid: dos *cotillas* viejas en veinte r[eale]s las dos.

cuartilla

cuartilla, quartilla, cuartillas, quartillas
n.º casos: 32

1. *f.* Medida de capacidad para áridos, cuarta parte de una fanega, equivalente a 1387 cl aproximadamente (*DLE* s.v. CUARTILLA, acep. 1).

1726, AL-Abla: mas tres *quartillas* de sal ▪ **1748**, AL-Abrucena: Tres *quartillas* de garbanzos ▪ **1750**, AL-Dalías: tres *cuartillas* de hauas ▪ **1785**, AL-Fiñana: de una fanega y tres *cuartillas* de abillas, ciento y cinco r[eale]s ▪ **1827**, AL-María: Una *cuartilla* de medir grano en seis reales ▪ **1852**, MD-Madrid: Seis arrobas, *cuartillas* lana para colchones ▪ **1858**, MD-Escorial: Una *cuartilla*, un medio celemin de medir grano ▪ **1868**, AL-Tabernas: Cuatro y *cuartilla* fanegas de maiz.

2. *f.* Medida de capacidad para líquidos, cuarta parte de la cántara o de una arroba (adaptado de *DLE* s.v. CUARTILLA, aceps. 2, 3).

1704, AL-Cuevas de Almanzora: Una calabaza enredada, cabida de una *cuartilla* ▪ **1825**, AL-Zurgena: Vna media *quartilla* de barro de medir azeite ▪ **1845**, MD-Ambite: Vna *cuartilla* de oja de lata en diez r[eale]s ▪ **1870**, MD-Aranjuez: Dos *cuartillas* para medir vino en dos escudos.

cuartillo

cuartillo, quartillo, cuartillos, quartillos
n.º casos: 28

1. *m.* Medida de capacidad para áridos, cuarta parte de un celemin, equivalente a 1156 ml aproximadamente (*DLE* s.v. CUARTILLO, acep. 1).

1835, AL-Fiñana: [...] medio celemin, *cuartillo* y medio *cuartillo* biejos para medir grano.

2. *m.* Medida de líquidos, cuarta parte de una azumbre, equivalente a 504 ml (*DLE* s.v. CUARTILLO, acep. 2).

1766, AL-Cuevas de Almanzora: un barrilico de vidrio de m[edi]o q[uarti]llo y zinco platts ▪ **1793**, MD-Madrid: ocho

vasos de a *cuartillo* ▪ **1835**, AL-Fiñana: Un *cuartillo* y medio *cuartillo* para medir vino, ambos de lata ▪ **1845**, MD-Ambite: Otra docena de pucheros de a *cuartillo* ▪ **1870**, MD-Aranjuez: Seis aceiteras de á tres *cuartillos* con pitorro [...] Una cafetera de caber cuatro *cuartillos* en un escudo.

3. *m.* Cuarta parte de un real (*DLE* s.v. CUARTILLO, acep. 3).

1750, MD-Madrid: Una caja de plata [...] vale diez y seis reales y *cuartillo* de plata de diez y siete cuartos cada uno ▪ **1784**, MD-Pinto: Veinte baras de cotton para para [sic] cortinas, a onze rr[eale]s y quartillo ▪ **1818**, MD-Madrid: Una salbilla de plata de peso treinta onzas y tres quartas, a diez y ocho r[eale]s y tres *quartillos* onza, quinientos setenta y seis.

E

escamilla

escamilla

n.º casos: 2

(§ *Escamilla y raspilla*)

f. Posiblemente, técnica con la que se urdía la tela formando ligamentos que simulaban la forma de escamas pequeñas (def. propia).

1806, MD-Madrid: Un jubon de seda azul de *escamilla* nuevo en cien r[eale]s ▪ **1815**, MD-Madrid: Un jubon de *escamilla* de color en 60.

escardillo

escardillo

n.º casos: 5

m. Azada pequeña para arrancar las malas hierbas de la tierra (*DLE* s.v. ESCARDILLO, acep. 1).

1785, AL-Fiñana: Un *escardillo* en dos r[eale]s y medio ▪ **1840**, AL-Fiñana: Vna azada, un azadon y un *escardillo* ▪ **1868**, AL-Tabernas: Un hacha. Un *escardillo*. Un legon viejo.

esportillo

esportillo, esportillos

n.º casos: 2

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

m. Capacho hecho de esparto, palma o paja que se emplea para transportar las provisiones a la casa (*DLE* s.v. ESPORTILLO).

1822, AL-Fiñana: Dos *esportillos* en 1 ▪ **1835**, AL-Fiñana: Un *esportillo* de paja de centeno en 1.

espumilla

espumilla

n.º casos: 2

f. Tejido de lino o seda muy ligero y delicado, semejante al crespón (*DLE*, *DHTT*).

1844, AL-Cuevas de Almanzora: Un paño de altar de *espumilla*. Otro id[em] de linon floreado ▪ **1844**, AL-Almería: Uno id[em] [pañuelo] de *espumilla* de olan.

esterilla

esterilla

n.º casos: 2

f. Galón o trencilla de hilo de oro o plata, ordinariamente muy estrecha (*Autoridades*, *DLE* s.v. ESTERILLA).

1784, MD-Pinto: Otro guardapies de *esterilla* con guarniz[ió]n en quarenta r[eale]s ▪ **1870**, MD-Aranjuez: Siete varas de *esterilla* dorada en un escudo y doscientas mil[esima]s.

estopilla

estopilla

n.º casos: 7

f. La parte más fina y delgada que queda en el rastrillo al pasar por él por segunda vez el lino o el cáñamo antes de hilarse. Se denominaba también así la tela más delgada y fina que la estopa, parecida al cambray, pero semejante en lo transparente a la gasa (*Autoridades* s.v. ESTOPILLA).

1789, MD-Escorial: vna mantilla de *estopilla* nueva en cinquenta r[eale]s ▪

1800, MD-Brunete: Un delantal de *estopilla* nuevo en doze [...] Otro de *estopilla* bueno en diez y ocho. Una mantilla de *estopilla* andada en veinte [...] Otros q[ua]tro [pañuelos] de *estopilla* buenos en veinte y seis ▪ **1834**, MD-Algete: otra almuada de *estopilla* en cinco r[eale]s ▪ **1840**, MD-Madrid: Una guarnicion de *estopilla* de colcha chica.

estufilla

estufilla

n.º casos: 1

(§ 6.3.1. *El contexto como elemento desambiguador*)

f. Manguito pequeño hecho de pieles finas, tales como las martas, con las que se abrigan las manos durante el invierno (*Autoridades* s.v. ESTUFILLA).

1724, MD-Madrid: Otro abanico ordinario biexo en seis. Vna *estufilla* ordinaria bien tratada en diez y ocho. Vn par de guantes nuevos, bordados de plata, en siete r[eale]s

F

faldillas

faldillas

n.º casos: 2

f. pl. Partes que cuelgan de la ropilla, jubón, casaca o el colete y que cubren el cuerpo de la cintura para abajo a modo de sobrefalda. Caían sobre el calzón, en el caso de los hombres, o sobre la saya o basquiña que vestían las mujeres (*Autoridades* s.v. FALDILLAS, *GTIRCE*).

1703, MD-Madrid: Un coletto de ante con seis *faldillas* q[ue] esta vien tratado ▪ **1750**, MD-Madrid: Una casaca de tafetan negra forrada en bocadillo y las *faldillas* en tafetan de porzelana en setenta y cinco r[eale]s.

felpilla

ferpillas

n.º casos: 1

f. Cierta especie de hilo o cordón de seda, con pelo similar a la felpa, utilizado para bordar o guarnecer vestidos u otras cosas (*Autoridades* s.v. FELPILLA).

1835, AL-Fiñana: Una basquiña de cubica mediada con dos *ferpillas* por bajo, en 50.

francesita

francesita

n.º casos: 1

(§ *Francesita*)

f. Prenda de vestir femenina (def. propia). También tela de algodón a cuadros, a menudo rayada (*DHTT* s.v. FRANCESILLA).

1789, MD-Escorial: vna *francesita* de duroy nueba en quarenta rr[eale]s. Otra vsada de estameña en catorce rr[eale]s.

G

gargantilla

gargantilla, gargantillas

n.º casos: 8

f. Collar ajustado que usaban las mujeres como adorno en la garganta y que suele estar hecho de metales o piedras preciosas como diamantes, rubíes, esmeraldas, perlas, corales o azabache (*Autoridades* s.v. GARGANTILLA, *GTIRCE*).

1715, MD-Madrid: Vna *gargantilla* correspondiente a lo de arriua con quarenta y quatro piezas ▪ **1736**, AL-Almería: una *gargantilla* o collar de perlas ▪ **1784**, MD-Pinto: Una *gargantilla* de aljofar menuda de tres bueltas [...] Una *gargantilla* y pendientes de piedras de Francia ▪ **1824**, AL-Fiñana: Una *gargantilla* con cruz y zarcillos correspondientes de color flor de romero en v[ein]te r[eale]s [...] Dos *gargantillas* y sarcillos de cristal en veinte r[eale]s [...] Dos *gargantillas* de coral fino en sesenta r[eale]s.

golilla

golilla

n.º casos: 2

f. Adorno de cartón forrado de tafetán u otra tela y de forma circular que rodea el cuello y sobre el cual se coloca una valona de gasa engomada o almidonada. Formaba parte de la vestimenta cortesana masculina hasta principios del siglo XVIII, hasta que pasó a ser un accesorio propio de ministros togados, abogados y alguaciles (*Autoridades* s.v. GOLILLA, *GTIRCE*).

1714, AL-Cuevas de Almanzora: Unas medias de estambre, nuevas. Una *golilla*. Una toalla con desilados y rapazejos ▪ **1715**, MD-Madrid: vn bestido de nobleza de *golilla* viejo.

granadillo

granadillo

n.º casos: 4

m. Árbol procedente de la India, cuya madera es muy maciza y de color amusco u oscuro (*Autoridades* s.v. GRANADILLO).

1721, AL-Almería: un san Antonio de bulto con la peana de *granadillo* ▪ **1724**, MD-Madrid: Vn espexo de a terzia con marco de *granadillo* en quarenta y cinco reales ▪ **1736**, AL-Almería: una cama de *granadillo* bronzada ▪ **1791**, AL-Vera: vna capa de paño, *granadillo* negro y buelta para hombre.

granillo

granillo

n.º casos: 1

(§ *Granillo*)

m. Labor menuda con que se solía tejer el lienzo para la confección de textiles, principalmente colchas, manteles y servilletas. Posiblemente deba su nombre a la similitud de la trama con los granos de la espiga del cereal, de ahí que algunas fuentes testimonien las formas ‘granillo de trigo’ o ‘granillo de cebada’. Es voz regional

documentada en la zona norte-centro peninsular (def. propia).

1815, MD-Madrid: una colcha con guarnicion blanca, bien tratada, en 160. Yd[em] otra blanca de *granillo* en 120.

gusanillo

gusanillo, gussanillo

n.º casos: 64

m. Cierta género de labor que se hace en los tejidos de lienzo y otras telas (*Autoridades* s.v. GUSANILLO).

1700, MD-Madrid: Mas dos paños de manos de *gussanillo* a medio traer ▪ **1701**, MD-Madrid: Una colcha de *gusanillo* pequeña muy usada ▪ **1726**, AL-Abla: mas dos tavlas de manteles de *gusanillo* mediadas, mas dos servilletas de *gusanillo* mediadas ▪ **1784**, MD-Pinto: Una colcha grande, labor de *gusanillo*, en quinze r[eale]s. Dos serbilletas sin hazer, labor de *gusanillo*, en diez r[eale]s ▪ **1805**, MD-Madrid: Quatro colchas, las tres blancas de *gusanillo* y hueso de melocoton, y la otra de coton guarnecida de muselina [...] Tres manteles de *gusanillo* casero, nuevos [...] Quatro tohallas de *gusanillo* nuevas ▪ **1818**, MD-Madrid: Dos rollos de manteleria de *gusanillo* casero de siete varas cada uno [...] Una docena de toallas de *gusanillo* con listas azules sin estrenar ▪ **1840**, MD-Madrid: Dos tablas de manteles de *gusanillo* á doce r[eale]s [...] Un mantel de *gusanillo* de algodón [...] Seis toallas de *gusanillo* en buen uso á cinco r[eale]s.

H

hiladillo

iladillo, yladillo, ilaillo

n.º casos: 12

m. El hilo que sale de la estopa de la seda y lo que no se puede torcer de ella, el cual se hila en la rueca como el lino, de ahí su nombre (*Autoridades* s.v. HILADILLO).

1736, AL-Almería: una colcha de *yladillo* y seda ▪ **1775**, AL-Escúllar: un jugon de picotillo de *iladillo* ▪ **1785**, AL-Fiñana: treynta y tres onzas de *ilaillo* ▪ **1832**, AL-Fiñana: vna corcha de *yladillo* carmesi [...] quatro fundas de *yladillo* pobladas ▪ **1835**, AL-Fiñana: Dos fundas de *iladillo* listadas con enchimiento de lana [...] Otro mullidor de *iladillo* mediado, bacido [...] Cuatro fundas de almoadas de *iladillo* mediadas con enchimiento de lana.

holandilla

olandilla

n.º casos: 24

f. Cierta especie de lienzo teñido y prensado, que sirve para forros de vestidos y otras cosas (*Autoridades* s.v. HOLANDILLA).

1703, MD-Madrid: Dos cortinas de brocatel encarnado y anteado fforrados en *olandilla* [...] Un capote de barragan plateado forrado con *olandilla* amusca ▪ **1711**, MD-Madrid: Un forro de *olandilla* encarnado para dosel que tendra asta doce baras. Dos baras de *olandilla* musa [sic] [...] Una senefa de *olandilla* encarnada en dos pedazos ▪ **1728**, MD-Madrid: Vna chupa de raso encarnado con flores, forrada de *olandilla* azul [...] Vna mantilla de damasco azul zelete antiguo con encajes, forrada en *olandilla* ▪ **1745**, AL-Fiñana: Vna mantilla de raso verde forrada en *olandilla* encarnada, viexa ▪ **1750**, MD-Madrid: Un tontillo grande de *olandilla* amarilla ▪ **1766**, AL-Cuevas de Almanzora: otra colcha de tafettan listado, forrada en *olandilla*, a mas de medio servir ▪ **1845**, MD-Ambite: Vna colcha de damasco azul forrada en *olandilla* ▪ **1852**, MD-Madrid: Cinco varas y cuarta de *olandilla* para almoadones.

horquilla

horquillas

n.º casos: 1

f. Pieza metálica o de otro material doblada en forma de U abierta o muy cerrada que se emplea para sujetar el pelo (*DLE* s.v. HORQUILLA, acep. 9, *DEA* s.v., acep. 1).

1890, MD-Alcalá: Dos *horquillas* de oro, una con perlas y otra idem.

horquillo

orquillos

n.º casos: 2

(§ *Horquillo*)

m. reg. Horca agrícola pequeña con púas, que ordinariamente sirve para colgar y descolgar las cosas que están altas, o para afianzar y asegurar alguna cosa en el suelo (*Autoridades* s.v. HORQUILLA, acep. 1, *DEA*, s.v.). En el *TLHA* se define como ‘tirabrasas’ y se documenta en Jaén.

1836, MD-Estremera: Un allegador en 4. Seis *orquillos* y tres palas en 18 ▪ **1845**, MD-Ambite: Vn rastro para llegar las parbas en diez r[eale]s. Diez *orquillos* andados en veinte r[eale]s.

husillo

usillo, usillos

n.º casos: 3

m. Tornillo grande con muescas espirales usado como fuerza principal de las prensas, como el que se usa en los lagares y en otras máquinas para apretar o subir grandes pesos (*Autoridades* s.v. HUSILLO, *DEA*).

1784, AL-Abrucena: Un *usillo* de estruxar ubas, trescientos reales ▪ **1832**, AL-Fiñana: vna mesa de *usillo* completa, vale 120 ▪ **1834**, AL-Escúllar: Un banco y otro con *usillos* en 25.

J

jerguilla

jerguilla, gerguilla

n.º casos: 4

f. Tela delgada de seda o lana, o mezcla de una y otra, que se parece en el tejido a la jerga (*Autoridades* s.v. XERGUILLA).

1703, MD-Madrid: Un cobertor para cama de *jerguilla* encarnada vien tratado ▪ **1705**, MD-Madrid: una cortina grande de *gerguilla* en veinte r[eale]s [...] una cortina de *jerguilla* de puerta en diez y seis r[eale]s [...] dos cortinas de *jerguilla* de puertas, manchadas.

jeringuilla

geringuillas

n.º casos: 1

f. Jeringa pequeña en la que se enchufa una aguja hueca de punta aguda cortada a bisel, y sirve para inyectar sustancias medicamentosas en tejidos u órganos (*DLE* s.v. JERINGUILLA).

1870, MD-Aranjuez: Diez y ocho *geringuillas* de cristal, á ciento cincuenta mil[esima]s una.

junquillo

junquillo

n.º casos: 1

m. Flor semejante al narciso, de color blanco o amarillo, con el tallo liso y parecido al junco (*Autoridades* s.v. JUNQUILLO).

1844, AL-Almería: Un *junquillo* de oro con nueve adarmes, ciento ochenta r[eale]s.

justillo

justillo, gustillo, justillos

n.º casos: 19

m. Vestido interior ajustado al cuerpo a modo de jubón, del que se diferencia en no tener mangas (*Autoridades* s.v. JUSTILLO).

1717, MD-Torrelaguna: vn *gustillo* de tela tasado en ziento y beinte ▪ **1791**, AL-Vera: vn *justillo* de trepe azul ▪ **1722**, MD-Colmenar de Oreja: Vn *justillo* bordado en doze rr[eale]s ▪ **1728**, MD-Madrid: Vna almilla y vn *justillo* de coton en diez y ocho reales ▪ **1765**, MD-Madrid: dos *justillos* de

cotanza en quize r[eale]s ▪ **1800**, MD-Navalafuente: Una camisa de coruña nueba y un *justillo* p[ar]a muger en cinquenta y cinco r[eale]s ▪ **1805**, MD-Madrid: Dos *justillos* de coruña guarnecidos en veinte r[eale]s.

L

lanilla

lanilla

n.º casos: 5

f. Tejido similar a la lana, más delgado y fino que la lamparilla, empleado para la confección de vestidos de verano y hábitos (*Autoridades* s.v. LANILLA, acep. 2).

1701, MD-Madrid: Una colgadura de cama con zinco cortinas y zielo de *lanilla* muy anttigua y dos cortinas con su zenefa de lo mismo ▪ **1724**, MD-Madrid: Otra casaca de *lanilla* color del carmen, buena ▪ **1792**, AL-Cuevas de Almanzora: otra basquiña bieja de *lanilla*, su valor diez r[eale]s ▪ **1845**, AL-Cuevas de Almanzora: Un vestido de *lanilla*.

M

manecilla

manecilla

n.º casos: 1

f. Equivalente a *manilla* en su significado de ‘mango’ (adaptado de *DLE* s.v. MANILLA, acep. 3).

1832, AL-Fiñana: vn braceró de cobre con *manecilla* y tarima, y vale 30.

manguito

manguito, manguitos

n.º casos: 8

m. Especie de manga abierta por ambos lados que sirve para abrigarse las manos metiéndolas cada una por su lado. Habitualmente hecha de martas u otras

pieles adobadas, aunque también las hay de pluma o seda para mayor adorno. Con este término se hace referencia asimismo a la media manga ajustada al brazo, que llega desde el codo a la muñeca (*Autoridades* s.v. MANGUITO). Como complemento o accesorio ornamental, el manguito servía para ambos sexos y se consideraba un objeto preciado debido a las ricas pieles con las que se hacían y los materiales y técnicas con los que se decoraban (*GTIRCE*). A menudo se documentan a juego con las *palatinas* o *paletinas*, una especie de estola o chal con la que las mujeres se cubrían los hombros y el escote y que servía como adorno y complemento para enriquecer el vestido (*GTIRCE* s.v. PALATINA).

1722, MD-Colmenar de Oreja: Vnos *manguitos* de droguete en dos rr[eale]s ▪ **1745**, MD-Madrid: Mas vn *manguito* y paletina de color de punzon de pluma en zien rs. Mas vn *manguito* y paletina de todos colores de pluma ▪ **1750**, MD-Madrid: Un juego de *manguito* y paletina de cartulina sobre blanco con varias flores encarnadas en doze r[eale]s. Otro *manguito* entorchado viejo ▪ **1852**, MD-Madrid: Un *manguito* de piel regalado por d[on]a Pilar Gruyer.

manilla

manillas

n.º casos: 16

(§ *Manilla*)

1. f. Adorno usado por las mujeres en las muñecas, compuesto de unas sartas que dan varias vueltas, de perlas, corales, granates u otras cuentas (*Autoridades* s.v. MANILLA).

1701, MD-Madrid: Unas *manillas* de aljofar y granates ▪ **1703**, MD-Madrid: Dos bueltas de *manillas* de aljofar menuda y granattes ▪ **1717**, MD-Torrelaguna: vnas *manillas* de perlas de boemia tasado en treinta y dos r[eale]s ▪ **1750**, MD-Madrid: Dos bueltas de *manillas* de alxofar de genero de medio rostrillo ▪ **1834**, MD-

Madrid: Vnas *manillas* de oro pesan una onza y un adarme.

2. *f. reg.* Aro de madera o metal del que se cuelgan las cortinas para que puedan correrse y descorrerse con facilidad (def. propia, adaptado de Venceslada).

1745, AL-Fiñana: Vna cortina de lienzo apretadillo con varetas y *manillas* de yerro ▪ **1766**, AL-Cuevas de Almanzora: zinq[uen]ta *manillas* de metal y hierro para cortinas en tres r[eale]s [...] dos atados con treintta y seis *manillas* de hierro y bronze p[ar]a cortina [...] una cortina de lienzo delgado con vareta de hierro y *manillas* p[ar]a la sala del valcon en v[ein]te y siete r[eale]s. Ytt[em] otra cortina de almarieta con vareta de hierro y *manillas* p[ar]a la puerta de la sala.

mantilla

mantilla, manttilla, mantillas

n.º casos: 62

(§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*)

f. Cobertura de bayeta, grana u otra tela con que las mujeres se cubrían antiguamente desde la cabeza hasta la cintura para abrigarse. Con el tiempo podía incluir una guarnición de tul o encaje como elemento decorativo hasta que acabaron por confeccionarse enteramente de telas lustrosas, por lo general de seda (def. propia, adaptado de *Autoridades* s.v. MANTILLA, acep. 1 y RAE, 1869).

1703, MD-Madrid: Una *mantilla* de paño encarnado con un encaje de porzelana por guarnizion, forrada en tafetan anteadado ▪ **1711**, MD-Madrid: Una *mantilla* p[ar]a ombros de baieta negra ▪ **1722**, MD-Colmenar de Oreja: Vna *mantilla* blanca de baieta ▪ **1724**, MD-Madrid: Vna *mantilla* de baieta guarnezida con liston de color de oro ▪ **1728**, MD-Madrid: Vna *mantilla* de baieta blanca con riuete negro [...] Vna *mantilla* de damasco azul zeleste antiguo con encajes forrada en olandilla ▪ **1745**, AL-Fiñana: Vna *manttilla* de raso verde forrada en

olandilla ▪ **1766**, AL-Cuevas de Almanzora: vna *manttilla* de bayeta de Ynglaterra poco vsada ▪ **1776**, AL-Fiñana: una *mantilla* de cristal ▪ **1800**, MD-Madrid: Una *mantilla* de estopilla andada en veinte ▪ **1815**, MD-Madrid: una *mantilla* de sarga de seda lisa [...] una *mantilla* de franela con cinta ▪ **1822**, AL-Fiñana: Una *mantilla* de mosolina mediada ▪ **1822**, AL-Tabernas: una *mantilla* bordada en sesenta [...] una *mantilla* fina de bayeta en veinte r[eale]s ▪ **1824**, AL-Fiñana: Una *mantilla*, tafetan de Francia con guarnicion de blonda en doscientos r[eale]s ▪ **1825**, AL-Zurgena: Vna *mantilla*, lana color negro en diez r[eale]s ▪ **1833**, AL-Abla: Una *mantilla* de clarin en 7. Otra id[em] mediada. Una *mantilla* de seda biexa ▪ **1834**, MD-Algete: una *mantilla* de franela con cinta de terciopelo ▪ **1834**, MD-Madrid: Vna *mantilla* de encage de blonda negra nueba ▪ **1858**, MD-Escorial: Dos *mantillas*, una de blondas con casco rayado y otra lisa con velo.

marquilla

marquilla

n.º casos: 1

f. PAPEL DE ~. Papel de tina, grueso, lustroso y muy blanco, que se emplea ordinariamente para dibujar (*DLE* s.v. PAPEL DE MARQUILLA)

1870, MD-Aranjuez: Tres id[em] [hojas de lata] marquilla, á doscientas.

medianillo

medianillo, medio anillo

n.º casos: 31

(§ (*Lienzo*) *apretadillo* y (*lienzo*) *medianillo*)

adj. reg. También LIENZO ~. Lienzo de consistencia más fina que el ordinario y especialmente empleado en Andalucía, con mayor presencia en la zona oriental (def. propia).

1721, AL-Almería: Ytt[em] doze tablas de manteles grandes nuevas de *medio anillo*.

Ytt[em] ciento y catorze servilletas de *medio anillo* nuevas de vara de largo [...] Ytt[em] seis paños de mano del mismo *medio anillo* nuevos con rapazejo [...] Ytt[em] seis savanas de *medio anillo* nuevas ▪ **1726**, AL-Abla: mas vna tohalla de *medianillo* con puntas ▪ **1736**, AL-Almería: ocho almohadas de lienzo *medianillo* [...] quatro sauanas de *medianillo* [...] una tabla de manteles de *medianillo* [...] dos dosenas de seruilletas de *medianillo* ▪ **1766**, AL-Cuevas de Almanzora: dos almoadas de *medianillo* con botones en las vocas [...] una camisa p[ar]a hombre de *medianillo* y mangas delgadas a medio serbir [...] unos calzoncillos de *medianillo* biexos en un r[eal] [...] vna camisa de muger de *medianillo* sin coser.

molinillo

molinillo, moliniño, molinillos

n.º casos: 7

m. Instrumento pequeño para moler, sea cual sea su uso, y especialmente el que se emplea para batir y disolver el chocolate. (*Autoridades* s.v. MOLINILLO).

1721, AL-Almería: dos chocolateras de n[umero] 87 con sus *molinillos*, la una grande y la otra mediana ▪ **1766**, AL-Cuevas de Almanzora: un *molinillo* p[ar]a chocolattera, nuebo, en un real ▪ **1791**, AL-Vera: vna chocolatera y *molinillo*, quatro rr[eale]s ▪ **1825**, AL-Zurgena: Vna chocolatera de cobre con su *moliniño* en seis r[eale]s ▪ **1832**, MD-Navas del Rey: un pesebre y *molinillo* ▪ **1833**, AL-Almería: Un *molinillo* de acemitre.

morillo

morillo, morillos, murillos

n.º casos: 6

m. El caballete de hierro que se pone en el hogar para sustentar la leña. Su nombre procede de las figurillas como cabezas de moros que se ponen en ellos o por estar siempre tiznados y negros como ellos,

según siente Covarr. (*Autoridades* s.v. MORILLO, acep. 2).

1700, MD-Madrid: Vn *morillo* de yerro tasado en dos reales ▪ **1715**, AL-Serón: Dos *morillos* y dos mazos, grande y un pequeño ▪ **1745**, AL-Fiñana: Vnos *murillos* de yerro ▪ **1766**, AL-Cuevas de Almanzora: vn *morillo* de hierro, una rasera y unas parrillas ▪ **1827**, Al-María: Un *morillo* de fierro en cuatro r[eale]s ▪ **1832**, AL-Fiñana: vn *morillo* de hierro vale 12.

mostacilla

mostacilla

n.º casos: 1

f. Abalorio de cuentecillas muy menudas (*DLE* s.v. MOSTACILLA, acep. 2).

1844, AL-Almería: Una cinta de *mostacilla* con pasadores de oro, cuarenta r[eale]s.

P

pajecillo

pagecillo

n.º casos: 1

(§ *Pajecillo*)

m. reg. Nombre con el que se denomina en Andalucía al bufete pequeño en el que se ponen las luces (Venceslada, *Autoridades* s.v. PAGECILLO).

1806, AL-Almería: un *pagecillo* en treinta.

palillo

palillo, palillos

n.º casos: 3

m. Varilla por la parte inferior aguda y por la superior redonda, con un agujerillo en medio donde se encaja la aguja para hacer media. Tiene poco más de un palmo de largo y se pone en la cintura para que se mantenga firme (*Autoridades* s.v. PALILLO).

1724, MD-Madrid: Vn *palillo* de ebano de hazer media con su casquillo y cadenilla de plata en doze reales ▪ **1789**, MD-Escorial:

vna cesta de azer media con su *palillo* en dos rr[eale]s.

palomilla

palomillas

n.º casos: 1

f. Posiblemente, tuerca con dos extensiones laterales en que se apoyan los dedos para darle vueltas (*DLE* s.v. PALOMILLA, acep. 11).

1832, MD-Madrid: Cuatro achas de pino con sus *palomillas* á cuatro r[eale]s.

parrilla

parrilla, parrillas

n.º casos: 26

f. Usado en plural. Instrumento empleado en la cocina para asar o tostar alguna cosa. Se compone de una rejuela de hierro cuadrada con un pie en cada una de las esquinas y en medio un mango para manejarla. Su nombre deriva de la similitud de su figura con el parral en que se sostienen las parras (*Autoridades* s.v. PARRILLAS, acep. 1).

1715, AL-Serón: unas *parrillas* y tres espeteras ▪ **1715**, MD-Madrid: Vnas trebedes, vnas *parrillas* y un braserito y un asador ▪ **1745**, AL-Fiñana: Vnas *parrillas* de yero viexas ▪ **1750**, AL-Dalías: las *parrillas* grandes de yerro valen seis rr[eale]s. Otras *parrillas* biejas y dos asadores valen vn rr[ea]l ▪ **1785**, AL-Fiñana: un colchon de *parrilla* en diez r[eale]s ▪ **1825**, AL-Zurgena: Vnas *parrillas* á medio vsar en cinco r[eale]s ▪ **1836**, MD-Estremera: Unas *parrillas* de fierro en 3 ▪ **1877**, MD-Madrid: un calentador de cobre, dos *parrillas* de yerro.

petillo

pettillo

n.º casos: 1

m. Pedazo de tela cortado en forma de triángulo que utilizan como adorno las mujeres en el pecho. Con el mismo nombre

se llama también la joya hecha en la misma figura (*Autoridades* s.v. PETILLO).

1717, MD-Torrelaguna: vn *pettillo* de bordado de oro y plata tasado en sesenta r[eale]s.

picotillo

picotillo

n.º casos: 1

m. Picote de seda de inferior calidad (*DHTT* s.v. PICOTILLO). Con *picote* se designa una tela áspera y basta de pelo de cabra, pero también un tipo de tela de seda muy lustrosa con la que se hacían vestidos (*DLE*, *DHTT* s.v. PICOTE).

1755, AL-Escúllar: It[em] un jugon de *picotillo* de iladillo.

puntilla, -ca

puntilla, puntillas, punticas

n.º casos: 11

(§ 6.2. *Significado transparente y lexicalizado de los sufijos -ico, -ito, -illo y § Puntilla*)

f. Encaje muy angosto hecho en puntas, ondas o picos, que se suele añadir y coser a la orilla de otro encaje ancho y con el que se adorna el borde de pañuelos, toallas, vestidos, etc. (*Autoridades* s.v. PUNTILLA, acep. 2, *DHTT*).

1700, MD-Madrid: Dos pares de calzoncillos ya mojados, los vnos con unas *puntillas* y los otros mas rejios [...] Vn pañuelo con *punticas* antiguas, biejo, tasado en dos reales ▪ **1750**, MD-Madrid: Vna bara de tafetan carmesi con *puntilla* blanca en diez r[eale]s ▪ **1779**, MD-Escorial: Un collar con una cruz de plata de cinta encarnada con *puntilla* en veinte ▪ **1844**, AL-Almería: Tres id[em] [pañuelos] de tul con *puntilla* ▪ **1852**, MD-Madrid: Doce varas de *puntilla* para un juego de cama ▪ **1855**, MD-Algete: ocho varas de *puntilla* ancha ▪ **1870**, MD-Aranjuez: Una cama de hierro pintada con tres colchones, un almohadon y dos almohadas, todo poblado de lana, tela de cuti con sus

correspond[ien]tes fundas blancas con *puntilla* ▪ **1890**, MD-Alcalá: Cinco pedazos de *puntilla* de hifo [sic] y algodón.

R

raspilla

raspilla

n.º casos: 1

(§ 6.5.2.1. *Regionalismos y voces poco documentadas* s.v. ESCAMILLA y RASPILLA)

f. Posiblemente, técnica con la que se urdía la tela formando ligamentos que simulaban una raspa (def. propia).

1822, AL-Tabernas: It[e]n otro [armador] de *raspilla* en veinte.

redecilla

redecilla, redecillas

n.º casos: 3

f. Prenda de malla, en forma de bolsa, y con cordones o cintas para recoger el pelo o adornar la cabeza (*DLE* s.v. REDECILLA, acep. 2).

1786, AL-Cuevas de Almanzora: una *redecilla* encarnada bordada ▪ **1789**, MD-Escorial: vna *redecilla* azul de cintas en sesenta y quatro [...] dos *redecillas* negras en veinte rr[eale]s.

rejilla

regilla

n.º casos: 1

f. Entramado de metal, madera u otro material que, generalmente enmarcado en un hueco, permite el paso del aire, la luz, la voz, etc. (*DLE* s.v. REJILLA, acep. 1).

1890, MD-Alcalá: Vn portamoneda de *regilla* de plata.

ropilla

ropilla

n.º casos: 10

f. Vestidura corta que se llevaba sobre el jubón y ceñida al cuerpo hasta la cintura

con mangas y brahones (“rosca o doblez que ceñía la parte superior del brazo” *DLE*), de las cuales penden regularmente otras mangas sueltas o perdidas (*GTIRCE*, *Autoridades* s.v. ROPILLA, acep. 2).

1700, MD-Madrid: Mas vna *ropilla* y calzon de paño de Segobia y mangas de raso liso a medio traer [...] Mas una *ropilla* y calzon de vayeta de Alconcher con mangas de tafetan a medio traer ▪ **1700**, MD-Madrid: un calzon y *ropilla* de paño de olanda [...] Vna *ropilla* con mangas de paño negro con sus franjas de raso ▪ **1703**, MD-Madrid: Un vestido de paño negro nuevo q se conpone de *ropilla* y calzones y mangas de rizo ▪ **1711**, MD-Madrid: Diez y siete vestidos de felpa berde lisa, guarnecidos de trencilla de plata falsa que se componen de *ropilla* con mangas perdidas y calzones, y algunos de ellos sin guarnicion y algunas de las d[ic]has mangas son de restaño con fleco ▪ **1714**, AL-Cuevas de Almanzora: Una *ropilla* negra y calzones de vaieta ▪ **1728**, MD-Madrid: Vna *ropilla* de fondo negra, con fleco de seda en las mangas, en treinta r[eale]s.

rostrillo

rostrillo

n.º casos: 1

m. Perla de aljófar en forma de grano pequeño e irregular (*GTIRCE* s.v. ROSTRILLO). El rostrillo ordinario solía contener 600 perlas la onza, pero según la cantidad de granos que contenía una onza recibía diferentes denominaciones: *rostrillo grueso* (500 granos la onza), *rostrillo menudo* (700 granos la onza), *medio rostrillo mejor* (850 granos la onza), *medio rostrillo* (1200 granos la onza) (*Autoridades* s.v. ROSTRILLO, acep. 2).

1750, MD-Madrid: Dos bueltas de manillas de alxofar de genero de *medio rostrillo* compuestas de mil trescientos y nouenta granos que yualmente se expresa en d[ic]ha hixuela, pesan veinte y cinco adarnes y terzia y salen de ochozientos y

setenta y siete granos y tres cuartos en onza y vale segun su calidad quinientos y sesenta y quatro r[eale]s de v[ello]n que es su valor.

S

soguilla

soguilla

n.º casos: 2

f. Trenza delgada hecha de pelo o de esparto (*DLE* s.v. SOGUILLA, acep. 1).

1721, AL-Almería: seis sillas, madera de pino, con los asientos de *soguilla*, a medio servir ▪ **1750**, AL-Dalías: las ocho sillas de *soguilla* mediadas en diez y seis r[eale]s.

sombrilla

sombrilla

n.º casos: 4

f. Quitasol (*DLE* s.v. SOMBRILLA).

1840, MD-Madrid: Una *sombrilla* de tafetan negro usada ▪ **1852**, MD-Madrid: Una *sombrilla* en buen uso, cuarenta [...] Una *sombrilla* de seda, color de rosa, regalo de d[on]a Dolores Navarro ▪ **1855**, MD-Algete: un paraguas y *sombrilla* en 50.

T

tenacillas

tenacillas

n.º casos: 4

(§ 6.3.1. *El contexto como elemento desambiguador*)

1. f. pl. Tenaza pequeña de muelle, que sirve para coger terrones de azúcar, dulces y otras cosas (*DLE* s.v. TENACILLAS, acep. 1).

1843, MD-Madrid: 1 *tenacillas* de plata p[ar]a tomar azucar para el café ▪ **1890**, MD-Alcalá: Vnas *tenacillas* para azucar, al parecer de plata.

2. f. pl. Instrumento, a manera de tenaza pequeña, que sirve para rizar el pelo o la

ropa (adaptado de *DLE* s.v. TENACILLAS, acep. 2).

1877, MD-Madrid: cuatro pares de *tenacillas* para rizar ropa.

3. f. pl. Instrumento a propósito para tener cogido el cigarrillo al tiempo de fumarlo (*DLE* s.v. TENACILLAS, acep. 5).

1890, MD-Alcalá: Vnas *tenacillas* de oro para fumar [...] Vnas *tenacillas* de plata y un limpia oídos de plata (sobre la interpretación de este ejemplo, vid. § 6.3.1. *El contexto como elemento desambiguador*).

tontillo

tontillo

n.º casos: 2

m. Especie de faldellín o enagua reforzada con aros de ballena o de otra materia que usan las mujeres para ahuecar las faldas. Termina en una tela que arrastra por el suelo a modo de guardapiés y se colocaba debajo de la basquiña (*Autoridades*, *GTIRCE* s.v. TONTILLO). En el pasado también llamado *guardainfante* (*Autoridades*), aunque el tontillo se diferencia de aquel por ser más ligero y menos voluminoso, razón por la que se empleaba para diario en la corte española el último tercio del XVII y principios del XVIII, mientras que el guardainfante, más incómodo, fue quedando en desuso (*GTIRCE*).

1745, MD-Madrid: Mas vn *tontillo* de tafetan azul andado en ziento y zinq[uen]ta r[eale]s ▪ **1750**, MD-Madrid: Un *tontillo* grande de olandilla amarilla.

toquilla

toquillas

n.º casos: 1

f. Banda ancha que servía para cubrir la cabeza o como adorno sobre el sombrero, hecha de tela delgada como tul, seda, velillo u holán y que podía formar distintas figuras según el fin para el que se use (*Autoridades*, *GTIRCE* s.v. TOQUILLA).

1840, MD-Madrid: Dos *toquillas* de tul.

tornillo

tornillo, tornillos

n.º casos: 4

1. *m.* Pieza cilíndrica o cónica, por lo general metálica, con resalte en hélice y cabeza apropiada para enroscarla (*DLE* s.v. TORNILLO, acep. 1).

1715, AL-Serón: Dos camas de pino y una de nogal con sus *tornillos*.

2. *m.* Instrumento con que se mantienen sujetas las piezas que se están trabajando, por medio de dos topes, uno fijo y otro móvil (*DLE* s.v. TORNILLO, acep. 2).

1711, MD-Madrid: Un *tornillo* para azer *zerilla* con dos zigüañas de yerro ▪ **1870**, MD-Aranjuez: Un *tornillo* grande de cerrajero en diez escudos [...] Dos *tornillos* pequeños en cuatro escudos.

trabilla

trauilla

n.º casos: 1

f. Tira de tela o de cuero que pasa por debajo del pie para sujetar los bordes inferiores del pantalón, del botín, de la polaina o de la calceta (*DLE* s.v. TRABILLA, acep. 2).

1705, MD-Madrid: quatro pares de calzetas, dos de pie y dos de *trauilla*, en quarenta r[eale]s.

trampilla

trampillas

n.º casos: 1

f. Portezuela que se levanta sobre goznes colocados en su parte superior (*DLE* s.v. TRAMPILLA, acep. 3).

1870, MD-Aranjuez: Un mostrador con *trampillas* en seis id[em].

trencilla

trencilla

n.º casos: 1

f. Galón trenzado de seda, algodón o lana, que sirve para adornos de pasamanería,

bordados y otras muchas cosas (*DLE* s.v. TRENCILLA, acep. 1).

1711, MD-Madrid: Diez y siete vestidos de felpa berde lisa, guarnecidos de *trencilla* de plata falsa.

V

vainica

bainicas

n.º casos: 1

f. Bordado que se hace especialmente en el borde de los dobladillos, sacando algunas hebras del tejido (*DLE* s.v. VAINICA). Alternó en el pasado con *vainilla* (*Autoridades* s.v. VAINICA O VAINILLA).

1751, MD-Brunete: Vna sabana de tres piernas con sus *bainicas* en beinte y dos r[eale]s.

varilla

varilla, barilla, varillas, barillas

n.º casos: 19

1. *f.* Barra pequeña de hierro, larga, delgada y redonda que se usa para colgar las cortinas (*Autoridades* s.v. VARILLA, acep. 2).

1711, MD-Madrid: En *barillas* de yerro de todos tamaños se pesaron ocho arrobas y treçe libras ▪ **1784**, MD-Pinto: Dos *barillas* de cortina ▪ **1836**, MD-Estremera: Doce *barillas* de cortinas en 24 ▪ **1858**, MD-Escorial: Seis cortinas de percal con cuatro *barillas* en sesenta y seis reales ▪ **1870**, MD-Aranjuez: Cuatro cortinas de balcon y alcoba con sus correspond[ien]tes *varillas*.

2. *f.* Cada una de las piezas largas y delgadas que forman la armazón de los abanicos, paraguas, quitasoles, etc. (*DLE* s.v. VARILLA, acep. 2).

1724, MD-Madrid: Otro abanico de lamina con *barillas* de concha en treinta y siete reales ▪ **1750**, MD-Madrid: Un abanico rico de Roma con *barillas* blancas embutidas de plata en ciento y veinte reales [...] Un

abanico de *barillas* blancas enreñadas en treinta r[eale]s ▪ **1843**, MD-Madrid: Un abanico de *varillas* de metal dorado con piedras finas.

3. *f. pl.* Bastidor rectangular en que se mueven los cedazos para cerner (*DLE* s.v. VARILLA, acep. 6).

1784, MD-Pinto: Una yntera y *barillas* para zerner ▪ **1809**, MD-Navalafuente: Vna

artesa para cozer con *varillas*, zedazos, pala, raydora ▪ **1832**, MD-Navas del Rey: zedazos, *varillas* y peso ▪ **1836**, MD-Estremera: Dos artesas con sus *barillas* en 28.

7. ANÁLISIS ESTADÍSTICO DE LOS DIMINUTIVOS *-ICO*, *-ITO* E *-ILLO*

En este capítulo se lleva a cabo el análisis estadístico a partir del recuento de los diminutivos transparentes.

La clasificación exhaustiva de los diminutivos se ha realizado con vistas al estudio cuantitativo que se realiza en este capítulo y cuyos resultados pretenden aportar nuevos datos sobre la evolución de los diminutivos *-ico*, *-ito* e *-illo* en Almería a partir de un corpus de inventarios de los siglos XVIII y XIX en contraste con un subcorpus de Madrid del mismo periodo. El estudio, de tipo contrastivo, pretende examinar desde el punto de vista cronológico el uso de los diminutivos en Almería, por un lado, y Madrid, por otro, durante el transcurso de dos centurias. Para el contraste diatópico, se analizarán las diferencias entre dos zonas a priori lingüísticamente alejadas entre sí en este periodo de tiempo.

Los datos que se han incluido en el análisis se especifican en § 7.1. *Corpus y datos analizados*. En este apartado se detalla la distribución de los datos en subcorpus que ha permitido el análisis contrastivo. Como se ha mencionado, el estudio morfológico que se plantea en esta tesis tiene como principal objetivo analizar la evolución de los diminutivos en el corpus de Almería en los siglos XVIII y XIX en contraste con el de Madrid. Este objetivo principal parte de diferentes hipótesis, las cuales se formulan y detallan en § 7.2. *Variables e hipótesis de estudio*, donde además se explican las variables que forman parte de nuestro análisis. La metodología específica de análisis, así como la estructura y las herramientas empleadas se mencionan en § 7.3. *Modelo estadístico empleado y estructura del análisis*.

El análisis cuantitativo propiamente dicho se aborda en § 7.4. *Resultados*. En esta sección se presentan los datos de manera independiente para el corpus de Almería, en primer lugar, y para el corpus de Madrid, a continuación, y se analiza la evolución diacrónica para cada provincia. Por último, se dedica un apartado a la comparación diatópica de ambos. Para cada uno de estos tres niveles de estudio se llevará a cabo una exploración del conjunto total de los datos a partir de un análisis estadístico con el que se pretende evaluar si las diferencias observadas en el uso de los diminutivos son significativas en estas regiones durante los siglos estudiados.

Finalmente, en la sección § 7.5. *Discusión*, se expone la interpretación lingüística de los resultados concluyentes de nuestro análisis, los cuales se contrastan con la bibliografía previa sobre el tema en las cuestiones relacionadas con la cronología y la distribución geográfica de los diminutivos.

7.1. Corpus y datos analizados

La mayoría de los trabajos enfocados en la distribución geográfica de los diminutivos desde una perspectiva histórica se han basado tradicionalmente en el análisis de textos literarios (§ 5.4. *Fuentes para el estudio de los diminutivos*). Esta investigación proyecta la creación de su propia fuente documental de estudio, la cual constituye un corpus de inventarios de bienes creado expreso para las provincias de Almería y Madrid de los

siglos XVIII y XIX –incluido a su vez en el corpus *ODE*– con el que poder extraer conclusiones sobre el uso de los diminutivos en estas zonas y durante este marco temporal. Se pretenden comprobar las afirmaciones de trabajos previos acerca de la evolución de los diminutivos en torno a algunos aspectos como los relacionados con la cronología o la influencia del origen geográfico. Los nuevos datos podrán precisar o mejorar el conocimiento que ya existe sobre la evolución de los diminutivos en el español, empleando esta vez documentación de tipo notarial bajo la metodología específica de la lingüística de corpus.

Para este análisis se ha hecho una selección rigurosa de todas las formas que incluyen sufijación diminutiva en el corpus, escogiendo únicamente aquellas palabras en la que los sufijos *-ico*, *-ito* e *-illo* cumplen con una función propiamente diminutiva (§ 6.4. *Diminutivos transparentes*). Los datos analizados se corresponden con los diminutivos transparentes procedentes del corpus de inventarios empleado en esta tesis y cuya lista completa de formas se puede consultar en § *Anexo IV. Relación de diminutivos transparentes incluidos en el análisis*. A lo largo de este capítulo, cuando se alude a los diminutivos de modo general, nos referimos únicamente a los que conservan tal carácter de disminución, esto es, los considerados transparentes.

Para la extracción de los datos, el modelo tecnológico del corpus *ODE* ha sido fundamental. Todo el procesamiento automático del corpus detallado en el Capítulo 4 de esta tesis tiene su razón de ser en la posterior recuperación de datos, de manera que se facilite la viabilidad de análisis lingüísticos como el que aquí se presenta. El etiquetado automático de los diminutivos ha sido revisado manualmente (§ 4.10. *Revisión manual de errores en el corpus*) con el objetivo de asegurar la fiabilidad y la utilidad del corpus para estudios morfosintácticos por parte de cualquier usuario. *ODE* permite recuperar y cuantificar en una sola búsqueda cruzada todos los diminutivos presentes en el corpus. En nuestro caso, se han buscado todas las formas con las terminaciones pertinentes en *-ico*, *-ito* e *-illo* y sus variantes de género y número que han sido etiquetadas como diminutivos indicando lo siguiente en el buscador:

```
[pos = "NC.+D|AQD.+" ] :: match.text_province = "Almería|Madrid" & match.text_cat = "inv" & match.text_century = "XVIII|XIX"101
```

Para salvar la búsqueda el resultado de la concordancia se ha volcado en un documento Excel donde se ha mantenido la información extralingüística de los datos: provincia, siglo y tipo de diminutivo que ha permitido clasificar los datos en diferentes subcorpus.

Hasta ahora, el estudio cualitativo de los diminutivos transparentes se ha llevado a cabo en el conjunto total de textos que integran el corpus. Para el análisis cuantitativo que en este apartado se aborda es necesario separar estos textos en diferentes grupos o subcorpus a efectos de comparabilidad de los resultados. De esta manera, se tiene en

¹⁰¹ Cualquier nombre (NC) o adjetivo (AQ) que contenga un diminutivo (D) teniendo en cuenta la siguiente información extratextual: perteneciente a Almería o Madrid en inventarios de bienes (inv) y en los siglos XVIII y XIX. La sintaxis de búsqueda con la que funciona el buscador de *ODE* responde al lenguaje CQL y se explica detenidamente en § 4.8. *Búsquedas en el corpus basadas en lenguaje CQL*.

cuenta la distinción esencial entre el corpus de estudio y el corpus de control con la que se diseñó el corpus documental de esta tesis y que permitirá la distribución geográfica y cronológica de los datos que más adelante se exponen. Una descripción detallada de cada uno de estos corpus puede verse en § 2.5.2. *Corpus de estudio de Almería* y § 2.5.3. *Corpus de control de Madrid*.

A efectos prácticos, tanto los inventarios de Almería como los de Madrid conforman un único corpus en su conjunto. Sin embargo, en el diseño de la investigación podemos hablar de hasta cuatro subcorpus diferenciados que actúan como unidades de análisis independientes dentro de este corpus común. Entender cuáles son sus partes y sus características es importante para considerar las implicaciones que estas tendrán en el desarrollo y el resultado del análisis estadístico.

En un primer nivel de análisis podemos hablar de dos corpus distintos: el llamado corpus de estudio, constituido por todos los inventarios de Almería, y el corpus de control, conformado por los inventarios de Madrid. A su vez, cada uno de estos corpus se divide en otros dos subcorpus correspondientes a cada uno de los siglos estudiados. En total, cuatro subcorpus diferentes cuyo objetivo al cruzar los datos de todos ellos es poder determinar la evolución de los diminutivos en las zonas y el marco temporal representado.

Llegados a este punto conviene recordar el tamaño de los subcorpus, pues el número total de palabras será fundamental para hallar las frecuencias relativas del uso de los diminutivos. Como se explicará más adelante, este dato servirá para poder comparar los recuentos obtenidos para cada sufijo salvando las diferencias de tamaño que existen entre muestras distintas, para lo que se hace necesario considerar el volumen de palabras.

	Siglo	N.º palabras	Total
Almería Corpus de estudio	XVIII	27 544	53 313
	XIX	25 769	
Madrid Corpus de control	XVIII	29 980	65 504
	XIX	35 524	
			118 817

Tabla 21: División de la muestra documental en subcorpus y número de palabras

Teniendo en cuenta la diferenciación entre corpus de estudio y corpus de control, el recuento total de los diminutivos en el corpus asciende a 235 casos en el corpus de Madrid y 137 en el corpus de Almería. Estas cifras representan los datos globales de la aparición de palabras con sufijos diminutivos en ambas provincias. Más adelante, se mostrará el desglose de estos datos por siglos para cada uno de los subcorpus (§ 7.4. *Resultados*), detallando las frecuencias observadas y los porcentajes de uso de cada diminutivo, con el fin de realizar un análisis estadístico que permita abordar, tanto desde

la vertiente diacrónica como diatópica, el fenómeno de la variación de los diminutivos en los siglos XVIII y XIX en las provincias de Almería y Madrid.

7.2. Variables e hipótesis de estudio

El estudio morfológico tiene como principal objetivo analizar la evolución de los diminutivos *-ico*, *-ito* e *-illo* en la provincia de Almería en contraste con la de Madrid en los siglos XVIII y XIX. La frecuencia de uso de los diminutivos se erige como la variable lingüística de interés en el corpus, la cual opera en unos contextos muy concretos determinados probablemente desde el punto de vista cronológico y geográfico. El análisis, por tanto, se plantea en torno al contraste de tres variables, cada una de ellas con distintos valores o categorías:

- a. Una variable lingüística (*diminutivo*) con tres posibles valores correspondientes a cada uno de los sufijos que son objeto de estudio: *-ico*, *-ito* e *-illo*.
- b. Una variable cronológica (*siglo*) con dos valores que se relacionan con el siglo de procedencia: XVIII y XIX.
- c. Una variable geográfica (*provincia*) con dos valores asociados al lugar de origen: Almería y Madrid.

Las variables cronológica y geográfica constituyen en nuestro análisis variables binarias con dos valores posibles cada una: por un lado, XVIII y XIX en el caso de la variable correspondiente al siglo; por otro, Almería y Madrid para los valores asociados a la provincia. Estas se identifican con las variables explicativas, es decir, aquellas que explican o predicen el comportamiento de la variable lingüística y que, en nuestro caso, adopta tres valores diferentes asociados con cada uno de los tres sufijos.

Todas las variables de estudio se agrupan así en diferentes categorías. Se trata, en definitiva, de variables categóricas que nos permiten agrupar los datos y analizar el impacto que las variables explicativas (el siglo y la provincia) tienen sobre el empleo de los diminutivos. Precisamente el manejo de variables categóricas es lo que fijará el empleo de una metodología concreta para determinar la relevancia estadística de la relación que existe entre variables.

El análisis trata de dar respuesta a varias preguntas de investigación en torno al uso de los diminutivos a partir de los datos y corpus que se han descrito. Respecto a la cronología, una de las cuestiones principales tratará de determinar la influencia del paso del tiempo sobre la distribución de los diminutivos. La bibliografía coincide en el ascenso de *-ito* a partir de los siglos XV-XVI en un proceso lento y gradual que culminará con el triunfo de *-ito* sobre *-illo* (§ 5.2. *Periodización y extensión geográfica de -ico, -ito e -illo*). ¿En qué momento se produce este cambio?, ¿es posible apreciar la difusión de *-ito* en los corpus de Almería y Madrid en este periodo?, ¿es el siglo XVIII un momento clave en la evolución de los diminutivos como lo es para otros fenómenos del español? Por otro lado, estudios previos señalan en el siglo XVII el uso de *-ico*

como un fenómeno dialectal propio de determinadas zonas peninsulares, entre ellas Almería (Calderón Campos, 2019c, p. 120). ¿Sigue siendo *-ico* un rasgo característico de esta región en los siglos XVIII y XIX?, ¿hay diferencias entre ambos periodos o, por el contrario, hay una estabilización en lo que respecta al empleo de los diminutivos? Y, por último, ¿en qué porcentaje podemos decir que aumenta o disminuye la frecuencia de uso de los diminutivos en el siglo XIX frente al XVIII?

En cuanto a la distribución geográfica, ¿son significativas las diferencias en el uso de los diminutivos entre el corpus de estudio de Almería y el corpus de control de Madrid?, ¿es la difusión de *-ito* igual en ambas zonas?, ¿es el lugar de origen un factor determinante para la preferencia de un sufijo frente a otro?, y si es así, ¿en qué medida?

Todas estas cuestiones se materializan en las hipótesis de investigación que se someterán a prueba en el análisis cuantitativo. Por un lado, desde el punto de vista cronológico, se probarán las diferencias en el uso de los diminutivos en uno y otro siglo para cada una de las provincias de manera separada; por otro, los datos obtenidos se contrastarán en el plano diatópico para cada siglo.

Desde el punto de vista cronológico, la hipótesis nula (H_0) indica la no relación entre las variables, es decir, que el uso de los diminutivos permanece sin cambios significativos en los siglos XVIII y XIX, mientras que la hipótesis alternativa (H_1) postula justamente lo contrario, esto es, que hay una diferencia significativa en el uso de los diminutivos en función del siglo en cada provincia (hipótesis cronológica).

Por otro lado, además de las posibles diferencias temporales, en esta tesis se comprueba si la relación que existe entre la provincia y el uso de los diminutivos es significativa (hipótesis diatópica). Se trata de analizar la distribución de los sufijos también desde una perspectiva diatópica: el objetivo es comprobar si las diferencias espaciales no afectan de manera particular al uso de los diminutivos en Madrid y Almería (H_0) o si, por el contrario, estas diferencias son significativas y en ellas deben tenerse en cuenta factores geográficos (H_1). La asociación entre la variable geográfica y lingüística se analizará para cada siglo de forma independiente.

De esta manera, nos acercamos al estudio de los diminutivos poniendo el foco en factores extralingüísticos como son el marco cronológico y el origen geográfico. La formulación de ambas hipótesis quedaría desglosada de la siguiente manera:

*fr. = frecuencia

a. Hipótesis cronológica

H_0 = fr. * *-ico* / *-ito* / *-illo* Almería XVIII = fr. XIX

H_1 = fr. *-ico* / *-ito* / *-illo* Almería XVIII $\neq$ fr. XIX

H_0 = fr. *-ico* / *-ito* / *-illo* Madrid XVIII = fr. XIX

H_1 = fr. *-ico* / *-ito* / *-illo* Madrid XVIII $\neq$ fr. XIX

b. Hipótesis diatópica

H_0 = fr. (-ico, -ito, -illo) Almería XVIII = fr. (-ico, -ito, -illo) Madrid XVIII

H_1 = fr. (-ico, -ito, -illo) Almería XVIII $\neq$ fr. (-ico, -ito, -illo) Madrid XVIII

H_0 = fr. (-ico, -ito, -illo) Almería XIX = fr. (-ico, -ito, -illo) Madrid XIX

H_1 = fr. (-ico, -ito, -illo) Almería XIX $\neq$ fr. (-ico, -ito, -illo) Madrid XIX

7.3. Modelo estadístico empleado y estructura del análisis

La anotación morfosintáctica del corpus (§ 4.5. *Anotación morfosintáctica (POS) en el corpus ODE*) permite que con un par de búsquedas al alcance de cualquier usuario se pueda obtener una cuantificación fiable del total de los diminutivos presentes en los inventarios elegidos. Sin embargo, más allá de ofrecer un estudio lingüístico a partir de la descripción de frecuencias absolutas y porcentajes, en esta tesis se hace un tratamiento estadístico más profundo de los datos que permita determinar de manera precisa y objetiva en qué medida los factores geográfico y cronológico explican la presencia de los diminutivos. El análisis tratará además de evaluar la significancia de las diferencias que existen entre las distintas muestras analizadas causadas por estos factores.

Para el análisis estadístico dispondremos los datos en forma de tabla de contingencia 3x2 en la que se muestran todos los recuentos de cada uno de los posibles valores de la variable lingüística. En nuestro caso, el fenómeno de variación se divide en tres (-ico, -ito e -illo), de modo que cada celda de la tabla recoge los casos de los tres diminutivos (columnas) agrupados en torno a los dos valores (filas) de la variable explicativa estudiada en cada caso, tal y como se sintetiza en la siguiente tabla de contingencia:

Sufijo Siglo	-ICO	-ITO	-ILLO
S. XVIII	44 (52.4 %)	17 (20.2 %)	23 (27.4 %)
S. XIX	4 (7.6 %)	14 (26.4 %)	35 (66 %)

Tabla 22: Tabla de contingencia. Frecuencias observadas de los diminutivos en Almería

La tabla recoge las frecuencias observadas de los diminutivos en Almería, por lo que la suma de todas ellas representa el total de los casos en el corpus –se representa así el fenómeno de estudio en su conjunto–. De esta manera, de cada diminutivo puede hallarse la proporción de uso o porcentaje en el corpus, representando la suma de todos los casos el 100 %.

Puesto que en todo momento se analizan los datos agrupados de variables categóricas, el test de Chi-cuadrado (χ^2) es la prueba estadística más apropiada en este caso (Brezina, 2018, p. 112; Stefanowitsch, 2020, p. 177). Esta prueba se utilizará para

determinar la relación existente entre las variables y derivada de ella se calcularán también otras medidas que permitan examinar el tamaño global del efecto (V de Cramer, razón de probabilidad), así como la relevancia estadística de cada uno de los valores observados de los diminutivos (residuos de Pearson). A continuación, se detalla la metodología específica llevada a cabo para el desarrollo del análisis estadístico.

De manera separada, se analiza por un lado el corpus de estudio y, por otro, el corpus de control. Para cada caso el análisis comienza por conocer cómo están distribuidos los datos en el corpus con el objetivo de detectar frecuencias anómalas que podrían distorsionar la interpretación de los resultados. Para ello realizaremos un sencillo estudio de la dispersión de los diminutivos en el corpus. En estadística, se entiende por *dispersión* la distribución del fenómeno de estudio a lo largo del corpus o de las diferentes partes que lo componen (Brezina, 2018, p. 47). Una concentración excesivamente alta de un sufijo diminutivo cualquiera en alguno de los textos podría estar revelando un uso anómalo de los mismos, el cual haría considerar este hecho como un rasgo particular del escribano que nada tendría que ver con el uso real del diminutivo en la región. Explorar los datos del corpus es importante para detectar posibles patrones de comportamiento extraños que resulten llamativos y que contrasten con la distribución que en un principio sería esperable. De esta manera, se hace indispensable reconocer cuál es la distribución de los diminutivos en cada uno de los subcorpus analizados para no dar lugar a conclusiones equivocadas.

Para el tratamiento estadístico de los datos se ha utilizado el programa RStudio (2021, versión 4.1.0). Para cada una de las provincias se someterán a prueba las hipótesis formuladas en el apartado anterior (§ 7.2. *Variables e hipótesis de estudio*). En primer lugar, se probará la significancia estadística de los datos a partir de la prueba de Chi-cuadrado de Pearson (χ^2) o, en su defecto, de *Log-likelihood* (LL) cuando esta sea la más indicada¹⁰². El valor de *p* obtenido, siempre que este sea igual o inferior al 5 % ($p < 0.05$), servirá para descartar con suficiente evidencia la hipótesis nula, así como para determinar si existen diferencias significativas entre las variables en contraste.

Las pruebas de significancia (χ^2 y LL) no aportan información sobre cómo de fuerte es la relación entre las variables. Para apreciar la magnitud de la asociación existen diferentes medidas que permiten cuantificar en términos estadísticos la fuerza o intensidad de esa asociación una vez que se ha demostrado que las diferencias entre las variables son significativas¹⁰³. En tablas de contingencia mayores de 2x2 o en las que el número de filas no coincide con el de columnas (como es nuestro caso), la V de Cramer resulta la medida más apropiada para calcular el tamaño del efecto (Desagulier, 2017, p. 183). Su cálculo servirá para cuantificar la fuerza global de la relación entre las variables reflejadas en la tabla de contingencia. Esta medida refiere un número entre 0 y

¹⁰² El 80% de las frecuencias esperadas en una tabla de contingencia deben ser mayores de 5 para que la prueba de χ^2 se considere apropiada (Brezina, 2018, p. 113). Este supuesto no se cumple en nuestro análisis en dos ocasiones, concretamente cuando en las tablas de contingencia se incluyen los diminutivos del siglo XIX en el corpus madrileño. En su lugar, el test de *Log-likelihood* o el test exacto de Fisher resultan los más indicados (Brezina, 2018, p. 113).

¹⁰³ Un listado de las pruebas existentes para cuantificar el tamaño del efecto especialmente aplicadas a la lingüística de corpus puede verse en Brezina (2018, p. 275).

1, donde el 0 se traduce en la total independencia de las variables entre sí y el 1 en su perfecta asociación. La mayor o menor cercanía a estos valores se interpreta de la siguiente forma tomando como referencia la escala de Kotrlik y Williams (2003, p. 5):

<i>Valores estimados de V</i>	<i>Interpretación de la asociación</i>
0.00 – 0.10	Insignificante
0.10 – 0.20	Débil
0.20 – 0.40	Moderada
0.40 – 0.60	Relativamente fuerte
0.60 – 0.80	Fuerte
0.80 – 1.00	Muy fuerte

Tabla 23: Escala interpretativa de la V de Cramer (Kotrlik y Williams, 2003)

Una vez calculada la fuerza global de la asociación, se calculará con los residuos de Pearson cuál de las frecuencias obtenidas influye en mayor medida en la significatividad de los datos. Este cálculo compara las frecuencias observadas de los diminutivos con sus frecuencias esperadas para determinar cuál de ellos es estadísticamente más notable según el contexto analizado¹⁰⁴. Así, tomando como datos de entrada la tabla de contingencia, se analiza qué celda tiene mayor peso en el total. Aquellos residuos que más se alejan del 0 reflejan un mayor efecto sobre los resultados (Desagulier, 2017, p. 183). Esta prueba se considera útil en nuestro estudio para evaluar cuál de todos los diminutivos es el que aporta el valor más significativo a la hora de establecer la dependencia de las variables. Nos basaremos en la visualización de los residuos en un gráfico de asociación para interpretar de forma más clara los resultados.

De manera complementaria a las pruebas estadísticas mencionadas, las diferencias en el uso de los diminutivos serán evaluadas de manera más precisa hallando la probabilidad de que aparezca *-ico*, *-ito* o *-illo* en un contexto determinado. Para ello se utilizará el cálculo de la razón de probabilidad (PR)¹⁰⁵, otra medida del tamaño del efecto que permitirá analizar la aparición de los diminutivos al comparar dos valores entre sí de una misma variante (Brezina, 2018, p. 115).

Por último, se destacan los aspectos más relevantes del análisis efectuado para cada una de las provincias desde la perspectiva cronológica y para cada uno de los siglos desde una perspectiva diatópica. Los resultados de las pruebas de significación estadística se complementan con un análisis sencillo a partir de las frecuencias relativas que servirá a modo de resumen para representar de forma gráfica y más clara las tendencias expresadas en el uso de los diminutivos de acuerdo con los contextos estudiados. Este análisis se complementa con el cálculo de los porcentajes de aumento o

¹⁰⁴ La frecuencia esperada (FE) es la frecuencia que cabría esperar si no hubiera relación entre las variables analizadas, es decir, en el caso de que la hipótesis nula fuera cierta (Brezina, 2018, p. 113). Su cálculo se obtiene a partir de las frecuencias observadas. Su cálculo se obtiene automáticamente en R a partir de los valores de χ^2 .

¹⁰⁵ Del inglés, *probability ratio*. También llamado *relative risk* o *risk ratio* (RR).

disminución de cada diminutivo, el cual permitirá ofrecer una mejor idea del panorama de evolución del fenómeno lingüístico en las provincias estudiadas. Estos porcentajes solo se calcularán para el contraste cronológico de los diminutivos en Almería y Madrid y no para el diatópico, puesto que es una medida para cuantificar de manera más evidente el cambio a lo largo del tiempo.

Durante todo el análisis, trataremos de apoyarnos siempre en la medida de lo posible en la representación visual de los datos mediante gráficos o figuras. Así, la proporción de los datos en las tablas de contingencia se refleja en forma de gráficos cuadrangulares (*vid.* Figura 79 y Figura 84), para los que se ha utilizado la herramienta Lancaster Stats Tools¹⁰⁶ (Brezina, 2022), mientras que con el programa RStudio se ha hecho el gráfico para mostrar los residuos de Pearson¹⁰⁷. Para visualizar las tendencias ascendentes y descendentes se emplean gráficos de líneas o de barras elaborados con Excel que facilitan la comparación, ya sea cronológica o diatópica, del uso de los diminutivos. Los mapas utilizados para representar la distribución geográfica de los diminutivos se han realizado con el programa de software libre Tableau Desktop (2023, versión, 2023.1). Con este mismo programa se han hecho también los gráficos de dispersión para cada uno de los corpus analizados (Almería y Madrid).

El modelo de análisis descrito se reproduce para cada una de las provincias estudiadas y, posteriormente, para el análisis diatópico. De esta manera el análisis se divide en dos grandes partes. La primera de ellas está dedicada al estudio cronológico, realizado de manera independiente para cada uno de los subcorpus: Almería, por un lado, y Madrid por otro. En estas secciones se presenta la dispersión de los datos, seguida de las pruebas de estadística inferencial nombradas. Por último, se sintetizan los aspectos más destacados y los resultados obtenidos se ratifican a modo de recapitulación con una breve comparación de las frecuencias relativas. Las pruebas de significación estadística se realizan para cada una de las variables en contraste, de modo que la estructura del análisis se repite para cada uno de los subcorpus analizados. El análisis en su conjunto realizado en esta tesis queda estructurado de la siguiente manera:

1. Análisis cronológico

- a. Corpus de estudio (Almería):
 - Dispersión de los datos.
 - Pruebas de significación estadística:
 - Chi-cuadrado (χ^2)
 - V de Cramer
 - Residuos de Pearson
 - Razón de probabilidad (PR)
 - Recapitulación y análisis descriptivo:
 - Frecuencias relativas
 - Porcentajes de aumento y disminución

¹⁰⁶ <http://corpora.lancs.ac.uk/stats/>

¹⁰⁷ Para elaborar el gráfico de asociación de los residuos se ha utilizado en R la función `assocplot()` del paquete ‘vcd’ (Desagulier, 2017, p. 184).

- b. Corpus de control (Madrid):
 - Dispersión de los datos.
 - Pruebas de significación estadística:
 - Chi-cuadrado (χ^2) y Log-likelihood (LL)
 - V de Cramer y razón de probabilidad (PR)
 - Residuos de Pearson
 - Recapitulación y análisis descriptivo:
 - Frecuencias relativas
 - Porcentajes de aumento y disminución

2. Análisis diatópico

- c. Pruebas de significación estadística:
 - Contraste Almería vs. Madrid en el siglo XVIII:
 - Chi-cuadrado (χ^2)
 - V de Cramer
 - Residuos de Pearson
 - Razón de probabilidad (PR)
 - Contraste Almería vs. Madrid en el siglo XIX:
 - Chi-cuadrado (χ^2) y Log-likelihood (LL)
 - V de Cramer y razón de probabilidad (PR)
 - Residuos de Pearson.
- d. Distribución geográfica:
 - Mapeado de los datos y comparación de frecuencias relativas.

7.4. Resultados

En este apartado se muestran los resultados obtenidos tras el recuento de los diminutivos transparentes. Se analizan de manera separada los datos del corpus de Almería, en primer lugar, y los del corpus de Madrid, seguidamente, para evaluar las diferencias cronológicas en una y otra región. Por último, estos resultados se contrastan para obtener un análisis diatópico.

7.4.1. Corpus de Almería

7.4.1.1. Dispersión de los datos en el corpus de Almería

Entendemos que un buen análisis sobre la evolución del fenómeno morfológico que nos interesa debería sustentarse en la distribución más o menos homogénea de los diminutivos a lo largo de los textos, de manera que no se encuentren aglutinados solo en unos pocos de ellos o en puntos geográficos muy concretos, sino más bien repartidos por la totalidad del corpus.

Para este estudio nos limitamos a un reconocimiento básico de la dispersión para cada siglo en el que se examina la concentración de diminutivos en todos los documentos del corpus que contienen, al menos, un caso. Este análisis da como resultado un gráfico para el siglo XVIII (Figura 77) y otro para el siglo XIX (Figura 78) en los que se muestra el total de documentos que contienen diminutivos y su distribución en forma de barras de frecuencia. En la parte superior se muestra el total de documentos en los que aparece algún diminutivo en el corpus de Almería (numerados con el código que identifica de manera interna a cada documento en el corpus), mientras que en la parte inferior se señala el municipio del que proceden. En cada uno de estos documentos puede haber ocurrencias de solo un tipo de diminutivo, de dos o de todos ellos (-ico, -illo o -ito). En cualquiera de los casos, la altura de la barra indica una frecuencia mayor o menor del conjunto de diminutivos.

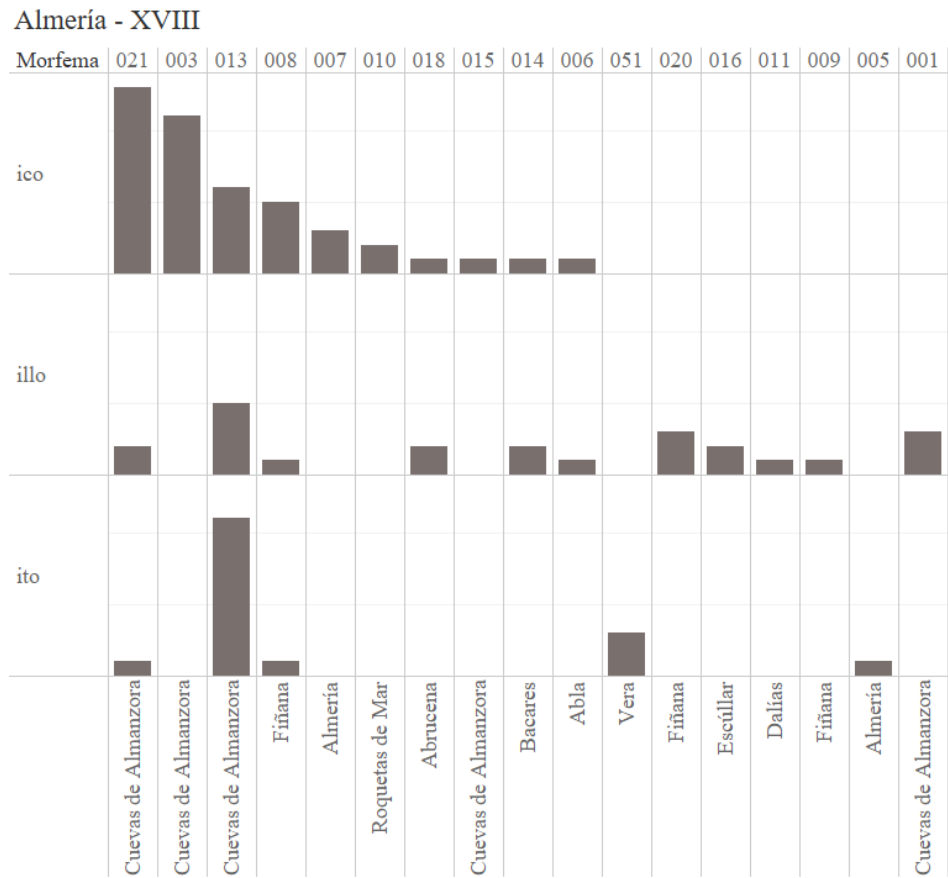


Figura 77: Dispersión de los diminutivos en el corpus de Almería (s. XVIII)

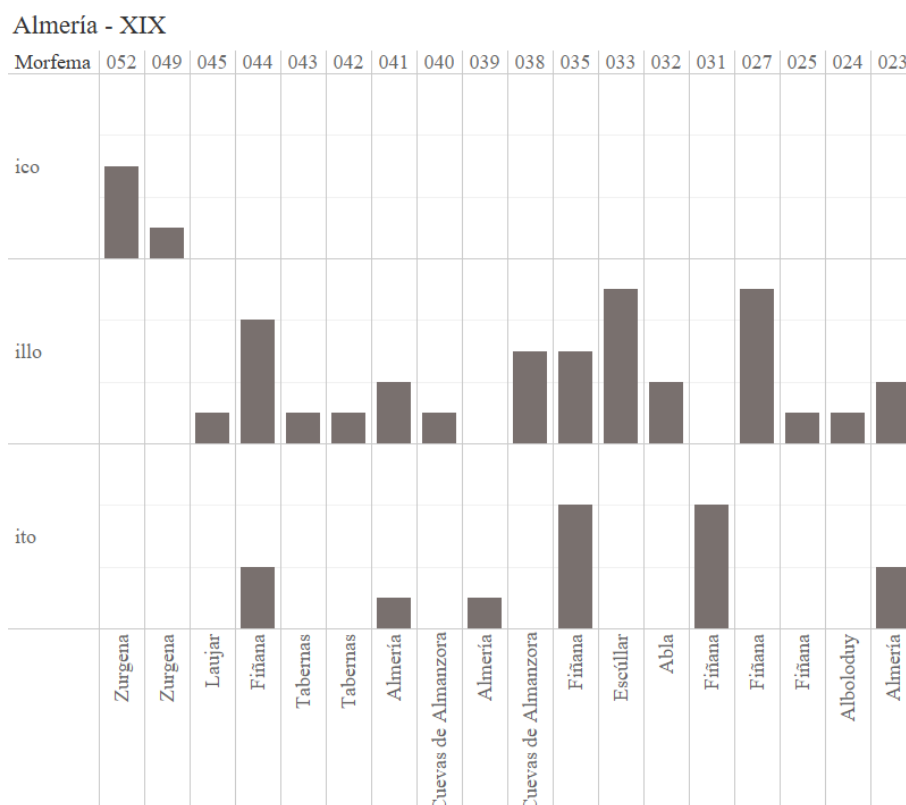


Figura 78: Dispersión de los diminutivos en el corpus de Almería (s. XIX)

La exploración de los diminutivos partiendo del análisis de su dispersión revela lo siguiente: de los 24 documentos que conforman el corpus de Almería del siglo XVIII, en 17 de ellos se localiza al menos un diminutivo. De entre ellos, *-ico* está presente en más de la mitad (7 en total). Sobresalen con una frecuencia más alta los pertenecientes a Cuevas de Almanzora, aunque cabe señalar que estos textos en particular tienen una extensión superior a la media, hecho que explica el mayor número de diminutivos en estos documentos.

En el siglo XIX (Figura 78), se localiza al menos un diminutivo en 18 documentos del total de 28 que constituyen este subcorpus. La ausencia de dispersión para el diminutivo *-ico* coincide con un descenso en el uso de esta variante. Esta distribución no debe interpretarse como un defecto de los datos, sino como una evidencia del declive que este diminutivo experimenta en este siglo (§ 5.2. *Periodización y extensión geográfica de -ico, -ito e -illo*; **Error! No se encuentra el origen de la referencia.**). En lo que respecta a los sufijos *-ito* e *-illo*, nada parece indicar una dispersión anómala; la distribución responde al comportamiento que para la misma época se espera de estos diminutivos.

7.4.1.2. Relación estadística entre diminutivo y siglo en el corpus de Almería

El análisis que se plantea evalúa en términos estadísticos la cronología de los diminutivos en Almería a lo largo de los siglos XVIII y XIX. El estudio confirmará la

existencia de diferencias notables en el uso de los diminutivos en este periodo, determinará el tamaño de esas diferencias o si estas son simplemente producto del azar. Se pretende probar la hipótesis de partida según la cual el uso de los diminutivos en Almería es desigual en los siglos XVIII y XIX y que formulábamos de la siguiente manera:

$$H_0 = \text{fr. -ico / -ito / -illo Almería XVIII} = \text{fr. XIX}$$

$$H_1 = \text{fr. -ico / -ito / -illo Almería XVIII} \neq \text{fr. XIX}$$

La significatividad estadística de los datos se evalúa a partir de la siguiente tabla de contingencia con el objeto de analizar la relación existente entre las variables estudiadas:

Sufijo Siglo	-ICO	-ITO	-ILLO
S. XVIII	44 (52.4 %)	17 (20.2 %)	23 (27.4 %)
S. XIX	4 (7.6 %)	14 (26.4 %)	35 (66%)
X-squared test (χ^2): 30.662, df = 2, p = 2.19729e-7 Cramer's V = 0.473			

Tabla 24: Tabla de contingencia. Distribución de los diminutivos por siglos (XVIII-XIX) en la provincia de Almería. Frecuencias observadas y porcentajes

El recuento de los diminutivos en el conjunto total de textos que conforman el corpus de Almería asciende a 137 formas con diminutivo. Junto a la frecuencia de cada diminutivo en el corpus, lo que habitualmente se denomina en términos estadísticos *frecuencia observada* (FO) o *frecuencia absoluta*, se incluyen entre paréntesis los porcentajes, calculados a partir del total de cada fila (se señalan en negrita los más relevantes para cada siglo). Los porcentajes permiten comparar más fácilmente la frecuencia de aparición de cada uno de los diminutivos en los siglos estudiados y entender cómo se distribuyen los datos en el corpus. Por ejemplo, se puede decir que en el siglo XVIII es clara la preferencia de *-ico* (52.4 %) en la región almeriense frente a *-ito* e *-illo*, lo que indica una probabilidad de aparición de *-ico* de más de la mitad de los casos. Por otro lado, en el XIX *-illo* es el diminutivo preferido en los inventarios almerienses, con una proporción del 66 %.

La información contenida en la Tabla 24 se representa visualmente en la Figura 79. Por un lado, cada tono de gris simboliza uno de los diminutivos, dispuestos horizontalmente; por otro, verticalmente se reflejan de manera proporcional las frecuencias de aparición en cada siglo. El tamaño de las distintas áreas rectangulares es directamente proporcional a la mayor o menor frecuencia de uso de cada uno de los valores de la variable lingüística. Así, la altura de cada área equivale a esta frecuencia en el corpus teniendo en cuenta la variable explicativa, mientras que la anchura de las columnas se corresponde con el tamaño de la muestra (Brezina, 2018, p. 109).

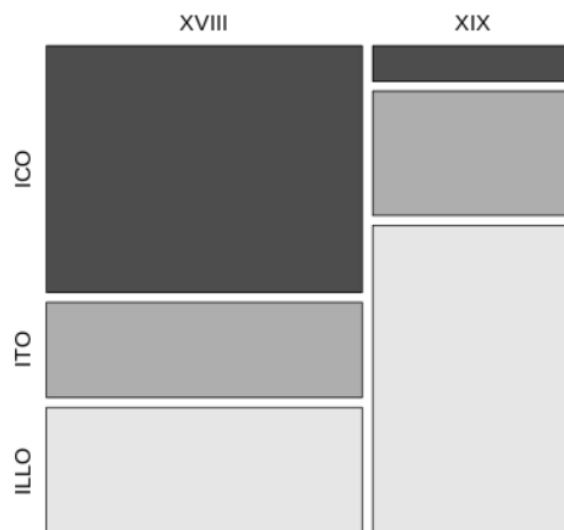


Figura 79: Distribución proporcional de los diminutivos en Almería en los ss. XVIII y XIX

La interpretación del cuadro deja ver que en el siglo XVIII el diminutivo más frecuente era *-ico*, tal y como muestra el tamaño mayor de esta área sombreada en un color más oscuro. Por el contrario, esta situación se invierte en el siglo XIX e *-ico* pasa a reflejar el menor porcentaje del total de los diminutivos. Los sufijos *-ito* e *-illo* reflejan en el XVIII un uso más o menos similar, mientras que en el XIX *-illo* alcanza una sobrada ventaja frente al resto.

El resultado obtenido en la prueba de Chi-cuadrado (χ^2) estima que la diferencia en el uso de los diminutivos entre uno y otro siglo es estadísticamente significativa ($p < 0.05$). A partir del valor de p es posible rechazar la hipótesis nula asumiendo que hay evidencias estadísticas suficientes para afirmar que los datos de las muestras analizadas no son fruto de la casualidad y tampoco se deben simplemente a un error de cálculo o de la recogida de los datos. Se acepta, por tanto, que las diferencias entre las frecuencias observadas no se deben al azar.

Por otro lado, el bajo valor de p establece que hay una asociación real entre las variables analizadas. La variable explicativa (el siglo) tiene un efecto sobre la variable lingüística (el tipo de diminutivo), lo que nos permite asegurar que existe una relación entre ambas y que esta es significativa en términos estadísticos. La medida del tamaño del efecto cuantifica la fuerza de esa asociación, siendo en este caso relativamente fuerte (V de Cramer = 0.47) de acuerdo con la escala de valores anteriormente descrita (§ 7.3. *Modelo estadístico empleado y estructura del análisis*, Tabla 23). Así, podemos concluir que la relación existente entre las variables es significativa y de una fuerza considerable.

A partir de estos datos, deducimos que el paso del tiempo influye de manera significativa y con relativa fuerza en el uso de los diminutivos en Almería. En consecuencia, el tiempo es un factor determinante que debe considerarse en la evolución de los diminutivos, ya que, concretamente en esta región, ambas variables se encuentran estrechamente relacionadas. Sin embargo, no todas las frecuencias de la Tabla 24 influyen de la misma manera en la relevancia estadística de los datos. En lo que sigue se

analizará cuál de todas las cifras es la que aporta un valor más significativo al resultado que se acaba de mostrar.

Basándonos en la diferencia entre las frecuencias observadas (FO) y las frecuencias esperadas (FE) obtenemos lo que en estadística se conoce como *residuos de Pearson*. Esta medida derivada de χ^2 se utilizará para comprobar cuál de los datos observables de los diminutivos en el corpus de Almería es más significativo. La mayor desviación entre las frecuencias observadas y esperadas advierte cuáles de los datos de entrada aportan los valores más relevantes. El cálculo de los residuos se refleja en la Tabla 26, cuya representación en el gráfico de asociación (Figura 80) será en la que nos basemos para una mejor interpretación de los resultados obtenidos.

	XVIII	XIX
ICO	29.43066	18.56934
ITO	19.00730	11.99270
ILLO	35.56204	22.43796

Tabla 25: Frecuencias esperadas de los diminutivos en Almería a partir de χ^2

	XVIII	XIX
ICO	2.6855917	-3.3809730
ITO	-0.4604176	0.5796337
ILLO	-2.1065266	2.6519703

Tabla 26: Residuos de Pearson para los diminutivos en el corpus de Almería

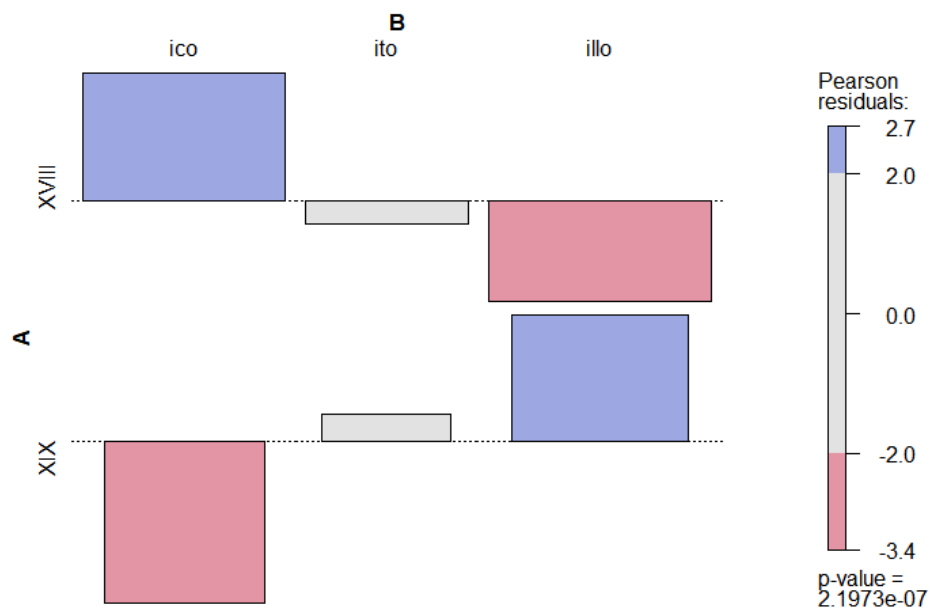


Figura 80: Asociación de las variables siglo/diminutivo en Almería basada en los residuos de Pearson

La línea de puntos o línea base en cada siglo representa el 0 y se traduce en una total independencia entre el siglo (eje A) y el diminutivo (eje B). Los diminutivos por debajo de la línea base reflejan una FO menor que su respectiva FE (residuos en negativo en la

Tabla 26), mientras que las casillas por encima de la línea base se relacionan con una frecuencia observada superior a la que cabría esperar (residuos en positivo)¹⁰⁸. Los valores que más se desvían del 0, tanto si lo hacen al alza como a la baja, reflejan una mayor discrepancia entre FE y FO que el resto, lo que se traduce también en una mayor relevancia estadística: son los que en el gráfico aparecen coloreados, puesto que superan los 2 puntos, ya sea por encima de la línea base (casillas en azul) o por debajo (casillas en rojo); se resaltan así los datos más significativos. Estas celdas destacan porque los residuos de Pearson (diferencia entre FE y FO) se alejan lo suficiente del 0 como para resultar relevantes estadísticamente. En color gris, en cambio, aparecen aquellos diminutivos en los que la diferencia entre los valores observados y los valores previstos no es tan significativa.

En nuestro análisis, los datos que juegan un papel más relevante corresponden a los diminutivos *-ico* e *-illo* en ambos siglos. Primeramente, la frecuencia de *-ico* en el XVIII superior a la esperada contrasta significativamente con la frecuencia obtenida en el XIX, muy por debajo de la esperada. De forma paralela, el diminutivo *-illo* experimenta el caso contrario y manifiesta una evolución inversamente proporcional a la de *-ico*. Por su parte, *-ito* no desempeña un papel relevante en el proceso de cambio que afecte a la significancia de los datos. Su uso aumenta ligeramente en el siglo XIX situándose por encima de la frecuencia esperada con respecto al siglo anterior.

El cálculo de los residuos de Pearson confirma estadísticamente que este sufijo es sobradamente el más utilizado en el XVIII, pero experimenta una importante caída después, en el XIX, cuando es relevado por *-illo*. Más novedoso es en cambio el resultado que extraemos en lo relativo a *-illo*: la diferencia observada en el uso de este diminutivo entre el siglo XVIII y XIX es también muy significativa desde el punto de vista estadístico.

En resumen, los datos de *-ico* e *-illo* (aquellos que superan los 2 puntos en los residuos), son los que más contribuyen a la no independencia entre la variable siglo y diminutivo. Son, por tanto, los que de forma más significativa permiten descartar la hipótesis nula y aceptar su alternativa: el paso del tiempo tiene un claro efecto en la distribución de los diminutivos en Almería, lo que nos permite afirmar que el uso de los diminutivos difiere significativamente entre el siglo XVIII y el XIX.

Los datos recién analizados reflejan en Almería una tendencia bastante clara y opuesta entre el uso de *-ico* e *-illo*: el aumento de uno hace que el otro disminuya y, al contrario. En concreto, en el XVIII la preferencia de *-ico* explica una menor frecuencia de *-illo*, mientras que la disminución de *-ico* en el XIX conlleva el aumento de *-illo* (Figura 80).

Una percepción más precisa de este cambio nos permite determinar la probabilidad de que ocurra determinado fenómeno lingüístico en cada uno de sus

¹⁰⁸ El tamaño de las celdas en el gráfico es proporcional a la diferencia entre las frecuencias (FO y FE). La altura de cada casilla equivale a su residuo de Pearson correspondiente (escala lateral derecha), mientras que la anchura es proporcional a la raíz cuadrada de los valores esperados (Desagulier, 2017, p. 184).

posibles contextos a partir del cálculo de la razón de probabilidad (PR)¹⁰⁹. Esta medida cuantifica el tamaño del efecto comparando entre sí dos valores de la variable lingüística. En nuestro caso, nos permitirá conocer cuál es la probabilidad de aparición de los diminutivos *-ico*, *-illo* e *-ito* en los siglos XVIII y XIX comparando estos ratios entre sí. Puesto que esta medida estadística determina una probabilidad entre dos valores, para nuestros datos existen las siguientes opciones de probabilidad:

$$\text{a. PR de -ico en s. XVIII vs. XIX} = \frac{0.523}{0.075} = \mathbf{6.97}$$

$$\text{PR de -ico en s. XIX vs. XVIII} = \frac{0.075}{0.523} = 0.14$$

$$\text{b. PR de -illo en s. XVIII vs. XIX} = \frac{0.273}{0.660} = 0.41$$

$$\text{PR de -illo en s. XIX vs. XVIII} = \frac{0.660}{0.273} = \mathbf{2.41}$$

$$\text{c. PR de -ito en s. XVIII vs. XIX} = \frac{0.202}{0.264} = 0.76$$

$$\text{PR de -ito en s. XIX vs. XVIII} = \frac{0.264}{0.202} = 1.2$$

Se destacan los resultados más relevantes. El diminutivo *-ico* es casi 7 veces más probable que aparezca en el siglo XVIII que en el XIX. Por el contrario, la probabilidad de que aparezca en el XIX es tan solo de 0.14 veces superior a la probabilidad de que aparezca en el XVIII, un valor muy pequeño en comparación, lo que quiere decir que su empleo en el XIX es mucho menos probable que en el XVIII, según los datos del corpus. Por su parte, el diminutivo *-illo* aumenta su probabilidad de aparición conforme pasa el tiempo: 2.4 veces superior en el XIX con respecto al siglo anterior. Estos datos reflejan nuevamente la tendencia opuesta que se ha señalado para los diminutivos *-ico* e *-illo*: los valores positivos de *-ico* en el XVIII contrastan con los valores negativos de *-illo* en este mismo siglo, y viceversa: en el XIX el descenso de *-ico* se traduce en un aumento de *-illo* de hasta 2.4 veces su probabilidad de aparición frente al siglo XVIII.

Por último, del diminutivo *-ito* se constata un aumento en el siglo XIX de 1.2 veces la probabilidad de que ocurra en el XVIII. La evolución de este diminutivo, en cambio, es la menos destacada con respecto al resto de los sufijos, tal y como ya se había revelado previamente (los residuos de *-ito* –en color gris en la Figura 81– fueron los menos significativos).

En conclusión, los resultados de la PR muestran que las probabilidades de encontrar *-ico* en Almería son superiores en el siglo XVIII frente al XIX: su uso va decreciendo de forma muy considerable; por el contrario, las probabilidades de *-illo*

¹⁰⁹ Calculada a partir de los porcentajes del total de cada fila (Tabla 24) sin multiplicar por 100 (Brezina, 2018, p. 115)

aumentan con el paso del XVIII al XIX. Asimismo, no se observan diferencias muy notables en el uso de *-ito* entre estos dos siglos.

7.4.1.3. Disminución de *-ico* y aumento de *-illo* en el corpus de Almería

Las pruebas realizadas hasta ahora han confirmado que existen diferencias significativas entre uno y otro siglo en el uso de los diminutivos mencionados. Asimismo, se ha demostrado que el paso del tiempo tiene un efecto considerable en el empleo de los diminutivos en esta provincia, particularmente sobre *-ico* e *-illo*, y se ha medido la intensidad de la asociación entre estas variables.

Del estudio de los diminutivos en Almería podemos destacar dos aspectos clave que sobresalen a la luz de los datos reflejados en el corpus y a los que se ha hecho referencia en repetidas ocasiones a lo largo de todo el análisis estadístico: por un lado, la preferencia por el diminutivo *-ico* en el siglo XVIII en esta región; por otro, la considerable disminución de este mismo sufijo en el XIX y, en consecuencia, el aumento de *-illo* en su lugar. Como resumen de todo lo expuesto, esta distribución se analiza a partir del contraste entre la frecuencia relativa (FR) de cada diminutivo –también llamada frecuencia normalizada–, calculada como resultado del cociente de la FO entre el número total de palabras de cada corpus y multiplicado por 10 000 como base de normalización¹¹⁰:

	-ICO	-ITO	-ILLO	TOTAL
S. XVIII	15.97	6.17	8.35	30.49
S. XIX	1.55	5.43	13.58	20.56

Tabla 27: Frecuencias relativas de los diminutivos en Almería (ss. XVIII-XIX)

La frecuencia relativa permite cotejar dos o más corpus (en este caso los relativos a los siglos XVIII y XIX) teniendo en cuenta la diferencia en el tamaño de las muestras para asegurar la comparabilidad. De lo contrario, el contraste de dos corpus que difieren en número de palabras podría dar como resultado una visión distorsionada de los datos que no se corresponde con la realidad. Esta medida permite además estimar promedios: el resultado obtenido para el sufijo *-ico* en el siglo XVIII (FR = 15.97) estaría indicando que, en un hipotético corpus de diez mil palabras, la frecuencia de aparición de voces con este diminutivo sería de 15.97 veces. Es decir, aproximadamente 16 palabras de cada diez mil en el corpus contendrán este sufijo, lo que supone una frecuencia bastante alta si se compara con la del resto de diminutivos en esta centuria.

¹¹⁰ El cálculo de la frecuencia relativa por millón de palabras es lo más común en lingüística de corpus. No obstante, en corpus más pequeños como este, lo más apropiado es utilizar una base de normalización cercana al tamaño del corpus estudiado. Utilizar cantidades demasiado grandes en relación con el tamaño real del corpus podría incrementar el número de casos de manera artificial, alterando, por lo tanto, el número de casos reales (Brezina, 2018, p. 43).

Este análisis es independiente al modelo estadístico que se ha desarrollado hasta ahora y en el que se han tenido en cuenta las frecuencias observadas, y no relativas, de los diminutivos, pero nos sirve para apreciar visualmente con mayor claridad las tendencias de uso de los diminutivos en el corpus de Almería en los siglos XVIII y XIX:

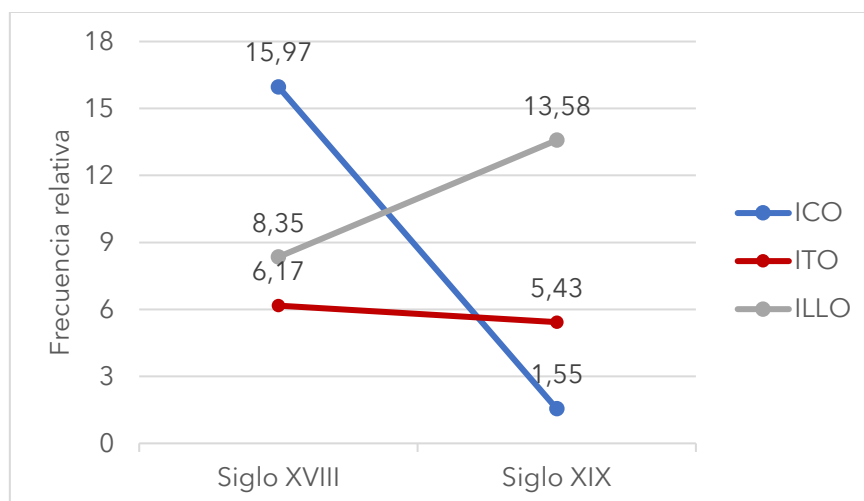


Figura 81: Distribución diacrónica de los diminutivos en Almería en los siglos XVIII-XIX

Es llamativa la evolución descendiente del diminutivo *-ico* en la provincia de Almería. En el siglo XVIII, el uso de este sufijo es un rasgo característico de la región, a la vista de su elevada frecuencia en contraposición con la de sus análogos *-ito* e *-illo*, cuyas frecuencias se distancian en buena medida de la de *-ico*. Esta situación cambiará en el siglo XIX, cuando se asiste a un cambio radical de los resultados. Este periodo destaca por el brusco descenso de *-ico*, invirtiéndose por completo la posición privilegiada de que gozaba en el XVIII: la frecuencia de aparición de *-ico* pasa a ser la más baja de los diminutivos en el corpus. Esta disminución conlleva necesariamente un aumento del resto del paradigma, especialmente de *-illo*, ya ligeramente superior a *-ito*, que experimentará un ascenso hasta convertirse en el diminutivo más habitual en el XIX. Por otra parte, *-ito* es el diminutivo con la distribución más estable a lo largo del marco temporal estudiado. Su frecuencia se mantiene en el XIX con unos valores similares a los que ya venía reflejando con anterioridad, aunque con un uso ligeramente inferior en este siglo con respecto al XVIII.

La proporción en la que aumenta o disminuye el uso de los diminutivos en Almería puede hallarse fácilmente a partir también de las frecuencias relativas como complemento al gráfico anterior. Los porcentajes de incremento y disminución permiten establecer en qué proporción aumentó o disminuyó determinado valor entre dos periodos concretos de tiempo¹¹¹ (Brezina, 2018, p. 230):

¹¹¹ La ecuación para obtener el porcentaje es la que sigue, donde el corpus 1 corresponde al corpus de periodo anterior y el corpus 2 al de periodo posterior (Brezina, 2018, p. 230):

$$\% \text{ aumento/disminución} = \frac{\text{FR en corpus 2} - \text{FR en corpus 1}}{\text{FR en corpus 1}} \times 100$$

	Porcentaje incremento/disminución
-ICO	-90.3 %
-ITO	-12 %
-ILLO	+62.6 %

Tabla 28: Porcentaje de incremento/disminución de los diminutivos en el s. XIX con respecto al s. XVIII en Almería

En dos casos (*-ico* e *-ito*) el empleo del sufijo disminuyó, mientras que solo en uno de ellos (*-illo*) este porcentaje aumentó. De acuerdo con lo ya expuesto, el cambio más drástico en la provincia de Almería es el que experimenta el sufijo *-ico* con una disminución del 90 % en el siglo XIX con respecto al XVIII. En el extremo opuesto, *-illo* obtuvo un aumento de los casos del 62 %, con lo que se convierte, en este periodo, en el diminutivo con la frecuencia de aparición más alta.

Las diferencias en el uso del diminutivo *-ito* en uno y otro siglo no eran especialmente significativas según el análisis de los residuos de Pearson para los diminutivos en Almería (§ 7.4.1.2. *Relación estadística entre diminutivo y siglo en el corpus de Almería*, Figura 80). Sin embargo, el porcentaje hallado para este diminutivo permite apreciar ahora con mayor claridad la magnitud de su evolución: si bien el cambio no es el más destacado frente al resto de diminutivos, su frecuencia de uso en el XIX descendió un 12 % con respecto al siglo anterior, por lo que tampoco esta cifra es insignificante para una visión completa y rigurosa de la evolución de los diminutivos en Almería en los siglos XVIII y XIX.

En definitiva, varios aspectos parecen clave del uso de los diminutivos en Almería según el análisis estadístico realizado previamente y que también se constata en el contraste de las frecuencias relativas: en primer lugar, la pujanza de *-ico* durante el siglo XVIII y su posterior descenso en el XIX. Por oposición a este hecho, se asiste al crecimiento de *-illo*, cuyo aumento de los casos está ligado a la caída de *-ico* en el siglo XIX. La línea ascendente y descendente (para *-illo* e *-ico* respectivamente) en el gráfico (Figura 81) reflejan visiblemente estas tendencias, mientras que *-ito*, por su parte, manifiesta una evolución más lineal con un ligero descenso.

A la vista de los resultados, las diferentes medidas estadísticas han permitido apreciar la fuerza del cambio en el uso de los diminutivos en el corpus de Almería: el análisis estadístico determinó la significancia de las diferencias observadas en el uso de los diminutivos y permitió cuantificar de manera objetiva el tamaño de esas diferencias. En el siglo XIX *-ico* e *-ito* disminuyen su frecuencia de uso (90.3 % y 12 % respectivamente), mientras que *-illo* se convierte en este siglo en el diminutivo por excelencia en la región, con un aumento de los casos del 62.6 %.

7.4.2. *Corpus de Madrid*

7.4.2.1. *Dispersión de los datos en el corpus de Madrid*

Del mismo modo que para el corpus de Almería, se incluyen los gráficos de dispersión de los datos para los documentos pertenecientes a la provincia de Madrid:

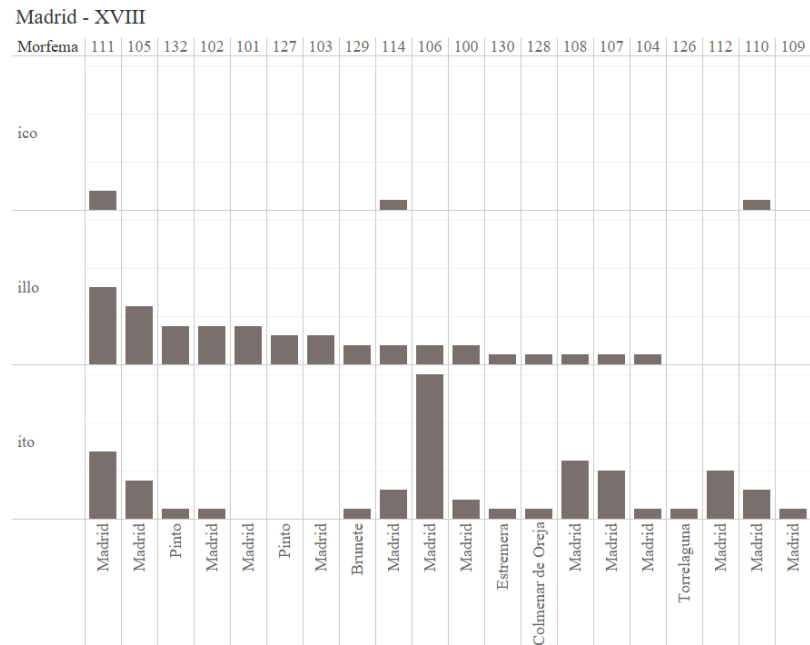


Figura 82: Dispersión de los diminutivos en el corpus de Madrid (s. XVIII)

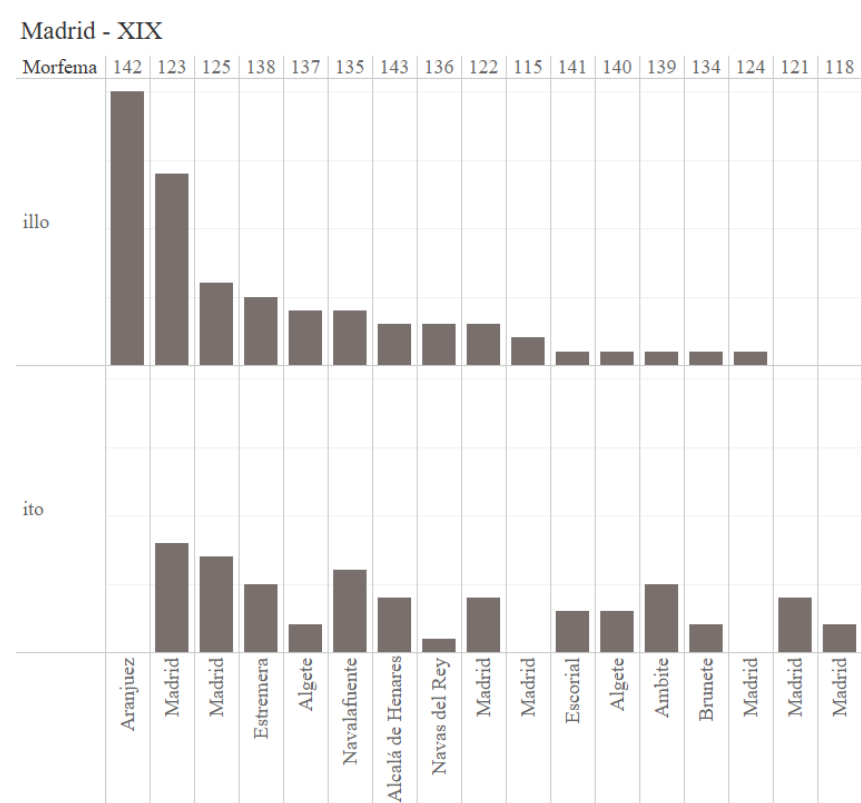


Figura 83: Dispersión de los diminutivos en el corpus de Madrid (s. XIX)

Los documentos del corpus de Madrid aparecen ordenados de mayor a menor según la frecuencia de aparición de *-illo*. En los textos del siglo XVIII, la aparición de *-illo* presenta una frecuencia similar en todos los documentos en los que está presente. Paralelamente, *-ito* manifiesta una distribución parecida y se encuentra también presente en la gran mayoría de los documentos. La dispersión, por tanto, es muy homogénea: los casos están muy repartidos y se distribuyen de manera regular a lo largo de todo el corpus (en 20 documentos del total de los 23 que componen el subcorpus se localizan diminutivos). El hecho de que *-ico* esté presente solo en tres documentos se relaciona con la baja frecuencia del diminutivo en esta centuria en el corpus madrileño y las pocas ocurrencias que hay no se concentran en un solo documento.

En el siglo XIX, la distribución de los diminutivos se reparte entre *-ito* e *-illo*, mientras que *-ico* desaparece por completo. Al igual que para el periodo anterior, el corpus ofrece una dispersión bastante uniforme de *-ito* e *-illo*: 17 de los 22 documentos cuentan con la presencia de alguno de estos dos diminutivos, o de los dos, lo que resulta más frecuente.

Los gráficos de dispersión muestran que los diminutivos están presentes en la mayoría de los documentos y localidades que componen el corpus de control y que, por consiguiente, no constituyen un rasgo particular solo de unos pocos documentos. La distribución de los datos parece ofrecer una buena representatividad del uso de los diminutivos en la región.

7.4.2.2. Relación estadística entre diminutivo y siglo en el corpus de Madrid

A continuación, determinaremos la relación que existe entre las variables de estudio con las pruebas de significación estadística pertinentes realizadas a partir de los datos obtenidos en el corpus de control dispuestos en una tabla de contingencia (Tabla 29). En lo que respecta al corpus de Madrid, se contabilizan un total de 235 formas con diminutivo. La hipótesis sometida a prueba es la siguiente:

$$H_0 = \text{fr. } -ico / -ito / -illo \text{ Madrid XVIII} = \text{fr. XIX}$$

$$H_1 = \text{fr. } -ico / -ito / -illo \text{ Madrid XVIII} \neq \text{fr. XIX}$$

Sufijo Siglo	-ICO	-ITO	-ILLO
S. XVIII	4 (3.6 %)	60 (54.6 %)	46 (41.8 %)
S. XIX	0 (0 %)	56 (44.8 %)	69 (55.2 %)
X-squared test = 7.8123, df = 2, p = 0.02012 Log likelihood test: 9.36, p = 0.009296633 Fisher exact test: p = 0.01391 Cramer's V = 0.1823			

Tabla 29: Tabla de contingencia. Distribución de los diminutivos en los siglos XVIII-XIX en la provincia de Madrid. Frecuencias observadas y porcentajes

De acuerdo con las frecuencias observadas en el corpus, la presencia de *-ico* en Madrid es apenas perceptible. El uso de los diminutivos en esta región está caracterizado más bien por la alternancia *-ito/-illo*, con mayor o menor presencia de uno u otro dependiendo del siglo. La relación entre frecuencia de los diminutivos en el corpus madrileño se visualiza de la siguiente manera:

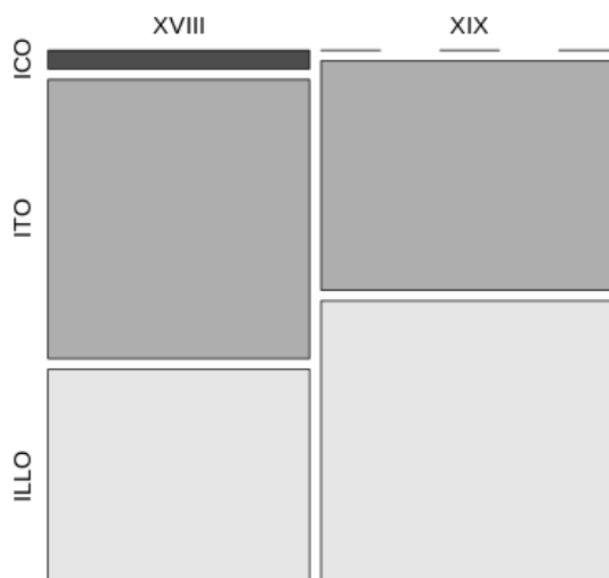


Figura 84: Distribución proporcional de los diminutivos en Madrid en los ss. XVIII y XIX

La figura representa una buena instantánea de lo que sería el uso de los diminutivos en Madrid durante los siglos XVIII y XIX. Las celdas de *-ito* e *-illo*, de proporciones similares a simple vista, reflejarían la pugna de estos dos diminutivos en la región, lo que se traduce en un uso más o menos igualado y continuado en el tiempo, aunque con algunas oscilaciones. Durante estos dos siglos el empleo de los diminutivos está caracterizado por la alternancia entre *-ito* e *-illo* sin que, por ahora, y a falta de más pruebas, podamos constatar en términos estadísticos que haya un predominio claro de uno sobre otro. El análisis que se realiza a continuación pretende demostrar precisamente la relevancia estadística de esas diferencias entre el siglo XVIII y XIX en la provincia de Madrid y estimar qué incidencia tiene la variable tiempo en el empleo de los diminutivos en esta región.

Para el contraste de hipótesis se utilizará en este caso el test de *log-likelihood*. Esta prueba estadística es preferible al test de χ^2 cuando las muestras son pequeñas (Desagulier, 2017, p. 209)¹¹². Tomando como base el valor de *p* obtenido en esta prueba (*p* = 0.00929), se acepta que las diferencias observadas entre las variables son estadísticamente significativas y la probabilidad de que los datos se deban al azar es mínima. Dado que *p* < 0.05, hay evidencia suficiente para poder rechazar la hipótesis

¹¹² La prueba de Chi-cuadrado (χ^2) resulta apropiada cuando en tablas de contingencia mayores de 2x2 al menos el 80 % de las frecuencias esperadas son mayores de 5 (Brezina, 2018, p. 113). Este supuesto se cumplía para los datos obtenidos en el corpus de Almería, pero no en este caso, donde dos de las FE en el corpus de Madrid no alcanzan esta cifra (véase más adelante la Tabla 30).

nula y aceptar que hay una relación entre las dos variables categóricas (siglo/diminutivo). Se concluye de esta manera que hay una diferencia real y significativa en el uso de los diminutivos en la provincia de Madrid entre los siglos XVIII y XIX.

El valor de p no informa sobre la fuerza de la asociación entre las variables, por lo que se toman los valores de la V de Cramer y la PR como medidas estadísticas para cuantificar el tamaño del efecto¹¹³. A la vista del resultado de V , la fuerza con la que se relacionan las variables indica una asociación débil entre ambas (V de Cramer = 0.18). Con el cálculo de PR podemos afinar este resultado y establecer cómo de grandes son las diferencias observadas entre el uso de los diminutivos en uno y otro siglo. La prueba de PR, recordamos, calcula la probabilidad de aparición que tiene un determinado fenómeno lingüístico en dos corpus diferentes al comparar la probabilidad en el corpus A con respecto al corpus B, y viceversa. Dado que en el siglo XIX la presencia de *-ico* es 0, las opciones de probabilidad para los diminutivos en el corpus de Madrid se reducen a las siguientes:

$$\text{a. PR de -illo en s. XVIII vs. XIX} = \frac{0.418}{0.552} = 0.75$$

$$\text{PR de -illo en s. XIX vs. XVIII} = \frac{0.552}{0.418} = \mathbf{1.32}$$

$$\text{b. PR de -ito en s. XVIII vs. XIX} = \frac{0.545}{0.448} = \mathbf{1.22}$$

$$\text{PR de -ito en s. XIX vs. XVIII} = \frac{0.448}{0.545} = 0.82$$

El diminutivo *-illo* tiene 1.32 veces más probabilidad de aparecer en el siglo XIX que en el XVIII. A la inversa, la probabilidad de que aparezca en el XVIII es 0.75 veces su probabilidad de aparecer en el XIX, lo que lo convierte con mayor probabilidad en el diminutivo más frecuente en este siglo. En el caso de *-illo*, aunque las diferencias no son tan marcadas como las que se registraban para Almería¹¹⁴, se aprecian también diferencias regionales importantes. En efecto, la posibilidad de encontrar *-illo* en el XIX es 1.32 veces más alta a la de que aparezca en el XVIII, por lo que sigue siendo una probabilidad superior que debe tenerse en cuenta. Por otro lado, *-ito* es más probable en el XVIII (1.22 veces más) que en el XIX y, al contrario, en el siglo XIX tan solo 0.82 veces su probabilidad de aparecer en el siglo anterior.

¹¹³ Puesto que la V de Cramer es una medida derivada de χ^2 , esta prueba no es la más indicada para calcular el tamaño del efecto (véase la nota anterior). Sin embargo, dado que las pruebas de Log-likelihood y de Chi-cuadrado arrojaron unos valores de p muy similares, consideramos que el análisis de V puede dar también resultados fiables para los datos que se testan en esta ocasión. Junto a esta medida se calcula también la PR (razón de probabilidad) como alternativa a la V de Cramer.

¹¹⁴ En esta provincia, la PR de *illo* en el s. XIX era casi el doble: 2.41 veces más probable que en el XVIII.

De acuerdo con los datos obtenidos, ambos diminutivos ofrecen valores estadísticos similares en el corpus de Madrid al reflejar probabilidades de aparición prácticamente iguales, pero totalmente contrarias según el siglo: el aumento de la probabilidad de uno de ellos supone el descenso de la probabilidad de su contrario. Esta circunstancia se visualiza claramente en el gráfico obtenido a partir de los residuos de Pearson (Figura 85), en el que se reflejan las diferencias entre las frecuencias observadas y esperadas de estos diminutivos en Madrid:

	XVIII	XIX
ICO	1.87234	2.12766
ITO	54.29787	61.70213
ILLO	53.82979	61.17021

Tabla 30: Frecuencias esperadas de los diminutivos en Madrid

	XVIII	XIX
ICO	1.5549260	-1.4586499
ITO	0.7738299	-0.7259168
ILLO	-1.0671823	1.0011058

Tabla 31: Residuos de Pearson para los diminutivos en el corpus de Madrid

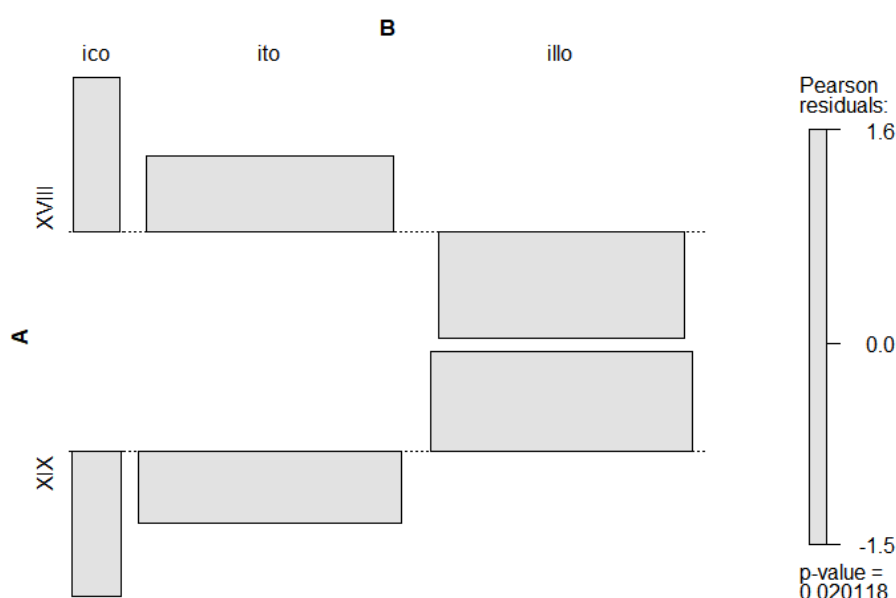


Figura 85: Gráfico de asociación (siglo/diminutivo en Madrid) basado en los residuos de Pearson

El gráfico refleja a la perfección el comportamiento contrario que revela el uso de los diminutivos en Madrid en los siglos XVIII y XIX. Los valores de *-ico* e *-ito* se sitúan en el siglo XVIII por encima de lo esperado (casillas por encima de la línea de puntos), mientras que en el siglo XIX la situación es la contraria y sus frecuencias de aparición están por debajo de lo esperado. Por su parte, la frecuencia de *-illo* en el XVIII es inferior a la esperada y se restablece en el XIX prácticamente en la misma proporción.

A pesar de que el gráfico no refleja relevancias estadísticas destacables entre el uso de los diminutivos (ninguno de los residuos sobrepasa los 2 puntos), las diferencias

de uso entre ambos siglos son considerables y manifiestan comportamientos totalmente opuestos.

7.4.2.3. Alternancia de -ito/-illo en el corpus de Madrid

El análisis estadístico de los datos obtenidos para el corpus de Madrid se concluye con una distribución más o menos proporcional de los diminutivos *-illo* e *-ito* en los siglos XVIII y XIX. En esta provincia, lo más llamativo de los datos mostrados es la disputa entre estos dos diminutivos en proporciones similares. La prueba de Chi-cuadrado concluía que las diferencias entre uno y otro siglo son significativas estadísticamente ($p < 0.05$), pero no con tanta evidencia como lo eran para el caso de Almería. La asociación entre las variables siglo y diminutivo era además débil ($V = 0.18$).

Complementaremos estos resultados con los obtenidos ahora a partir de las frecuencias relativas, las cuales nos permiten ofrecer una visualización más clara de la evolución de los diminutivos en Madrid en el periodo estudiado (Figura 86).

	-ICO	-ITO	-ILLO	TOTAL
s. XVIII	1.33	20.01	15.34	36.69
s. XIX	0	15.76	19.42	35.18

Tabla 32: Frecuencias relativas de los diminutivos en Madrid (ss. XVIII-XIX)

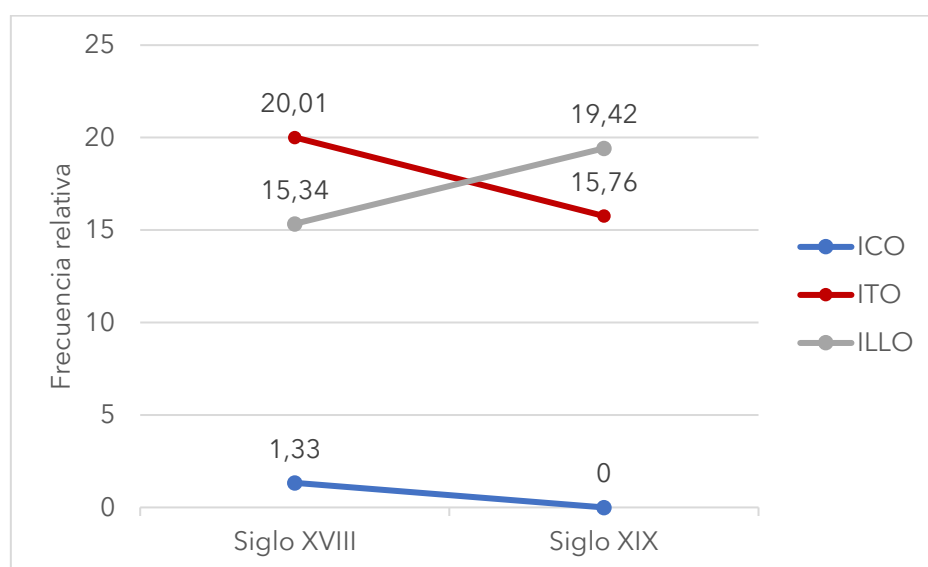


Figura 86: Distribución diacrónica de los diminutivos en Madrid en los siglos XVIII-XIX

La distribución de los diminutivos en el corpus de control a partir de las frecuencias relativas refleja una situación bien distinta de la que se señaló previamente en Almería. Si el sufijo *-ico* gozaba de vitalidad en aquella provincia en el siglo XVIII, en Madrid obtiene la cota más baja de aparición con un uso muy ocasional. Su presencia en el

XVIII se da exclusivamente en la capital, concretamente, se localiza en cuatro ocasiones en tres documentos. Este uso, ya de por sí bastante reducido, llega hasta su total desaparición en el XIX, periodo para el que no se registra ningún caso. El sufijo juega por tanto un papel residual prácticamente inapreciable en la provincia madrileña durante estas dos centurias. No obstante, este uso casi aislado implica una frecuencia alta del resto de sufijos, cuya distribución refleja una tendencia bastante reveladora.

En lo que respecta al empleo de *-ito* e *-illo*, sus frecuencias de aparición parecen reflejar pautas de comportamiento inversamente proporcionales: una frecuencia más elevada de *-ito* conlleva una disminución de *-illo* y viceversa, el aumento de *-illo* supone el descenso de *-ito*. Esta circunstancia es claramente apreciable en la Figura 84 con la peculiar distribución entrecruzada de las líneas de tendencia de estos dos diminutivos.

La comparación cronológica de los resultados nos revela que su empleo osciló durante este periodo con frecuencias cercanas entre sí. De manera individual para cada siglo, los porcentajes de *-ito* e *-illo* en el XVIII (54.6 % y 41.8 % respectivamente) y en el XIX (44.8 % y 55.2 %) ya indicaban unas proporciones más o menos similares sin una ventaja destacada de uno con respecto a otro. En Madrid, ambos diminutivos compiten en frecuencia de aparición y su uso se basará en la fluctuación. De manera más precisa, el porcentaje de aumento y disminución¹¹⁵ nos permite medir la frecuencia de cambio de estos diminutivos entre los dos periodos de tiempo estudiados:

	Porcentaje incremento/disminución
-ICO	-100 %
-ITO	-21.2 %
-ILLO	+26.6 %

Tabla 33: Porcentaje de incremento/disminución de los diminutivos en el s. XIX con respecto al s. XVIII en Madrid

El resultado ofrece el porcentaje obtenido en el siglo XIX con respecto a las frecuencias del XVIII. Por tanto, en el siglo XIX la presencia de *-ico* en Madrid llega a desaparecer por completo con una caída del 100 % de los casos. Por otro lado, mientras que *-ito* registra en este mismo siglo una disminución del 21.2 % de los casos, *-illo* manifiesta la tendencia contraria y aumenta su frecuencia de aparición en un 26.6 %, lo que refleja de forma más clara la diferencia entre el uso de *-ito* e *-illo* en Madrid en este siglo.

7.4.3. Análisis comparativo entre subcorpus

Hasta ahora se ha hecho una descripción de los datos recogidos en cada una de las provincias de Almería y Madrid a lo largo de dos siglos, considerando cada siglo y

¹¹⁵ La ecuación que se ha utilizado para el cálculo del porcentaje se señala en la nota número 111.

provincia como subcorpus independientes. Las tablas y gráficos de las páginas precedentes permiten explorar el corpus en su conjunto y ofrecen una visión general de la distribución de los datos en su perspectiva diacrónica. En lo que sigue, el objetivo es interpretar los datos mostrados hasta ahora haciendo un estudio comparativo que aborde los planos diacrónico y diatópico al mismo tiempo, situando el foco más en este último.

Para este análisis contrastivo se abordará, en primer lugar, la significancia estadística de los datos obtenidos analizando las diferencias entre ambas provincias en cada centuria. La distribución de cada uno de los diminutivos en Almería y Madrid se compara de forma conjunta a lo largo de estos dos siglos, primero en el siglo XVIII y seguidamente en el XIX, contrastando las siguientes hipótesis de partida:

$H_0 = \text{fr. (-ico, -ito, -illo) Almería XVIII} = \text{fr. (-ico, -ito, -illo) Madrid XVIII}$

$H_1 = \text{fr. (-ico, -ito, -illo) Almería XVIII} \neq \text{fr. (-ico, -ito, -illo) Madrid XVIII}$

$H_0 = \text{fr. (-ico, -ito, -illo) Almería XIX} = \text{fr. (-ico, -ito, -illo) Madrid XIX}$

$H_1 = \text{fr. (-ico, -ito, -illo) Almería XIX} \neq \text{fr. (-ico, -ito, -illo) Madrid XIX}$

Para finalizar, y a modo de recapitulación, se ofrecerá una distribución geográfica de los diminutivos a partir del mapeado de los datos y el cálculo de la frecuencia relativa.

7.4.3.1. Contraste diatópico Almería vs. Madrid en el siglo XVIII

Se someten a prueba los datos obtenidos en el siglo XVIII en Almería y Madrid. En la tabla de contingencia, la variable explicativa se corresponde esta vez con la provincia de origen:

Sufijo Provincia	-ICO	-ITO	-ILLO
Almería	44	17	23
Madrid	4	60	46
X-squared test = 62.654, df = 2, p = 2.482551E-14 Cramer's V = 0.568			

Tabla 34: Tabla de contingencia. Contraste diminutivo/provincia en el s. XVIII

La siguiente figura muestra un primer acercamiento a la distribución de los diminutivos en ambas provincias en el siglo XVIII. El uso de los sufijos está claramente marcado por las diferencias en el mayor uso de *-ico* en Almería e *-ito* en Madrid, y al revés: un uso ocasional de *-ico* en Madrid y el menor empleo de *-ito* en Almería.

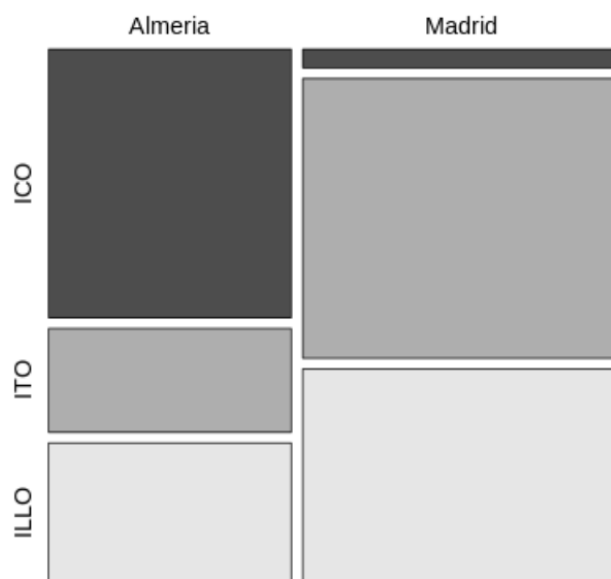


Figura 87: Proporción de los diminutivos en Almería y Madrid en el s. XVIII

La prueba estadística aplicada a los datos del siglo XVIII constata que las diferencias observadas en el empleo de los diminutivos entre Almería y Madrid en este periodo son significativas ($p < 0.05$). El bajo valor de p permite probar la hipótesis nula de partida y rechazarla en favor de su alternativa, según la cual se acepta que el uso de los diminutivos en el siglo XVIII difiere en ambas provincias. En términos estadísticos, estas diferencias son importantes y tienen un efecto relativamente fuerte en la asociación de las variables (V de Cramer = 0.56). Las variables *provincia* y *diminutivo* se hallan muy relacionadas y hay evidencias estadísticas para afirmar que el origen geográfico ejerce una fuerza determinante en el empleo de los diminutivos.

En el siglo XVIII, los valores que contribuyen en mayor medida al tamaño del efecto son los correspondientes a los diminutivos *-ico* e *-ito*, estos son los datos más significativos según se desprende del análisis de los residuos de Pearson. Particularmente, el corpus refleja diferencias considerables en cuanto al uso de estos dos diminutivos. Los resultados obtenidos para ambos en esta medida se alejan considerablemente del cero, lo que permite probar la dependencia de las variables.

	Almería	Madrid
ICO	20.78351	27.21649
ITO	33.34021	43.65979
ILLO	29.87629	39.12371

Tabla 35: Frecuencias esperadas de los diminutivos en el s. XVIII

	Almería	Madrid
ICO	5.092573	-4.450211
ITO	-2.829915	2.472958
ILLO	-1.258029	1.099345

Tabla 36: Residuos de Pearson para los diminutivos en el s. XVIII

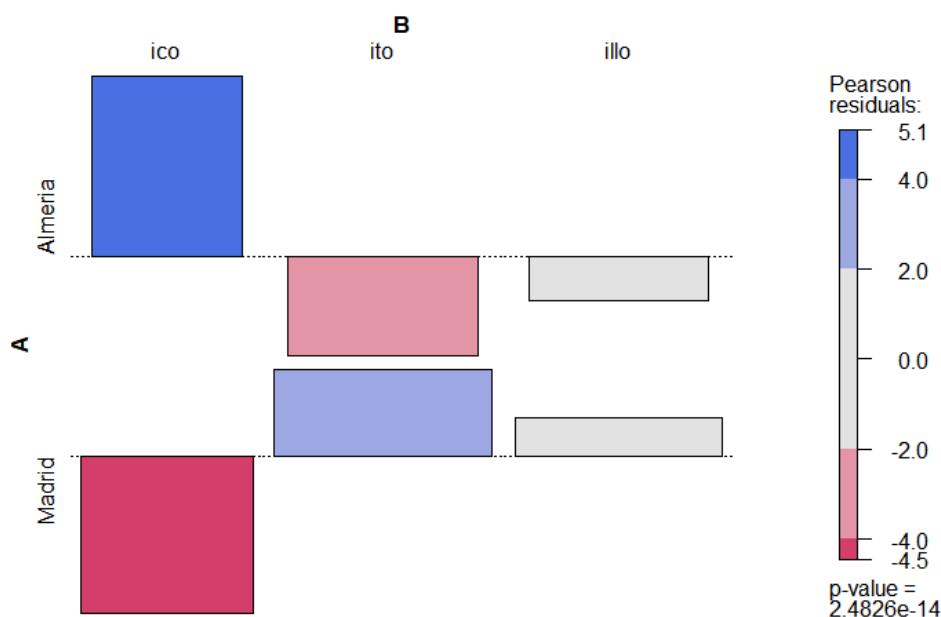


Figura 88: Gráfico de asociación (provincia/diminutivo en el s. XVIII) basado en los residuos de Pearson

El gráfico resultante refleja una distribución inversa de los diminutivos en Almería y Madrid: los diminutivos cuya presencia se da por encima de las frecuencias esperadas en Almería son los mismos que en Madrid se sitúan por debajo de lo que cabría esperar, y al revés. Aunque la presencia de *-illo* refleja tendencias opuestas entre una y otra provincia, las diferencias entre zonas no son de las más representativas para el uso de los sufijos en este siglo (se trata de los residuos más cercanos al cero sombreados en gris). Por lo tanto, este diminutivo no desempeña un papel tan relevante en la distribución de los diminutivos en el siglo XVIII como sus análogos.

Los diminutivos *-ito* e *-ico*, y de manera particular este último, son los que aportan los valores más significativos (marcados en el gráfico en rojo y azul), es decir, son los que reflejan una mayor discrepancia entre los valores observados y los esperados en el siglo XVIII. De entre estos, la frecuencia de *-ico* está asociada en mayor grado al lugar de origen, por lo que su aparición está fuertemente vinculada con la zona geográfica (Almería en este caso). Por el contrario, *-ico* no se vincula con Madrid, donde su frecuencia de aparición está muy por debajo de la esperada. En menor medida, pero también muy significativos, son los resultados de *-ito*. Su frecuencia en Madrid se halla por encima de lo esperado, mientras que en Almería tiene lugar el comportamiento contrario.

De manera más exacta es posible determinar en qué proporción difiere el uso de los diminutivos en una y otra provincia en el siglo XVIII calculando la razón de probabilidad (PR) de cada diminutivo en uno y otro corpus. Esta prueba probabilística permite afinar en qué medida es mayor o menor la presencia de cada diminutivo en Almería con respecto a Madrid, y viceversa. Se señalan en **negrita** las probabilidades que con más interés nos sirven para evaluar las principales diferencias en el uso de los diminutivos entre estas dos provincias en el siglo XVIII:

$$a. \text{ PR de -ico en Almería vs. Madrid (s. XVIII)} = \frac{0.523}{0.036} = \mathbf{14.52}$$

$$\text{PR de -ico en Madrid vs. Almería (s. XVIII)} = \frac{0.036}{0.523} = 0.06$$

$$b. \text{ PR de -ito en Almería vs. Madrid (s. XVIII)} = \frac{0.202}{0.545} = 0.37$$

$$\text{PR de -ito en Madrid vs. Almería (s. XVIII)} = \frac{0.545}{0.202} = \mathbf{2.69}$$

$$c. \text{ PR de -illo en Almería vs. Madrid (s. XVIII)} = \frac{0.273}{0.418} = 0.65$$

$$\text{PR de -illo en Madrid vs. Almería (s. XVIII)} = \frac{0.418}{0.273} = \mathbf{1.53}$$

Coincidiendo con los datos más significativos que se destacaban en color en la Figura 88, la diferencia proporcional más alta corresponde al empleo del diminutivo *-ico*. Su uso en Almería en el siglo XVIII es de 14.5 veces superior a Madrid, donde la probabilidad de que aparezca *-ico* es tan solo de 0.06 veces la probabilidad de que lo haga en Almería.

Por su parte, *-ito* e *-illo* tienen mayores probabilidades de aparecer en Madrid que en Almería (2.7 y 1.5 veces más, respectivamente). En primer lugar, la presencia de *-ito* en Madrid es 2.7 veces más probable que su presencia en Almería, donde, por el contrario, en el siglo XVIII su probabilidad de aparición es tan solo de 0.37 veces su presencia en Madrid. En segundo lugar, con respecto a *-illo*, la presencia en Madrid es 1.5 veces más probable que en Almería en este mismo periodo. Si no es de la más significativa, de nuevo se refleja el uso contrario de este diminutivo en una y otra provincia.

En definitiva, según los datos del corpus, del uso de los diminutivos en el siglo XVIII destaca el significativo uso de *-ico* en Almería en contraste con el de Madrid. Por otro lado, en Madrid, *-illo* e *-ito* aparecen con más probabilidad que en Almería, más este último que *-illo*. Las diferencias evidencian la asociación relativamente fuerte que existe en este siglo entre la provincia y el uso de los diminutivos: la aparición de un diminutivo u otro se vincula de manera estrecha con la procedencia geográfica.

7.4.3.2. Contraste diatópico Almería vs. Madrid en el siglo XIX

En este apartado se repite el análisis estadístico anterior, esta vez para los datos obtenidos en Almería y Madrid en el siglo XIX:

Sufijo Provincia	-ICO	-ITO	-ILLO
Almería	4	14	35
Madrid	0	56	69
X-squared test = 13.381, df= 2, p = 0.001243 Log likelihood test: 13.88, p = 0.0009697316 Fisher exact test: p = 0.001147991 Cramer's V = 0.274			

Tabla 37: Tabla de contingencia. Contraste diminutivo/provincia en el s. XIX

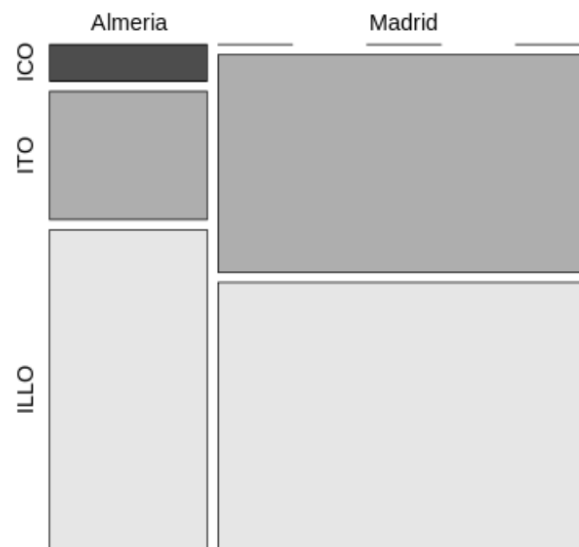


Figura 89: Proporción de uso de los diminutivos en Almería y Madrid en el s. XIX

Con respecto al siglo anterior, el siglo XIX se caracteriza por un uso ocasional de *-ico*, incluso en Almería, mientras que *-ito* e *-illo* compiten en frecuencia de uso. En este siglo las diferencias en el uso de los diminutivos entre Almería y Madrid vuelven a ser significativas ($p < 0.05$), aunque esta vez las variables se asocian entre sí con una intensidad moderada (V de Cramer = 0.27). La prueba para el contraste de hipótesis permite por tanto rechazar la hipótesis nula y asumir, con evidencias estadísticas, que las diferencias observadas no son fruto del azar y que en el siglo XIX el uso de los diminutivos está supeditado al paso del tiempo con una fuerza moderada.

Las diferencias proporcionales entre el uso de los diminutivos entre Almería y Madrid en el siglo XIX se interpretan mejor a partir de las medidas de PR obtenidas para los sufijos *-illo* e *-ito*. Con respecto a *-ico*, su ausencia es total en Madrid, por lo que para este periodo las opciones son las siguientes:

$$a. \text{ PR de } -illo \text{ en Almería vs. Madrid (s. XIX)} = \frac{0.660}{0.552} = \mathbf{1.19}$$

$$\text{PR de } -illo \text{ en Madrid vs. Almería (s. XIX)} = \frac{0.552}{0.660} = 0.83$$

$$b. \text{ PR de -ito en Almería vs. Madrid (s. XIX)} = \frac{0.264}{0.448} = 0.58$$

$$\text{PR de -ito en Madrid vs. Almería (s. XIX)} = \frac{0.448}{0.264} = \mathbf{1.69}$$

En Almería sobresale el uso de *-illo*, superior 1.19 veces a la probabilidad de que aparezca en Madrid. Por otro lado, en Madrid, *-ito* es 1.69 veces más probable que en Almería.

A partir de la asociación entre las variables provincia/diminutivo, las diferencias entre las frecuencias observadas y esperadas obtenidas en los residuos de Pearson (Figura 90) reflejan que en el siglo XIX el uso de los diminutivos en Almería y Madrid destaca estadísticamente por el uso que se hace del sufijo *-ico* en ambas zonas. El gráfico nos permite determinar cuál de todos los datos obtenidos en este siglo es el más significativo en la asociación de las dos variables estudiadas.

	Almería	Madrid
ICO	1.191011	2.808989
ITO	20.842697	49.157303
ILLO	30.966292	73.033708

Tabla 38: Frecuencias esperadas de los diminutivos en el s. XIX

	Almería	Madrid
ICO	2.5739024	-1.6760038
ITO	-1.4988229	0.9759628
ILLO	0.7248695	-0.4720008

Tabla 39: Residuos de Pearson para los diminutivos en el s. XIX

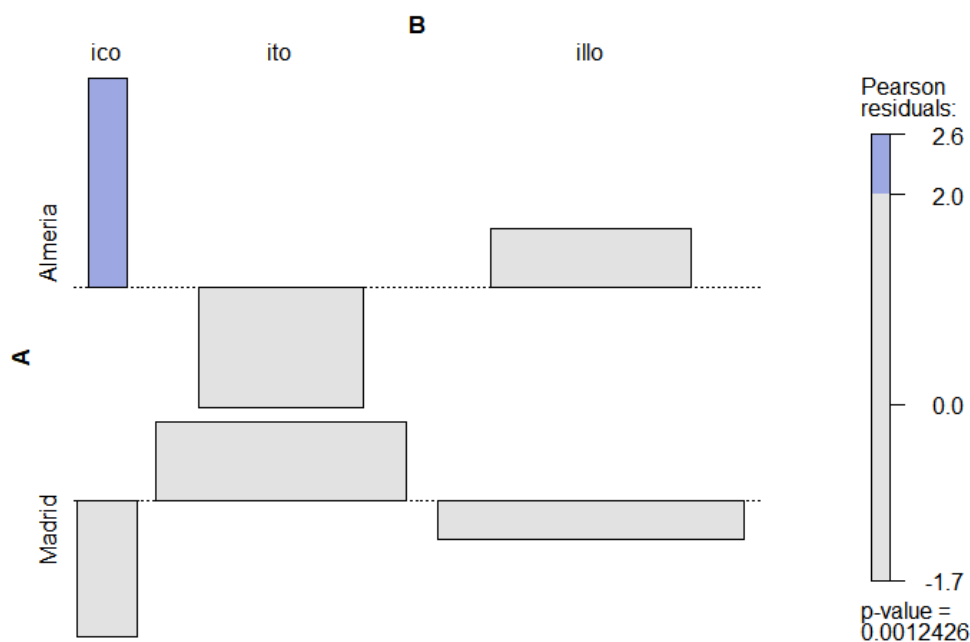


Figura 90: Gráfico de asociación (provincia/diminutivo en el s. XIX) basado en los residuos de Pearson

Nuestro corpus refleja diferencias considerables en el uso de *-ico* entre Almería y Madrid en el siglo XIX, a pesar de que el uso de este diminutivo en Almería es ya bastante bajo en este periodo con respecto al siglo anterior. En Madrid, *-ico* se sitúa muy por debajo de lo esperado (recordemos $FO = 0$), mientras que en Almería lo hace por encima. Este último dato resulta para este siglo el más significativo y es el que más parece depender de la variable *provincia* en el siglo XIX.

El resto de los diminutivos reflejan de nuevo tendencias opuestas en una y otra zona geográfica. Con respecto a *-ito*, el comportamiento es el mismo que en el siglo XVIII, por debajo de lo esperado en Almería y por encima en Madrid, aunque las diferencias en el XIX son menos significativas que en el siglo anterior. El uso de *-illo* sigue reflejando trayectorias divergentes entre Almería y Madrid, pero en el XIX la dirección de *-illo* es la contraria al siglo XVIII: en este siglo *-illo* registraba una frecuencia superior a la esperada en Madrid, mientras que en el XIX esta tendencia se cumple solo para el caso de Almería. En Madrid, en cambio, es *-ito* el que registra una frecuencia superior a la esperada y no *-illo*.

En definitiva, en el siglo XIX, a pesar de que las diferencias en el uso de los diminutivos son significativas, no están tan marcadas como en el siglo anterior. El empleo de estos en Almería y Madrid refleja siempre comportamientos contrarios: donde uno aumenta, el otro disminuye.

7.4.3.3. Distribución geográfica

Una vez probada la significancia de los datos, a modo de síntesis reuniremos la información mostrada hasta ahora apoyándonos en diferentes visualizaciones.

En el mapa se recopilan de un vistazo y de manera conjunta los resultados más significativos que hasta ahora se han ido mostrando en Madrid, a la izquierda en la figura, y en Almería, a la derecha. Los gráficos de sectores reflejan la proporción en la que en cada siglo se emplean los diminutivos en los diferentes municipios, de modo que para algunos puntos es posible que los textos documenten el uso de un solo diminutivo, de dos de ellos o de los tres.

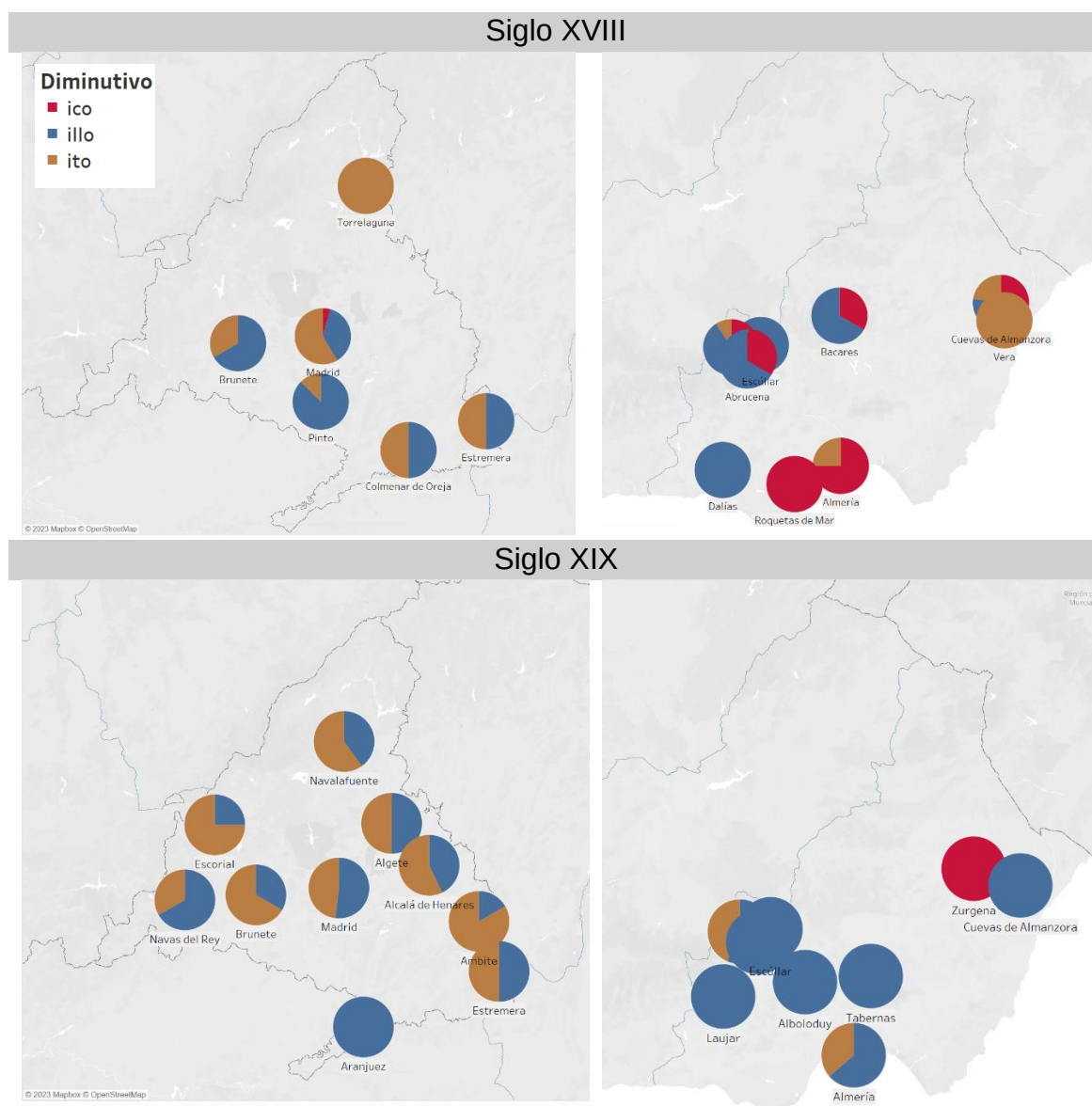


Figura 91: Localización geográfica de los diminutivos en los corpus de Madrid y Almería en los siglos XVIII-XIX

El uso del diminutivo *-ico* es característico exclusivamente de Almería, donde durante el siglo XVIII conserva un fuerte carácter dialectal en contraste con lo que ocurre en el corpus de control. Durante este siglo, el diminutivo alcanzó la cota más alta de frecuencia y se constata su uso bien repartido por distintos municipios de la provincia, incluida la capital. La situación de *-ico* en Almería en el XIX contrasta notablemente con la del XVIII. En el transcurso de un siglo, este diminutivo redujo considerablemente su extensión y prácticamente desaparece en el siglo XIX, donde se localiza en un solo punto de la zona más oriental de la provincia (concretamente en el municipio de Zurgena). El descenso de *-ico* contrasta con el predominio de *-illo*, también por encima de *-ito*, aunque el uso de estos dos diminutivos oscila en algunas localidades (Fiñana y Almería). Es llamativa la presencia de *-illo* en Almería durante el siglo XIX.

Por otro lado, las oscilaciones que destacábamos entre *-ito* e *-illo* a propósito del análisis de las FR en el corpus de Madrid (Figura 86) se reflejan visiblemente en el

mapa con la convivencia de ambos diminutivos en un amplio número de municipios en esta región. Sobre todo, en el siglo XIX se aprecia la presencia casi igualada de ambos diminutivos en algunos de los puntos.

En lo que sigue trataremos de precisar más estos datos con cifras concretas a partir de la comparación de las frecuencias relativas de cada diminutivo. Los siguientes gráficos de barras ofrecen una más clara visualización del contraste en el uso de los diminutivos entre Almería y Madrid en cada siglo:

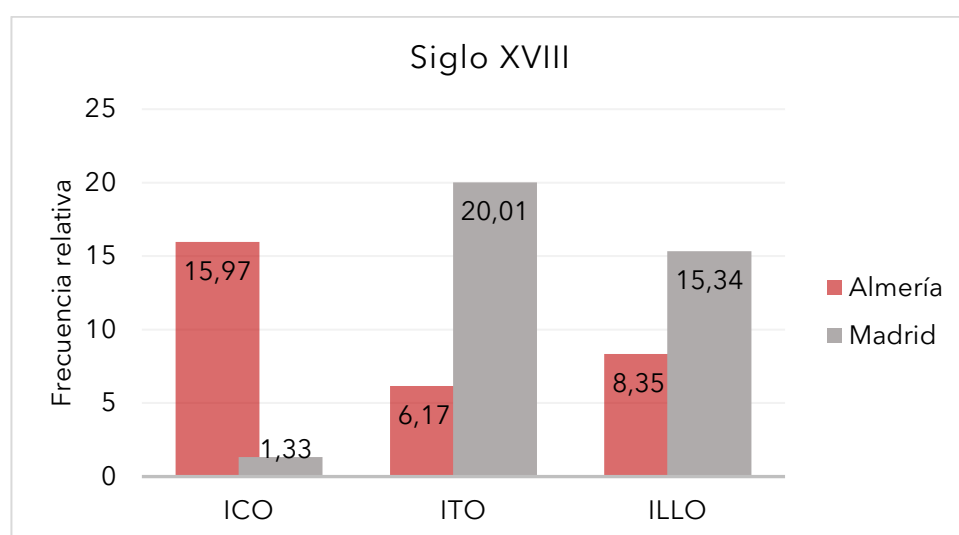


Figura 921: Frecuencias de aparición de los diminutivos por provincia en el siglo XVIII

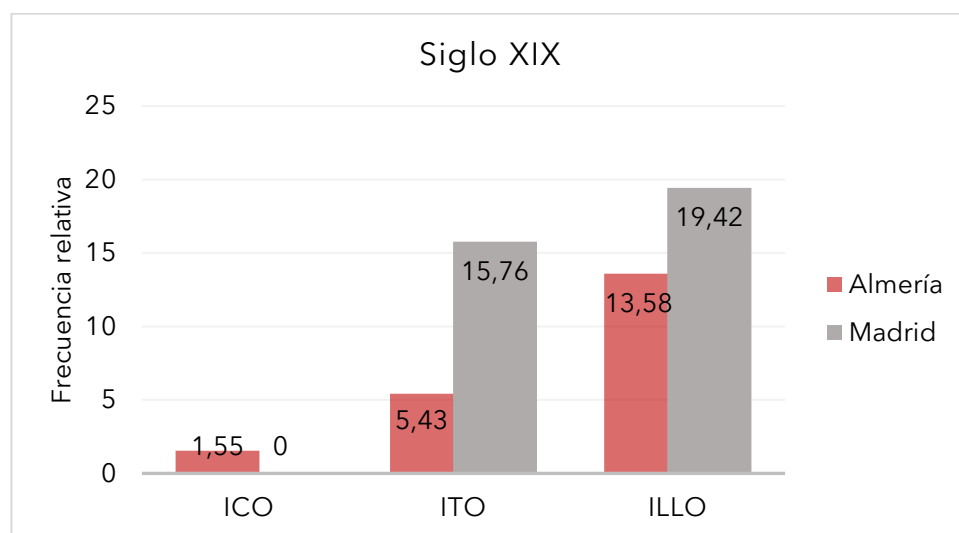


Figura 93: Frecuencia de aparición de los diminutivos por provincia en el siglo XIX

A la vista de los gráficos, se evidencia una distribución muy distinta del uso de los diminutivos en Almería y Madrid que va cambiando con el tiempo. El contraste entre ambas provincias se hace más notable incluso en el siglo XVIII que en el XIX.

En el siglo XVIII el uso de *-ico* es significativamente superior en Almería, lo que le confiere un marcado carácter dialectal respecto del uso madrileño. Es el diminutivo preferido en Almería si se compara con las frecuencias obtenidas para el

resto de los sufijos. Por el contrario, en Madrid, si bien *-ito* alcanza la cota más alta, no está tan claro que esto le otorgue sobrada ventaja con respecto a *-illo*, con quien compite en frecuencia de aparición.

En el siglo XIX el orden de frecuencia de los diminutivos es similar en ambas provincias. Con respecto a *-ico*, su presencia en Madrid es nula, pero en Almería también cae considerablemente y se convierte en la menor de las frecuencias registradas para los tres diminutivos en esta región. Desciende a valores similares a los que tenía Madrid en el XVIII, donde ya era muy residual. En cuanto a *-illo*, en el siglo XIX este es el diminutivo que registra las frecuencias más altas en ambas provincias, aunque en Madrid le sigue de cerca *-ito* con una frecuencia de 15.76. Las diferencias son más apreciables en el caso de *-ito*. Almería y Madrid difieren especialmente en el uso de este diminutivo, que posee una frecuencia de aparición mucho mayor en Madrid (15.76) en comparación a la que se recoge en Almería (5.43).

En base a estos datos, varios son los aspectos que destacamos. En primer lugar, *-ico* es el diminutivo característico de Almería al menos hasta el siglo XVIII. En el XIX, su disminución se resuelve con la supremacía de *-illo*, lo que lo convierte en el diminutivo con más presencia en esta región. Por otro lado, en Madrid no es tan evidente el predominio de un diminutivo sobre otro. Los sufijos *-ito* e *-illo* alcanzan frecuencias similares en ambos siglos, aunque *-ito* sobresale en el XVIII e *-illo* en el XIX. Quedó demostrado que estas diferencias eran significativas en términos estadísticos, aunque con una fuerza de asociación débil.

El análisis diatópico nos lleva a concluir que el patrón dialectal de Almería se va fijando a mayor velocidad, el cual se encuentra ya prácticamente establecido en el XIX. Mientras tanto, en Madrid, la alternancia *-ito/-illo* no se ha resuelto de forma definitiva, por lo que no se puede hablar todavía de una distribución clara a favor de uno de estos diminutivos. En resumen, Almería establece en una época más temprana una nueva tendencia dialectal, de manera que asistimos a una reorganización del empleo de los diminutivos a mayor velocidad que en Madrid, donde, por otra parte, no hay indicios, con los datos del corpus analizado, de que la alternancia *-illo* e *-ito* se resuelva definitivamente a favor de *-ito* ni siquiera en el siglo XIX.

7.5. Discusión

En este capítulo se ha analizado la presencia de los diminutivos *-ico*, *-ito* e *-illo* en un corpus de inventarios de los siglos XVIII y XIX de Almería en contraste con un subcorpus de Madrid. Determinar la evolución de los diminutivos es el objetivo central de esta tesis, para lo que el análisis estadístico ha ayudado a expresar con rigor y precisión la información cuantitativa y contrastarla con la de estudios previos sobre el tema. A raíz de las preguntas de investigación mediante las que pretendíamos detectar patrones de convergencia y divergencia entre dos modalidades lingüísticamente diferenciadas (§ 7.2. *Variables e hipótesis de estudio*) se han obtenido las respuestas que se muestran a continuación en torno a la cronología y la distribución geográfica de los diminutivos.

En cuanto a las diferencias diatópicas, el uso de los diminutivos en Almería y Madrid refleja tendencias opuestas en muchos casos. En el siglo XVIII las diferencias son aun más destacadas que las que se observan un siglo después. El espacio geográfico incide sobre la elección del tipo de diminutivo con una intensidad relativamente fuerte en el siglo XVIII ($V = 0.56$) y moderada en el XIX ($V = 0.27$), por lo que la asociación entre la provincia y los diminutivos empleados se manifiesta con distinta fuerza en cada siglo. En el XVIII, ambas provincias se diferencian sobre todo por el uso de *-ico* (en Almería su probabilidad de aparición es de 14.5 veces mayor que en Madrid) y de *-ito* (superior en Madrid 2.7 veces). Sin embargo, las conclusiones más determinantes de este estudio tienen que ver con los resultados obtenidos en el siglo XIX y no tanto con los del XVIII, puesto que este siglo refleja la tendencia que ya se venía observando en etapas anteriores en lo que respecta al uso de los diminutivos en las provincias estudiadas.

En los siglos XVI y XVII, *-ico* ya se distinguía por su cierto carácter de rusticidad y su uso regional (Náñez Fernández, 2006, p. 183). La contienda *-ico/-ito* se fue resolviendo muy gradualmente a favor de este último, que fue ganando popularidad, mientras que *-ico* redujo su uso exclusivamente a algunas regiones determinadas como Aragón, Murcia y Andalucía oriental (Lapesa, 2014, p. 334). En el siglo XVII, Almería se encontraba ya entre las quince provincias españolas con una frecuencia de *-ico* superior a la media, únicamente sobrepasada, en este orden, por León, Murcia, Navarra, Albacete, Teruel y Zamora (Calderón Campos, 2019c, p. 121). Esta situación se mantiene también en el XVIII, cuando seguirá vigente el predominio de *-ico* en Almería. Según nuestros datos, su uso representa el 52.4 % del total de los diminutivos, lo que lo convierte en el más frecuente en la región con sobrada ventaja respecto al siguiente (*-illo* con un 27.4 %).

En el corpus madrileño, el uso de los diminutivos en el XVIII también se corresponde con lo que señala la bibliografía previa. En este periodo *-illo* pierde el predominio del que había gozado desde los orígenes del castellano y es sobrepasado por *-ito*. Los factores que impulsaron este cambio (detallados en § 5.2. *Periodización y extensión geográfica de -ico, -ito e -illo*) harán que *-ito* se convierta en adelante en el diminutivo más empleado según constata Náñez Fernández a partir del estudio de textos literarios. De manera particular, este autor señala la prevalencia de *-ito* en Madrid, cuyo uso se identifica “como un rasgo del lenguaje popular, madrileño y doméstico” (Náñez Fernández, 2006, p. 230).

A partir de documentación de corpus, el estudio de Paredes García (2023) sobre el uso de los diminutivos en el corpus CODEA+ 2015 resulta especialmente valioso para comparar los datos que en esta tesis se presentan. Este estudio señala que es en el siglo XVIII cuando *-ito* sobrepasa a *-illo*, invirtiéndose así la tendencia que hasta el momento había caracterizado el uso de estos dos sufijos. Solo para documentos madrileños, este autor refiere una frecuencia de aparición de *-illo* del 21.8 % frente al 30.3 % de *-ito*, lo que se traduce en la prevalencia de *-ito* en esta provincia, tal y como se confirma en nuestro corpus. Sin embargo, los datos de ODE para Madrid pueden matizar en algunos aspectos este resultado. En primer lugar, en lo relativo al tamaño y la representatividad de las muestras. Los porcentajes mencionados reflejan el uso de *-ito*

en el conjunto total de textos de Madrid incluidos en el corpus CODEA+ 2015, desde el año 1200 hasta 1800. Ciertamente, los casos de *-ito* son notablemente superiores en el siglo XVIII, pero los porcentajes se aplican a los 435 documentos madrileños de que consta el corpus, de los cuales tan solo 19 pertenecen al siglo XVIII y proceden de la capital y de los municipios de Alcalá de Henares, Daganzo y Valdemoro¹¹⁶. En *ODE*, donde el corpus madrileño del siglo XVIII supera al CODEA+ 2015 en volumen de palabras y también en el número de municipios representados¹¹⁷, las diferencias entre el uso de *-illo* e *-ito* son aun mayores. Se constata igualmente el predominio de este último frente a *-illo*, sin embargo, la diferencia de tamaño entre ambas muestras permite apreciar mejor la magnitud del cambio *-illo* > *-ito* en este siglo en Madrid, ya que en un corpus mayor las frecuencias de estos diminutivos reflejan un uso aun más diferenciado que el que se refleja en CODEA+ 2015 (54.6 % de *-ito* frente a un 41.5 % para *-illo* en *ODE*).

De esta manera, el siglo XVIII supone un punto de inflexión en el uso de los diminutivos en Madrid, pero cabe preguntarse si es a partir de este momento cuando *-ito* sigue aumentando su uso hasta convertirse en el diminutivo más frecuente, posición que ocupa actualmente. ¿Es este siglo el periodo en el que se invierte de manera definitiva la tendencia en el uso de los sufijos?, ¿puede darse el proceso de cambio *-illo* > *-ito* por culminado? En este sentido, los datos de *ODE* son reveladores y aportan continuidad al estudio de los diminutivos en Madrid a partir de la documentación del CODEA+ 2015 para poder responder a estas preguntas. La superioridad de *-ito* sobre *-illo* se cumple en *ODE* para el siglo XVIII, pero no para el XIX, como se expondrá más adelante.

La bibliografía previa sitúa el inicio del creciente uso de *-ito* a partir de la segunda mitad del siglo XV (González Ollé, 1962; Pharies, 2002) o siglo XVI (Paredes García, 2023), momento a partir del cual no haría más que aumentar gradualmente hasta que logró sobreponerse a *-illo*. Paredes García (2023, p. 125) sitúa este hecho en la segunda mitad del siglo XVIII, mientras que Lapesa lo retrasa hasta el siglo XIX (2014, p. 334). Dada esta horquilla de variación, parece lógico pensar que el cambio de *-illo* a *-ito* no se produjo al mismo tiempo ni se extendió a la misma velocidad en todas las zonas, por lo que la datación puede variar significativamente en función de los datos tenidos en cuenta. Las pruebas estadísticas en Madrid se resolvieron a favor de la significancia de los datos: las diferencias en el uso de los diminutivos entre el XVIII y el XIX son significativas ($p < 0.05$). Sin embargo, en Madrid la elección del diminutivo no está fuertemente relacionada con el siglo ($V = 0.18$). La asociación de estas variables es débil, por lo que no se puede afirmar que la variable tiempo incida con la misma intensidad en Madrid que en Almería. A la luz de los datos de *ODE*, el siglo XVIII supone el inicio de un cambio de paradigma en Madrid en cuanto al uso de los

¹¹⁶ Estos documentos suman un total de 5 700 palabras aproximadamente. Cifras obtenidas en una búsqueda propia directamente sobre el sitio web con los filtros activados de provincia y siglo. <http://corpuscodea.es/index-2015.html> [fecha de consulta: 02/06/2023].

¹¹⁷ *ODE* incluye documentación de hasta 13 municipios de Madrid, de los cuales 7 corresponden al siglo XVIII, entre los que se incluye la capital. La relación de municipios madrileños, así como el tamaño del corpus, puede consultarse con detalle en § 2.5.3. *Corpus de control de Madrid*.

diminutivos. Efectivamente, el sufijo *-ito* se impone en este siglo, pero lejos de estabilizarse, vuelve a descender en el siglo XIX, concretamente un 21 %. Este dato es bastante revelador si se tiene en cuenta que, de acuerdo con la bibliografía, *-ito* es el diminutivo por excelencia en el mundo hispánico desde el siglo XVIII hasta la actualidad. Las frecuencias de *-ito* en nuestro corpus en el XVIII (54.6 %) están lejos de alcanzar las cifras en torno al 80 % que otros autores han registrado para la variedad madrileña actual (Reynoso Noverón, 2001; Malaver y Paredes García, 2020), por lo que en el periodo estudiado (ss. XVIII-XIX) el uso de estos dos sufijos está todavía bastante equilibrado en uno y otro siglo según lo que refleja la documentación de corte notarial analizada.

La rivalidad entre *-illo/-ito* se extenderá en el tiempo de manera gradual, llegando a ocupar también el siglo XIX. Durante un lapso de dos siglos la forma innovadora convive con la antigua, haciendo que ambos diminutivos coexistan con frecuencias más o menos similares, sin una clara ventaja de uno sobre otro. En el siglo XIX ambos diminutivos alternan y la tendencia aventajada que *-ito* había alcanzado en el siglo anterior no se prolonga en el tiempo. Los datos analizados reflejan que en la provincia de Madrid, el siglo XIX responde a un periodo en el que el proceso de cambio no habría concluido todavía, sin que se pueda confirmar una clara ventaja de *-ito* sobre *-illo*.

Por el contrario, la velocidad con la que se da el cambio de paradigma en el uso de los diminutivos en Almería es mucho mayor. En esta provincia, la evolución cronológica de los sufijos refleja que el cambio de tendencia en el transcurso de un siglo es mucho más acentuado y con frecuencias muy dispares entre los diminutivos: el empleo de *-ico* cae en el XIX más de un 90 % mientras que *-illo* asciende un 62.6 %. Los escasos testimonios de *-ico* en el corpus quedan reducidos en el siglo XIX a la zona más oriental de la provincia, tal vez influida por su cercanía con Murcia, donde también el uso de *-ico* es un rasgo caracterizador de esta región en el siglo XVIII (Vivancos Mulero, 2012). En la provincia de Almería, y en contraposición a lo que ocurre en Madrid, los datos del siglo XIX apuntan ya a lo que será la distribución actual de los diminutivos en el español andaluz: *-illo* es el diminutivo que destaca proporcionalmente en Andalucía (NGLE § 9.11, p. 631). Si se comparan los datos de Almería con los de las provincias limítrofes, en el español actual *-illo* abunda especialmente en la zona central y oriental, seguido de lejos por *-ico* en las provincias de Granada, Jaén y Almería, según datos del ALEA (Uritani y Berrueta, 1985). En las provincias orientales, el uso de *-ico* es un rasgo dialectal que convive con el uso general de *-illo*, aunque aquel tiene mayores connotaciones afectivas (Ariza, 1998, p. 359) y se relaciona con los niveles socioculturales más bajos, como sucede en el habla de Granada (Manjón-Cabeza, 2012, p. 73).

Por último, cabe plantearse el papel que juega el siglo XVIII en los mismos términos en que lo hace Sánchez Lancis (2012) como frontera temporal del español, y más concretamente, en lo que atañe a la evolución de los diminutivos en las provincias de Almería y Madrid de acuerdo con los datos de ODE analizados.

Para la periodización del español hablado en Madrid, el siglo XVIII marca una frontera en cuanto a que supone una ruptura con el paradigma anterior: *-ito* acaba por

imponerse a *-illo*, iniciando así un cambio de tendencia que acaba con la clara supremacía de *-illo*. Pero este siglo no define un uso estable y predominante de *-ito*, sino más bien el comienzo de un periodo de fluctuaciones entre *-ito* e *-illo* que parece alargarse en el tiempo. El cambio que se produce en este periodo justifica suficientemente la importancia del siglo XVIII, que en el caso de Madrid estaría también ligada a determinados factores externos. Como capital, se ha señalado la influencia de Madrid en la difusión de términos léxicos (Vázquez Balonga, 2019), pero también de otros fenómenos pertenecientes a diferentes planos de la lengua que se irían extendiendo de arriba a abajo en la escala social (Sánchez-Prieto Borja, 2019, p. 107). En lo relacionado con el uso de los diminutivos, Madrid también será determinante para la difusión de *-ito*, dado que en esta provincia su uso es proporcionalmente mayor con respecto a otras regiones (Paredes García, 2023).

Por otro lado, en Almería, el siglo XVIII supone la continuación de la tendencia anterior: *-ico* es el diminutivo marcado dialectalmente y el que aparece con más frecuencia, algo que ya era notorio en el español auri secular. El siglo XVIII forma parte de un periodo previo al cambio de tendencia que implicará el aumento significativo de *-illo* en Andalucía, incluso en aquellas provincias, como Almería, en las que coexiste el uso regional de *-ico*. Este cambio ocurre a mayor velocidad de lo que se desarrolla en la variedad madrileña, pues en el siglo XIX hay un predominio claro de *-illo* frente al resto de sufijos, lo que revela ya en este periodo un patrón bastante cercano a la distribución actual de los diminutivos en esta región.

En definitiva, se confirma que el siglo XVIII actúa como frontera temporal en la evolución de los diminutivos, pero de manera muy distinta en las provincias analizadas, según los datos reflejados en los inventarios. En Madrid supone el inicio de una fase fundamental dentro del proceso de cambio *-illo* > *-ito* (*-ito* sobrepasa por primera vez a *-illo*), mientras que en Almería representa el final de un periodo de estabilidad (*-ico* continúa predominando en la región). En este sentido, la periodización del uso de los diminutivos difiere en base a la variedad dialectal que se analiza, lo que demuestra, una vez más, que los estudios realizados sobre determinados fenómenos del español en general muestran una visión de conjunto que varía sustancialmente cuando el foco se sitúa en zonas geográficas o variedades concretas. Los corpus regionales que aquí se han analizado nos permiten fijar con mayor precisión la evolución de los diminutivos *-ico*, *-ito* e *-illo* entre el español clásico y el español actual. La distribución que se ha descrito a partir de ODE aporta nuevos datos sobre cómo se produjo este cambio en las provincias de Almería y Madrid, siendo siempre conscientes de la limitación que supone tratar de expresar en términos cuantitativos la evolución de un fenómeno como el que aquí se trata.

La configuración documental del corpus analizado, segmentado en corpus de estudio y corpus de control, se ve justificada en el hecho de que las frecuencias de uso de los diminutivos en una zona o en un momento concreto no adquieren valor alguno si no se contraponen con las de otras zonas, épocas o tipologías textuales. Incluso el uso aislado de un diminutivo particular no se entiende sin el contraste con sus análogos, los cuales constituyen también alternativas posibles dentro del espectro de variación lingüística. Esta afirmación, en principio evidente, se ratifica en que en muy pocas

ocasiones los inventarios manifiestan el uso de uno solo de los diminutivos, –y cuando lo hacen no se debe, con toda probabilidad, a la inexistencia de esa forma–, del mismo modo que un hablante actual no emplea una única variante en todos los contextos. En la mayoría de los inventarios alternan dos y hasta tres formas, algo que se hacía especialmente visible en los gráficos de dispersión de los diminutivos en uno y otro corpus (Figura 77, Figura 78, Figura 82 y Figura 83). Creemos conveniente destacar este hecho, ya que subraya la importancia y el valor que tienen los estudios de tipo contrastivo para muchos de los fenómenos lingüísticos en nuestra lengua y, de manera especial, para la evolución de los diminutivos a partir de corpus con documentación fielmente datada.

A modo de síntesis, los principales resultados obtenidos en el análisis estadístico y que sirven para condensar todo lo mencionado en este capítulo se enumeran junto al resto de las conclusiones derivadas de la tesis doctoral en el apartado § 8. *CONCLUSIONES*.

8. CONCLUSIONES

Esta tesis doctoral se ha propuesto examinar el proceso de cambio de los sufijos diminutivos *-ico*, *-ito* e *-illo* en el español almeriense de los siglos XVIII y XIX, prestando especial atención a la evolución cronológica y a la distribución geográfica del fenómeno objeto de estudio. Con la intención de completar el análisis diatópico contrastivo, se ha analizado de forma paralela un corpus de inventarios de Madrid que ha servido para la comparación de los datos de Almería. Las conclusiones que se han alcanzado con esta investigación se vinculan con cada uno de los objetivos específicos que nos planteábamos al inicio y que se dividen en dos bloques: los objetivos encaminados a la creación del corpus documental (objetivos del 1 al 5) y los objetivos considerados propiamente lingüísticos (objetivos del 6 al 9).

De acuerdo con los objetivos de carácter tecnológico, se han seleccionado en total 97 inventarios de bienes de los siglos XVIII y XIX procedentes de Almería y Madrid que constituyen el corpus documental, fragmentado en corpus de estudio (52 documentos) y corpus de control (45 documentos) (objetivo 1). El conjunto completo de los documentos se ha transcrito siguiendo el estándar de edición XML-TEI (objetivo 2), para posibilitar su publicación en línea y la posterior explotación del corpus. El procesamiento informático de los textos ha contemplado las fases de normalización, anotación morfosintáctica y lematización de los textos con la herramienta TEITOK (objetivo 3). Esta metodología se ajusta a la desarrollada en el corpus *ODE*, del que forma parte el corpus de esta tesis. Todo el proceso de elaboración del corpus, desde su edición digital en XML-TEI (§ Capítulo 3. *EDICIÓN DIGITAL DEL CORPUS EN XML-TEI*) hasta su tratamiento lingüístico en TEITOK (§ Capítulo 4. *CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK*), se ha descrito de manera pormenorizada, con especial atención al funcionamiento de TEITOK y a la toma de decisiones que caracterizan el corpus *ODE*. La descripción del proceso sirve como guía y establece los principios necesarios para la edición y tratamiento de nuevos textos en el corpus (objetivo 4).

Por último, la configuración del corpus garantiza la fiabilidad en la adjudicación de etiquetas morfosintácticas y lemas, así como la recuperación de información mediante un sistema de búsquedas basado en el lenguaje informático CQL. Su uso y el de las expresiones regulares que permiten afinar las consultas en *ODE* se detalla de forma pormenorizada en § 4.8. *Búsquedas en el corpus basadas en lenguaje CQL* (objetivo 5). El procesamiento informático de los textos ha permitido comprobar que las tareas de codificación llevadas a cabo de forma automática deben ser revisadas a fin de conseguir un corpus de calidad con el menor número de fallos posibles, para lo que resulta ineludible la revisión manual de errores derivados de los procesos automáticos. Para este estudio, la revisión manual ha sido especialmente importante en lo que respecta al etiquetado morfosintáctico (véase § 4.10. *Revisión manual de errores en el corpus*).

En cuanto a los objetivos lingüísticos, el estudio de los diminutivos ha conllevado una serie de decisiones metodológicas previas. Para el análisis ha sido necesario seleccionar de entre todas las palabras con sufijación diminutiva en el corpus

solo aquellas en las que el sufijo conlleva un auténtico valor diminutivo. Se ha distinguido entre voces con significado transparente y voces lexicalizadas en base a unos criterios que permiten realizar esta clasificación de manera rigurosa y extensible a otros corpus diacrónicos (objetivo 6). Partiendo de los conceptos de ‘voces transparentes’ y ‘voces opacas’ que establece la *NGLE*, los criterios de clasificación se han regido principalmente por la inclusión de las palabras en el diccionario y por el significado que adquieren las voces en la época de los textos y que puede no estar contemplado en las obras lexicográficas (§ 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*). De este proceso de clasificación se puede concluir que el contexto sirve como elemento desambiguador en un amplio número de casos, facilitando la distinción entre el valor transparente y opaco del diminutivo (§ 6.3.1. *El contexto como elemento desambiguador*). Sin embargo, dado que es evidente que el proceso de lexicalización se lleva a cabo gradualmente, en otros muchos casos esta distinción ha supuesto un costoso ejercicio de reflexión y de análisis de cada ejemplo concreto (§ 6.3.2. *Casos dudosos y problemas para la cuantificación*). Al respecto se pueden mencionar las palabras *azadilla*, *bacinilla*, *banquillo*, *braserilla*, *cencerrica*, *cortinilla*, *cucharilla*, *mesilla*, *platillo* y *puntilla*. En algunas de ellas esta dificultad ha estado motivada por la variación que presenta el diminutivo (como ocurre en *azadilla/azadica* o *mesilla/mesita/mesica*), característica que hace pensar en una fase previa a la fijación definitiva del término. En otras, a pesar de la variación y del valor lexicalizado que se incluye en los diccionarios, los casos del corpus se identifican con un uso claramente transparente del término asociado con la disminución de tamaño (como en *platillo*, *cucharilla* o *banquillo*).

En total, se han obtenido 372 formas diminutivas (141 bases léxicas) en las que el sufijo se relaciona principalmente con el significado objetivo de disminución, frente a los 769 casos (100 bases léxicas) que han desarrollado valores específicos.

Con respecto al valor semántico de las formas lexicalizadas (objetivo 7), el corpus evidencia que la mayoría de las voces en las que el sufijo ha experimentado un cambio de significado se corresponden con nombres de telas o tejidos (*albornocillo*, *algodoncillo*, *apretadillo*, *batistilla*, *beatilla*, *bocadillo*, *borloncillo*, *hiladillo*, *holandilla*, *escamilla*, *espumilla*, *estopilla*, *jerguilla*, *lanilla*, *medianillo* y *picotillo*), así como con las labores o adornos que se aplicaban a las telas (*bolillo*, *canutillo*, *confitillo*, *cordoncillo*, *felpilla*, *granillo*, *gusanillo*, *puntilla*, *vainica*), en las que el sufijo se asocia con el significado de ‘cierto tipo de’, que encierra además otras connotaciones relacionadas con la calidad, el valor o el grosor del material. A este conjunto de palabras le siguen otros grupos de voces lexicalizadas pertenecientes al campo de la vestimenta (*almilla*, *bolsillo*, *calzoncillo(s)*, *capotillo*, *chaquetilla*, *coletillo*, *cotilla*, *golilla*, *justillo*, *manguito*, *mantilla*, *petillo*, *ropilla* o *tontillo*), de las joyas (*almendrilla*, *arillo*, *broquelillo*, *cintillo*, *gargantilla*, *horquilla*, *manilla* o *mostacilla*) y de una gran variedad de objetos domésticos entre los que se encuentran útiles de labranza (*escardillo*, *esportillo*, *horquillo*), herramientas de trabajo (*cepillo*, *husillo*, *palomilla*, *tornillo*), instrumentos para tejer o relacionados con la costura (*almohadilla*, *bastillo*, *bolillo*, *palillo*), utensilios de cocina (*botillo*, *cocinilla*, *cuartilla*, *cuartillo*, *molinillo*, *morillo*, *paletilla*, *parrilla*, *tenacillas*), mobiliario del hogar (*banquillo*, *bufetillo*, *camilla*,

pajecillo) o materiales para su construcción (*granadillo* y *soguilla*). En casi todas ellas el diminutivo mantiene su alusión al tamaño menor, pero también impera el significado ‘tipo de’ o ‘especie de’ para designar un nuevo objeto destinado a un uso específico distinto al habitual (§ 6.5.2. *Características de los diminutivos lexicalizados*).

El análisis de todas estas voces lexicalizadas ha servido para elaborar un glosario (§ 6.5.3.4. *Glosario*) que incluye la definición y los ejemplos de cada palabra, así como información adicional que proporciona el uso de estas palabras en el corpus (objetivo 8). Se combinan en las entradas del glosario las definiciones de distintas fuentes, tanto de diccionarios de uso general como de obras lexicográficas de uso específico (§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*) y se amplía esta información con la que proporciona el corpus de inventarios, utilizado como fuente primaria.

Los mismos ejemplos del corpus han permitido mejorar la información lexicográfica existente, así como documentar otras voces de las que no hay constancia en los diccionarios (*albornocillo*, *algodoncillo*, *borloncillo*, *escamilla*, *raspilla* y *apretadillo*). Para algunos objetos como *bocadillo* y *botillo*, el corpus constata otros usos además de los que refieren las fuentes lexicográficas: los ejemplos de corpus revelan que el lienzo de *bocadillo* no solo se emplea para elaborar cortinas y visillos, como indica *Autoridades*, sino también para la confección de otras muchas más prendas (tocas, pañuelos, delantales, toallas, camisas, etc.). Por otra parte, un *botillo* era usado no solo como recipiente para contener líquidos, sino también sal; y un *esportillo* podía elaborarse de paja, además de esparto o palma. En otros casos, los diccionarios restringen el uso de determinadas palabras: *coletillo*, según el *DLE*, es un corpiño “usado por las serranas de Castilla”, pero del que *ODE* atestigua un uso mucho más extendido y general sin limitaciones geográficas claras; al igual que *capotillo*, “un traje que se usa poco o nada” según *Autoridades*, pero que se registra con cierta frecuencia en el corpus, especialmente en los inventarios de Madrid.

La consulta de los corpus complementarios ha ayudado a rastrear la difusión de algunas voces escasamente documentadas en las fuentes lexicográficas y de reducido uso en nuestro corpus, lo que nos ha llevado a considerarlas regionalismos: *bastillo*, *batistilla*, *horquillo*, *granillo*. En algunos casos, el carácter regional se vincula únicamente con alguna de las acepciones del término, como ocurre con *manilla* y *francesita*. De las voces que no han sido recogidas en los diccionarios o de las que apenas se tiene información (*albornocillo*, *algodoncillo*, *borloncillo*, *escamilla*, *raspilla* y *apretadillo*) se confirma que su extensión se reduce a zonas concretas, lo que les otorga a algunas un cierto carácter dialectal (§ 6.5.2.1. *Regionalismos y voces poco documentadas*).

Por último, a partir de los ejemplos de corpus asistimos también a la evolución semántica de algunas palabras, objetos que han cambiado su uso, como *mantilla*: empleada mayoritariamente como prenda de abrigo y posteriormente como elemento decorativo en actos solemnes. Se destaca asimismo la entrada de *bufetillo/bufetico/bufetito*. Las distintas acepciones que se ofrecen son fruto del análisis exhaustivo de los ejemplos que proporciona *ODE*, cotejados también con los de otros corpus complementarios. Los inventarios permiten documentar magníficamente la evolución semántica del término, que estuvo ligada a su vez a la variación morfológica

del diminutivo. El estudio léxico de esta palabra refleja los distintos usos a los que se destinó el objeto en la época, algunos de ellos no descritos lexicográficamente (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). Este caso ejemplifica de manera clara la dificultad que ha supuesto diferenciar entre diminutivos transparentes y lexicalizados, más si cabe cuando se trata de fuentes históricas.

A partir del recuento de los diminutivos transparentes se ha realizado un análisis estadístico que permite estudiar la distribución diatópica y cronológica de los sufijos *-ico*, *-ito* e *-illo* durante los siglos XVIII y XIX en Almería y Madrid (objetivo 9), aspecto central en los objetivos de carácter lingüístico de esta tesis y al que se dedica el Capítulo 7. A partir del estudio estadístico de los datos, cuyos principales resultados se mencionaron exhaustivamente en la discusión (§ 7.5. *Discusión*), se extraen las siguientes conclusiones:

1. Las diferencias diatópicas entre Madrid y Almería son significativamente mayores en el siglo XVIII, sobre todo en lo relativo al uso de *-ico* (14.5 veces superior en Almería) e *-ito* (2.7 veces superior en Madrid), de acuerdo con la razón de probabilidad calculada para cada diminutivo en una y otra provincia. En este siglo el origen geográfico incide significativamente en el impacto de los resultados con una intensidad relativamente fuerte sobre la elección del diminutivo ($V = 0.56$).
2. El uso de los sufijos en Almería y Madrid tanto en el siglo XVIII como en el XIX refleja tendencias opuestas: en el XVIII aumenta la frecuencia de *-ico* en Almería, mientras que en Madrid es apenas perceptible, y lo contrario ocurre con *-ito*, el sufijo más frecuente en Madrid y el menos usado en Almería en esta centuria; en el XIX, donde las diferencias no son tan destacadas, de nuevo *-illo* e *-ito* tienen preferencias contrarias, el primero en Almería y el segundo en Madrid. Para ello ha sido reveladora la prueba de los residuos de Pearson, en la que claramente se ve el uso opuesto de los diminutivos en una y otra región en ambos siglos (p. 342 y p. 345).
3. En lo que respecta a la cronología del uso de los diminutivos, el sufijo *-ico* conserva en el siglo XVIII un fuerte carácter dialectal que lo hace exclusivo de determinadas regiones. Es característico en Almería, donde alcanza las cotas más altas de aparición en este siglo, seguido de lejos por *-illo* e *-ito*. Por otra parte, en Madrid en el siglo XVIII se produce el avance de *-ito* sobre *-illo*, aunque las diferencias entre uno y otro no son tan destacadas (§ 7.4.2.3. *Alternancia de -ito/-illo en el corpus de Madrid*). Estas observaciones a partir de los datos de corpus coinciden con la bibliografía previa sobre la distribución cronológica de los diminutivos.
4. El proceso de cambio lingüístico que hará de *-ito* el diminutivo más frecuente en el español desde el siglo XVIII en adelante, no parece fijado del todo durante

este periodo de manera definitiva en Madrid. En contra de lo esperable, en el siglo XIX volverá a experimentar un retroceso, poco destacado, con respecto a *-illo*. A la vista de este resultado, se interpreta que en esta provincia el cambio ocurre de forma más lenta y gradual, dadas las fluctuaciones que afectan al uso de ambos diminutivos, sin que tampoco en el siglo XIX se pueda confirmar la clara prevalencia de *-ito* sobre *-illo*.

5. Los datos cuantitativos mostrados nos han permitido detallar algunos aspectos sobre la periodización de la evolución de los diminutivos en los siglos XVIII y XIX en dos zonas diferenciadas lingüísticamente como Madrid y Almería. En primer lugar, se destaca la importancia de reescribir la historia de algunos fenómenos del español desde el estudio de fuentes no literarias, y más concretamente, de aquellas que puedan ser fielmente datadas y vinculadas a una determinada zona geográfica. En segundo lugar, se incide en la necesidad de ampliar los estudios diacrónicos en torno al uso del diminutivo con el empleo de técnicas estadísticas apropiadas y avanzadas que permitan conocer con mayor profundidad la variación lingüística y la incidencia de los factores que intervienen en ella. Solo a partir de fuentes correctamente datadas y que puedan ser contrastadas desde el punto de vista histórico se podrá ahondar en los procesos de cambio lingüístico y conocer con mayor precisión su difusión y consolidación en las distintas épocas y regiones de habla hispánica.

8.1. Futuras vías de investigación

El corpus creado para esta tesis doctoral se ha demostrado eficaz para el análisis de los diminutivos que aquí se ha planteado. De manera particular, la anotación morfosintáctica se revela como una función especialmente clave y necesaria para los estudios de carácter morfológico a partir de corpus diacrónicos. Asimismo, el método estadístico desarrollado se ha mostrado útil para obtener resultados concluyentes en torno a la evolución de los diminutivos en los siglos y lugares seleccionados. Sin embargo, en futuras investigaciones podrían precisarse más los resultados ampliando el corpus documental.

La ampliación del corpus supone aumentar la red de puntos de ambas provincias con el objetivo de obtener una muestra representativa de toda la provincia en su conjunto, incluyendo más poblaciones rurales y no solo urbanas con las que priorizar los textos pertenecientes a pequeños municipios. El incremento del volumen del corpus ofrecerá la posibilidad de complementar el estudio provincial que se ha hecho con un estudio de tipo intradialectal que permita aumentar la granularidad de los datos y establecer diferencias diatópicas entre los municipios de una misma región. En el caso de Almería, será posible detectar áreas de influencia en una región caracterizada dialectalmente. Con respecto a Madrid, se ha demostrado que los primeros casos de *-ito* en la provincia proceden precisamente de textos producidos en los pueblos (Paredes García, 2023, p. 132), por lo que parece necesario un estudio que tenga en cuenta el

contraste del habla urbana y rural en la difusión del cambio de los diminutivos. La gran mayoría de los estudios diacrónicos se han centrado en el análisis de textos emanados desde las capitales de provincia, mientras que las zonas rurales han quedado más desatendidas. En este sentido, son necesarias nuevas investigaciones que ayuden a comprender la dirección de los procesos de cambio, para lo que es ineludible tener en cuenta también el habla de las pequeñas poblaciones.

Junto a la creación de una red de municipios más densa sería también conveniente seleccionar documentos de otros años dentro de los mismos siglos y afinar la periodización que se ha establecido en esta tesis. Aunque la división amplia por siglos (XVIII y XIX) ya ofrece diferencias significativas, en trabajos futuros se podría complementar fijando bloques temporales más reducidos. Un análisis contrastivo con cortes cronológicos adicionales permite fragmentar aun más los datos con el objetivo de determinar con mayor precisión patrones o cambios variacionales en el uso de los diminutivos en periodos más concretos.

Por otra parte, la documentación ya existente de Almería y Madrid se presta a la comparación diatópica de nuevas regiones representadas en el corpus *ODE*. Actualmente, este corpus recoge inventarios del siglo XVIII procedentes de todas las provincias andaluzas, por lo que no se descarta la posibilidad de hacer un estudio interprovincial que analice el uso de los diminutivos en el español dieciochesco de Andalucía y que trate de establecer, por ejemplo, diferencias significativas entre el oriente y occidente.

En cuanto a nuevos estudios contrastivos desde una perspectiva diacrónica, el *CorLexIn* nos servirá como corpus de control para el siglo XVII. A los inventarios de este corpus repartidos por gran parte de la geografía peninsular se podrán sumar los aquí analizados para los siglos XVIII y XIX, añadiendo así un episodio previo al estudio de los diminutivos y analizando el fenómeno durante el avance de tres siglos completos considerados clave para su evolución. En primera instancia, del *CorLexIn* interesan principalmente los textos que conciernen a las provincias objeto de nuestro estudio: Almería y Madrid, pero, en segundo término, este análisis se podrá extender al de otras provincias utilizando también los datos que proporciona *ODE*. La comparabilidad textual para futuros análisis tanto desde el punto de vista diacrónico como diatópico queda asegurada utilizando uno y otro corpus.

Por último, cabe destacar que las características tecnológicas y la metodología de elaboración del corpus de esta tesis se fundamentan en la máxima explotación de los datos que contiene y que van más allá de los objetivos lingüísticos en torno al uso de los diminutivos que han centrado esta investigación. En la selección y edición de los textos se han tenido en cuenta muy diversos aspectos a fin de que el corpus se muestre útil para el estudio de otras variables y otros intereses. Con especial atención se han revisado la normalización y anotación morfosintáctica, no solo para garantizar la calidad del presente estudio, sino con la intención de ofrecer un corpus fidedigno con el que analizar otros aspectos léxicos, fonéticos y morfológicos. Algunas de las posibilidades que ofrecen los inventarios ya se mostraban a raíz de los ejemplos de distintos tipos de búsquedas en *ODE* (§ 4.8.3. *Casos prácticos de búsquedas en el corpus ODE* y § 4.9. *Interfaz de consulta del corpus ODE*). En estos apartados se adelantaban solo algunas

de las cuestiones que muy probablemente merezcan en el futuro nuestra atención e interés. En definitiva, se asegura que el corpus seguirá siendo el punto de partida para el análisis de nuevas investigaciones, para lo que se constituye como un recurso inagotable del que queda sobradamente demostrado su enorme potencial.

SUMMARY AND CONCLUSIONS

This study aims to examine the process of change of the diminutive suffixes *-ico*, *-ito* and *-illo* in Almerian Spanish in the 18th and 19th centuries, paying special attention to the chronological evolution and geographical distribution. In order to complete the contrastive diatopic analysis, a corpus of inventories from Madrid has been analysed, which has been used to compare the data from Almería. The conclusions reached in this study are linked to each of the specific objectives we set out at the beginning, which are divided into two blocks: the objectives aimed at the creation of the corpus (objectives 1 to 5) and the objectives considered strictly linguistic (objectives 6 to 9).

In accordance with the technological objectives, 97 inventories of goods of 18th and 19th century from Almeria and Madrid have been selected to constitute the documentary corpus, fragmented into a study corpus (52 documents) and a control corpus (45 documents) (objective 1). The complete set of documents has been transcribed following the XML-TEI standard (objective 2), to enable its online publication and the subsequent optimization of the corpus. The computer processing of the texts included the phases of normalization, morphosyntactic annotation and lemmatization of the texts with the TEITOK tool (objective 3). This methodology is in line with that developed in the *ODE* corpus, of which the corpus of this thesis forms part. The whole process of elaboration of the corpus, from its digital edition in XML-TEI (§ Chapter 3. 3. *EDICIÓN DIGITAL DEL CORPUS EN XML-TEI*) to its linguistic processing in TEITOK (§ Chapter 4. 4. *CONFIGURACIÓN DEL CORPUS LINGÜÍSTICO EN TEITOK*), has been described in detail, with special attention to the functions of TEITOK and the decision-making that characterizes the *ODE* corpus. The description of the process serves as a guide and establishes the principles necessary for the editing and processing of new texts in the corpus (objective 4).

Finally, the corpus configuration ensures reliability in the allocation of morphosyntactic tags and lemmas, as well as information retrieval of searches based on the CQL language. Its use and the use of regular expressions to refine queries in *ODE* are detailed in § 4.8. *Búsquedas en el corpus basadas en lenguaje CQL* (objective 5). The computer processing of the texts has shown that the coding tasks carried out automatically must be revised to achieve a high-quality corpus with as few errors as possible, for which the manual revision of errors derived from the automatic processes is essential. For this study, manual review has been particularly important in morphosyntactic tagging (§ 4.10. *Revisión manual de errores en el corpus*).

Regarding the linguistic objectives, the study of diminutives has involved some methodological decisions. For the analysis it has been necessary to select from among all the words with diminutive suffixation in the corpus only those in which the suffix carries an authentic diminutive value. A distinction has been made between voices with a transparent meaning and lexicalized voices based on criteria which allow this classification to be carried out in a rigorous manner and which can be extended to other diachronic corpora (objective 6). Starting from the concepts of ‘transparent voices’ and ‘opaque voices’ established by the *NGLE*, the classification criteria have been governed mainly by the inclusion of the words in the dictionary and by the meaning acquired by

the voices at the time of the texts and which may not be contemplated in the lexicographical works (§ 6.3. *Criterios para la desambiguación de diminutivos transparentes y lexicalizados*). From this classification process, it can be concluded that the context serves as a disambiguating element in many cases, facilitating the distinction between the transparent and opaque value of the diminutive (§ 6.3.1. *El contexto como elemento desambiguador*). However, since the process of lexicalization is carried out gradually, in many other cases this distinction has involved an exercise of reflection and analysis of each specific example (§ 6.3.2. *Casos dudosos y problemas para la cuantificación*). In this respect, we can mention the words *azadilla*, *bacinilla*, *banquillo*, *braserilla*, *cencerrica*, *cortinilla*, *cucharilla*, *mesilla*, *platillo* and *puntilla*. In some of them, this difficulty has been caused by the variation in the diminutive (as in *azadilla/azadica* or *mesilla/mesita/mesica*), a characteristic which suggests a previous stage to final stabilization of the term. In other cases, despite the variation and the lexicalized value included in the dictionaries, the cases in the corpus are identified with a clearly transparent use of the term associated with diminution in size (as in *platillo*, *cucharilla* or *banquillo*).

In total, 372 diminutive forms have been obtained (141 lexical bases) in which the suffix is mainly related to the objective meaning of diminution, compared to 769 cases (100 lexical bases) which have developed specific values.

With regard to the semantic value of the lexicalised forms (objective 7), the corpus shows that most of the words in which the suffix has developed a change of meaning correspond to the names of fabrics or textiles (*albornocillo*, *algodoncillo*, *apretadillo*, *batistilla*, *beatilla*, *bocadillo*, *borloncillo*, *hiladillo*, *holandilla*, *escamilla*, *espumilla*, *estopilla*, *jerguilla*, *lanilla*, *medianillo* and *picotillo*), as well as with the work or ornaments that were applied to fabrics (*bolillo*, *canutillo*, *confitillo*, *cordoncillo*, *felpilla*, *granillo*, *gusanillo*, *puntilla*, *vainica*), in which the suffix is associated with the meaning of ‘a certain type of’, which also carries other connotations related to the quality, value or thickness of the material. This group of words is followed by other groups of lexicalized words belonging to the field of clothing (*almilla*, *bolsillo*, *calzoncillo(s)*, *capotillo*, *chaquetilla*, *coletillo*, *cotilla*, *golilla*, *justillo*, *manguito*, *mantilla*, *petillo*, *ropilla* or *tontillo*), of jewellery (*almendrilla*, *arillo*, *broquelillo*, *cintillo*, *gargantilla*, *horquilla*, *manilla* or *mostacilla*) and of a wide variety of domestic objects including farming tools (*escardillo*, *esportillo*, *horquillo*), work tools (*cepillo*, *spindle*, *palomilla*, *tornillo*), knitting or sewing tools (*almohadilla*, *bastillo*, *bolillo*, *palillo*), kitchen utensils (*botillo*, *cocinilla*, *cuartilla*, *cuartillo*, *molinillo*, *morillo*, *paletilla*, *parrilla*, *tenacillas*), household furniture (*banquillo*, *bufetillo*, *camilla*, *pajecillo*) or building materials (*granadillo* y *soguilla*). In almost all of them, the diminutive maintains its allusion to the smaller size, but the meaning ‘type of’ or ‘kind of’ also prevails to designate a new object destined for a specific use different from the usual one (§ 6.5.2. *Características de los diminutivos lexicalizados*).

The analysis of all these lexicalized words has been used to compile a glossary (§ 6.5.3.4. *Glosario*) which includes the definition and examples of each word, as well as additional information provided by the corpus (objective 8). The glossary entries combine definitions from different sources, both from general dictionaries and from

lexicographical specific works (§ 6.5.3.2. *Fuentes utilizadas para la elaboración del glosario*) and this information is supplemented by the information provided by the corpus of inventories used as the primary source.

The same corpus examples have made it possible to improve the existing lexicographical information, as well as to document other words not included in dictionaries (*albornocillo*, *algodoncillo*, *borloncillo*, *escamilla*, *raspilla* and *apretadillo*). For some objects such as *bocadillo* and *botillo*, the corpus shows other uses in addition to those referred to in the lexicographical sources: the corpus examples show that the linen of *bocadillo* is not only used to make curtains and net curtains, as *Autoridades* indicates, but also for making many other garments (headdresses, handkerchiefs, aprons, towels, shirts, etc.). On the other hand, a *botillo* was used not only as a container to hold liquids, but also salt; and an *esportillo* could be made of straw, as well as esparto grass or palm. In other cases, dictionaries restrict the use of certain words: *coletillo*, according to the *DLE*, is a bodice “worn by the women of Castilla”, but *ODE* attests to a much more widespread and general use without clear geographical limitations; as does *capotillo*, “a costume worn little or not at all” according to *Autoridades*, but which is recorded with some frequency in the corpus, especially in the inventories of Madrid.

Consultation of the complementary corpora has helped us to trace the diffusion of some words scarcely documented in lexicographical sources and of limited use in our corpus, which has led us to consider them as regionalisms: *bastillo*, *batistilla*, *horquillo*, *granillo*. In some cases, the regional character is only linked to one of the meanings of the term, as is the case with *manilla* and *francesita*. Of the words that have not been included in the dictionaries or for which hardly any information is available (*albornocillo*, *algodoncillo*, *borloncillo*, *escamilla*, *raspilla* and *apretadillo*), it is confirmed that their extension is limited to specific areas, which gives some of them a certain dialectal character (§ 6.5.2.1. *Regionalismos y voces poco documentadas*).

Finally, from the corpus examples, we can also see the semantic evolution of some words, objects that have changed their use, such as *mantilla*: mainly used as a garment to wear as a coat and later as a decorative element in solemn ceremonies. Also noteworthy is the entry for *bufetillo/bufetico/bufetito*. The different meanings offered are the result of an exhaustive analysis of the examples provided by *ODE*, also cross-checked with those of other complementary corpora. The inventories provide an excellent documentation of the semantic evolution of the term, which in turn was linked to the morphological variation of the diminutive. The lexical study of this word reflects the different uses to which the object was put at the time, some of them not described lexicographically (§ 6.5.2.2. *La variación del diminutivo en voces opacas: el caso particular de «bufetillo»*). This case clearly exemplifies the difficulty of differentiating between transparent and lexicalised diminutives, even more so when it comes to historical sources.

From the count of the transparent diminutives, a statistical analysis has been carried out to study the diatopic and chronological distribution of the suffixes *-ico*, *-ito* and *-illo* during the 18th and 19th centuries in Almería and Madrid (objective 9), a central aspect in the linguistic objectives of this thesis (Chapter 7). On the basis of the

statistical study of the data, whose main results were mentioned exhaustively in the discussion (§ 7.5. *Discusión*), we can conclude that:

1. The diatopic differences between Madrid and Almería are significantly greater in the 18th century, especially with regard to the use of *-ico* (14.5 times higher in Almería) and *-ito* (2.7 times higher in Madrid), according to the likelihood ratio calculated for each diminutive in both provinces. In this century, geographical origin significantly affects the impact of the results with a relatively strong intensity on the choice of diminutive ($V = 0.56$).
2. The use of suffixes in Almería and Madrid in both the 18th and 19th centuries reflects opposite trends: in the 18th century the frequency of *-ico* increases in Almería, while in Madrid it is barely perceptible, and the opposite happens with *-ito*, the most frequent suffix in Madrid and the least used in Almería in this century; in the 19th century, where the differences are not so outstanding, again *-illo* and *-ito* have opposite preferences, the former in Almería and the latter in Madrid. The test of Pearson's residuals has been revealing in this respect, in which the opposite use of diminutives in both regions in both centuries is clearly seen (p. 342 and p. 345)
3. As far as the chronology of the use of diminutives is concerned, the suffix *-ico* conserves a strong dialectal character in the 18th century, which makes it exclusive to certain regions. It is characteristic in Almería, where it reached the highest levels of occurrence in this century, followed by *-illo* and *-ito*. On the other hand, in Madrid in the 18th century, there is an advance of *-ito* over *-illo*, although the differences between one and the other are not so outstanding (§ 7.4.2.3. *Alternancia de -ito/-illo en el corpus de Madrid*). These observations from the corpus data coincide with previous literature on the chronological distribution of diminutives.
4. The process of linguistic change that will make *-ito* the most frequent diminutive in Spanish from the 18th century onwards, does not seem to be completely fixed during this period in a definitive way in Madrid. Contrary to what might be expected, in the 19th century it will again experience a decline, not very significant, with respect to *-illo*. In view of this result, it can be interpreted that in this province the change occurs more slowly and gradually, given the fluctuations that affect the use of both diminutives, without the clear prevalence of *-ito* over *-illo* being confirmed in the 19th century.
5. The quantitative data shown have allowed us to detail some aspects of the periodisation of the evolution of diminutives in the 18th and 19th centuries in two linguistically differentiated areas such as Madrid and Almería. Firstly, the importance of rewriting the history of some Spanish phenomena from the study of non-literary sources, and more specifically, of those which can be faithfully

dated and linked to a specific geographical area, is highlighted. Secondly, it stresses the need to extend diachronic studies on the use of the diminutive with the use of appropriate and advanced statistical techniques which allow us to gain a deeper knowledge of linguistic variation and the incidence of the factors involved in it. Only on the basis of correctly dated sources that can be contrasted from a historical point of view will it be possible to delve deeper into the processes of linguistic change and to know with greater precision its diffusion and consolidation in the different periods and regions of the Spanish-speaking world.

REFERENCIAS BIBLIOGRÁFICAS

- Alcalá Venceslada, A. (1999). *Vocabulario andaluz*. El Mundo.
- ALEA = Alvar, M. (1960-1973). *Atlas lingüístico y etnográfico de Andalucía*. Universidad de Granada.
- Aleza Izquierdo, M. (2016). El diminutivo en el español de Santo Domingo. *Círculo de Lingüística aplicada a la comunicación* 67, 3-56.
- Aleza Izquierdo, M. (2017). Sobre los americanismos léxicos con sufijos diminutivos en el «Diccionario de americanismos». *Bulletin of Spanish Studies*, 94(5), 711-743. <https://doi.org/10.1080/14753820.2017.1296247>
- Alonso, A. (1961). Noción, emoción, acción y fantasía en los diminutivos. En A. Alonso, *Estudios lingüísticos: temas españoles* (pp. 161-189). Gredos.
- Amaro Martos, I. (2018). El mercado textil en la España del siglo XVIII. En A. Holguera Cabrera, E. Prieto Ustio y M. Uriondo Lozano (coords.), *Coleccionismo, mecenazgo y mercado artístico: su proyección en Europa y América* (pp. 681-699). Universidad de Sevilla, Secretariado de Recursos Audiovisuales y Nuevas Tecnologías. <https://hdl.handle.net/11441/90603>
- Anthony, L. (2015). A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2), 141-161. <https://doi.org/10.17250/khisli.30.2.201308.001>
- Arbeteta, L. (1998). *La joyería española de Felipe II a Alfonso XIII en los museos estatales*. Editorial Nerea.
- Ariza, M. (1998). El sufijo -ico. En C. García Turza, F. González Bachiller y J. J. Mangado (coords.), *Actas del IV Congreso Internacional de Historia de la Lengua Española (La Rioja, 1997)* (pp. 351-360). Arco Libros.
- Arrabal Rodríguez, P. (2020). Edición de un corpus digital de inventarios de bienes. *Procesamiento del Lenguaje Natural*, 65, 67-74.
- Arrabal Rodríguez, P. (2022). TEITOK, a visual solution for XML/TEI encoding: editing, annotating and hosting linguistic corpora. *RIDE*, 15. <https://ride.i-de.de/issues/issue-15/teitok/?hilite=arrabal>
- Arrabal Rodríguez, P. (2023). Un cofrecito o un cofre pequeño: la expresión del tamaño en un corpus madrileño de los siglos XVIII y XIX. *Études romanes de Brno*, 44, 31-45. <https://doi.org/10.5817/ERB2023-1-3>
- Arrabal Rodríguez, P. (en prensa). Los sufijos -ico, -ito, -illo en inventarios de bienes de los siglos XVIII y XIX: significado transparente y lexicalizado del diminutivo. En M. Calderón Campos e I. González Sopeña (eds.), *Scripta manent. Diacronía*

del español, documentación archivística y Humanidades digitales. Peter Lang: Fondo Hispánico de Lingüística y Filología.

- Arrabal Rodríguez, P. y Calderón Campos, M. (2022). Fluctuación del vocalismo átono: variación diastrática en el corpus «Oralia diacrónica del español (ODE)». En J. M.^a García Martín, M. Maiseyenko, F. Ruiz Fernández, N. Campos Carrasco y B. Gutierrez Santaella (eds.), *La historia de la lengua, la dialectología y el concepto de cambio lingüístico en el pensamiento de Eugenio Coseriu* (pp. 341-363). Peter Lang (Romance Literatures and Cultures).
- Autoridades = Real Academia Española (1726-1739). *Diccionario de Autoridades*. <https://apps2.rae.es/DA.html>
- Bartolomé Marcos, L. y Junquera Martínez, A. (2018). Cosas de montañeses (y II): sustantividad y léxico, *Altamira. Revista del Centro de Estudios Montañeses*, LXXXIX, 97-135.
- Bastardín Candón, T. (2015). Léxico de la vida cotidiana en las particiones de bienes del Archivo Histórico Provincial de Cádiz (siglo XVIII). En T. Bastardín Candón y M.^a M. Barrientos Márquez (eds.), *Lengua e historia en el Archivo Histórico Provincial de Cádiz* (pp. 39-56). Universidad de Cádiz.
- Bertolotti, V. y Coll, M. (2014). *Retrato lingüístico del Uruguay. Un enfoque histórico sobre las lenguas en la región*. Facultad de Información y Comunicación- Comisión Sectorial de Educación Permanente.
- Bertolotti, V. y Company Company, C. (2022). Corpus diacrónicos del español en las Américas. En G. Parodi, P. Cantos-Gómez y C. Howe (eds.), *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 45-58). Routledge.
- Blasco Mateo, E. (2002). La lexicalización y las colocaciones. *Lingüística española actual*, 24(1), 35-62.
- Bravo-García, E., Mancera Rueda, A. y Martín Aizpuru, L. (2021). Presentación del proyecto Cuba y Andalucía en el siglo XIX: estudio de los lazos lingüísticos y culturales desde las Humanidades Digitales. *Asociación Argentina de Humanidades Digitales (AAHD)*, Vol. 1(2), 31-49.
- Brezina, V. (2018). *Statistics in corpus linguistics. A practical guide*. Cambridge University Press.
- Brezina, V. (2022). Lancaster Stats Tools online. Lancaster University [software]. <http://corpora.lancs.ac.uk/stats/toolbox.php>

- Brezina, V., Weill-Tessier, P., y McEnery, A. (2020). *LancsBox* (versión 5.x) [software]. <http://corpora.lancs.ac.uk/lancsbox>
- BVMC = *Biblioteca Virtual Miguel de Cervantes*. Alicante, España: Universitat d'Alacant. Recuperado de: <https://www.cervantesvirtual.com/>
- Cabrera Lafuente, A. (2005). Los tejidos como patrimonio: investigación y exposición. *Bienes culturales: Revista del Instituto del Patrimonio Histórico Español*, 5, 5-20.
- Calderón Campos, M. (2014). Muestras de oralidad en el «Corpus diacrónico del español del Reino de Granada» (siglo XVIII). *Oralia: análisis del discurso oral*, 17, 117-145.
- Calderón Campos, M. (2015). *El español del reino de Granada en sus documentos (1492-1833): oralidad y escritura*. Peter Lang.
- Calderón Campos, M. (2018a). Las ‘declaraciones de esencia’ del siglo XVIII: un tipo textual para el estudio de la terminología anatómica. *Dynamis*, 38(2), 427–52.
- Calderón Campos, M. (2018b). Andalucismos en el corpus del reino de Granada. En D. Corbella Díaz, A. Fajardo, y J. Langenbacher-Liebgott (eds), *Historia del léxico español y humanidades digitales* (pp. 317-339). Peter Lang.
- Calderón Campos, M. (2019a). La edición de corpus históricos en la plataforma TEITOK. El caso de Oralia diacrónica del español. *Chimera*, 6, 21-36.
- Calderón Campos, M. (2019b). Los corpus del español clásico y moderno: entre la filología y la lingüística computacional. *Revista de lingüística teórica y aplicada*, 57(2), 41-64.
- Calderón Campos, M. (2019c). La configuración de la variedad meridional en el reino de Granada. En V. Codita Petrea (ed.), *La configuración histórica de las normas del castellano* (pp. 109-134). Tirant Humanidades.
- Calderón Campos, M. (en prensa). Corpus, size and tagging. Methodological strategies for research on the history of -ito, -illo and -ico in Spanish. En M. Calderón Campos y G. Vaamonde (eds.), *Linguistic Corpora and Big Data in Spanish and Portuguese*. De Gruyter, Serie: Humanidades Digitales y Big Data (Gold Open Access).
- Calderón Campos, M. y Díaz Bravo, R. (2021). An online corpus for the study of historical dialectology: «Oralia diacrónica del español». *Digital Scholarship in the Humanities*, Vol. 36, 30-48.
- Calderón Campos, M. y García-Godoy, M.^a T. (2009). El «Corpus Diacrónico del Español del Reino de Granada (CORDEREGRA)». En A. Enrique-Arias (ed.),

Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus (pp. 229-249). Iberoamericana/Vervuert.

- Calderón Campos, M. y García-Godoy, M.^a T. (2023). «Y el dicho testigo dijo su dicho». Gramaticalización y etiquetado de *dicho* en el corpus *Oralia diacrónica del español* (ODE). En P. Giménez Eguíbar, M.^a C. Grande López, V. J. Marcet Rodríguez, L. Martín Aizpuru y R. Sánchez Romo (eds.), *Despertar palabras, renacer historias. Estudios lingüísticos en homenaje a M.^a Nieves Sánchez González de Herrero* (pp. 155-165). Ediciones Universidad de Salamanca. <https://doi.org/10.14201/0AQ0349>
- Calderón Campos, M. y Vaamonde, G. (2020). «Oralia diacrónica del español»: un nuevo corpus de la edad moderna. *Scriptum digital*, 9, 167-189.
- Callebaut, S. (2011). *Entre sistematización y variación: el sufijo diminutivo en España y en Hispanoamérica* [Tesis doctoral]. Universiteit Gent. Repositorio institucional: https://libstore.ugent.be/fulltxt/RUG01/001/786/647/RUG01-001786647_2012_0001_AC.pdf
- Cambridge Sketch Engine (2012). Using Corpus Query Language (CQL) (1.3). Cambridge University Press. <https://www.cambridge.org/sketch/help/userguides/CQL%20Help%201.3.pdf>
- Campos Souto, M. (2016). Léxico del futuro para la lengua del pasado. En R. Cotelo García (coord.), *Entre dos coordenadas: la perspectiva diacrónica y diatópica en los estudios léxicos del español* (pp. 33-72). Cilengua. Fundación San Millán de la Cogolla.
- Campos Souto, M. (2018). Las bases documentales del «NDHE». En D. Corbella, A. Fajardo y J. Langenbacher-Liebgott (eds.), *Historia del léxico español y Humanidades digitales* (pp. 19-45). Peter Lang.
- Cano Aguilar, R. (coord.) (2004). *Historia de la lengua española*. Ariel.
- Caravedo, R. (1999). *Lingüística del corpus. Cuestiones teórico-metodológicas aplicadas al español*. Serie Josse de Kock (dir.). *Gramática española: enseñanza e investigación. Apuntes metodológicos*. Ediciones Universidad de Salamanca.
- Castillo Lluch, M. y Diez del Corral Areta, E. (2018). Fondos documentales hispánicos en Suiza: una exploración de conjunto. *Scriptum digital*, 7, 95-105.
- Castillo Lluch, M. y Diez del Corral Areta, E. (eds.) (2019). *Reescribiendo la historia de la lengua española a partir de la edición de documentos*. Peter Lang.
- Catalán, D. (1989). *Las lenguas circunvecinas del castellano*. Paraninfo.

- Christ, O. (1994). A modular and flexible architecture for an integrated Corpus Query System. *Proceedings of COMPLEX'94. 3rd Conference on computational lexicography and text research*, Budapest, Hungary, 23-32.
- Clavería Nadal, G. (2003). Procesos de lexicalización con sufijos diminutivos en los nombres de plantas. En F. Sánchez Miret (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románicas*, Vol. 3 (pp. 69-82). Max Niemeyer Verlag.
- Clavería Nadal, G. (2019). Historia del léxico. En E. Ridruejo (ed.), *Manual de lingüística española* (pp. 133-166). De Gruyter.
- Company Company, C. (2002). Gramaticalización y dialectología comparada. Una isoglosa sintáctico-semántica del español. *Dicenda. Cuadernos de Filología Hispánica*, 20, 39-71.
- Company Company, C. (2007). *El siglo XVIII y la identidad lingüística de México*. Universidad Nacional Autónoma de México / Academia Mexicana de la Lengua.
- Company Company, C. (2012). El español del siglo XVIII. Un parteaguas lingüístico entre México y España. En M.^a T. García-Godoy (ed.), *El español del siglo XVIII. Cambios diacrónicos en el primer español moderno* (pp. 255-291). Peter Lang.
- Contreras, J. M. y Suñer, A. (2004). Los procesos de lexicalización. En I. Zabala Unzu, E. Pérez Gaztelu y L. Gràcia y Solé (eds), *Las fronteras de la composición en lenguas románicas y en vasco* (pp. 47-108). Universidad de Deusto.
- CORHIBER = Torruella, J. y Kabatek, J. (eds.) *Portal de Corpus Históricos Iberorrománicos*. Recuperado de <http://www.corhiber.org/>
- Coseriu, E. (1965). Discussion. [En Monge 1965]. En G. Straka (ed.), Vol. 1.
- Davies, M. y Parodi, G. (2022). Constitución de corpus crecientes del español. En G. Parodi, P. Cantos-Gómez y C. Howe (eds.), *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 13-32). Routledge.
- DCECH = Corominas, J. y Pascual, J. A. (1984). *Diccionario crítico etimológico castellano e hispánico*. Gredos.
- DEA = Seco, M.; Andrés, O. y Ramos, G. (1999). *Diccionario del español actual*. Aguilar.
- De Benito Moreno, C. (2019). Los corpus del español desde la perspectiva del usuario lingüista. *Scriptum digital*, 8, 1-21.

- Desagulier, G. (2017). *Corpus linguistics and Statistics with R. Introduction to quantitative methods in Linguistics*. Springer.
- DHTT = Dávila Corona, R. M.^a; Durán Pujol, M., y García Fernández, M. (2004). *Diccionario histórico de telas y tejidos: castellano-catalán*. Junta de Castilla y León, Consejería de Cultura y Turismo.
- Díaz Bravo, R. (2015). Herramientas computacionales aplicadas al estudio de la historia de la lengua española. En J. Sánchez Méndez, M. de La Torre y Viorica Codita (eds.), *Temas, problemas y métodos para la edición y el estudio de documentos hispánicos antiguos* (pp. 377-394). Tirant Humanidades.
- DLE = Real Academia Española (2022). *Diccionario de la lengua española*, 23.^a ed., [versión 23.6 en línea]. <https://dle.rae.es>
- Eberenz, R. (1991). Castellano antiguo y español moderno: reflexiones sobre la periodización en la historia de la lengua. *Revista de filología española*, 71, fasc. 1-2, 79-106.
- Eberenz, R. (2009). La periodización de la historia morfosintáctica del español: propuestas y aportaciones recientes. *Cahiers d' études hispaniques medievales*, 32, 181-201.
- Echenique Elizondo, M.^a T. y Martínez Alcalde, M.^a J. (2005). *Diacronía y gramática histórica de la lengua española*. Tirant lo Blanch.
- Eddington, D. (2017). Dialectal variation in Spanish diminutives: a performance model. *Studies in Hispanic and Lusophone Linguistics*, 10(1), 39-66.
- Egido, M.^a C. (2016). Joyas femeninas: un aporte sobre léxico del español colonial americano. *Anuario de Letras. Lingüística y Filología*, Vol. 4(2), 99-151.
- Egido, M.^a C. (2019). La variación diatópica: un reto para el estudio de la Morfosintaxis histórica del español. En V. Codita Petrea y M. de la Torre (eds.), *Tendencias y perspectivas en el estudio de la morfosintaxis histórica hispanoamericana* (pp. 127-151). Iberoamericana/Vervuert.
- Enguita Utrilla, J. M. (1984). Notas sobre los diminutivos en el espacio geográfico aragonés. *Archivo de filología aragonesa*, Vol. 34-35, 229-250.
- Enguita Utrilla, J. M. y Lagüéns Gracia, V. (2019). Los inventarios de bienes aragoneses y su interés para la lexicografía histórica. *Archivo de filología aragonesa*, Vol. 75, 15-40.
- Enrique-Arias, A. (2009). *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Iberoamericana/Vervuert.

- Fábregas, A. (2017). ¿Son algunos interfijos morfemas apreciativos? *ELUA: Estudios de Lingüística*, 31, 135-50.
- Fernández Alcaide, M. (2021). Representación de las hablas andaluzas en los corpus del español y revisión de sus rasgos sintácticos. *Verba: Anuario Galego de Filoloxía*, 48, 1-25. <https://doi.org/10.15304/verba.48.6739>
- Fichero general* = Real Academia Española. *Fichero general de la lengua española*. <https://apps2.rae.es/fichero.html>
- Fontanella de Weinberg, M.^a B. (1962). Algunas observaciones sobre el diminutivo en Bogotá. *Boletín del Instituto Caro y Cuervo*, 17(3), 557-73.
- Fontanella de Weinberg, M.^a B. (1987). *El español bonaerense. Cuatro siglos de evolución lingüística (1580-1980)*. Hachette.
- Gago-Jover, F. y Pueyo Mena, F. J. (2022). *Old Spanish Textual Archive. Manual de consulta. Versión 2.0*. <http://hispanicseminary.org/docs/OSTA-manual.pdf>
- García-Godoy, M.^a T. (ed.) (2012). *El español del siglo XVIII. Cambios diacrónicos en el primer español moderno*. Peter Lang.
- García-Godoy, M.^a T. (2018). El discurso directo en el «Corpus diacrónico del reino de Granada» («CORDEREGRA», 1492-1833). Nuevos datos sobre el pronombre «usted». En M.^a L. Arnal Purroy, R. M.^a Castañer Martín, J. M.^a Enguita Utrilla, V. Lagüéns Gracia y M.^a A. Martín Zorraquino (coords.) *Actas del X Congreso Internacional de Historia de la Lengua Española (Zaragoza, 2015)*, Vol. 1 (pp. 663-677). Institución Fernando el Católico.
- García Pérez, R. (2009). El diminutivo en la novela de tesis: el ejemplo de «Doña Perfecta» de Galdós. *Boletín de Filología*, 44(1), 83-95.
- García Sánchez, J. J. (2015). La variedad sufijal, especialmente la diminutiva, en la toponimia conquense. En J. M. García Martín (dir.), *Actas del IX Congreso Internacional de Historia de la Lengua Española (Cádiz, 2012)*, Vol. 2 (pp. 2239-2250). Iberoamericana/Vervuert.
- Gasso-Villar, E. (2014). El diminutivo en el español coloquial de Córdoba: descripción de valores semánticos y efectos sociopragmáticos [Tesis doctoral]. Universidad Nacional de Córdoba. Repositorio institucional: <https://rdu.unc.edu.ar/handle/11086/1591>
- Gil I. y Torres, M. (2011). Tratamiento lexicográfico de los derivados en -illo e -ito y el proceso de lexicalización. En J. Pazó, I. Gil y M.^a Á. Cano (eds.), *Teoría morfológica y morfología del español* (pp. 253-266). Ediciones Universidad Autónoma de Madrid.

- Girón Alconchel, J. L. (2004). Cambios gramaticales en los Siglos de Oro. En R. Cano (coord.), *Historia de la lengua española* (pp. 859-894). Ariel.
- Gómez Seibane, S. y Ramírez Luengo, J. L. (2007): *El castellano de Bilbao en el siglo XVIII: materiales para su estudio*. Universidad de Deusto.
- González-Espresati García-Medall, C. (2015). *Problemas de morfopragmática del diminutivo en español* [Tesis doctoral]. Universidad de Valladolid. Repositorio institucional: <http://uvadoc.uva.es/handle/10324/16500>
- González Ollé, F. (1962). *Los sufijos diminutivos en castellano medieval*. Gómez.
- González Ollé, F. (1965). Primeros testimonios de algunos sufijos diminutivos en castellano y algunos datos para su historia. En G. Straka, *Actes du X^e Congrès International du Linguistique et Philologie romanes*, (Estrasburgo, 1962), Vol. I (pp. 547-552). Klincksieck.
- González Ollé, F. (2006). Origen de -ito, con una revisión histórica de otros sufijos diminutivos románicos. En E. Cásanova i Herrero y X. Terrado i Pablo (eds.), *Studia in honorem Joan Coromines* (pp. 157-177). Pagès.
- González Salgado, J. A. (2007). La sufijación diminutiva en la toponimia extremeña. En F. Hermoso Ruiz (coord.), *VIII Congreso de Estudios Extremeños: libro de actas* (pp. 1969-1982). Diputación Provincial de Badajoz.
- González Sopena, I. (2022). Documentación notarial extremeña del siglo XVII en «Oralia Diacrónica del Español (ODE)»: el léxico de la vida cotidiana a través de inventarios de bienes pacenses. *Romanica Olomucensia*, 34(1), 13-30.
- Gordón Peral, M. D. (coord.) (2010). *Toponimia de España. Estado actual y perspectivas de la investigación*. De Gruyter.
- GTIRCE = Tejada Fernández, M. (2006). *Glosario de términos de la indumentaria regia y cortesana en España: siglos XVII y XVIII*. Universidad de Málaga.
- Guzmán Riverón, M. y Sáez Rivera, D. M. (eds.) (2016). *Márgenes y centros en el español del siglo XVIII*. Tirant Humanidades.
- Hasselrot, B. (1957), *Etudes sur la formation diminutive dans les langues romanes*. Almqvist and Wiksell.
- Herradón Figueroa, M.^a A. (2019). Joyería femenina en España en el siglo XVIII. Las manillas. En J. Rivas Carmona, I. J. García Zapata (coords.), *Estudios de platería. San Eloy 2019* (pp. 285-300). Universidad de Murcia. <https://doi.org/10.6018/editum.2797>

- Hummel, M. (1997). Para la lingüística de vuestro diminutivo: los diminutivos como apreciativos. *Anuario de Estudios Filológicos*, XX, 191-210.
- Hummel, M. (2015). The semantics and pragmatics of Romance evaluative suffixes. En P. O. Müller, I. Ohnheiser, S. Olsen y F. Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, Vol. 2 (pp. 1528-1545). De Gruyter Mouton.
- Isasi Martínez, C., Martín Aizpuru, L., Pérez Isasi, S., Pierazzo, E. y Spence P. (2020). *Edición digital de documentos antiguos: marcación XML-TEI basada en los criterios CHARTA*. Editorial Universidad de Sevilla.
- Janssen, M. (2012). NeoTag: a POS tagger for grammatical neologism detection. *Proceedings of the 8th International Conference on Language Resources and Evaluation, (LREC 2012), ELRA*, Estambul, Turquía, 2118-2124.
- Janssen, M. (2016). Teitok: Text-Faithful annotated corpora. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Eslovenia, 4037-4043.
- Janssen, M. (2018). Teitok as a tool for dependency grammar. *Procesamiento del Lenguaje Natural*, 61, 185-188.
- Janssen, M., Ausensi, J. y Fontana, J. M. (2017). Improving POS tagging in Old Spanish using TEITOK. *Proceedings of the NoDaLiDa 2017 workshop on Processing Historical Language*, Gotemburgo, Suecia, 2-6.
- Junquera Martínez, A. (2020). No dar puntada sin hilo: bordados y guarniciones del siglo XVII. *Estudios Humanísticos. Filología*, 42, 39-57.
- Kabatek, J. (2013). ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania*, 77, 8-28.
- Kabatek, J. (2016). *Lingüística de corpus y lingüística histórica iberorrománica*. De Gruyter.
- Kilgariff, A. y Kosem, I. (2012). Corpus tool for lexicographers. En S. Granger y M. Paquot (eds.) *Electronic Lexicography* (pp. 31-56). Oxford University Press.
- Kornfeld, L. M. (2021). Appreciative morphology. En A. Fábregas, V. Acedo-Matellán, G. Armstrong, M.^a C. Cuervo, I. Pujol Payet, *The Routledge handbook of Spanish morphology* (pp. 269-284). Routledge.
- Kotrlik, J. W. y Williams, H. A. (2003). The incorporation of effect size in Information Technology, Learning, and Performance Research. *Information Technology, Learning, and Performance Journal*, Vol. 21 (1).

- Laguna Álvarez, M.^a A. (2017). *Voces de la vida cotidiana en protocolos notariales de Burgos (siglo XVIII)* [Tesis doctoral]. Universidad de Burgos. Repositorio institucional: <http://hdl.handle.net/10259/5384>
- Lapesa, R. (2014). *Historia de la lengua española*. Gredos.
- Latorre, F. (1956). Diminutivos, despectivos y aumentativos en el siglo XVII. *Archivo de Filología Aragonesa*, VIII-IX, 105-120.
- Lázaro Mora, F. A. (1976). Compatibilidad entre lexemas nominales y sufijos diminutivos. *Boletín del Instituto Caro y Cuervo*, 31(1), 41-57.
- Lázaro Mora, F. A. (1999). La derivación apreciativa. En I. Bosque y V. Demonte (dirs.), *Gramática descriptiva de la lengua española. Entre la oración y el discurso. Morfología*, Vol. 3 (pp. 4645-4682). Espasa Calpe.
- León-Castro Gómez, M. (2020a). Variación en el empleo del diminutivo en el habla urbana de Sevilla: estudio sociolingüístico sobre el corpus PRESEEA. En D. Repede y M. León-Castro Gómez (coords.), *Patrones sociolingüísticos del español hablado en la ciudad de Sevilla* (pp. 93-116). Peter Lang.
- León-Castro Gómez, M. (2020b). El empleo del diminutivo en la ciudad de Sevilla: perspectivas sociolingüística y pragmática. *Lengua y habla*, 24, 112-131.
- Liria, P. (1974). «Libro de buen amor»: los diminutivos con el sufijo -illo en boca de Trotaconventos. *Romance Notes*, 16(1), 222-227.
- López González, A. M.^a (2005). /-ito/, /-illo/, /-ico/: diminutivos en la radio local de Almería. En M.^a L. Carrió Pastor (coord.), *Perspectivas interdisciplinares de la lingüística aplicada* (pp. 235-244), Vol. 3. Asociación Española de Lingüística Aplicada, AESLA.
- Mackowiack, S. (2018). Distribución geográfica de los diminutivos en España. En J. P. Cabanilles Gomar (ed.), *Jóvenes plumas del hispanismo: nuevos retos y enfoques de la investigación filológica* (pp. 205-223). Fundación Ramón Menéndez Pidal.
- Malaver, I. (2018). Igual, iguales e igualitos: diminutivos en el habla de Caracas. *Oralia: Análisis del discurso oral*, 21(2), 271-293.
- Malaver, I. (2021). Estudio sociopragmático del diminutivo en Granada. En M.^a M. Soto Melgar y A. Zholobova (eds.), *El español de Granada* (pp. 373-399). Peter Lang.
- Malaver, I. y Paredes García, F. (2020). Convergences and divergences in the use of the diminutive in Medellin, Caracas and Madrid. *Spanish in Context*, 17(2), 317-340.

- Manjón-Cabeza Cruz, A. (2012). Los diminutivos en el corpus PRESEEA de Granada. En J. A. Moya Corral y E. Waluch de la Torre (coords.), *Español hablado. Estudios sobre el corpus PRESEEA-Granada* (pp. 55-76). Instituto de Estudios Ibéricos e Iberoamericanos de la Universidad de Varsovia.
- Manjón-Cabeza Cruz, A. (2016). Los sufijos -ico, -ito, -illo en el español de Granada, España. *Spanish in Context*, 13(3), 420-435.
- Martín Rubio, A. (2023). *Las declaraciones de cirujanos en la historia de la lengua española. Edición digital y estudio lingüístico (1505-1834)* [Tesis doctoral]. Universidad de Granada. Repositorio institucional: <https://hdl.handle.net/10481/85058>
- Martín Zorraquino, M.^a A. (2012). Los diminutivos en español: aspectos morfológicos, semánticos y pragmáticos. Los valores estilísticos de los diminutivos y la teoría de la cortesía verbal. En L. Luque Toro, J. F. Medina Montero y R. Luque Colautti (eds.), *Léxico español actual III* (pp. 123-140). Università Ca' Foscari di Venezia.
- Mayoral Hernández, R. (2022). Estudio de la variación (morfo)sintáctica mediante corpus. En G. Parodi, P. Cantos-Gómez y C. Howe (eds.), *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 137-149). Routledge.
- Medina Domínguez, N. (2022, mayo 16). Evolución de la sufijación apreciativa en el habla rural del siglo XX al siglo XXI: los sufijos -ito e -illo en el «Atlas lingüístico de la Península Ibérica» y en el «Corpus Oral y Sonoro del Español Rural» [Comunicación en congreso]. XII Congreso Internacional de Historia de la Lengua Española, León, España.
- Miguel Borge, M. (2016). Inventario de bienes de la fortaleza de Villalpando en 1633. Estudio del léxico del lujo. *Anuario del Instituto de Estudios Zamoranos Florián de Ocampo*, 31, 335-357.
- Miguel Borge, M. (2018). Entre delantales y mandiles en la Tierra de Campos del siglo XVII. En E. Álvarez García y M.^a V. Barrio Corral (eds.), *Estudios panhispánicos: lingüística teórica y aplicada* (pp. 81-94). Universidad de León.
- Molina Díaz, F. (2008). Variantes diacrónicas y diatópicas de los diminutivos en la toponimia. En C. Company Company y J. G. Moreno de Alba (coords.), *Actas del VII Congreso Internacional de Historia de la Lengua Española (Mérida, Yucatán, 2006)*, Vol. 2 (pp. 1429-1440). Arco Libros.
- Mondéjar, J. (2006). De la antigüedad y naturaleza de las hablas andaluzas [Discurso del nuevo doctor honoris causa D. José Mondéjar]. Universidad de Málaga. https://www.uma.es/publicadores/gabinete_r/wwwuma/disc_josemondejar_honoris.pdf

- Monge, F. (1965). Los diminutivos en español. En G. Straka, *Actes du X^e Congrès International du Linguistique et Philologie romanes*, (Estrasburgo, 1962), Vol. I (pp. 137-147). Klincksieck.
- Monge, F. (1988). Diminutivos: cuantificación, subjetividad, especialización. En J. Lüdtke (ed.), *Das sprachtheoretische Denk Eugenio Coserius in der Diskussion*, Vol. 2 (pp. 131-140). Gunter Narr Verlag.
- Montero Curiel, P. (2008). El proceso de lexicalización del diminutivo «ardilla». *Anuario de Estudios Filológicos*, 31, 117-131. <http://hdl.handle.net/10662/788>
- Montes Giraldo, J. J. (1972). Funciones del diminutivo en español: ensayo de clasificación. *Boletín del Instituto Caro y Cuervo*, 27(1), 71-88.
- Morala, J. R. (1986). Toponimia y diacronía: los sufijos diminutivos en León. *Lletres asturianas: Boletín Oficial de l' Academia de la Llingua Asturiana*, 19, 65-82.
- Morala, J. R. (2012a). Léxico e inventarios de bienes en los Siglos de oro. En G. Clavería Nadal, M. Freixas, M. Prat Sabater y J. Torruella (coords.): *Historia del léxico: perspectivas de investigación* (pp. 199-218). Iberoamericana.
- Morala, J. R. (2012b). El proyecto *CorLexIn*. En D. Corbella, J. Dorta, A. Fajardo Aguirre, L. Izquierdo, J. Medina López y A. Nelsi Torres (eds.): *Lexicografía hispánica del siglo XXI: nuevos proyectos y perspectivas. Homenaje al Profesor Cristóbal Corrales Zumbado* (pp. 421-439). Arco/Libros.
- Morala, J. R. (2014). El *CorLexIn*, un corpus para el estudio del léxico histórico y dialectal del Siglo de Oro. *Scriptum Digital*, 3, 5-28.
- Morala, J. R. (2015). Los inventarios de bienes y el léxico del siglo XVII en el AHP de Cádiz. En T. Bastardín y M.^a M. Barrientos (ed.), *Lengua y cultura en el Archivo Histórico Provincial de Cádiz* (pp. 147-174). Universidad de Cádiz.
- Morala, J. R. (2019). Fórmulas para la desambiguación léxica en documentación del Siglo de Oro. En M.^a C. Cazorla Vivas, M.^a Á. García Aranda y M.^a P. Nuño Álvarez (eds.), *Lo que hablan las palabras. Estudios de lexicología, lexicografía y gramática en honor de Manuel Alvar Ezquerro* (pp. 387-401). Axac.
- Moreno Fernández, F. (2015). Hablar madrileño. En A. M.^a Cestero Mancera, I. Molina Martos y F. Paredes García (coords), *Patrones sociolingüísticos de Madrid* (pp. 3-9). Peter Lang.
- Moreno Fernández, F. (2020). *Variedades de la lengua española*. Routledge.

- Muela Ezquerro, J. (1979). Estudio estilístico de algunos sufijos apreciativos en «Narciso», de Germán Sánchez Espeso: Premio Nadal 1978. *Cuadernos de investigación filológica*, 5, 59-78.
- Náñez Fernández, E. (1952). El diminutivo en «La Galatea». *Anales Cervantinos*, II, 269-285.
- Náñez Fernández, E. (1954). El diminutivo en Cervantes. *Anales Cervantinos*, IV, 3-79.
- Náñez Fernández, E. (1997). Amado Alonso y el diminutivo. *Cauce*, 20-21, 173-182.
- Náñez Fernández, E. (2006). *El diminutivo: historia y funciones en el Español Clásico y Moderno*. Universidad Autónoma de Madrid.
- Náñez Fernández, E. (2010). «Señoritos», no «señoricos», en el «Quijote». Pequeña crónica de una ¿errata, manipulación, simple descuido? *Anales Cervantinos*, XLII, 355-357.
- Narbona Jiménez, A., Cano Aguilar, R. y Morillo Velarde-Pérez, R. (2011). *El español hablado en Andalucía*. Universidad de Sevilla.
- NGLE = Real Academia Española, Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española*. Espasa.
- Nieuwenhuijsen, D. (2016). Notas sobre la aportación del análisis estadístico a la lingüística de corpus. En J. Kabatek (ed.), *Lingüística de corpus y lingüística histórica iberorrománica* (pp. 215-237). De Gruyter.
- NTLLE = Real Academia Española. *Nuevo Tesoro Lexicográfico de la lengua española*. Banco de datos. Recuperado de <http://ntlle.rae.es/ntlle/SrvltGUILoginNtlle>
- Ortiz Cruz, D. (2019). Aproximación morfosintáctica a los inventarios de bienes aragoneses del siglo XVIII. *Alazet*, 31, 113-142.
- Oxygen = Syncro Soft (2017). *Oxygen XML Editor* (versión 19.1) [software]. <https://www.oxygenxml.com/>
- Paquot, M. y Gries, S. (eds.) (2020). *A Practical Handbook of Corpus Linguistics*. Springer.
- Paredes García, F. (2011). Variación en el uso del diminutivo en el habla de Madrid: avance de un estudio sociolingüístico. En A. M.^a Cestero, I. Molina y F. Paredes (eds.), *La lengua, lugar de encuentro. Actas del XVI Congreso Internacional de la ALFAL, Alcalá de Henares, 6-9 de junio de 2011* (pp. 3709-3719). Universidad de Alcalá.

- Paredes García, F. (2015). Funciones subjetivadoras del diminutivo en el habla de Madrid. En A. M.^a Cestero Mancera, I. Molina Martos y F. Paredes García (eds.), *Patrones sociolingüísticos de Madrid* (pp. 84-110). Peter Lang.
- Paredes García, F. (2023). Evolución de las formas de diminutivo en español. La historia del cambio «-illo > -ito». En B. Almeida Cabrejas y P. Sánchez-Prieto Borja (coords.), *Varia lección de la lengua española. Estudios sobre el corpus CODEA* (pp. 97-135). Tirant lo Blanch.
- Parodi, G., Cantos-Gómez, P. y Howe, C. (eds.) (2022). *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics*. Routledge.
- Pato, E. y Felíu Arquiola, E. (2021). Morphological variation in the Spanish-speaking world. En A. Fábregas, V. Acedo-Matellán, G. Armstrong, M.^a C. Cuervo, I. Pujol Payet, *The Routledge handbook of Spanish morphology* (pp. 68-80). Routledge.
- Pedraza Gracia, M. J. (1999). Lector, lecturas, bibliotecas...: el inventario como fuente para su investigación histórica. *Anales de documentación*, 2, 137-158.
- Perdiguero, H. (2012). Variación léxica en protocolos notariales de Castilla en el siglo XVII. *Cuadernos del Instituto Historia de la Lengua*, 7, 333-345.
- Perdiguero, H. (2015). Léxico de bienes domésticos en documentos notariales del siglo XVII: sustantivos con sufijos diminutivos y variación diatópica. En J. M. García Martín (dir.), *Actas del IX Congreso Internacional de Historia de la Lengua Española*, Tomo II (pp. 1565-1578). Iberoamericana/Vervuert.
- Pérez Toral, M. (2017). A vueltas con el léxico textil inventariado en el Siglo de Oro. *Anadiss, número hors-série. In honorem doctor honoris causa Johannes Kabatek*, 89-108. <https://www.diacronia.ro/en/indexing/details/A26270>
- Pesqueira, D. (2006). Hacia una caracterización de las formas diminutivas en el español del siglo XVI. En J. L. Girón Alconchel y J. J. Bustos Tovar (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua española*, Vol. 1 (pp. 999-1010). Arco Libros.
- Pharies, D. (2002). *Diccionario etimológico de los sufijos españoles*. Gredos.
- Pharies, D. (2007). *Breve historia de la lengua española*. The University of Chicago Press.
- Pierazzo, E. (2016). *Digital scholarly editing: theories, models and methods*. Routledge.
- Puche Lorenzo, M. A. (2012). Léxico de la vida cotidiana en la Murcia áurea. *Cuadernos del Instituto de Historia de la Lengua*, 7, 343-360.

- Puche Lorenzo, M. A. (2018). Léxico dialectal / léxico especializado en los inventarios de bienes. En M.^a L. Arnal Purroy, R. M.^a Castañer Martín, J. M.^a Enguita Utrilla, V. Lagüéns Gracia y M.^a A. Martín Zorraquino (coords.) *Actas del X Congreso Internacional de Historia de la Lengua Española (Zaragoza, 2015)*, Vol. 2 (pp. 1655-1670). Institución Fernando el Católico.
- Puente Gutiérrez, R. (2000). Las marcas en Lexicografía. La marca de materia «Agricultura» en el Diccionario de la Lengua de la Real Academia. *Boletín de la Real Academia Española*, 80(280), 309-332.
- Quirós García, M. (2011). Léxico e inventarios de bienes en el Bilbao del Siglo de Oro. *Oihenart*, 26, 423-353.
- Ramírez Luengo, J. L. (2005). Variación gramatical y tipos textuales: el diminutivo en la época de «Don Quijote». *Revista de Estudios de Ciencias Sociales y Humanidades*, 13, 29-34.
- Ramírez Luengo, J. L. (2006). Una nota de sociolingüística histórica: el diminutivo en el español uruguayo del siglo XIX. *Res Diachronicae*, 5, 39-45.
- Ramírez Luengo, J. L. (2013): *Una descripción del español de mediados del siglo XVIII. Edición y estudio de las cartas de M. Martierena del Barranco (1757-1763)*. Axac.
- Red CHARTA (2013, abril). *Criterios de edición de documentos hispánicos (Orígenes - siglo XIX)* [2^a versión, abril de 2013; 1^a versión, febrero de 2011]. <https://www.redcharta.es/criterios-de-edicion/>
- Reynoso Noverón, J. (2001). *Los diminutivos en español. Un estudio de dialectología comparada* [Tesis doctoral]. Universidad Nacional Autónoma de México. Repositorio institucional: <https://repositorio.unam.mx/contenidos/75789>
- Reynoso Noverón, J. (2002). El diminutivo en el español actual: de la gramática a la pragmática. En M.^a T. Echenique Elizondo y J. Sánchez Méndez (eds.), *Actas del V Congreso Internacional de Historia de la Lengua Española (Valencia, 2000)* (pp. 935-943). Gredos.
- Reynoso Noverón, J. (2005). Procesos de gramaticalización por subjetivización: El uso del diminutivo en el español. En D. Eddington, *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 79-86). Cascadilla Proceedings Project.
- Ridruejo, E. (ed.). (2019). *Manual de lingüística española*. De Gruyter.
- Rodríguez Manzano, M. (2019). *Estudio filológico de testamentos e inventarios de bienes de difuntos mexicanos (siglos XVI-XVIII)* [Tesis doctoral]. Universidad de Sevilla. Repositorio institucional: <https://hdl.handle.net/11441/89415>

- Rodríguez Molina, J. y Octavio de Toledo y Huerta, Á. (2017). La imprescindible distinción entre texto y testimonio: el CORDE y los criterios de fiabilidad lingüística. *Scriptum digital*, 6, 5-68.
- Rojo, G. (2010). Sobre codificación y explotación de corpus textuales: otra comparación del «Corpus del español» con el CORDE y el CREA. *Lingüística*, 24, 11-50.
- Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge.
- RStudio (2021). *The R Foundation for Statistical Computing* (versión 4.1.0) [software]. <https://posit.co/download/rstudio-desktop/>
- Samper Hernández, M. (2019). Aproximación al uso del diminutivo en Las Palmas de Gran Canaria: datos sociolingüísticos a partir de muestras orales de actuación. En A. Cabedo Nebot y A. Hidalgo Navarro (eds.), *Pragmática del español hablado: hacia nuevos horizontes* (pp. 433-445). Universitat de València.
- Sánchez Lancis, C. (2009). Corpus diacrónicos y periodización del español. *Cahiers d'études hispaniques médiévales*, 32, 159-180.
- Sánchez Lancis, C. (2012). Periodización y cambio gramatical: el siglo XVIII, ¿frontera temporal del español? En M.^a T. García Godoy (ed.), *El español del siglo XVIII: cambios diacrónicos en el primer español moderno* (pp. 21-51). Peter Lang.
- Sánchez Lancis, C. (2022). Corpus diacrónicos del español de España. En G. Parodi, P. Cantos-Gómez y C. Howe (eds.), *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 33-44). Routledge.
- Sánchez Méndez, J. P. (2020). La periodización de la historia del español americano desde la evolución morfosintáctica. En E. Bravo-García y M. Fernández Alcaide (eds.), *El español de América: morfosintaxis histórica y variación* (pp. 151-184). Tirant Humanidades.
- Sánchez-Prieto Borja, P. (2019). Madrid en la configuración del español moderno. En M. Castillo Lluch y E. Díez del Corral Areta (eds.), *Reescribiendo la historia de la lengua española a partir de la edición de documentos* (pp. 85-112). Peter Lang.
- Sánchez-Prieto Borja, P., Torrens Álvarez, M.^a J. y Vázquez Balonga, D. (coords.) (2022). *La lengua de Madrid a lo largo del tiempo*. Universidad de Sevilla.
- Seco, M. (1970). *Arniches y el habla de Madrid*. Alfaguara.
- Serigos, J. (2022). Métodos de análisis estadísticos basados en corpus del español. En G. Parodi, P. Cantos-Gómez y C. Howe (eds.), *Lingüística de corpus en español*.

The Routledge Handbook of Spanish Corpus Linguistics (pp. 529-544). Routledge.

Silva Almanza, M. (2012). Diminutivos e ideología de género en El Habla de Monterrey corpus PRESEEA. En L. Rodríguez Alfano y M.^a E. Flores Treviño (eds.), *Estudios lingüístico-discursivos en distintas emisiones lingüísticas reales. Aplicaciones concretas al corpus de El habla de Monterrey* (pp. 145-165). Universidad Autónoma de Nuevo León.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Soler Espiauba, D. (1996). El español de los sentimientos. ‘Españolitos’, ‘curritos’, ‘amiguetes’, ‘coleguillas’ y demás ‘gentes de mal vivir’ (o la expresividad en el sistema español de sufijación. *Cuadernos Cervantes de la lengua española*, 8, 28-34.

Stala, E. (2014). Nombres de telas en el «Waaren-Lexicon in Zwölf Sprachen» de Ph.A. Nemnich (1797). *Revista de Investigación Lingüística*, 17, 191-220.

Stefanowitsch, A. (2020). *Corpus linguistics. A guide to the methodology*. Language Science Press.

Tableau Desktop (2023). Salesforce Company. (versión 2023.1) [software]. <https://www.tableau.com/products/desktop>

TEI = Text Encoding Initiative Consortium (ed.). (2018). *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (versión 3.4.0.) Recuperado de <http://www.tei-c.org/Guidelines/P5/>

TEITOK = Janssen, M. (2014 - Act.). *TEITOK: a Tokenized TEI environment* [software]. <http://teitok.corpuswiki.org/>

Terreros y Pando, E. de (1786). *Diccionario castellano con las voces de ciencias y artes y sus correspondientes en las tres lenguas francesa, latina é italiana*. Imprenta de la viuda de Ibarra, Hijos y Compañía. <https://apps.rae.es/ntlle/SrvltGUISalirNtlle>

Tirapu León, I. (2014). *Semántica de la sufijación valorativa en español oral contemporáneo* [Tesis doctoral]. Universidad de Salamanca. Repositorio institucional: <http://hdl.handle.net/10366/125966>

TLHA = Alvar Ezquerra, M. (2000). *Tesoro léxico de las hablas andaluzas*. Arco Libros.

Torres Martínez, M. (2019). Apuntes sobre el léxico de un inventario decimonónico de la Real Maestranza de Caballería de Ronda. *Nueva revista de filología hispánica*, 67(2), 447-471.

- Torres Montes, F. (1990). Los sufijos diminutivos del habla rural malagueña en el XVIII. *Analecta malacitana*, vol. 13(1), 69-76.
- Torruella, J. (2017). *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*. Peter Lang.
- Trapero, M. (1995). *Para una teoría lingüística de la toponimia. (Estudios de toponimia canaria)*. Universidad de Las Palmas de Gran Canaria.
- Uritani, N. y Berrueta de Uritani, A. (1985). Los diminutivos en los atlas lingüísticos españoles *Lingüística Española Actual*, Vol. 7(2), 203-236.
- Vaamonde, G. (2015a). P. S. Post Scriptum: dos corpus diacrónicos de escritura cotidiana. *Procesamiento del Lenguaje Natural*, 55, 57-64.
- Vaamonde, G. (2015b). Limitaciones en el uso de corpus diacrónicos del español. Nuevas aportaciones desde el proyecto de investigación Post Scriptum. *E-Aesla*, 1.
- Vaamonde, G. (2016). *Guía para la edición digital de textos en P.S. Post Scriptum*. Centro de Lingüística de Universidade de Lisboa. Recuperado de http://ps.clul.ul.pt/files/Manual_PS.pdf
- Vaamonde, G. y Magro, C. (2016). *Manual de Edición en P.S. Post Scriptum. Edición modernizada, Anotación POS, Anotación Sintáctica*. Centro de Lingüística da Universidade de Lisboa.
- Valeri, M.^a J. (2010). Los sufijos apreciativos -ito, -ico y -on en cartas escritas en Mérida colonial. *Lingua Americana*, 26, 15-33.
- Varela Merino, E. (2009). *Los galicismos en el español de los siglos XVI y XVII*. Vol. I. CSIC, Consejo Superior de Investigaciones Científicas.
- Varela Ortega, S. (1996). *Fundamentos de morfología*. Síntesis.
- Vázquez Balonga, D. (2019). Nuevos datos para la historia del léxico: aportaciones de los inventarios de bienes en la ciudad de Madrid en el siglo XVIII. En M. Castillo Lluch y E. Díez del Corral Areta (eds.), *Reescribiendo la historia de la lengua española a partir de la edición de documentos* (pp. 319-334). Peter Lang.
- Viña Brito, A. y Corbella, D. (2018). «Corpus documental de las islas Canarias»: un nuevo reto en Humanidades digitales. *Anuario de Estudios Atlánticos*, 65, 1-17.
- Vivancos Mulero, M.^a E. (2012). El sufijo -ico/-iquio como caracterizador dialectal del español murciano (siglo XVIII). En M.^a T. García-Godoy (ed.), *El español del siglo XVIII: cambios diacrónicos en el primer español moderno* (pp. 313-332). Peter Lang.

Zacarías Ponce de León, R. (2008). Morfemas apreciativos del español: entre la flexión y la derivación. *Núcleo*, 25, 221–37.

Zuluaga Ospina, A. (1970). La función del diminutivo en español. *Boletín del Instituto Caro y Cuervo*, 25(1), 23-48.

Corpus

Biblia Medieval = Enrique-Arias, A. (dir.). *Biblia Medieval*. Recuperado de <https://www.bibliamedieval.es/BM/>

CdE = Davies, M. *El corpus del español*. Recuperado de <https://www.corpusdelespanol.org/>

CDH = Instituto de Investigación Rafael Lapesa de la Real Academia Española (2013). *Corpus del Diccionario histórico de la lengua española (CDH)*. Recuperado de <http://web.frl.es/CNDHE>

CHARTA = *Corpus Hispánico y Americano en la Red: Textos Antiguos*. Recuperado de <http://www.corpuscharta.es/>

CODEA+ 2015 = Grupo de Investigación Textos para la Historia del Español [GITHE]: *CODEA+ 2015 (Corpus de documentos españoles anteriores a 1800)*. Recuperado de <https://doi.org/10.37536/CODEA.2015>

CODEA+ 2022 = Grupo de Investigación Textos para la Historia del Español [GITHE]: *CODEA+ 2022 (Corpus de documentos españoles anteriores a 1900)*. Recuperado de <https://www.corpuscodea.es/>

CODHECUN = Bravo-García, E., Mancera Rueda, A. y Martín Aizpuru, L. (dirs.). *Corpus Documental y Hemerográfico de la Cuba del Novecientos (CODHECUN)*. Recuperado de <http://cuba19.us.es/>

CORDE = Real Academia Española. *Corpus diacrónico del español*. Banco de datos. Recuperado de <http://corpus.rae.es/cordenet.html>

CORDIAM = Academia Mexicana de la Lengua. *Corpus diacrónico y diatópico del español de América (CORDIAM)*. Recuperado de <https://cordiam.org/>

CORDICan = Corbella, D. y Viña Brito, A. (2022-). *Corpus Documental de las Islas Canarias (CORDICan)*. Recuperado de <https://www.ull.es/corpora/cordican/>

CorLexIn = Morala Rodríguez, J. R. *Corpus léxico de inventarios (CorLexIn)*. Recuperado de <http://web.frl.es/CORLEXIN.html>

- CORPES XXI* = Real Academia Española. *Corpus del Español del Siglo XXI (CORPES)*. Banco de datos. Recuperado de <https://www.rae.es/corpes/>
- COSUIZA* = *Corpus de documentos hispánicos de Suiza*. Recuperado de <https://grafila.unil.ch/cosuiza/>
- CREA* = Real Academia Española. *Corpus de referencia del español actual*. Banco de datos. Recuperado de <https://www.rae.es/banco-de-datos/crea>
- EsTenTen18* = *Spanish Web corpus 2018 (esTenTen18)*. Recuperado de <https://www.sketchengine.eu/estenten-spanish-corpus/>
- Exploring Medieval Mary Magdalene*. Text editions. Recuperado de <https://digital-editing.fas.harvard.edu/>
- Léxico hispanoamericano de Peter Boyd-Bowman 1493-1993* = Harris-Northall, H. y Nitti, J. J. (eds.). Hispanic Seminary of Medieval Studies. Recuperado de <https://textred.spanport.lss.wisc.edu/>
- ODE* = Calderón Campos, M. y García Godoy, M. T. (dirs.) *Oralia diacrónica del español (ODE)*. Recuperado de <http://corpora.ugr.es/ode/>
- OSTA* = Gago Jover, F. y Pueyo Mena, J. *Old Spanish Textual Archive*. Recuperado de <http://osta.oldspanishtextualarchive.org/>
- Post Scriptum* = CLUL (ed.) (2014). *P.S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna*. Recuperado de <http://ps.clul.ul.pt>.
- Sketch Engine* = Kilgarrieff, A., Baisa, V., Bušta, J.; Jakubíček, M.; Kovář, V.; Michelfeit, J.; Rychlý, P.; Suchomel, V. (2014). *The Sketch Engine: ten years on. Lexicography*, 1: 7-36. Recuperado de <http://www.sketchengine.eu>

ÍNDICE DE TABLAS Y FIGURAS

Tablas

Tabla 1: Número de palabras en el corpus	46
Tabla 2: Relación de documentos de Almería con título, signatura, fecha y municipio	53
Tabla 3: Relación de documentos del corpus de Madrid	58
Tabla 4: Relación de caracteres reservados en XML	67
Tabla 5: Etiquetado morfosintáctico de adjetivos y nombres en el corpus ODE	111
Tabla 6: Búsqueda CQL por forma original (<i>form</i>)	143
Tabla 7: Búsqueda CQL por forma expandida (<i>fform</i>)	143
Tabla 8: Búsqueda CQL por forma normalizada (<i>nform</i>)	144
Tabla 9: Búsqueda CQL por etiqueta gramatical (<i>pos</i>)	144
Tabla 10: Búsqueda CQL por lema (<i>lemma</i>)	144
Tabla 11: Principales expresiones regulares empleadas en búsquedas CQL	151
Tabla 12: Recuento de formas diminutivas en el corpus documental	242
Tabla 13: Bases en el corpus a las que se añade un sufijo diminutivo	243
Tabla 14: Palabras con variación en el uso del diminutivo	244
Tabla 15: Distribución de diminutivos según la categoría gramatical	244
Tabla 16: Recuento de diminutivos transparentes y lexicalizados en el corpus	249
Tabla 17: Frecuencia de bufete y sus derivados con diminutivo en el siglo XVII (<i>CorLexIn</i>)	278
Tabla 18: Frecuencia de bufete y sus derivados con diminutivo en el siglo XVIII (<i>ODE</i>)	278
Tabla 19: Abreviaturas y signos empleados en el glosario	289
Tabla 20: Claves bibliográficas empleadas en las definiciones del glosario	290
Tabla 21: División de la muestra documental en subcorpus y número de palabras	315
Tabla 22: Tabla de contingencia. Frecuencias observadas de los diminutivos en Almería	318
Tabla 23: Escala interpretativa de la V de Cramer (Kotrlík y Williams, 2003)	320
Tabla 24: Tabla de contingencia. Distribución de los diminutivos por siglos (XVIII-XIX) en la provincia de Almería. Frecuencias observadas y porcentajes	325
Tabla 25: Frecuencias esperadas de los diminutivos en Almería a partir de χ^2	327
Tabla 26: Residuos de Pearson para los diminutivos en el corpus de Almería	327
Tabla 27: Frecuencias relativas de los diminutivos en Almería (ss. XVIII-XIX)	330
Tabla 28: Porcentaje de incremento/disminución de los diminutivos en el s. XIX con respecto al s. XVIII en Almería	332
Tabla 29: Tabla de contingencia. Distribución de los diminutivos en los siglos XVIII-XIX en la provincia de Madrid. Frecuencias observadas y porcentajes	334
Tabla 30: Frecuencias esperadas de los diminutivos en Madrid	337
Tabla 31: Residuos de Pearson para los diminutivos en el corpus de Madrid	337
Tabla 32: Frecuencias relativas de los diminutivos en Madrid (ss. XVIII-XIX)	338
Tabla 33: Porcentaje de incremento/disminución de los diminutivos en el s. XIX con respecto al s. XVIII en Madrid	339
Tabla 34: Tabla de contingencia. Contraste diminutivo/provincia en el s. XVIII	340
Tabla 35: Frecuencias esperadas de los diminutivos en el s. XVIII	341
Tabla 36: Residuos de Pearson para los diminutivos en el s. XVIII	341
Tabla 37: Tabla de contingencia. Contraste diminutivo/provincia en el s. XIX	344
Tabla 38: Frecuencias esperadas de los diminutivos en el s. XIX	345
Tabla 39: Residuos de Pearson para los diminutivos en el s. XIX	345

Figuras

Figura 1: Delimitación geográfica y tamaño del corpus documental.....	46
Figura 2: Acceso a los documentos a partir de la búsqueda en <i>ODE</i>	48
Figura 3: Mapa de municipios almerienses incluidos en el corpus.....	49
Figura 4: Mapa de municipios de Madrid representados en el corpus.....	54
Figura 5: Sintaxis básica de XML.....	65
Figura 6: Visualización en <i>ODE</i> del texto marcado con el elemento <code><add> </add></code>	66
Figura 7: Aviso de error en el programa <i>Oxygen</i> . Elemento <code><add> </add></code> mal cerrado.....	69
Figura 8: Texto superfluo por repetición.....	83
Figura 9: Elemento <i>sic</i> por falta de concordancia gramatical.....	84
Figura 10: Elemento <i>sic</i> por pérdida de <i>-d-</i> intervocálica.....	84
Figura 11: Elemento <i>sic</i> por omisión del escribano.....	84
Figura 12: Fragmento de texto con reproducción literal.....	85
Figura 13: Texto subrayado.....	85
Figura 14: Texto marcado en negrita.....	85
Figura 15: Texto omitido en la transcripción por deterioro del manuscrito.....	86
Figura 16: Texto tachado por el escribano imposible de leer.....	87
Figura 17: Texto omitido por el escribano.....	87
Figura 18: Texto omitido por razones editoriales.....	87
Figura 19: Adición en el margen izquierdo.....	88
Figura 20: Adición por encima de la línea.....	88
Figura 21: Adición por debajo de la línea.....	88
Figura 22: Texto tachado por el escribano.....	89
Figura 23: Texto tachado.....	89
Figura 24: Restitución con sobrescrito.....	90
Figura 25: Restitución con tachado y adición por encima de la línea.....	90
Figura 26: Restitución con raspado y sobrescrito.....	90
Figura 27: Cambio de mano.....	91
Figura 28: Firma y rúbrica.....	91
Figura 29: Conjetura segura por daños físicos del soporte.....	92
Figura 30: Conjetura segura por la encuadernación del documento.....	92
Figura 31: Conjetura no segura por grafía dudosa.....	92
Figura 32: Conjetura no segura de palabra incomprensible.....	92
Figura 33: Transcripción en XML-TEI de un inventario del corpus.....	96
Figura 34: Funcionamiento de TEITOK mediante el sistema de herencias.....	101
Figura 35: Resultado del archivo XML tras la tokenización y procesamiento lingüístico completo (normalización, anotación morfosintáctica y lematización).....	102
Figura 36: Resultado de las búsquedas normalizada (izquierda) y lematizada (derecha) de la forma <i>tinaja</i> . Concordancias ordenadas en base a la forma original transcrita.....	117
Figura 37: Resultado de la consulta en el <i>CDH</i> del lema <i>tijera</i> . Ejemplos del s. XVI.....	134
Figura 38: Visualización en TEITOK de un documento XML-TEI con su cabecera, texto e imagen del facsímil (edición semipaleográfica).....	140
Figura 39: Visualización en TEITOK de la edición normalizada e imagen del facsímil.....	141
Figura 40: Visualización en TEITOK de la edición normalizada en la que se muestra la etiqueta morfosintáctica y lema de cada palabra.....	141
Figura 41: Resultado de la búsqueda <i>cama de</i> en el corpus <i>ODE</i>	145
Figura 42: Búsqueda combinada. Resultado de la forma <i>fueron</i> del verbo ser.....	150

Figura 43: Resultado de búsqueda <i>case-insensitive</i> para la voz <i>cocina</i>	152
Figura 44: Resultado de la búsqueda <i>ser</i> + participio	154
Figura 45: Resultado de la búsqueda de la secuencia <i>grande</i> + sustantivo.....	154
Figura 46: Resultado de las alternativas <i>latón</i> / <i>azófar</i> en una sola consulta	155
Figura 47: Búsqueda combinada de variantes morfológicas del diminutivo	156
Figura 48: Resultado de la búsqueda en la que se excluyen los casos de <i>haber</i> auxiliar	158
Figura 49: Resultado de una búsqueda con elementos opcionales en la estructura <i>bufete</i> + cualquier cosa + <i>de</i> + sustantivo	160
Figura 50: Búsqueda para rescatar todos los tokens en el corpus	162
Figura 51: Interfaz de búsqueda simple en el corpus <i>ODE</i>	163
Figura 52: Búsqueda simple de la estructura “vecino de esta villa”	164
Figura 53: Interfaz de búsqueda avanzada del corpus <i>ODE</i>	165
Figura 54: Constructor de etiquetas en la interfaz de búsqueda.....	166
Figura 55: Ejemplos de laísmo como resultado de búsqueda por etiqueta gramatical.....	166
Figura 56: Documentos correspondientes a dotes. Búsqueda avanzada por ‘título’	167
Figura 57: Búsqueda por rangos cronológicos, anteriores o posteriores a una fecha concreta .	168
Figura 58: Búsqueda de documentos pertenecientes al municipio de Madrid	168
Figura 59: Ejemplo de búsqueda comparada en el corpus <i>ODE</i>	170
Figura 60: Distribución de las variantes <i>zafa</i> , <i>jofaina</i> y <i>palangana</i> . Búsqueda comparada.....	170
Figura 61: Búsqueda en el mapa. Distribución diatópica de las formas <i>delantal</i> y <i>devantal</i>	171
Figura 62: Búsqueda genérica. Documentos del Archivo Histórico de Protocolos de Madrid.	172
Figura 63: Búsqueda con lenguaje <i>XPath</i> . Casos de restituciones marcadas con <subst>	173
Figura 64: Distribución de palabras en el corpus por siglo y tipo textual	173
Figura 65: Nube de palabras en <i>ODE</i> con los 500 sustantivos más frecuentes en el corpus	174
Figura 66: Trigramas más frecuentes de la combinación <i>sustantivo</i> + <i>de</i> + <i>sustantivo</i>	175
Figura 67: Selección del token en el texto	177
Figura 68: Plantilla de edición de tokens en TEITOK	177
Figura 69: Función <i>lookup</i> para la palabra <i>dicho</i>	178
Figura 70: Búsquedas en el corpus de la forma <i>cobre</i> para la corrección de errores	179
Figura 71: Edición múltiple de tokens mediante selección de casos.....	180
Figura 72: Comprobación de errores en las etiquetas morfosintácticas	182
Figura 73: Corrección del lema mediante búsqueda lematizada. Resultado para *[lemma = "viur"].....	183
Figura 74: Corrección de lemas mal asociados a la forma <i>ir</i> a partir de una búsqueda	184
Figura 75: Extracto del resultado de la búsqueda de formas acabadas en <i>-er/-ir</i> ordenada según la frecuencia por lema	185
Figura 76: Resultados de la búsqueda ‘bufete/bufetillo de estrado/de luces’ en <i>ODE</i>	276
Figura 77: Dispersión de los diminutivos en el corpus de Almería (s. XVIII).....	323
Figura 78: Dispersión de los diminutivos en el corpus de Almería (s. XIX)	324
Figura 79: Distribución proporcional de los diminutivos en Almería en los ss. XVIII y XIX .	326
Figura 80: Asociación de las variables siglo/diminutivo en Almería basada en los residuos de Pearson	327
Figura 81: Distribución diacrónica de los diminutivos en Almería en los siglos XVIII-XIX...	331
Figura 82: Dispersión de los diminutivos en el corpus de Madrid (s. XVIII)	333
Figura 83: Dispersión de los diminutivos en el corpus de Madrid (s. XIX)	333
Figura 84: Distribución proporcional de los diminutivos en Madrid en los ss. XVIII y XIX...	335
Figura 85: Gráfico de asociación (siglo/diminutivo en Madrid) basado en los residuos de Pearson	337

Figura 86: Distribución diacrónica de los diminutivos en Madrid en los siglos XVIII-XIX....	338
Figura 87: Proporción de los diminutivos en Almería y Madrid en el s. XVIII.....	341
Figura 88: Gráfico de asociación (provincia/diminutivo en el s. XVIII) basado en los residuos de Pearson	342
Figura 89: Proporción de uso de los diminutivos en Almería y Madrid en el s. XIX.....	344
Figura 90: Gráfico de asociación (provincia/diminutivo en el s. XIX) basado en los residuos de Pearson	345
Figura 91: Localización geográfica de los diminutivos en los corpus de Madrid y Almería en los siglos XVIII-XIX	347
Figura 921: Frecuencias de aparición de los diminutivos por provincia en el siglo XVIII.....	348
Figura 93: Frecuencia de aparición de los diminutivos por provincia en el siglo XIX.....	348

ANEXOS

Anexo I. Ejemplo completo de cabecera etiquetada

Documento MD1711I0105: «Inventario de los bienes que quedaron tras fin y muerte de María Sánchez de Lázaro» (ODE – Madrid, 1711)¹¹⁸.

```
<TEI xmlns:tei="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Inventario de los bienes que quedaron tras fin y muerte de María Sánchez
de Lázaro</title>
        <editor id="CORDEREGRA">CORDEREGRA: Corpus diacrónico del español
del reino de Granada. 1492-1833</editor>
        <funder>Junta de Andalucía/FEDER: P09-HUM 4466, 2009-2014</funder>
        <respStmt>
          <resp id="transcription">Pilar Arrabal Rodríguez</resp>
        </respStmt>
        <respStmt>
          <resp id="revision">Pilar Arrabal Rodríguez</resp>
        </respStmt>
        <respStmt>
          <resp id="standardization">Pilar Arrabal Rodríguez</resp>
        </respStmt>
        <respStmt>
          <resp id="annotation">Pilar Arrabal Rodríguez</resp>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <publisher>UGR, Universidad de Granada</publisher>
        <pubPlace>Granada</pubPlace>
        <distributor>HUM-278. Grupo de Investigaciones Histórico-Lingüísticas y
Dialectales. UGR-Junta de Andalucía</distributor>
      </publicationStmt>
      <sourceDesc>
        <msDesc>
          <msIdentifier>
            <country>España</country>
            <settlement>Madrid</settlement>
            <institution>Archivo Histórico de Protocolos de Madrid</institution>
            <repository>Protocolos notariales del escribano Gabriel de
Nevares</repository>
            <idno>AHPM T.14527</idno>
          </msIdentifier>
          <msContents>
            <summary>Se recogen los bienes dejados por D.a María Sánchez de Lázaro
tras su muerte clasificados según el material de los objetos. Termina la lista con algunos
```

¹¹⁸ Enlace directo a la visualización del documento en el corpus:
<http://corpora.ugr.es/ode/index.php?action=file&id=MD1711I0105>

de los bienes empeñados. </summary>
 <msItem class="original">
 <p><locus>323r-323v, 324v-325v, 326v-327r, 329r-329v,
 330v</locus></p>
 </msItem>
 </msContents>
 <additional>
 <adminInfo>
 <availability>
 <p>Para hacer uso de las imágenes es necesaria la autorización del
 AHPM. La publicación de la imagen cumple con los permisos de publicación
 concedidos por el archivo</p>
 </availability>
 </adminInfo>
 <surrogates>
 <cit><bibl>facsimil digital guardado en formato JPEG</bibl></cit>
 </surrogates>
 </additional>
 </msDesc>
 </sourceDesc>
 </fileDesc>
 <encodingDesc>
 <classDecl>
 <taxonomy>
 <category id="inv">
 <catDesc>inventarios de bienes</catDesc>
 </category>
 <category id="cer">
 <catDesc>certificados médicos</catDesc>
 </category>
 <category id="dec">
 <catDesc>declaraciones de testigos</catDesc>
 </category>
 <category id="oth">
 <catDesc>otros</catDesc>
 </category>
 </taxonomy>
 </classDecl>
 </encodingDesc>
 <profileDesc>
 <settingDesc>
 <setting>
 <name type="place" subtype="Madrid" geo="40.4167047 -
 3.7035825">España, Madrid, Madrid</name>
 <date when="1711" when-custom="XVIII">1711</date>
 </setting>
 </settingDesc>
 <textClass>
 <catRef target="inv"/>
 </textClass>

```
<creation><name>NA</name></creation></profileDesc>
<revisionDesc><change who="xmltokenize" when="2019-08-05">tokenized using
xmltokenize.pl</change></revisionDesc>
</teiHeader>
```

Anexo II. Ejemplo completo de texto etiquetado

Texto perteneciente a la cabecera del § Anexo I.

Documento MD1711I0105: «Inventario de los bienes que quedaron tras fin y muerte de María Sánchez de Lázaro» (ODE – Madrid, 1711)¹¹⁹.

```
<text>
  <body>
    <pb n="323r" facs="DSC_0275a.JPG" id="e-1"/>
    <p id="p-1"><lb/> Dos pailas de cobre grandes con asas
    </p>
    <p id="p-2">
      <lb/> <add place="left">Cobre</add> y la una de ellas con dos ganchos en
    </p> <add place="left">y yerro</add> ella para azer zerilla.
      <lb/> Dos ollas de cobre con asas y picos, <lb/> una maior q otra.
      <lb/> Cuatro cucharas, las tres de cobre y la otra <lb/> de yerro.
      <lb/> Un perol pequeño de cobre con asas.
      <lb/> Un perolillo de cobre para colar.
      <lb/> Tres peroles de cobre con asas.
      <lb/> Una olla grande de cobre con asas y pico.
      <lb/> Un barquillo pa colar zera.
      <lb/> Una cuchara mediana de cobre.
      <lb/> Un yerro de azer zerilla.
      <lb/> Un torno de madera de alamo y ne<lb/>gro con sus manixas de yerro para
    </p> <lb/> blanquear zera.
    </p>
    <p id="p-3">
      <lb/> <add place="left">Madera</add>
      <lb/> Un tablero de pino de dos tablas con <lb/> sus pies para bruñir la zera.
      <lb/> Dos brunideras de madera de nogal.
      <lb/> Un tablero mediano de nogal.
      <lb/> De una tabla pa bruñir enzerado.
      <lb/> Tres yerros de bruñir enzerado.
      <lb/> Dos tornos con sus pies de pino pa azer zerilla.
    </p> <pb n="323v" facs="DSC_0276a.JPG" id="e-27"/>
      <lb/> Un tornillo para azer zerilla con dos <lb/> zigüañas de yerro.
      <lb/> Una fuente de pino para azer zerilla.
      <lb/> Un arillo de yerro con quarenta y ocho <lb/> escarpías para trabaxar con
    su peso en <lb/> medio.
      <lb/> Un bañador de pino con sus tramos.
      <lb/> Una antipara de yerro de dos pedazos <lb/> de chapa.
      <lb/> Dos cuchillos de madera para zezernar las <lb/> belas.
      <lb/> Un sello de bronze para zellar la zera.
      <lb/> Una canal con su acanalor de yerro.
      <lb/> Cinco clabos de yerro de remechar.
      <lb/> Unas tenasas de yerro.
```

¹¹⁹ Enlace directo a la visualización del documento en el corpus:
<http://corpora.ugr.es/ode/index.php?action=file&id=MD1711I0105>

<lb/> Unas trebedes de yerro grandes.

<lb/> Un oreo para la zera que se compone de seis <lb/> perchas, con pies de pino y su clabazon.

</p>

<p id="p-4">

<gap reason="editorial" id="e-46"/>

</p>

<pb n="324v" facs="DSC_0278a.JPG" id="e-47"/>

<p id="p-5">

<lb/> <gap reason="editorial" id="e-49"/>

<lb/> En llantas, piquetas, bolanderas, buges, clabos <lb/> de carro y aros para ellos y grapas de yerro, <lb/> se pesaron veinte y quatro arrobas de dicho <lb/> yerro.

<lb/> En barillas de yerro de todos tamaños se pesa<lb/>ron ocho arrobas y treçe libras.

<lb/> Una rexa de yerro de arar que peso nue<lb/>be libras.

<lb/> En barretas de azero de Mondragon se pesa<lb/>ron cuatro arrobas y ocho libras.

<lb/> En azero de Milan, en barras grandes y chicas, <lb/> se pesaron veinte y una arrobas y seis libras.

<lb/> En clabos gemales, birotos y escarpias de peso se <lb/> pesaron ciento y catorçe libras.

<lb/> Veinte y nueve cascós para sartenes del nume<lb/>ro diez y ocho.

<lb/> Nobenta y nueve cascós para sartenes de yerro <lb/> del número doçe.

<lb/> Cinco cascós de yerro grandes del número tres.

<lb/> Veinte y cinco cascós de sartenes de a seis.

<lb/> Tres almuasas ordinarias.

<lb/> Quatro almuasas finas.

<lb/> Tres brusas de zerda finas y otra ordinaria.

<lb/> Cinco braserillos de yerro con sus pies.

<lb/> Onze batideras de yerro.

<pb n="325r" facs="DSC_0279a.JPG" id="e-75"/>

<lb/> De çucharas de yerro con sus cabos, çien do<lb/>çenas todas yguales.

<lb/> Seis tenasas de yerro para lumbre.

<lb/> Doçe garabactos con quatro garfios cada uno.

<lb/> Veinte y dos zençerros medianos de yerro.

<lb/> Otro senzerro grande de yerro.

<lb/> Treinta y nueve <sic>senzerrillas</sic> de yerro.

<lb/> Dos paletillas de yerro para guebos.

<lb/> Diez arrendaderas de yerro grandes.

<lb/> Mas catorze arrendaderas de yerro grandes.

<lb/> Veinte y un arrendaderas de yerro medianas.

<lb/> Diez y seis arrendaderas chicas de yerro.

<lb/> Diez y nueve argollas para lamparas.

<lb/> Setenta y quatro colgaderos grandes de yerro.

<lb/> Quatro colgaderos grandes de yerro.

<lb/> Veinte y quatro armellas de yerro grandes.

<lb/> Mas treinta y quatro armellas medianas.

<lb/> Seis lebantaderos para cortinas de yerro.

<lb/> Cinco doçenas y media de sortixas de yerro.

<lb/> Quatro frenos bocados de yerro.

<lb/> Dos carlancas de yerro para perros.
 <lb/> Doçe doçenas de sortixas de yerro.
 <lb/> Catorze clauijas para coches.
 <lb/> Treinta y quatro clabijas mas pequeñas de yerro.
 <lb/> Onze orejas para cubos de yerro.
 <lb/> Una piqueta de yerro con su pico.
 <pb n="325v" facs="DSC_0280a.JPG" id="e-102"/>
 <lb/> Cinquenta y ocho bolanderas chicas de yerro.
 <lb/> Cinquenta pares de zibicas de yerro.
 <lb/> Sesenta y una pesoneras de yerro pa coches.
 <lb/> Quatro sortixones de yerro pa calezas.
 <lb/> Diez gosnes grandes de yerro.
 <lb/> Nobenta y siete gosnes medianos de yerro.
 <lb/> Sesenta y quatro gosnes, los quarenta de a ochabo <lb/> y los veinte y quatro de cuarto.
 <lb/> Treinta y quatro pernios con sus embras y machos.
 <lb/> Tres aldabones de yerro para atae.
 <lb/> Ducientos y sesenta y seis clabos de fixas chicos.
 <lb/> Doçe doçenas y ma de sortixas de bronze.
 <lb/> Ciento y veinte y dos clabos de fixas.
 <lb/> Cien abuxuelas de yerro.
 <lb/> Ducientos y setenta medias abujuelas.
 <lb/> Saetines, ochocientos.
 <lb/> Quinientos y sesenta clabos, tabaques de yerro.
 <lb/> Trecientos y quarenta clabos de a dos.
 <lb/> Ducientos y setenta clamones de yerro.
 <lb/> Diez armellas chicas de yerro.
 <lb/> Dos pares de estribos de yerro.
 <lb/> Diez cascós medianos de yerro pa sartenes.
 <lb/> Veinte y nueve libras de erraduras nuevas.
 <lb/> Diez cadenas de rostal de yerro.
 <lb/> Catorze cadenas para mulas de yerro.
 <lb/> Siete trabones de yerro.
 <lb/> Doçe cadenas pa poços de yerro.
 <lb/> Dos garruchas grandes pa obra.
 <lb/> Cinqta eses de yerro pa cabezadas.
 <lb/> Treinta y çinco millares y mo de clabos de <lb/> media chilla de ala de mosca viscainos.
 <pb n="326v" facs="DSC_0282a.JPG" id="e-134"/>
 <lb/> Catorze doçenas de tragaçetes de pino.
 <lb/> Seis doçenas y ma de baras largas pa lexones.
 <lb/> Tres cajas moldes pa morteretes de zera de cobre.
 <lb/> Una paletilla de yerro pa raer zera de la paila.
 <lb/> Un mostrador con tablero de nogal mui antiguo <lb/> y gastado con cinco cajones de pino viejo.
 <lb/> La armadura de la tienda donde estantes las co<lb/>sas de yerro que quedan <sic>ybentariadas</sic>.
 <lb/> Tres cuñas de yerro p<ex>ar</ex>a partir leña.
 <lb/> Una acha pa partir leña.
 </p>

<p id="p-6">
 <lb/> <add place="left">Madera</add> Cinco sillas poltronas de baqueta con clabazon <lb/> dorada.
 <lb/> Una papelera de nogal con onze gabectas con <lb/> su pie de pino.
 <lb/> Una papelera, forma de escritorio, de nogal con <lb/> tapa y dentro doce gabectas y su puertezuela.
 <lb/> Un bufete de nogal de bara de largo con <lb/> su cajón.
 <lb/> Un escaparate de caoba con cinco bidrios <lb/> en la fachada ordinarios y en el un costado tres, <lb/> con su pie cubierto de palo santo.
 <lb/> Un tocador cubierto de palo santo con su es<lb/>pejo de tercia en cuadro.
 <lb/> Un bufetico de luzes cubierto de palo so.
 <lb/> Una arquita de zipres de quarta de alto y <lb/> media bara de largo.
 <lb/> Una arca de nogal de una bara de largo.
 <pb n="327r" facs="DSC_0283a.JPG" id="e-162"/>
 <lb/> Un cofre de bara y media de largo cubierto <lb/> de enzerado viejo.
 <lb/> Otro cofre chico, maleta, cubierto de enzerado.
 <lb/> Otro pa malecta de tres quartas de largo de enze<lb/>rado viejo.
 <lb/> Otro cofre viejo sin serradura.
 <lb/> Una arca de pino de bara y quarta de largo.
 <lb/> Una mesa de pino cubierta de badana.
 <lb/> Cinco cajones de pino sin zerraduras.
 <lb/> Una cruz enbutida con santos y su peana de <lb/> pino de media bara de alto.
 <lb/> Una cruz de a terzia de alto y en ella un santísi<lb/>mo Christo de marfil.
 <lb/> Tres cofres viejos de <sic>difrentes</sic> tamaños.
 <lb/> Una arca vieja de dos baras de largo.
 <lb/> Una papelera con su red de pino de tres cuartas de <lb/> alto y la red de yerro.
 <lb/> Una cama de pino de seis tablas con sus pies.
 <lb/> Una arquita de pino de una tercia.
 <lb/> Un bufete de poco menos de bara de largo y dos <lb/> tercias de ancho y la cara d el de <sic>difreentes</sic> pie<lb/>dras, como son: agacta, profido, lapizlafide; con <lb/> adornos de plata y sus molduras de media caña <lb/> doradas de molido de bronze y los pies d el <lb/> de palo santo acartelados, cubiertos y enbutidos <lb/> por la parte de afuera y las caras de dha pedreria <lb/> con sus yerros lisos y labrados.
 </p>
 <pb n="329r" facs="DSC_0287a.JPG" id="e-190"/>
 <gap reason="editorial" id="e-191"/>
 <p id="p-7">
 <add place="left"><hi rend="underlined">Yerro</hi></add>
 <lb/> Primeramte, se pesaron cinco arrobas y ma de <lb/> diferentes barillas de yerro de diferentes tamaños.
 <lb/> Seis cortinas de gasa mui biejas y tres pedazitos </p>
 <p id="p-8">
 <lb/><add place="left">Ropa</add> de lo mismo.
 <lb/> Una colcha de Yndias labrada con ylo de pita <lb/> con su fleco <sic>alrredor</sic> de seda.
 <lb/> Un candado grande de yerro con su llabe doble.
 <lb/> Nuebe baras y ma de raso musco antiguo con <lb/> flores blancas.

<lb/> Tres baras en dos pedazos de picocete de media se<lb/>da, musco.
 <lb/> En tres pedazos, onze baras de damasco carne<lb/>sid usado.
 <lb/> Un forro de olandilla encarnado para dosel que <lb/> tendra asta doce
 baras.
 <lb/> Dos baras de olandilla <sic>musa</sic>.
 <lb/> En tres pedazos de olandilla bieja, asta dos baras.
 <lb/> Dos baras de raso liso negro.
 <lb/> Media bara de sarga musca.
 <lb/> Quatro senefas de tafetan doble, las tres berdes <lb/> y la otra carmesi.
 <lb/> Un pedazo de picocete musco biejo de bara de largo.
 <lb/> Unas mangas de raso liso musco foradas <del rend="cancelled">en</del>
 <pb n="329v" facs="DSC_0288a.JPG" id="e-215"/>
 <lb/> En dos pedazos, quatro paños de a bara y cuarta <lb/> de raso musco
 antiguo.
 <lb/> Media bara de raso negro.
 <lb/> Un pedazo de brocatel de bara y cuarta.
 <lb/> Cinco pedaços de olandilla encarnada.
 <lb/> Quatro baras de restaño encarnado en pedazos.
 <lb/> En tres pedaços de raso musco y blanco, quatro bs.
 <lb/> Un pedaço de tafetan de asta una bara, carne donzel.
 <lb/> Una casaca de tafetan doble negra para ombre.
 <lb/> Un guardapiés de raso azul con flores, forado, en <lb/> tafetan de caña.
 <lb/> Quatro pedazos de felpa encarnada mui biejos <lb/> y goteados de zera.
 <lb/> Un jubon de muger de raso liso negro con <lb/> botones de azabache.
 <lb/> Una casaca pa muger de raso liso negra vieja.
 <lb/> Una basquiña de lamparilla musca vieja.
 <lb/> Un jubon pa muger de sempiterna viejo.
 <lb/> Una basquiña de peldefiebre vieja musca.
 <lb/> Una mantilla pa ombros de baieta negra.
 <lb/> Cinco baras de albornoncillo negro.
 <lb/> Una senefa de olandilla encarnada en dos peda<add
 place="above">z</add>os</add>.
 <lb/> Una tualla de gasa encarnada de plata de ba<lb/>ra y media.
 <lb/> Dos cortinas de barragan pa silla de manos guar<lb/>necidas de
 flequesillo azul.
 <lb/> Unas medias de seda encarnada biejas.
 <lb/> Diez y siete vestidos de felpa berde lisa, guar<lb/>necidos de trencilla de
 plata falsa que se com<lb/>ponen <pb n="330r" facs="DSC_0289a.JPG" id="e-246"/>
 de ropilla con mangas perdidas y ca<add place="above">l</add>zones, <lb/> y algunos
 de ellos sin guarnicion y algunas de las dhas <lb/> mangas son de restaño con fleco.
 <lb/> Una capa de paño negro apolillada.
 <lb/> Dos pares de <sic>clazones</sic>, los unos de peldefiebre musco <lb/>
 bien tratados y los otros de ormesi listados viejos.
 <lb/> Un caparazon de silla de xineta, el coxin de ta<lb/>fetan encarnado y las
 faldas de gasa blanca <lb/> con rapasejos de plata falsa, mui usado.
 <lb/> Ocho caxas de ormesi encarnado: otras ocho de <lb/> tafetan sencillo y
 quatro de gasa encarnada y <lb/> blanca, todo desecho y manchado, al parecer auer
 <lb/> sido almua<add place="above">das</add> que se an desecho.
 <lb/> Una alfombra turca de asta seis baras de largo <lb/> y lo que corresponde
 de ancho, mui usada.

<lb/> Un manto de muger de seda usado.

</p>

<p id="p-9">

<lb/> <add place="left">Ropa <lb/> blanca</add>

<lb/> Seis colchones, los quatro grandes y los dos pe<lb/>queños de terliz y
 lienzo, poblados de lana, viejos.

<lb/> Dos pedazos de terliz biejo.

<lb/> Veinte baras de lienzo de la coruña angosto.

<lb/> Diez y seis baras de lienzo casero de bara de ancho.

<lb/> Quatro sabanas de lienzo mui usadas.

<lb/> Tres cortinas biejas de lienzo.

<lb/> Veinte y siete baras de cotonía en pieza.

<lb/> Seis baras de gusanillo fino en un pedazo.

<lb/> Quatro baras de gusanillo ordinario.

<lb/> Cinco baras de terliz de <sic>Yngalatera</sic> listado.

<lb/> Vete y ocho baras de gasa apeñascada en un pedazo.

<lb/> Una tabla de manteles reales.

<pb n="330v" facs="DSC_0290a.JPG" id="e-277"/>

<lb/> Una tabla de manteles reales mui fina.

<lb/> Una tabla de manteles mediana de cuadri<lb/>tos de bara y media de
 largo.

<lb/> Dos seruilletas de gusanillo, la una nue<lb/>ba y la otra usada.

<lb/> Una sobrecama de gasa listada forada <lb/> en lienzo.

<lb/> Ocho sabanas biejas y mui remendadas.

<lb/> Una cortina bieja y mui remendada.

<lb/> Tres pares de calzetras mas otro, biejas.

<lb/> Seis seruilletas alemaniscas biejas.

<lb/> Una tualla bieja.

<lb/> Dos almuadas biejas.

<lb/> Un abantal de lienzo pa la cozina.

<lb/> Dos paños de angulema q tienen a bara <lb/> y quarta de largo cada uno.

<lb/> Una tualla de lienzo pintado.

<lb/> Una senefa de olandilla encarnada.

</p>

<p id="p-10">

<lb/> <add place="left">Enpeño</add> Dos sabanas q estan en empeño.

<lb/> Una pierna de sabana mojada.

<lb/> Dos almuadas de santiago nuevas.

<lb/> Nuebe almuadas biejas.

<lb/> Una pierna de savana mojada.

<lb/> Unas enaguas de bocadillo con desilados.

<lb/> Una camisa pa muger bieja.

</p>

</body>

</text>

Anexo III. Catálogo de etiquetas XML-TEI utilizadas en ODE

ETIQUETAS	ATRIBUTOS	VALORES
pb <i>Page begining</i> Comienzo de una nueva página del manuscrito	n Número de folio.	Formado por uno o más dígitos seguidos de las letras r (recto) o v (vuelto) en minúscula y sin espacios: <pb n="16r"/> Si el nº de folio está conjeturado se incluye entre corchetes: <pb n="[16r]"/> Si el documento no tiene foliación se indica la abreviatura sf (sin foliación) <pb n="sf"/>
	facs Vincula la transcripción a la imagen.	Deber ser idéntico al nombre de la imagen en archivo JPG sin variar caracteres en mayúscula o minúscula: <pb facs="DSC_031.JPG"/>
add <i>Addition</i> Contiene letras, palabras o frases añadidas por el autor del manuscrito.	place Indica la posición de la adición en el manuscrito original.	above <add place="above"> adición por encima de la línea.
		below <add place="below"> adición por debajo de la línea.
		left <add place="left"> adición en el margen izquierdo.
		right <add place="right"> adición en el margen derecho.
		top <add place="top"> margen superior por encima de la caja de escritura.
		bottom <add place="bottom"> margen inferior de la caja de escritura.
		inLine <add place="inLine"> adición sobrescrita, encima de una palabra que ha sido borrada. Solo en el caso de sustituciones dentro de la etiqueta <subst> y combinada con <del>
del <i>Deletion</i> Texto que ha sido eliminado por el escribano mediante tachado, borrado o raspado y que es posible leer.	rend Señala el tipo de cancelado, especificado en el valor.	cancelled señala cancelaciones mediante tachado, borrado o raspado. <del rend="cancelled">
		overwritten fragmento eliminado mediante texto que ha sido sobrescrito. <del rend="overwritten">

supplied Conjetura Texto proporcionado por el transcriptor por cualquier motivo (bien por daño del soporte, encuadernación o ilegibilidad). La etiqueta debe contener la palabra entera como unidad mínima y no solo una parte de ella. No se conjeturan caracteres sueltos.	cert Indica la certeza con el que el transcriptor realiza la conjetura. El grado de certeza se señala en el valor del atributo.	high El grado de certeza es alto. El transcriptor proporciona una solución, aunque en el manuscrito no se lea con claridad debido a una grafía dificultosa, mala encuadernación, etc. <supplied cert="high">lebrillo</supplied> low El grado de certeza es bajo. El transcriptor no puede asegurar que lo transcrito sea lo que aparece en el manuscrito, bien porque carece de sentido o bien porque es una lectura incierta. <supplied cert="low">cartaros</supplied>
gap Omisión Indica un hueco en el que no es posible transcribir nada, bien porque el texto es ilegible o porque ha sido omitido en el original o porque, por razones editoriales, no interesa la transcripción. Esta etiqueta será siempre un elemento vacío.	reason Indica la razón del vacío, especificada en el valor del atributo.	illegible <gap reason="illegible"/> Señala un vacío de texto causado por: - daños en el soporte original. - grafía dificultosa. - cancelado o tachado del escribano que es imposible leer. - encuadernación que impide la lectura omitted <gap reason="omitted"/> Señala un hueco presente en el manuscrito original debido a despistes u olvidos del escribano. editorial <gap reason="editorial"/> Señala un fragmento de texto que no se transcribe por no resultar relevante lingüísticamente.
hi <i>Highlighted</i> Marca una palabra o frase marcada gráficamente diferente del resto del texto que la circunda.	rend Indica la manera en la que el texto ha sido resaltado, especificado en el valor del atributo.	underlined Subrayado: <hi rend="underlined"> bold Negrita: <hi rend="bold"> italic Cursiva: <hi rend="italic"> Es posible combinar dos opciones separando ambos valores con una

		coma (i.e. texto subrayado y en negrita) <hi rend="bold, underlined">
foreign Palabra o frase perteneciente a otra lengua distinta de la del resto del texto en que se incluye.	lang Indica la lengua especificada en el valor del atributo.	la Latín: <foreign lang="la"> Ej: <i>ad litem; in facie ecclesiae</i>
		fr Francés: <foreign lang="fr">
		en Inglés: <foreign lang="en">
seg	type	signature Encierra la firma. Ej.: <seg type="signature">Francisco de Torres</seg>
ETIQUETAS SIN ATRIBUTO		
p <i>Paragraph</i> Nuevo párrafo.	texto aquí Encierra un párrafo completo. Aparecerá después de cada punto y aparte, siempre que se quiera indicar visualmente que empieza un párrafo nuevo.	
lb <i>Line beginning</i> Inicio de línea.	<lb/> Elemento vacío. Es importante dejar espacios entre la etiqueta y la palabra que empieza la línea. Ej.: vecino <lb/> de esta villa Por el contrario, no se dejan espacios cuando el final de línea coincide con rotura de palabra. Ej.: me<lb/>nester	
sic Texto aparentemente incorrecto o inexacto.	<sic>otro</sic> mujer	
subst <i>Substitution</i> Agrupa una o más cancelaciones con una o más adiciones en una sola intervención en el texto.	<subst> </subst> Incluye las etiquetas <del> y <add> o <gap/> para indicar las cancelaciones y restituciones originales hechas por el escribano. Dentro de <subst> se añade <del> con el texto tachado si es posible leerlo o con <gap> como elemento vacío si no lo es. A continuación, se indica la restitución con <add>: Ej. <subst><del rend="overwritten">ll</del> <add place="inLine">y</add>endo</subst> Nótese que en los casos de tachado combinados con una adición, el atributo de <add> será siempre @place con el valor 'inLine'.	
surplus	de de esta villa de <surplus>de</surplus> esta villa	

Texto considerado redundante generalmente por despistes del escribano que repite una palabra.	
q <i>Quoted.</i> Contiene un fragmento que se marca como literal.	Ej.: una firma que dice <code><q>licenciado Viana</q></code>
handShift <i>Cambio de mano</i>	<code><handShift/></code> Elemento vacío. Marca el principio de texto atribuible a otra mano distinta respecto a la anterior o a un cambio de estilo de escritura, tinta o letra.
<code><!-- --></code> Comentarios de uso interno solo visibles en lenguaje XML.	<code><!-- esto es un comentario - - ></code> No se pueden utilizar dentro de etiquetas. No se ven en la edición del documento. Esta etiqueta no tiene una etiqueta de cierre.
ATAJOS EN TEITOK	
<code> </code> Desarrollo de abreviaturas	Ej.: vezo vezino En TEITOK es preferible usar la doble barra vertical <code> </code> para la expansión de abreviaturas. Se suprime así la etiqueta <code><ex></code> proporcionada por TEI para esta misma cuestión. A la izquierda de la doble barra debe aparecer la abreviatura original; a la derecha debe indicarse la expansión de esa abreviatura proporcionada por el editor. La expansión debe ser continuación exacta de lo que está a la izquierda. La expansión de abreviaturas <code> </code> es combinable con la normalización (<i>vid.</i> abajo)
<code> =</code> Normalización	Es posible añadir la normalización de una palabra directamente desde la edición en XML. Para ello se utiliza barra vertical seguida del signo igual. Ej. vezino =vecino Se puede combinar con la expansión de abreviaturas, pero no se deben añadir espacios antes ni después de las plecas y el signo = Ej. pco publico =público

Anexo IV. Relación de diminutivos transparentes incluidos en el análisis

A

alfombra

alfombrita

n.º casos: 1

1877, MD-Madrid: una *alfombrita* de moqueta vieja, pequeña.

almendra

almendritas

n.º casos: 1

1815, MD-Madrid: otro par de pendientes de lazo con tres *almendritas* de diamantes y granos de arjofar engastados en oro.

arca

arquilla, arquita, arquitas

n.º casos: 14

1700, MD-Madrid: vna *arquilla* de pino bieja tasada en quatro reales ▪ **1711**, MD-Madrid: una *arquita* de zipres de quarta de alto y media bara de largo ▪ **1747**, MD-Madrid: mas dos *arquitas* de pino pequeñas con sus zerraduras y llaues ambas ▪ **1748**, AL-Fiñana: Una *arquilla* pequeña en dos r[eale]s ▪ **1765**, MD-Madrid: Una *arquita* de cedro en cien r[eale]s ▪ **1766**, AL-Cuevas de Almanzora: Ytt[em] vna *arquita* pequeña de albercoquero con zerradura y llaue.

arcón

arconzillo

n.º casos: 1

1720, MD-Pinto: vn *arconzillo* pequeño viexo.

aro

arillo

n.º casos: 1

1711, MD-Madrid: un *arillo* de yerro con quarenta y ocho escarpas.

artesa

artesilla

n.º casos: 2

1809, MD-Navalafuente: vna *artesilla* mondonguera en cinco r[eale]s ▪ **1870**, AL-Tabernas: una *artesilla* para yeso.

asa

asitas

n.º casos: 1

1750, MD-Madrid: seis calabazitas de Indias pequeñas con pies y *asitas* de plata.

azada

azadica, azadilla

n.º casos: 2

1714, AL-Cuevas de Almanzora: una *azadica* pequeña ▪ **1845**, MD-Ambite: vna *azadilla* para destallar olivos.

azadón

azadonzillo, azoncillo

n.º casos: 2

1704, AL-Cuevas de Almanzora: un *azadonzillo* biejo ▪ **1834**, MD-Algete: Yd[em] un podon, un *azoncillo* y una azada de yerro.

B

bacina

bazinica

n.º casos: 1

1793, MD-Madrid: Ytem una *bazinica* de cobre en veinte r[eale]s vellon.

bancal

vancalicos

n.º casos: 1

1809, AL-Zurgena: quatro *vancalicos* con quatro olivos, algunos arboles frutales y un parral.

bandeja

bandejitas

n.º casos: 1

1855, MD-Algete: dos *bandejitas*, unos frascos y dos figuras.

banco

banquillo, banquillos

n.º casos: 15

1722, MD-Colmenar de Oreja: vn *banquillo* de pino en quatro reales ▪ **1800**, MD-Brunete: una cama de cinco tablas con sus *banquillos* de pino ▪ **1801**, MD-Madrid: un tablado de cama con sus *banquillos* y respaldo en sesenta r[eale]s ▪ **1832**, MD-Navas del Rey: una banqueta y dos *banquillos* ▪ **1858**, MD-Escorial: un *banquillo* de labar en quatro reales ▪ **1870**, MD-Aranjuez: un tablado de cama con sus *banquillos* en setenta y un escudos ▪ **1877**, MD-Madrid: un baul forrado de piel con sus *banquillos*.

barril

barrilico, barrilitos

n.º casos: 3

1766, AL-Cuevas de Almanzora: un *barrilico* de vidrio de m[edi]o q[uarti]llo y zinco plattos [...] Ytt[em] vna frasquera pequeña viexa con nueve *barrilitos* de cristal en quarentta r[eale]s [...] Un espeziro y tres *barrilitos* de vidrio de a medio q[uarti]llo.

baúl

baulico

n.º casos: 1

1745, AL-Fiñana: un *baulico* pequeño tachonado y forrado, con su llave.

bigornia, -o

vigornillo

n.º casos: 1

1870, MD-Aranjuez: un *vigornillo* en un escudo.

blusa

blusecitas

n.º casos: 1

1840, MD-Madrid: dos *blusecitas* blancas.

bocal

bocalito

n.º casos: 1

1715, MD-Madrid: mas un pomito de xptal con su *bocalito* de plata.

bolso

bolsillo

n.º casos: 2

1822, AL-Fiñana: um *bolsillo* de picote ▪ **1852**, MD-Madrid: un *bolsillo* de terciopelo para la mano.

bote

botecico

n.º casos: 1

1786, AL-Cuevas de Almanzora: un *botecico* de oja de latta.

botón

botonzillos

n.º casos: 1

1705, MD-Madrid: otras quatro [almohadas] de *botonzillos*.

brasero, -a

braserillo, braserilla, braserillos, braserito

n.º casos: 6

1711, MD-Madrid: Cinco *braserillos* de yerro con sus pies ▪ **1715**, MD-Madrid: Vnas trebedes, vnas parrillas y un *braserito* y un asador, todo de hierro ▪ **1840**, MD-Madrid: Una [sic] almirez con su mano. Una *braserilla* ▪ **1843**, MD-Madrid: 1

braserillo p[ar]a fumar de plata con mangos de madera ▪ **1844**, AL-Cuevas de Almanzora: Un *braserillo* de azofar ▪ **1870**, MD-Aranjuez: Una cafetera con *braserillo* en tres escudos.

broche

brochecitos

n.º casos: 1

1728, MD-Madrid: otros *brochecitos* pequeños de plata.

C

cabecero

cavecerillo

n.º casos: 2

1844, AL-Cuevas de Almanzora: una cama de matrimo con *cavecerillo*. ▪ **1845**, AL-Cuevas de Almanzora: una cama de *cavecerillo* pintada.

cacerola

cacerolitas

n.º casos: 1

1793, MD-Madrid: ytem dos *cacerolitas* de oja de lata.

cachorro

cachorrillo

n.º casos: 1

1870, MD-Aranjuez: un *cachorrillo* en ochocientas m[ilesima]s.

cadena

cadenilla, cadenillas, cadenitas

n.º casos: 4

1703, MD-Madrid: Una espada ancha con guarnizion hordinaria y puno de platta en forma de *cadenilla* ▪ **1724**, MD-Madrid: Vn palillo de ebano de hazer media con su casquillo y *cadenilla* de plata ▪ **1750**, MD-Madrid: Un atado de dices de plata que se compone de [...] dos engastes y un hueso

con sus *cadenillas* de plata ▪ **1843**, MD-Madrid: Un collarcito de oro de *cadenitas* con un camafeo.

caja

caxica, cajica, cagita, cajita

n.º casos: 4

1728, MD-Madrid: Otra *cajita* de plata pequeña que pesa dos onzas ▪ **1745**, AL-Fiñana: Vna *caxica* de Dios engastada en platta ▪ **1766**, AL-Cuevas de Almanzora: Ytt[em] una *cagita* de madera pintada en dorado ▪ **1786**, AL-Cuevas de Almanzora: una *cajica* de carton con dos pares de botones de platta.

cajón

caxoncito, cajoncitos

n.º casos: 2

1750, MD-Madrid: Una caxa para muger [...] hechura de *caxoncito* ▪ **1877**, MD-Madrid: Un neceser en forma de armario pequeñito para guardar alhajas, tiene seis *cajoncitos*.

calabaza

calabazitas

n.º casos: 1

1750, MD-Madrid: seis *calabazitas* de Indias pequeñas con pies y asitas de plata.

caldera

calderilla, calderita

n.º casos: 4

1726, AL-Abla: mas vna *calderilla* de ylar seda ▪ **1751**, MD-Brunete: vna *calderilla* en diez r[eale]s ▪ **1793**, MD-Madrid: Ytem una *calderilla* en cuatro r[eale]s ▪ **1800**, MD-Brunete: una *calderita* de cobre nueva.

caldero

calderito

n.º casos: 1

1845, MD-Ambite: un *calderito* en treinta r[eale]s.

camisa

camisita

n.º casos: 1

1728, MD-Madrid: Vna *camisita* de un niño nueva en dos r[eale]s.

campana

campanilla, campanillas

n.º casos: 3

1750, MD-Madrid: Un atado de dices de plata que se compone de una cornetilla con tres cascabeles, una *campanilla*, dos castañas [...] ▪ **1809**, MD-Navalafuente: vna *campanilla* con su zinto en ocho r[eale]s ▪ **1844**, AL-Cuevas de Almanzora: un collar con cinco *campanillas*.

canasta

canastica

n.º casos: 1

1786, AL-Cuevas de Almanzora: una *canastica* con tres oblillicos de hilo blanco.

cántaro

cantarico, cantaricos

n.º casos: 2

1766, AL-Cuevas de Almanzora: Ytt[em] tres morteros y dos *cantaricos* de hecha [sic] azeite todo de barro ▪ **1786**, AL-Cuevas de Almanzora: un *cantarico* p[ar]a el azeite.

capa

capita

n.º casos: 2

1766, AL-Cuevas de Almanzora: una *capita* de pañete azul biexa en quattro r[eale]s ▪ **1840**, MD-Madrid: una *capita* escocesa.

cazo

cacillo, cazito, cacillos, cacitos

n.º casos: 9

1705, MD-Madrid: el *cazito* de azofar ▪ **1801**, MD-Madrid: Un cazo, una sarten,

cacillo de azofar y otro de yerro ▪ **1815**, MD-Madrid: un *cazito* chico de azofar ▪ **1836**, MD-Estremera: Una requesonera y tres *cacitos* pequeños en 8 ▪ **1843**, MD-Madrid: 3 cucharones soperos o de *cacillo* con 35 onzas de plata. 2 *cacillos* calados p[ar]a el azucar con 9 onzas de plata. 2 *cacillos* de plata para la mostaza con 1 ½ onza.

cazuela

cazolita

n.º casos: 1

1715, MD-Madrid: una *cazolita* de plata con su tapadera de lo mismo.

cedazo

zedazillo

n.º casos: 1

1720, MD-Pinto: Dos varrenon (sic) grandes de *zedazillo*, vn rallo y vn embudo.

cencerra

zenzerrica, senzerrillas

n.º casos: 2

1711, MD-Madrid: Otro senzerro grande de yerro. Treinta y nueve *senzerrillas* de yerro ▪ **1714**, AL-Cuevas de Almanzora: Una *zenzerrica*. Dos cadenas para mulas. Un collar de hierro para una mula.

cerco, -a

cerquillo, zerquilla

n.º casos: 4

1766, AL-Cuevas de Almanzora: vna porzelana con *zerquilla* de platta sobredorada y dentro vna pintura de N[uest]ra S[eñor]a en doze r[eale]s ▪ **1870**, MD-Aranjuez: Un embudo sin *cerquillo* en cuatroc[ien]tas mil[esima]s. Diez id[em] de embutir en quinientas id[em]. Seis id[em] pequeños sin *cerquillo* en seiscientas milésimas. Uno id[em] con *cerquillo* en quinientas mil[esima]s.

cerdo

serdillos, cerdicos, zerdicas, cerdito

n.º casos: 4

1775, AL-Escúllar: seis *serdillos* destetados en cuarenta y cinco r[eale]s ▪ **1803**, AL-Zurgena: tres *cerdicos* en v[ei]nte y seis r[eale]s. It[em] tres *zerdicas* a v[ei]nte y cinco ▪ **1809**, MD-Navalafuente: Vn *cerdito* en cien r[eale]s.

chapa

chapita

n.º casos: 1

1750, MD-Madrid: Un engarze de una santa Beronica de Jaen guarnezida de plata con chrystal y en el reverso una *chapita* de plata con un Jesus que es cuadrado, con cordonzillo, en quinze reales.

chico

chiquito, chiquita, chiquitos, chiquitas

n.º casos: 14

1832, AL-Fiñana: quatro cuadros *chiquitos* con cristales [...] otra [arca] *chiquita* de nogal [...] vn tonel *chiquito* ▪ **1835**, AL-Fiñana: Otra [arca] mas *chiquita* bieja con cerradura y llabe [...] Un cazo *chiquito* biejo de cobre [...] Cuatro laminas *chiquitas* en 4 ▪ **1844**, AL-Almería: Siete pañuelos para el cuello *chiquitos* ▪ **1858**, MD-Escorial: otro [pañuelo] para el cuello de lana dulce, *chiquito* ▪ **1882**, AL-Fiñana: Dos vasos *chiquitos* [...] Otras [trébedes] *chiquitas* en veinte y cinco centimos.

chispa

chispilla

n.º casos: 1

1766, AL-Cuevas de Almanzora: Ytt[em] un zintillo de oro con una esmeralda y una *chispilla* de diamantte en sesentta r[eale]s.

cofre

cofreçico, cofrezito, cofrecito

n.º casos: 3

1722, MD-Colmenar de Oreja: Vn *cofrezito* pequeño en dos reales y m[edi]o ▪ **1726**,

AL-Abla: vn *cofreçico* mediano con su llave y zeradura ▪ **1800**, MD-Brunete: Un *cofrecito* con su llave en diez.

colchón

colchonzillo

n.º casos: 1

1701, MD-Madrid: Un *colchonzillo* pequeño muy biejo.

collar

collarcito

n.º casos: 1

1843, MD-Madrid: Un *collarcito* de oro de cadenitas con un camafeo.

concha

conchita, conchitas

n.º casos: 2

1836, MD-Estremera: un platillo de espaviladeras y una *conchita* que todo peso setenta y cinco onzas ▪ **1877**, MD-Madrid: Dos floreros hechos de marisco ó sea de *conchitas* y caracoles.

contador

contadorzito

n.º casos: 1

1715, MD-Madrid: Vn *contadorzito* de una quarta de alto, de concha.

copa

copilla, copillas, copitas

n.º casos: 5

1853, AL-Fiñana: Una *copilla* en 4 [...] Cinco *copillas* de cristal en 4 ▪ **1877**, MD-Madrid: tres botellas con sus tapones de cristal, once *copitas* ▪ **1882**, AL-Fiñana: Una *copilla* de metal en veinticinco centimos ▪ **1896**, AL-Laujar: Una docena de *copillas* de vino.

corneta

cornetilla

n.º casos: 1

1750, MD-Madrid: Un atado de dices de plata que se compone de una *cornetilla* con tres cascabeles una campanilla, dos castañas, una mano de texon, una lengua de bibora, un jarro, dos engastes y un hueso con sus cadenillas de plata, todo en ciento y treinta y dos r[eale]s.

corona

coronita

n.º casos: 1

1765, MD-Madrid: Una *coronita* y tres potencias de plata treinta r[eale]s.

cortina

cortinita, corttinillas, cortinillas

n.º casos: 3

1701, MD-Madrid: Dos *corttinillas* de ttafettan con una zenefa uiejas ▪ **1728**, MD-Madrid: Una *cortinita* y otro paño viejo en vn real ▪ **1855**, MD-Algete: Yd[em] siete *cortinillas* de muselina andadas en 14.

cuadro

quadrico, quadritos, cuadros

n.º casos: 8

1711, MD-Madrid: Una tabla de manteles mediana de *cuadritos* de bara y media de largo ▪ **1714**, AL-Cuevas de Almanzora: Un *quadrico* con su marco, viejo ▪ **1747**, MD-Madrid: Mas dos *quadritos* de estampas de N[uest]ra S[eño]ra con sus marquitos dorados ▪ **1784**, MD-Pinto: Dos *quadritos* pintados en chrystal en veinte rr[eale]s ▪ **1858**, MD-Escorial: vn espejo, marco dorado y tres *cuadritos* en diez.

cuchara

cucharilla, cucharita, cucharillas, cucharitas

n.º casos: 11

1770, MD-Estremera: Una *cucharitta* de azofar en dos r[eale]s ▪ **1786**, AL-Cuevas de Almanzora: una cajica de carton con dos pares de botones de platta, un limpiadientes y clavete de lo mismo y una *cucharita* de

metal ▪ **1793**, MD-Madrid: una espumadera y dos *cucharitas* de yerro [...] una *cucharilla* grande en doce ▪ **1819**, AL-Alboloduy: Seis *cucharillas* de yerro ▪ **1834**, MD-Madrid: Vna docena de *cucharitas* de plata para café ▪ **1834**, MD-Algete: dos candiles, unas tenazas y unas un candelero y una *cucharilla* en veinte y seis r[eale]s ▪ **1843**, MD-Madrid: 22 *cucharitas* y 4 tenedores pequeños [...] 1 caliz con patena sobre ostia y *cucharilla* de plata [...] 1 cucharon ovalado 6 *cucharitas* para café.

D

daga

daguilla

n.º casos: 1

1786, AL-Cuevas de Almanzora: una *daguilla*.

dosel

doselito

n.º casos: 1

1715, MD-Madrid: Vn *doselito* de cama y en el un santísimo xpto cruzificado pequeño de marfil.

E

embudo

embudillo, embudico, embudillos

n.º casos: 5

1714, AL-Cuevas de Almanzora: Un *embudico* pequeño de oxa de lata ▪ **1786**, Cuevas de Almanzora: un *embudico* y una alcuza de oja de latta ▪ **1832**, MD-Navas del Rey: un *embudillo*, un real ▪ **1833**, AL-Abla: Un *embudillo* de lata ▪ **1870**, MD-Aranjuez: Una geringa de embutir con dos *embudillos* en tres escudos.

engaste

engasticos

n.º casos: 1

1750, MD-Madrid: Un brazalete de tumbaga [...] y alrededor de d[ic]ho brazalete ay seis *engasticos* de oro.

escriptorio

escriptorito, escriptorico

n.º casos: 3

1715, MD-Madrid: Vn *escriptorito* de caova de una bara de largo y poco mas de media de ancho [...] Vn *escriptorito* de zedro de poco mas de bara de largo con ocho nauetas ▪ **1747**, MD-Madrid: Mas vn *escriptorico* de box de tres q[uar]tas de largo.

espejo

espegillo

n.º casos: 1

1704, AL-Cuevas de Almanzora: Un *espegillo* pequeño biejo.

estera

esterita

n.º casos: 1

1715, MD-Madrid: Otra *esterita* de estrado pequeña.

F

fino

finita

n.º casos: 1

1890, MD-Alcalá: Vna cadena de reloj de oro de señora, *finita*.

fleco

flequesillo

n.º casos: 1

1711, MD-Madrid: Dos cortinas de barragan p[ar]a silla de manos guarnecidas de *flequesillo* azul.

frasco

frasquillos

n.º casos: 1

1890, MD-Alcalá: dos *frasquillos* pequeños de cristal, otro idem para mostaza.

G

gordo

gorditos

n.º casos: 1

1724, MD-Madrid: Otros quatro escotes de encajes mojados *gorditos* en diez y seis reales.

grano

granicos

n.º casos: 1

1750, MD-Madrid: Dos sortijas de oro, la una sin piedra y la otra guarnezida con tres piedras blancas y falsas y dos *granicos*.

H

haba

abillas

n.º casos: 2

1785, AL-Fiñana: Catorze cel[emine]s *abillas* ▪ **1834**, AL-Escúllar: de una fanega y tres cuartillas de *abillas*, ciento y cinco r[eale]s.

hacha

açhica

n.º casos: 1

1714, AL-Cuevas de Almanzora: Una *açhica* pequeña p[ar]a cortar leña.

harnero

arnerillo, arnerillos

n.º casos: 3

1834, MD-Algete: una criba, un *arnerillo* y un arnero grande en diez y siete r[eale]s ▪

1870, MD-Aranjuez: Cuatro *arnerillos* para arroz en un escudo y doscientas milésimas ▪ **1882**, AL-Fiñana: un *arnerillo* de lata en id[em].

herramienta

erramentillas

n.º casos: 1

1784, AL-Abrucena: una azuela, un zepillo, un escoplo y otras *erramentillas*, quarenta r[eale]s.

hierro

yerrezillos

n.º casos: 1

1770, MD-Madrid: Vn cajon en el qual habia diferentes *yerrezillos* uiejos.

J

jamuga

jamuguillas, jamugillas, gamugillas, amuguillas

n.º casos: 4

1750, AL-Dalías: las *jamuguillas* de madera para cavalleria valen siete r[eale]s ▪ **1784**, AL-Abrucena: Unas *jamugillas*, quatro r[eale]s ▪ **1834**, AL-Escúllar: Unas *gamugillas* en 6 ▪ **1882**, AL-Fiñana: Dos pares *amuguillas* para barcinar en cinquenta centimos.

jarra

jarrita

n.º casos: 1

1766, AL-Cuevas de Almanzora: una *jarrita* de vidrio de quattro asas.

jícara

jicarica

n.º casos: 1

1786, AL-Cuevas de Almanzora: un basico de cristal y una *jicarica* con un poco almidon.

jubón

jugoncillo (sic)

n.º casos: 1

1774, AL-Bacares: Un *jugoncillo* encarnado de bayeta, seis.

L

lámina

laminitas, laminicas

n.º casos: 2

1724, MD-Madrid: Dos *laminitas* de Nuestra Señora, bronzeadas ▪ **1736**, AL-Almería: dos *laminicas* papel con marcos negros.

lámpara

lamparilla, lamparillas

n.º casos: 2

1840, MD-Madrid: Una *lamparilla* de cristal ▪ **1870**, MD-Aranjuez: Cuatro *lamparillas* doradas p[ar]a gas Mille.

lechona

lechoncica

n.º casos: 1

1766, AL-Cuevas de Almanzora: una *lechoncica* en diez y seis r[eale]s.

libro

libritos

n.º casos: 1

1855, MD-Algete: cuatro *libritos* de yglesia, a dos r[eale]s.

M

mangote

mangotillos

n.º casos: 1

1745, AL-Fiñana: Vnos *mangottillos* de raso verdes y encargado, viexos.

manta

manttica

n.º casos: 1

1745, AL-Fiñana: Vna *manttica* azul de pelo de seda mediada.

mantilla

manttillitta

n.º casos: 2

1701, MD-Madrid: Una *manttillitta* de ombros de sempiterna encarnada bordada [...] Una *manttillitta* con vn encaje negro biejo.

mantellina

manttellica

n.º casos: 1

1786, AL-Cuevas de Almanzora: una *manttellica* blanca de lienzo delgado.

máquina

maquinilla

n.º casos: 3

1870, MD-Aranjuez: Una *maquinilla* de hacer filetes de seis juegos en quince escudos [...] Una *maquinilla* de volver hojas de canelon en dos escudos ▪ **1877**, MD-Madrid: una *maquinilla* para hacer café.

marco

marquitos, marquittos

n.º casos: 3

1701, MD-Madrid: Diez uittelas con sus *marquittos* ▪ **1715**, MD-Madrid: dos vitelas de s[a]n Franc[isc]o y s[a]n Antonio, con sus *marquitos* negros y molduras doradas ▪ **1747**, MD-Madrid: Mas dos quadritos de estampas de N[uest]ra S[eño]ra con sus *marquitos* dorados.

medalla

medallitas

n.º casos: 1

1765, MD-Madrid: Nuebe *medallitas* de plata en veinte y ocho r[eale]s.

medallón

medalloncito

n.º casos: 2

1834, MD-Madrid: Otras tres [medallas] portuguesas y otra de olanda con un *medalloncito* ▪ **1840**, MD-Madrid: Un *medalloncito* con cadena de oro, ciento sesenta r[eale]s.

melindre

melindrillo

n.º casos: 1

1786, AL-Cuevas de Almanzora: un perrico con su *melindrillo* p[ar]a el cuello.

mesa

mesilla, mesita, mesica, mesitas

n.º casos: 16

1700, MD-Madrid: Vna *mesilla* bieja con su cajon, tasada en quatro reales. Otra *mesilla* bieja en tres reales ▪ **1714**, AL-Cuevas de Almanzora: Una *mesica*, madera de pino, pequeña ▪ **1715**, MD-Madrid: Vna *mesita* de nogal con sus pies torneados cubiertos de badana ▪ **1745**, AL-Fiñana: Ottra *mesica* pequeña de serbal, pequeña ▪ **1748**, AL-Roquetas: Vna *mesica* ▪ **1765**, MD-Madrid: una papelera con su *mesita* de nogal en nobenta r[eale]s ▪ **1766**, AL-Cuevas de Almanzora: una *mesita* charolada en diez y ocho r[eale]s ▪ **1791**, AL-Vera: otra *mesita* mas pequeña ▪ **1809**, MD-Navalafuente: Vna mesa seis r[eale]s. Otra *mesita* de tres pies en real y m[edi]o ▪ **1834**, MD-Algete: Yd[em] una *mesita* con cajon y tapeta en diez rr[eale]s ▪ **1836**, MD-Estremera: Dos *mesitas* con juego de damas, muy estropeadas ▪ **1843**, MD-Madrid: Una *mesita* pequeña de tocador.

muleto

muletilla

n.º casos: 1

1775, AL-Escúllar: una *muletilla* de nueve meses, pelo negro.

N

nudo

nuditos

n.º casos: 4

1745, MD-Madrid: vna paletina de *nuditos* de color de rosa en sesenta r[eale]s ▪ **1750**, MD-Madrid: Vn peto y escusali de gasa blanca con sus guarniziones de *nuditos* en treinta reales [...] Otro dicho de gasa con *nuditos* en veinte y ocho reales [...] Una paletina blanca de *nuditos* y otra de maltas en veinte y quatro reales.

O

olla

ollicas

n.º casos: 1

1766, AL-Cuevas de Almanzora: Ytt[em] dos *ollicas*, tres cazuelas de barro de la tierra.

orón

oronicos

n.º casos: 1

1774, AL-Bacares: Dos *oronicos* pequeños, cinco r[eale]s.

orza

orcilla, orzica, orzitas

n.º casos: 3

1714, AL-Cuevas de Almanzora: Una *orzica* para tener azeite, cauida de quatro a[rrobas] ▪ **1766**, AL-Cuevas de Almanzora: Ytt[em] dos *orzitas* de barro y dos rebolbedores ▪ **1834**, AL-Escúllar: Una *orcilla* pequeña en 2.

ovillo

oblillicos

n.º casos: 1

1786, AL-Cuevas de Almanzora: una canastica con tres *oblillicos* de hilo blanco.

P

pájaro

pajarito

n.º casos: 1

1890, MD-Alcalá: Vn alfiler con un *pajarito* de oro.

paleta

paletilla, paletillas

n.º casos: 3

1711, MD-Madrid: Dos *paletillas* de yerro para guebos [...] Una *paletilla* de yerro p[ar]a raer zera de la paila ▪ **1809**, MD-Navalafuente: Dos *paletillas* en tres r[eale]s.

palo

palillos

n.º casos: 1

1843, MD-Madrid: Seis sillas de *palillos* al respaldo, asientos amarillos.

pantalón

pantaloncitos

n.º casos: 1

1840, MD-Madrid: Cuatro pares de *pantaloncitos* á seis r[eale]s.

pañó

pañico

n.º casos: 1

1736, AL-Almería: Yt[em] un peinador, tohalla y *pañico* de nabajas.

pañuelo

pañuelitos

n.º casos: 1

1855, MD-Algete: Yd[em] cuatro *pañuelitos* blancos de la mano en 16.

peana

peañita

n.º casos: 2

1700, MD-Madrid: mas vn s[a]n Ant[oni]o con su *peañita*. Mas vna pintura de N[uest]ra S[eño]ra del Populo pintura en tabla con su *peañita* pequeña.

pedazo

pedazitos

n.º casos: 1

1711, MD-Madrid: Seis cortinas de gasa mui biejas y tres *pedazitos* de lo mismo.

pequeño

pequeñito, *pequeñita*, *pequeñitos*, *pequeñitas*, *pequeñica*

n.º casos: 12

1715, MD-Madrid: Vna benera *pequeñita* de rubies y diamantes ▪ **1717**, MD-Torrelaguna: vna ymagen *pequeñita* de N[uest]ra S[eño]ra la Soledad ▪ **1728**, MD-Madrid: Diez cucharas de plata, las dos *pequeñitas* y las demas medianas ▪ **1784**, AL-Abrucena: otra sarten *pequeñica* ▪ **1836**, MD-Estremera: un tenedor suelto con dos cucharas *pequeñitas* ▪ **1843**, MD-Madrid: 1 sortija con quince brillantes *pequeñitos* ▪ **1877**, MD-Madrid: Un neceser en forma de armario *pequeñito* para guardar alhajas, tiene seis cajoncitos [...] cinco planchas de yerro, otra *pequeñita*.

perla

perlitas

n.º casos: 2

1834, MD-Madrid: Vnos adornos de oro para cabeza con *perlitas* y camafeo, pesan onza y media ▪ **1843**, MD-Madrid: 3 sortijas y dos alfileres de oro con *perlitas*.

perol

perolillo, *perolito*, *perolitos*

n.º casos: 4

1711, MD-Madrid: Un *perolillo* de cobre para colar ▪ **1806**, AL-Almería: un *perolito* pequeño de azofar viejo [...] un *perolito* azofar p[ar]a almidon ▪ **1843**, MD-Madrid: Dos *perolitos* de azofar.

perro

perrico

n.º casos: 1

1786, AL-Cuevas de Almanzora: un *perrico* con su milindrillo p[ar]a el cuello.

pieza

piezecicas

n.º casos: 1

1786, AL-Cuevas de Almanzora: seis *piezecicas* de zinta de colores de seda.

pila

pililla, *pilita*, *pilica*, *pilicas*

n.º casos: 5

1703, AL-Cuevas de Almanzora: Una *pililla* de platta para agua vendita pequena ▪ **1714**, AL-Cuevas de Almanzora: Una *pilica* varro de Ubeda ▪ **1786**, MD-Madrid: dos *pilicas* p[ar]a las jarras ▪ **1836**, MD-Estremera: dos candeleros, una campanilla pequeña y una *pilita* ▪ **1890**, MD-Alcalá: una *pililla* de agua bendita de porcelana.

pileta

piletica

n.º casos: 1

1736, AL-Almería: [...] seis tenedores, una saruilla y *piletica*, to[do] de plata.

pipote

pipotillo

n.º casos: 1

1835, AL-Fiñana: Un *pipotillo* de cinco a[rrobas] a dos r[eale]s y m[edi]o cada una.

plato

platillo, platillos, platitos, plasticos

n.º casos: 29

1714, AL-Cuevas de Almanzora: Tres *plasticos* pequeños finos azules ▪ **1715**, MD-Madrid: Doze *platillos* trincheros de plata ▪ **1750**, MD-Madrid: Una caja de plata lisa cuadrada, prolongada con *platillo* en la tapa ▪ **1766**, AL-Cuevas de Almanzora: doze *platillos* de china blancos con flores ▪ **1784**, MD-Pinto: Una jicara de china y *platillo* en quinze r[eale]s. Otra jicara y *platillo* de lo mismo en 15 ▪ **1809**, MD-Navalafuente: Tres *platitos* en dos reales y m[edi]o ▪ **1843**, MD-Madrid: 1 marcerina gemela, *platillo* figura de barco con dos tapas de plata [...] Nueve jicaras grandes con sus *platillos* de china [...] Una romana con *platillo* [...] Dos salseras con sus *platillos* de china ▪ **1833**, AL-Abla: Cuatro *platillos* de Balencia ▪ **1840**, MD-Madrid: Cuatro candeleros de plaque. Un *platillo* y unas espavilad[era]s de lo mismo ▪ **1877**, MD-Madrid: Un servicio de china para café con adornos dorados compuesto de doce tazas con sus *platillos* ▪ **1890**, MD-Alcalá: diez gicaras para chocolate, cuatro *platitos* encarnados.

pomo

pomito

n.º casos: 1

1715, MD-Madrid: mas un *pomito* de xptal con su bocalito de plata.

puerta

portizillas

n.º casos: 1

1703, MD-Madrid: Un cajon, tocante al ofizio de pastelero, y sirue de mostrador y tiene dos cajones en alto y dos *portizillas* en lo bajo [...]

puntilla

puntillas

n.º casos: 3

1844, AL-Almería: Un par de id[em] [aretes] con *puntillas* [...] Tres id[em] [sortijas] con *puntillas* de diamantes [...] Otro id[em] [alfiler] con *puntillas* de diamante.

Q

queso

quesezito

n.º casos: 1

1791, AL-Vera: tres tinajas para azeite y vn *quesezito* pequeño.

R

rambla

ranblilla

n.º casos: 2

1785, AL-Almería: Yt[em] en la *ranblilla* de Benites, en tierra de la propia testamentaria, tres olibos en setenta y siete rs. Yt[em] cuatro holibos en la *ranblilla* de Benites.

rejón

rejoncillos

n.º casos: 1

1870, MD-Madrid: Una red de codornices con sus *rejoncillos* en tes (sic) escudos.

retal

retalicos

n.º casos: 1

1786, AL-Cuevas de Almanzora: otros dos *retalicos* de lienzo nuevo.

romana

romanilla, romanita, romanica

n.º casos: 6

1700, MD-Madrid: Mas una *romanilla* pequeña de onzas y libras ▪ **1715**, MD-Madrid: Vna *romanilla* de hierro ▪ **1720**,

MD-Pinto: Vna *romanilla* pequeña ▪ **1775**, AL-Cuevas de Almanzora: Yt[em] una *romanica* vieja en ocho r[eal]es ▪ **1836**, MD-Estremera: Una *romanilla* pequeña maltratada en 7 ▪ **1845**, MD-Ambite: otra *romanita* de dos a[rrobas] completas.

rosa

rosilla, rositas

n.º casos: 2

1750, MD-Madrid: una cruz y *rosilla* pasador de oro ▪ **1834**, MD-Madrid: Otros dhos [zarcillos] de *rositas* en sesenta rr[eale]s u[ello]n.

S

sábana

sabanica, sabanillas

n.º casos: 2

1745, AL-Fiñana: Vna *sabanica* de lienzo tramado de dos varas ▪ **1836**, MD-Estremera: Dos *sabanillas* de altar muy maltratadas.

saco

saquillos

n.º casos: 1

1870, MD-Aranjuez: Dos *saquillos* de tachuelas en 2.

sartén

sartencita, sartencitas

n.º casos: 2

1809, MD-Navalafuente: Vna *sartencia* chica en cinco ▪ **1845**, MD-Ambite: dos *sartencias* de añadir en siete r[eale]s.

serón

seronzillos

n.º casos: 1

1704, AL-Cuevas de Almanzora: Dos *seronzillos* terreros.

sierra

sierrecillas

n.º casos: 1

1870, MD-Aranjuez: Ciento veinte *sierrecillas* de quinque.

silla

sillicas

n.º casos: 2

1714, AL-Cuevas de Almanzora: Ocho *sillicas* de esparto y madera ▪ **1748**, AL-Roquetas de Mar: cinco *sillicas*.

T

tabla

tablillas, ttablillas

n.º casos: 3

1701, MD-Madrid: Una armadura de cama de *ttablillas* muy antiguas de nogal ▪ **1766**, AL-Cuevas de Almanzora: una cama de pino con cinco *tablillas* en doze r[eale]s [...] Ytt[em] una cama de pino con siete *tablillas*, diez r[eale]s.

taburete

tauretillos

n.º casos: 1

1848, AL-Almería: dos *tauretillos* p[ar]a los pies de igual madera y forro.

tarima

tarimilla

n.º casos: 1

1824, AL-Fiñana: Un brasero con paleta y *tarimilla* en cuarenta y cuatro r[eale]s.

tartera

tartteritta

n.º casos: 1

1770, MD-Estremera: Una *tartteritta* de china esquinada con su tapa en cattorce r[eale]s.

taza

tacillas, tacicas, tazicas

n.º casos: 5

1714, AL-Cuevas de Almanzora: Tres *tazicas* finas azules ▪ **1766**, AL-Cuevas de Almanzora: tres *tacicas* de vidrio con asas p[ar]a almibar en zinco r[eale]s [...] zinco *tazicas* de cristal con asas para almibar en medio real ▪ **1806**, AL-Almería: It[em] veinte y dos *tacillas* pa almivar [...] It[em] ocho *tacillas* ped[erna]l.

telar

telarillo

n.º casos: 1

1834, AL-Escúllar: Un *telarillo* de cintas en 10.

tenaza

tenacillas

n.º casos: 1

1870, MD-Aranjuez: Unas *tenacillas* en trescientas mil[esima]s.

tinaja

tenajita

n.º casos: 1

1834, MD-Algete: una *tenajita* en quinze r[eale]s.

toalla

thoallitas

n.º casos: 1

1766, AL-Cuevas de Almanzora: tres *thoallitas* p[ar]a limpiar la nabaja.

tubo

tubillos

n.º casos: 1

1774, AL-Bacares: Dos *tubillos* chicos, un real.

U

urna

vrnita

n.º casos: 3

1715, MD-Madrid: vna *vrnita* de palo santo perfilada de marfin (sic) ▪ **1724**, MD-Madrid: Vna *vrnita* de san Anttonio de platta en veinte y quatro r[eale]s ▪ **1858**, MD-Escorial: una *vrnita* de madera fina con puerta de cristal.

V

valona

baloncillas

n.º casos: 5

1824, AL-Fiñana: Dos sabanas de lienzo casero con *baloncillas* en doscientos cuarenta r[eale]s. Dos de bretaña contrahecha mediadas con *baloncillas* en ciento y sesenta r[eale]s. Dos mediadas sin *baloncillas* en cien r[eale]s [...] Un paño de servicio con *baloncillas* en cuatro r[eale]s ▪ **1882**, AL-Fiñana: Otras dos [almohadas] con *baloncillas* moteadas en dos.

vaso

basito, vasillos, basitos, vasitos, basico

n.º casos: 5

1728, MD-Madrid: dos *vasitos* de platta con sus asas, el vno mas chico que el otro ▪ **1786**, AL-Cuevas de Almanzora: un *basico* de cristal y una jicarica con un poco almidón ▪ **1848**, AL-Almería: [...] copas chicas de cristal, seis *basitos* de id[em] ▪ **1870**, MD-Aranjuez: Ciento sesenta y ocho id[em] de id[em] [roscas de metal de regadera] p[ar]a quinqués y *vasillos* ▪ **1890**, MD-Alcalá: Vn *basito* de plata pequeño.

viejo

viejecilla

n.º casos: 1

1728, MD-Madrid: Otra toalla *viejecilla* en tres reales.

Z

zafa

zafita

n.º casos: 2

1766, AL-Cuevas de Almanzora: vna *zafita*, barro de Genoba, azul y blanca, en tres

r[eale]s. Ytt[em] otra *zafita* blanca y pagiza, barro de Alcora, en dos r[eale]s.

zaranda

zarandilla

n.º casos: 1

1868, AL-Tabernas: Un cucharón, dos jarros y una *zarandilla*.

