

Remarks on the semantics and paradigmaticity of NN compounds (pre-final version)

Author: Jesús Fernández-Domínguez

Source: *The Mental Lexicon*, Volume 15, Issue 1, Oct 2020, p. 79 - 100

DOI: <https://doi.org/10.1075/ml.00015.fer>

Address for correspondence:

Jesús Fernández-Domínguez

Department of English and German Philologies

Campus de Cartuja, s/n – 18071 Granada (Spain)

Phone: +34 958 241000 ext. 20253. Fax: +34 958 243 678.

Email: [jesusferdom@ugr.es](mailto:jesusferdom@ugr.es)

## Abstract

English Noun+Noun compounding has garnered the attention of morphologists due to characteristics that involve its semantics (Bauer & Tarasova, 2010; Jackendoff, 2009), degree of productivity (Bauer, Beliaeva, & Tarasova, 2019; Maguire, Wisniewski, & Storms, 2010) and possible paradigmatic nature (Bagasheva, in press; Boyé & Schalchli, 2016). This article addresses the above questions from an inclusive perspective with the aim bringing together various unsolved issues in the study of this morphological process. The experiment exploits data from the BNC Sampler and the *Oxford English Dictionary*, from which morphological and semantic information is extracted in order to fathom the contribution of nominal root compounding to the lexicon. The results show that the paradigmatic nature, semantics and high productivity of NN compounding are all closely interrelated, which explains certain characteristics associated to the current status of this process and its role in the enrichment of the lexicon.

**Keywords:** compounding, corpus-based, derivational paradigms, modifiers, productivity, semantics, word-formation paradigms.

### 1. Antecedents

More than half a century after the publication of morphological milestones like Jespersen (1942) or Hatcher (1960), it is becoming increasingly difficult to make even a modest contribution to the understanding of Noun+Noun (NN) compounding. A wealth of standpoints has tackled the theoretical and practical complexities behind this process,

the entirety of which can be hardly covered in full in an article of standard length. Sensitive areas for NN compounding are those concerning its semantics (Bauer & Tarasova, 2010; Jackendoff, 2009; Schäfer, 2018), degree of productivity (Bauer, Beliaeva, & Tarasova, 2019; Maguire, Wisniewski, & Storms, 2010) and possible paradigmatic nature (Bagasheva, in press; Boyé & Schalchli, 2016; Radimský, in press).

The present article sets out on the following objectives:

- i) To look into the meaning diversity of NN compounding in order to detect pros and cons in the classification of semantic roles and their cross-categorial overlap.
- ii) To discover connections between the productivity values of NN compounding, its alleged paradigmatic nature and its role in the enrichment of the lexicon.

This introduction is followed by a description of the methodology (Section 2) and an examination of unsettled matters on compounding in the light of empirical evidence (Section 3), after which the conclusions are presented in Section 4.

## 2. Data Preparation

The data in this paper comes from the BNC Sampler, a c. 2 million-word subcorpus of the BNC, preferred here for its fine-grained manual POS-tagging and 50/50 balance between written and spoken texts. The data was extracted with *WordSmith Tools* (Scott, 2012) as follows:

- i) Retrieval of all NN combinations, which gives a list of 18,483 concordances. This procedure adopts an all-inclusive policy which does not consider the distinction

compound versus phrase (Bauer, 1998: 68-69). Entries are then alphabetically arranged by their left-hand member.

- ii) Data sampling aiming at around 100 initial types for the sake of manageability, by picking out 1 unit of every 184. Lemmatization is carried out by unifying spelling and inflectional variants after which, based on the left-hand nouns of the initial 100, all other formations with the same left-hand constituent are retrieved. The plural form of a unit is kept if that is the only word-form in the corpus, e.g. *vitamin supplements* (1). This generates sets of formations with a shared left-hand member (N1-clusters).
- iii) Removal of irrelevant units, e.g. those including numerals (*million pounds*), titles (*Mr Chairman*), proper names (*Bridge Appeal Fund*) or potential synthetic compounds (*adrenalin flow*). The resulting list amounts to 514 entries.
- iv) Semantic analysis based on Levi's (1978) *Recoverably Deletable Predicates* (RDPs). This is a twofold inspection which assigns a Primary RDP for the reading considered most probable based on the unit's occurrence in context, plus a Secondary RDP for cases which could not be disambiguated through the concordance contexts. This model contemplates the following nine predicates (three of which are reversible, differentiated by a subscript <sub>1</sub> or <sub>2</sub>): CAUSE<sub>1</sub>, CAUSE<sub>2</sub>, HAVE<sub>1</sub>, HAVE<sub>2</sub>, MAKE<sub>1</sub>, MAKE<sub>2</sub>, USE, BE, IN, FOR, FROM and ABOUT (see Levi, 1978: 75-118). HAVE is one of the reversible predicates: in *family house* the first constituent possesses the second one ('a house which the family has'), so the suitable analysis would be under HAVE<sub>1</sub>; in *chocolate éclair* the directionality is the opposite ('an éclair which has chocolate'), so the relevant predicate is HAVE<sub>2</sub>. Therefore, in analyzing a unit like *crocodile mug* attention is paid to the co-text in the corpus: "Do you really want to go on using that

mug? [...] I like it. And my crocodile mug as well” and, based on this, the two most natural RDPs seem to be HAVE<sub>2</sub> and BE. Any RDP which does not fit the corpus evidence for a formation’s meaning (even if it *could*) is hence disregarded from this analysis. All further semantic analyses in this article resort to Hansen, Hansen, Neubert, & Schentke (1982).

v) Check-up of attestedness of NN units in the *Oxford English Dictionary* (OED).

Occurrence in the OED was considered positive when a unit was listed with the same meaning as in the BNC Sampler, be it as a lemma or as a sub-entry of another lemma. All NN formations were preserved regardless of their (non-)attestedness in the OED, as this information was later used for further inspection (see Section 3.2).

The above generates a 514-entry record of NN compounds, each accompanied by its corpus frequency, most likely semantic predicate(s) and information of OED attestedness, e.g. *cotton jacket* ( $N=3$ ; RDP1 MAKE<sub>2</sub>; RDP2 HAVE<sub>2</sub>).

### 3. The Misfits of NN Compounding

The approaches to the underlying mechanisms of compounding have been many and heterogeneous in nature, and have developed from virtually all theoretical and applied perspectives (see Libben, 2006; Lieber & Štekauer, 2009 for thorough reviews). This section, which is a compromise between detailedness and moderation, concentrates on the form and meaning (3.1), and paradigmatic nature of NN compounding (3.2).

#### 3.1. The Categorization of Compound Semantics

One central topic in NN compounding is the flexible meaning connection tying its constituents. Presuming that, in a construction (1), speakers are familiarized with the morphosemantic features of X and Y, the goal is to bridge the semantic gap  $\mathfrak{R}$  in order to achieve the meaning Z (see Guevara & Scalise, 2009: 107).

(1) [ X  $\mathfrak{R}$  Y ]Z

Numerous models have been devised which comprise the whole variety of meaning relationships between the head and the modifier<sup>1</sup>. The scope, motivation and purpose of such proposals are diverse, and they range from early models (Bloomfield, 1933; Hatcher, 1960; Jespersen, 1942), to works with a transformational background (Allen, 1978; Levi, 1978; Selkirk, 1982), meaning-oriented approaches (Downing, 1977; Jackendoff, 2009, 2016; Lieber, 2004; Štekauer, 2009, 2016), and recent miscellaneous proposals (Bourque, 2014; Mattiello & Dressler, 2018; Pepper, forthcoming; Schäfer, 2018). The study of compound semantics initially targeted heads (Allen, 1978; Hatcher, 1960; Jespersen, 1942), although it has gradually redefined the role of modifiers, thus disclosing their decisive effect on the meaning of the compound (Baayen, 2010; Breban & Kolkman, 2019; Levin, Glass, & Jurafsky, 2019; Štekauer, 2016; Warren, 1984). On the experimental side, it has been shown that the more compounds that share a given constituent, the more probable it is to have a variety of relations attached to it (Gagné, 2001, 2009; Gagné & Spalding, 2004; Libben, 2006).

A great share of the disagreement in this area has arisen over the granularity of such semantic relations. While some have proposed catalogues of at most ten relations (Hatcher, 1960; Jespersen, 1942), other sets fluctuate between twelve and fifty

(Bourque, 2014; Levi, 1978; Ryder, 1994), and in a few cases the combinability of functions may generate endless readings (Jackendoff, 2016; Lieber, 2004; Štekauer, 2009). Expectedly, the degree of specificity comes at the cost of the number of relations, so the choice is between a lower number of wide-ranging functions, or a higher number of more precise functions.

In the case of Levi (1978), categorial overlap and meaning indeterminacy have been often cited as causing the difficulty in unambiguously assigning RDPs. The following aims at detecting problematic areas of meaning through a twofold semantic inspection which delves into the vagueness caused by the lack of an ACTION. Table 1 shows the number of NN formations under each RDP as a primary and as a secondary reading:

[TABLE 1 ABOUT HERE]

These results point to regions of high and of low output. 100% of the units were assigned a primary reading (because every unit in the study has at least one conceivable meaning; see Section 2) and, more interestingly, 30.73% were assigned a secondary one. Put differently, for 30.73% of the data it was not possible to select just one RDP as representative of the unit's meaning. This value offers numerical proof of the long-discussed semantic indeterminacy of NN formations (Jackendoff, 2016; Maguire et al., 2010; Štekauer, 2005: 241-263). Overall, predicates whose values are high for a primary reading display correspondingly high values for their secondary reading, both concerning high (ABOUT, IN, FOR) and low rates (FROM, CAUSE, USE), though in differing proportions. Exceptions are BE and IN, the former because it is much more dominant as

a primary than as a secondary reading, the latter due to the contrary<sup>2</sup>. Figure 1 replicates the above visually:

[FIGURE 1 ABOUT HERE]

Figure 1 reveals the behavior of each RDP as well as intersections and affinities between semantic categories, where areas of low incidence and areas of high incidence can be distinguished. Among the former there are the predicates CAUSE, FROM and USE, where cross-categorical ambiguity is virtually null, and suggests that the NN units under these three RDPs only rarely have a second possible interpretation. These are precise kinds of predicates whose meaning paraphrases are hardly equivocal, e.g. *traffic noise* can be defined as ‘noise which is caused by traffic’ and is hence assigned the predicate CAUSE. It would seem forced to paraphrase *traffic noise* through any other ancillary reading, e.g. ‘noise which comes from traffic’ (FROM), or ‘traffic which produces noise’ (MAKE<sub>1</sub>). The same reasoning applies to (2)-(4):

			RDP1	RDP2
(2)	(a)	bomb explosions	CAUSE <sub>1</sub>	?
	(b)	traffic noise	CAUSE <sub>1</sub>	?
(3)	(a)	exam results	FROM	?
	(b)	milk product	FROM	?
(4)	(a)	brush strokes	USE	?



(b) keyhole surgery USE ?

As for high incidence, two areas are worth pinpointing in Figure 1. One is found in the semantic macro-region of BE, MAKE and HAVE (top right-hand area), the other in that of IN, ABOUT and FOR (bottom left-hand area). In both, we are faced with units with two possible readings, although the order of the primary and secondary RDP varies for each formation. For example, *cotton jacket* may be analysed as MAKE<sub>2</sub> ('a jacket which is made of cotton') or HAVE<sub>2</sub> ('a jacket which has cotton among other materials'), and *crocodile mug* as HAVE<sub>2</sub> ('a mug which has a crocodile depiction on it') or BE ('a mug which is a crocodile'). The most prominent meaning in this macro-region is composition/constituency: the head of the NN unit is an artifact and the modifier is some material. If the artifact (head) contains exclusively the material denoted by the modifier, then the reading of the unit requires the predicate MAKE (e.g. *rubber gloves* 'gloves which are made of rubber'); if the artifact contains the material denoted by the modifier but also other materials, then the reading of the unit requires the predicate HAVE (e.g. *chocolate éclair* 'éclair which has chocolate, but also milk, flour, egg, butter, etc.'). These features make the notional boundary between BE, MAKE and HAVE blurry, and seem to lay the task of RDP allocation more on the analyst's criterion than on the meaning of the NN units itself.

As Section 2 described, concordances were used for meaning disambiguation, but even this information did not always clear up the unit's definitive reading. As a token, (5) reproduces the three concordance lines for *cotton jacket*, from which it is impossible to gather if the most suitable predicate should be MAKE (i.e. 'a jacket which is made up entirely of cotton') or HAVE (i.e. 'a jacket which is made of various materials, one of

which is cotton’). In these cases, the most logical reading was selected, often by discarding illogical ones, e.g. ‘a jacket which is about cotton’ (ABOUT), ‘a jacket which is placed in cotton’ (IN) or ‘a jacket whose purpose is cotton’ (FOR).

- (5) (a) I could see her chest move in and out inside the light *cotton jacket*  
(b) A light *cotton jacket* hung over the back of her chair  
(c) XLR woven multicoloured *cotton jacket*, £49.99, worn with a black skirt...

The second macro-region in Figure 1 corresponds to ABOUT, FOR and IN, where again two possible readings can be afforded for each unit, e.g. *health association* as ABOUT (‘an association concerned with health’) or FOR (‘an association whose aim is health’), and *school coat* as FOR (‘a coat which is used for school’) or IN (‘a coat which is used while at school’). The main meaning intersections here are topic-purpose, topic-location and location-purpose, and are, like above, not easily told apart based on concordance lines. One example is *army plan*, which stands vaguely between ABOUT (‘a plan which is about the army’) and IN (‘a plan which is prepared/implemented in the army’):

- (6) (a) The *army plan* envisages a corps of about 150,000 US troops...  
(b) US *army plan* to reduce European forces by half

Tables 2 and 3 display formations belonging to the macro-regions BE-MAKE-HAVE and ABOUT-FOR-IN:

[TABLE 2 ABOUT HERE]

[TABLE 3 ABOUT HERE]

In order to evaluate the role of modifiers in the above, a possible correlation between the meaning of the modifier and that of the compound should be considered. Figure 2 shows the distribution of NN units across RDPs and one of five possible meanings for the modifier: Activity, Agent, Location, Manner or Substance.

[FIGURE 2 ABOUT HERE]

Out of the five possible meanings of modifiers, Substance (254) stands out, while Location (119), Activity (102), Agent (35) and Manner (1) follow. The semantics therefore of 50% modifiers in the sample is a Substance, e.g. *almond oil, brake drums, dress party, laser disk, nappy pin, phone bill*, etc. Figure 2 shows that Substance is mostly repeatedly projected onto the predicates ABOUT (*dress party, parachute regiment, seed companies, water company*) and MAKE (*chocolate bar, concrete driveway, cotton cloth, PVC liners, steel rods*). A parallel situation is found for Location and the RDP IN: *bank debt, county road, forest peoples, school lunch, shore crab*, etc. The rest of modifier-RDP connections are less evident, although we may pinpoint that between Activity and USE, and that between Agent and HAVE. The only modifier with the meaning Manner occurs for the predicate FOR (*DIY materials*).

All the above suggests that an analysis which considers NN formations in a vacuum will generally offer a broad picture of their semantic macro-region, but not a definitive

account of their meaning and referent<sup>3</sup>. Use of concordance lines has shown that even the corpus co-text is sometimes insufficient for an entirely reliable RDP allocation. It is not the case that, for instance, *home stories* should be analyzed as either ABOUT or IN; rather, it is that *home stories* potentially means *both* ABOUT and IN, and it is upon actual use that its semantics are narrowed down and a specific and contextual meaning materializes. *Home stories* may be employed meaning ‘stories which are about home’ on one particular occasion, and then meaning ‘stories which happen(ed) at home’ on a different occasion, by the same language user. Let us remember that NN compounding lies among the least transparent word-formation processes in English, and that the lack of an ACTION makes it rather easy to produce NN formations but also complex to interpret them (Štekauer, 2005: 62).

One neglected area has been that of compounds which, being non-lexicalised and consisting of simplex nouns, display a certain degree of figurative meaning. Figurative readings are explicitly left out in some models (Hatcher, 1960; Levi, 1978; Selkirk, 1982), but are in fact present in a good number of formations and may be argued to stand on a par with lexical morphology with good reason (see Lipka, 1990). NN formations with figurative meaning are shown in (7):

- (7) (a) crocodile deity
- (b) mandolin pendant
- (c) ring road

One challenge here is where the limit lies between figurativeness and non-figurativeness. While some NN formations are clearly metaphorical, e.g. (7), the meaning of others is not always readily apparent. Compare (8) and (9):

- (8) (a) family circle  
(b) peak workloads  
(c) seed pearls
- (9) (a) cash prizes  
(b) lager beer  
(c) vitamin supplements

Both sets are open to an interpretation and under Levi's BE with readings such as 'a circle which is the family' or 'supplements which are vitamins'. However, the degree of figurativeness in (8) is absent from (9), as these are formations genuinely falling under BE and retain an original 'essive/appositional' sense. In cases like (8), the load of figurativeness is certainly not carried by the head, whose interpretation is literal, but it is found in the modifier or in  $\mathfrak{R}$ . One problem is that the different nature of (8) and (9) is not captured by BE, and even checking the context does not clarify the degree of literal versus figurative meaning:

- (10) (a) Women are not only responsible for the care, education and development of their children, but also for the maintenance of the *family circle*.  
(b) [...] to work for periods of duty which only cover the *peak workloads*.

- (c) The Queen’s bedcover is of blue silk and quilted with *seed pearls*.

Metonymy and metaphor in word-formation have been tackled largely by onomasiologists (Blank, 1997; Dirven & Verspoor, 1998; Koch, 2001; Štekauer, 2009, 2016), although formal models have also engaged with them (Jackendoff, 2010, 2016; Lieber, 2004, 2016). Jackendoff’s (2010: 438) Parallel Architecture for example introduces the function ‘SIMILAR (X, Y)’ which, for a compound XY, may be glossed as ‘a Y which is in some way similar to an X’. Even if this function does not specify the precise element of comparison/contrast, it allows going beyond the literal interpretation of the compound. (11) is the application of ‘SIMILAR (X, Y)’ to *mandolin pendant* (‘a pendant whose shape is similar to a mandolin’):

- (11)  $\text{mandolin}_1 \text{ pendant}_2 = [\text{PENDANT}_2^\alpha; [\text{SIMILAR}(\alpha, \text{MANDOLIN}_1)]]$

The analysis of *mandolin pendant* under Štekauer’s (2009, 2016) approach is presented in (12). Note that the underlying breakdown is close to Jackendoff’s but, while Štekauer emphasizes the semantic roles of the constituents, Jackendoff focuses on the meaning link  $\mathfrak{R}$ :

- |      |          |                              |         |
|------|----------|------------------------------|---------|
| (12) | mandolin |                              | pendant |
|      | Patt     | <-(State)- <sup>PURP</sup> - | Pat     |
|      | SUBST    |                              | SUBST   |

Interpretation: Patient of explicitly unexpressed State having a similarity to a Pattern

Other cognitive models have explained metaphorical formations through the speakers' knowledge of frames and domains. From such perspectives, figurative formations are interpreted by "[...] cognitive strategies such as knowledge of conceptual frames, command of conceptual blending, and metaphor, and upon our cultural knowledge of our material and non-material environment" (Dirven & Verspoor, 1998: 59; see Blank, 1997; Koch, 2001: 17-22). It seems clear, at any rate, that the semantic versatility of NN compounds transcends verbatim interpretations, and that room should be made for some degree of metaphoricity in models of compound semantics.

### *3.2. The Paradigmaticity of Compounding*

Studying the formal and semantic features triggered by a common modifier or head naturally leads to the notion of paradigm, currently the center of meticulous attention within the morphological community. A number of works are at present exploring the notion of paradigm and, more specifically, whether its relevance for inflection can be extrapolated to derivation (Bauer, 1997, 2019; Bonami & Strnadová, 2018; Boyé & Schalchli, 2016; Štekauer, 2014) and to compounding (Bagasheva, 2014, in press; Radimský, in press).

Complexities arise here because of the formal features of NN compounding. One problem lies in how to bring together a set of NN units under the same paradigm. Given several NN units, which component should fulfill for NN compounding the function that the affix performs for affixation? Mattiello & Dressler for instance stress the role of analogy, in that the inclusion of a compound within a family depends on criteria like a shared constituent, a shared constituent position, a common degree of

opacity/transparency, or identical syntactic and semantic patterns (2018: 68-69; see Bauer, Lieber, & Plag, 2013: 524). Another explanation can be sought in right-headedness, endocentricity, non-argumental nature, etc., but these morphological and semantic features seem too general for these purposes. In line with the recent morpho-semantic tradition, three options emerge as possible *paradigmatic links* for compounding paradigms: the modifier,  $\mathfrak{R}$ , or the head.

Earmarking the modifier as the paradigmatic link of NN compounding amounts to setting up clusters like those in Section 3.1. The aim here is finding common semantic traits in a set of NN units which substantiates their inclusion under the same template. For (13), and based on the schema [ X  $\mathfrak{R}$  Y ]Z, one solution is a coarse-grained connection, such as ‘a Z which is somehow related to funerals’, which makes it possible to embrace any NN formation premodified by *funeral*, thus stressing what the paradigm members share in order not to exclude any of them:

- (13) (a) funeral account  
(b) funeral ceremonies  
(c) funeral cortege  
(d) funeral expenses  
(e) funeral procession  
(f) funeral service

The second possible paradigmatic link is  $\mathfrak{R}$ . Deciding on  $\mathfrak{R}$  involves bestowing the paradigm’s cohesion on a purely semantic component, which may challenge the long-established assumption that both formal and semantic change must occur from input to



output in word-formation; what Kastovsky calls “morphological-semantic type[s]” (1986: 597; see Schmid, 2015; Štekauer, 2009). While it seems evident that the mere concatenation of two (or more) nouns counts as formal modification, cases of apparent internal inflection may challenge this assumption, e.g. genitives compounds (Bauer, 1998: 78; Bauer & Tarasova, 2010), or units with pluralized left-hand members (Bauer et al., 2013: 443). The second requisite, that the meanings of the input and output lexemes should diverge, is fulfilled if we understand that *funeral service* or *mandolin pendant* are hyponyms of *service* and *pendant*, respectively.

The role of paradigmatic link may finally be allotted to the head, which results in paradigms like (14), for *shirt*, and (15), for *friend*:

- (14) (a) boy’s shirt  
(b) cotton shirt  
(c) country shirt  
(d) men’s shirt  
(e) sweat shirt

- (15) (a) animal friend  
(b) business friend  
(c) childhood friend  
(d) gardening friend  
(e) lady friend  
(f) pen friend  
(g) school friend

(h) women friends

Semantically speaking, the units in these paradigms are co-hyponyms. Co-hyponymy has been classically allocated to primary NN compounding, and is recurrently cited as inherent to lexemes created by this process (Allen, 1978; Ryder, 1994).

The question is which of these three possible paradigmatic links (modifier,  $\mathfrak{R}$ , head) most suitably contributes to a coherent notion of paradigm of NN compounds. While the three imply broad morphosemantic features and generalisations, each offers valuable insights for a synchronic analysis of NN compounding. The paradigmaticity of a process should point out the commonalities of its members as well as afford a wide view of word-formation by stressing the “correlation between potential, possible and actual words” (Bagasheva, in press). This seems good reason to embrace the three above features as variables in the paradigmaticity of NN compounding, although in different degrees. This said, my view is that the decisive bond in NN formations is  $\mathfrak{R}$ , as it connects the compatible semantic features of the two compound constituents, along the lines of Allen’s (1978: 92-94) *Variable R Condition*. With  $\mathfrak{R}$  fixed as the paradigmatic link, the modifier and the head can play a role in fine-tuning the nature of each paradigm, but they will be subordinated to the logical-semantic connection between them.

It is also possible to think, like Booij (2010), in terms of schemas where one of the constituents is lexically specified, this fixed constituent being the modifier (*funeral account, funeral ceremonies, funeral cortege, etc.*) or the head (*animal friend, business friend, childhood friend, etc.*). In that connection, Bauer et al. (2019: 51-53) speak of productivity rates for specific compound schemas, thus offering the opportunity for

computations of broad semantic categories (see Section 3.1). One application of such compound schemas is the direct comparison of productivity rates for the same noun in head versus modifier position, an otherwise tricky task<sup>4</sup>. Four nouns have been here selected for a pilot test: *army*, *client*, *crocodile*, and *water*. These are representative of the corpus data in including nouns which are frequent modifiers (*army*, *water*), infrequent modifiers (*forest*, *life*), as well as assorted semantic roles like abstract notion (*life*), location (*forest*), material (*water*), personal/entity (*army*). For each noun, the type frequency ( $V$ ), token frequency ( $N$ ) and hapax legomena ( $n_1$ ) have been retrieved in modifier (Mod) and head (H) position, and are shown in Table 4:

[TABLE 4 ABOUT HERE]

The values point to a numerical dominance of modifier clusters, as their frequency figures mostly outnumber those of head clusters in the cases of *army*, *forest* and *water*. This in principle indicates that modifier clusters have a higher lexical richness and productivity potential than head clusters, which becomes especially manifest in  $n_1$  data (hapaxes are taken to reveal high productivity; see Baayen, 2009: 903-904).

Raw corpus values have been largely processed for the measurement and assessment of productivity values, traditionally in affixational processes (see Bauer et al., 2019; Plag, 1999). These frequencies have been here combined towards the productivity ratings in Table 5, where schema-based paradigms open up a possibility for the application of productivity models, in this case the type-token ratio ( $V/N$ ), and Baayen's *productivity in the narrow sense* ( $P$ ) and *hapax-conditioned degree of productivity* ( $P^*$ ).  $P$  measures the probability of finding new coinages and employs the formula  $P = n_1 / N$ ,

while  $P^*$  computes the contribution of a given process to all the units of its lexical category through the formula  $P^* = n_1 / h_t$  (where  $h_t$  = total number of hapaxes in a given category). The  $V/N$  ratio is intended to quantify lexical richness more than morphological productivity, but can be used in conjunction with  $P$  and  $P^*$  for an overview of frequency tendencies. The models' estimates all range between 0 and 1 (see Baayen, 2009 for the rationale behind these formulae):

[TABLE 5 ABOUT HERE]

Globally speaking, *forest* displays the highest productivity rates (it ranks first in four out of the six measures), although this must be contextualized because of its extremely low frequencies (see Table 4), which may cause the models to malfunction. The results can be interpreted in at least two ways: by focusing on modifier versus head values, the outcome is almost of total balance, with the head values dominating in  $V/N$  and in  $P$  (3 nouns to 1), and the modifier values being dominant in  $P^*$  (3 nouns to 1). The second option pays attention to the results for each noun individually, in which case the head position turns out to be more productive for all four nouns (and also for the totals in the last row). Remarkably, the modifier position appears as most productive according to  $P^*$ , in theory the most robust formula, and the one which is entirely based on hapaxes. The interpretation must be a prudent one owing to corpus size (2 million words is deemed small for Baayen's computations), but makes it clear that schema-based measurements are feasible and offer a new path to the study of productivity in compounding (see Baayen, 2010; Maguire et al., 2010).

Productivity has been assessed also through lexicographical tools, although such results must be relativized due to the very nature and purposes of dictionary compilation (Baayen, 2009: 910-911; Plag, 1999: 96-98). Figure 3 displays the number of attested and unattested NN formations in the OED for each predicate:

[FIGURE 3 ABOUT HERE]

Figure 3 evidences almost complete homogeneity of attestedness across categories, as around 50% of the units in every predicate have been located in the OED. This is so irrespective of the predicates' semantic vagueness (Table 1) and frequency values. One largely discussed matter is the connection between language frequency and listedness in the lexicon which, being challenging for derivational processes in general, seems to be particularly complex for NN compounding due to its extremely high productivity (Baayen, 2010; Bauer, 1998; Gagné, 2001; Jackendoff, 2016). Table 6 summarizes the information about OED attestation and corpus frequencies, here sorted for three frequency ranges: units with frequency 1, units with frequency 2 and units with frequency 3 or higher:

[TABLE 6 ABOUT HERE]

The highest rate of unattested units (62.46%) is found among hapaxes, as could be expected from the volatile nature that characterizes frequency-1 formations. This is the case of units across all modifier clusters and semantic categories, e.g. *army radio* (IN), *bank debt* (ABOUT), *family reasons* (ABOUT), *poll result* (CAUSE), *steel pipe* (MAKE) or

*telegraph system* (FOR). There are however hapaxes which do occur in the OED (37.53%), and these are present in all RDPs as well, e.g. *almond oil* (MAKE), *axe head* (HAVE), *defence team* (FOR), *family life* (IN), *information room* (FOR) or *nursery school* (ABOUT). Various cases can be found of semantically analogous formations where one variant is listed in the OED but the other one is not: *army officer* (unattested) and *army captain* (attested), *bank debt* (unattested) and *bank loan* (attested), and *health department* (unattested), or *health centre* (attested). The opposite is true for units with a frequency of 3 or higher, most of which are attested in the OED: *county council* (43), *poll tax* (16), *client server* (12), *phone number* (11), *diamond ring* (9), *information system* (8), *motor car* (7), etc. Exceptions to this are *health authorities* (26), *seed companies* (5), *defence budget* (5), *mandolin pendant* (4), and *county councilor* (4), among others.

The above has offered an empirical background for the postulation of paradigms of NN compounds by focusing on sets of analogous formations and by calculating their degree of productivity in various ways. Three possible paradigmatic links (head, modifier,  $\mathfrak{R}$ ) have offered reasonable evidence for existence of such paradigms, although it has been shown that these techniques must be taken with caution in that they exploit corpus and lexicographical data, but do not *per se* look at the underlying semantics of nouns in modifier or head position. It can be maintained that, besides frequency information, the semantic versatility of nouns in modifier versus head position is a significant factor which should not be neglected in the shaping of compound paradigms.

#### 4. Conclusions

Using corpus and lexicographical information, this article has tackled phenomena relevant to NN compounding, in relation to its loose semantic potential (Bauer & Tarasova, 2010), high productivity degrees (Bauer et al., 2019; Maguire et al., 2010) and paradigmatic aspects (Bagasheva, in press; Boyé & Schalchli, 2016). In keeping with its objectives, the paper's findings can be summarized as follows:

- i) The potential meanings of a NN compound remain latent until one of them is narrowed down by use of the unit in context, at which point the pragmatic-semantic nuances are marked out for disambiguation. An analysis via concordances has been shown to reduce the degree of vagueness, as opposed to context-free explorations, where meaning remains fuzzy (see Figure 1). As in other works on compound semantics, the role of the situational context must be emphasized for a complete appreciation of the referent (Downing, 1977; Gagné & Spalding, 2004; Koch, 2001; Štekauer, 2009; Levin et al., 2019). Pragmatics, in this sense, is reaffirmed as the *deus ex machina* of meaning interpretation in NN compounding.
- ii) There exist semantic macro-regions in NN compounds which stand superordinate over traditional semantic relations, encompass broad kinds of meanings, and display sections of overlap and sections of strict separation. This argument has been illustrated by way of Levi (1978), but other meaning catalogues are expected to exhibit similar kinds of macro-regions. The nature of such regions takes us back to the unresolved relation between granularity levels in word-formation semantics, metonymy and lexical semantics, which is the focus of a number of works, recent and in the pipeline (Bagasheva, 2014; Bourque, 2014: 147-216; Mattiello & Dressler, 2018; Pepper, forthcoming).

iii) The structuring of NN compounds into paradigms has arisen as a promising path for the organization and feeding of the lexicon (see Figure 3 and Table 6 for the relationship between low-frequency units and OED attestedness). This paradigmaticity is substantiated by the contribution of the head, the modifier and  $\mathfrak{R}$ , which bind NN units together, formally and semantically. In particular, that the meaning in  $\mathfrak{R}$  is not formalized entails not only not a weakness, but a boost for the creation of NN units, since “[o]mitting markers is positively *non sequitur*” (Beard, 1995: 51).

The above does not exhaust the unanswered questions on nominal compounding. Issues that deserve further attention include NN synthetic compounds whose head is derived by conversion (Bauer et al., 2013: 482; Jackendoff, 2016: 25-26), or an adequate inspection of how the customary non-modal, non-tensed semantic reading of NN formations favors some meanings over others (Zimmer, 1971).

#### List of abbreviations

H	Head
$h_t$	Total number of hapaxes in a given category
Mod	Modifier
$N$	Token frequency
$n_1$	Hapax legomena
OED	<i>Oxford English Dictionary</i>
$P$	Productivity in the narrow sense



$P^*$	Hapax-conditioned degree of productivity
RDP	Recoverably Deletable Predicate
$V$	Type frequency

## References

- Allen, M. R. (1978). *Morphological investigations*. Doctoral dissertation, University of Connecticut, Storrs, CT.
- Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling, & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp.899-919). Berlin: Mouton de Gruyter.
- Baayen, R. H. (2010). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. *Linguistische Berichte Sonderheft, 17*, 383-402.
- Bagasheva, A. (2014). Frame semantics, metaphonymy and compound verbs in English. In G. Rundblad, A. Tytus, O. Knapton, & C. Tang (Eds.), *Selected papers from the 4th UK Cognitive Linguistics conference* (pp.1-17). London: UK Cognitive Linguistics Association.
- Bagasheva, A. (in press). Paradigmaticity in word-formation and compounds: Where from and where to? In J. Fernández-Domínguez, A. Bagasheva & C. Lara-Clares (Eds.), *Paradigmatic relations in word-formation*. Leiden: Brill.
- Bauer, L. (1997). Derivational paradigms. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp.243-256). Dordrecht: Kluwer.
- Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics, 2*, 65-86.
- Bauer, L. (2019). Notions of paradigm and their value in word-formation. *Word Structure, 12*(2), 153-175.

- Bauer, L., & Tarasova, E. (2010). The meaning link in nominal compounds. *SKASE Journal of Theoretical Linguistics*, 10, 1-18.
- Bauer, L., Beliaeva, N., & Tarasova, E. (2019). Recalibrating productivity: Factors involved. *Zeitschrift für Wortbildung / Journal of Word Formation*, 3(1), 44-80.
- Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Beard, R. (1995). *Lexeme-morpheme base morphology: A general theory of inflection and word formation*. Albany, NY: SUNY Press.
- Blank, A. (1997). Outlines of a cognitive approach to word-formation. In *Proceedings of the 16th International Congress of Linguists*. Paper No. 0291. Oxford, U.K.: Pergamon. *romanischen Sprachen*. Tübingen: Niemeyer.
- Bloomfield, L. (1933). *Language*. Chicago, IL: The University of Chicago Press.
- Bonami, O., & Strnadová, J. (2018). Paradigm structure and predictability in derivational morphology. *Morphology*, 29(2), 167-197.
- Booij, G. E. (2010). Compound construction: Schemas or analogy? A Construction Morphology perspective. In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding* (pp.93-108). Amsterdam & Philadelphia, PA: John Benjamins.
- Bourque, Y. S. (2014). *Toward a typology of semantic transparency: The case of French compounds*. University of Toronto PhD dissertation.
- Boyé, G., & Schalchli, G. (2016). The status of paradigms. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp.206-234). Cambridge: Cambridge University Press.
- Breban, T., & Kolkmann, J. (2019) (Eds.). *English Language and Linguistics*, Volume 23, No. 4 - Special issue: Different perspectives on proper noun modifiers.

- Dirven, R., & Verspoor, M. (1998). *Cognitive exploration of language and linguistics*. Amsterdam: John Benjamins.
- Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 53, 810-842.
- Fernández-Domínguez, J. (2009). *Productivity in English word-formation. An approach to N+N compounding*. Bern: Peter Lang.
- Gagné, C. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(1), 236-254.
- Gagné, C. (2009). Psycholinguistic perspectives. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp.255-271). Oxford: Oxford University Press.
- Gagné, C., & Spalding, T. (2004). Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain & Language*, 90(1-3), 478-486.
- Guevara, E., & Scalise, S. (2009). Searching for universals in compounding. In S. Scalise, E. Magni & A. Bisetto (Eds.), *Universals of language today* (pp.101-128). Dordrecht: Springer.
- Hansen, B., Hansen, K., Neubert, A., & Schentke, M. (1982). *Englische Lexikologie. Einführung in Wortbildung und lexikalische Semantik*. Leipzig: VEB Verlag Enzyklopädie.
- Hatcher, A. G. (1960). An introduction to the analysis of English noun compounds. *Word*, 16, 356-373.
- Jackendoff, R. (2009). Compounding in the Parallel Architecture and Conceptual Semantics. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp.105-129). Oxford: Oxford University Press.

- Jackendoff, R. (2010). *Meaning and the lexicon: The Parallel Architecture, 1975–2010*. Oxford: Oxford University Press.
- Jackendoff, R. (2016). English noun-noun compounds in Conceptual Semantics. In P. ten Hacken (Ed.), *The semantics of compounding* (pp.15-37). Cambridge: Cambridge University Press.
- Jespersen, O. (1942). *A Modern English grammar on historical principles. Part VI, morphology*. London and Copenhagen: Munksgaard.
- Kastovsky, D. (1986). The problem of productivity in word formation. *Linguistics*, 24, 585-600.
- Koch, P. (2001). Bedeutungswandel und Bezeichnungswandel: Von der kognitiven Semasiologie zur kognitiven Onomasiologie. *Zeitschrift für Literaturwissenschaft und Linguistik*, 121, 7-36.
- Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York, NY: Academic Press.
- Levin, B., Glass, L., & Jurafsky, D. (2019). Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds. *Linguistics* 57, 3, 429-471.
- Libben, G. (2006). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp.3-22). Cambridge: Cambridge University Press.
- Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge: Cambridge University Press.
- Lieber, R. (2016). *English nouns. The ecology of nominalization*. Cambridge: Cambridge University Press.

- Lieber, R., & Štekauer, P. (2009). *The Oxford handbook of compounding*. Oxford: Oxford University Press.
- Lipka, L. (1990). Metaphor and metonymy as productive processes on the level of the lexicon. In W. Bahner, J. Schildt, & D. Viehweger (Eds.), *Proceedings of the Fourteenth International Congress of Linguists* (pp.1207-1210). Berlin: Akademie-Verlag.
- Maguire, P., Wisniewski, E. J., & Storms, G. (2010). A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory*, 6(1), 49-73.
- Mattiello, E., & Dressler, W. U. (2018). The morphosemantic transparency/opacity of novel English analogical compounds and compound families. *Studia Anglia Posnaniensia*, 53, 67-114.
- Oxford English Dictionary*. Available at: <http://www.oed.com>. Last accessed 13 February 2020.
- Pepper, S. (forthcoming). *The typology of binominal lexemes. Noun-noun compounds and their functional equivalents*. Doctoral dissertation, University of Oslo.
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Berlin & New York, NY: Mouton de Gruyter.
- Radimský, J. (in press). A paradigmatic approach to compounding. In J. Fernández-Domínguez, A. Bagasheva & C. Lara-Clares (Eds.), *Paradigmatic relations in word-formation*. Leiden: Brill.
- Ryder, M. E. (1994). *Ordered chaos: The interpretation of English noun-noun compounds*. Berkeley, CA: University of California Press.
- Schäfer, M. (2018). *The semantic transparency of English compound nouns*. Berlin: Language Science Press.

- Schmid, H.-J. (2015). The scope of word-formation research. In P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (Eds.), *Word-formation. An international handbook of the languages of Europe* (vol. I, pp.1-21). Berlin & Boston, MA: De Gruyter.
- Scott, M. (2012). *WordSmith Tools version 6*. Stroud: Lexical Analysis Software.
- Selkirk, E. O. (1982). *The syntax of words*. Cambridge, MA: MIT Press.
- Štekauer, P. (2005). *Meaning predictability in word formation. Novel, context-free naming units*. Berlin & New York, NY: John Benjamins.
- Štekauer, P. (2009). Meaning predictability of novel context-free compounds. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (pp.272-297). Oxford: Oxford University Press.
- Štekauer, P. (2014). Derivational paradigms. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology* (pp.354-369). Oxford: Oxford University Press.
- Štekauer, P. (2016). Compounding from an onomasiological perspective. In P. ten Hacken (Ed.), *The semantics of compounding* (pp.54-68). Cambridge: Cambridge University Press.
- Warren, B. (1984). The functions of modifiers of nouns. *Quaderni di Semantica*, 5(1), 111-122.
- Zimmer, K. E. (1971). Some general observations about nominal compounds. *Working Papers on Language Universals*, 5, 1-21.

## Notes

---

<sup>1</sup> The terminology in this field has referred to  $\mathfrak{R}$  variously as *type* (Jespersen 1942, Hatcher 1960), *predicate* (Levi 1978), *class* (Warren 1978), *relation* (Downing 1977, Bourque 2014), or *function* (Jackendoff 2009). We hereafter speak of *predicate* and *relation*, with no underlying implication.

<sup>2</sup> Levi (1978) situates FOR, IN and HAVE<sub>1</sub> as the most productive RDPs, and CAUSE, MAKE, FROM and HAVE<sub>2</sub> as the least productive. Such view is overall in agreement with this experiment, although it should be noted that Levi’s approach “[...] incorporates neither the maximal nor the minimal degree of generalization possible but rather an optimal degree [of RDP application]” (1978: 85).

<sup>3</sup> See Štekauer (2005) for experimental work on the interpretation of context-free naming units.

<sup>4</sup>  $\mathfrak{R}$ -based productivity computations have been carried out for example in Fernández-Domínguez (2009), where several productivity models are operated on Levi’s (1978) RDPs. The results evidence that cross-categorial overlap and inaccurate values are common across RDPs, which is why an  $\mathfrak{R}$ -based alternative has been disregarded for the present approach.

## Tables and figures

### Table 1

#### *Results of Primary and Secondary RDPs*



	Primary RDP	Secondary RDP
BE	42	3
MAKE	42	7
HAVE	60	18
CAUSE	12	1
FOR	73	62
ABOUT	158	32
IN	114	31
FROM	2	4
USE	11	-
Total	514	158

Table 2

*Macro-region BE-MAKE-HAVE*

Unit	Primary RDP	Secondary RDP
cotton cloth	MAKE <sub>2</sub>	HAVE <sub>2</sub>

cotton jacket	MAKE <sub>2</sub>	HAVE <sub>2</sub>
cotton trousers	MAKE <sub>2</sub>	HAVE <sub>2</sub>
strawberry jam	MAKE <sub>2</sub>	HAVE <sub>2</sub>
crocodile mug	HAVE <sub>2</sub>	BE
diamond multichip	HAVE <sub>2</sub>	BE

Table 3

*Macro-region ABOUT-FOR-IN*

	Primary RDP	Secondary RDP
business report	ABOUT	FOR
health association	ABOUT	FOR
information office	ABOUT	FOR
army plan	ABOUT	IN
county chairman	ABOUT	IN
home stories	ABOUT	IN
defence industries	FOR	ABOUT
health spa	FOR	ABOUT
school coat	FOR	IN
army newspaper	IN	ABOUT
family illness	IN	ABOUT

Table 4

*Corpus frequencies for nouns in modifier and head position*

	<i>V</i>	<i>N</i>	<i>n</i> <sub>1</sub>

	Mod	H	Mod	H	Mod	H
<i>army</i>	20	4	38	5	13	3
<i>forest</i>	5	3	6	3	4	3
<i>life</i>	7	15	13	38	3	10
<i>water</i>	33	10	64	18	21	7

Table 5

*Productivity models for nouns in modifier and head position*

	<i>V/N</i>		<i>P</i>		<i>P*</i>	
	Mod	H	Mod	H	Mod	H
<i>army</i>	0.526	0.8	0.342	0.6	0.035	0.008
<i>forest</i>	0.833	1	0.666	1	0.010	0.008
<i>life</i>	0.538	0.394	0.230	0.263	0.008	0.027
<i>water</i>	0.515	0.555	0.328	0.388	0.057	0.019
Total	0.603	0.687	0.391	0.563	0.028	0.015

Table 6

*Attested and unattested units per frequency range*

Frequency	Unattested	Attested
-----------	------------	----------

---

1	228	62.46%	137	37.53%
2	36	45%	44	55%
3+	17	25.75%	49	74.24%
Total	281	54.99%	230	45.01%

---

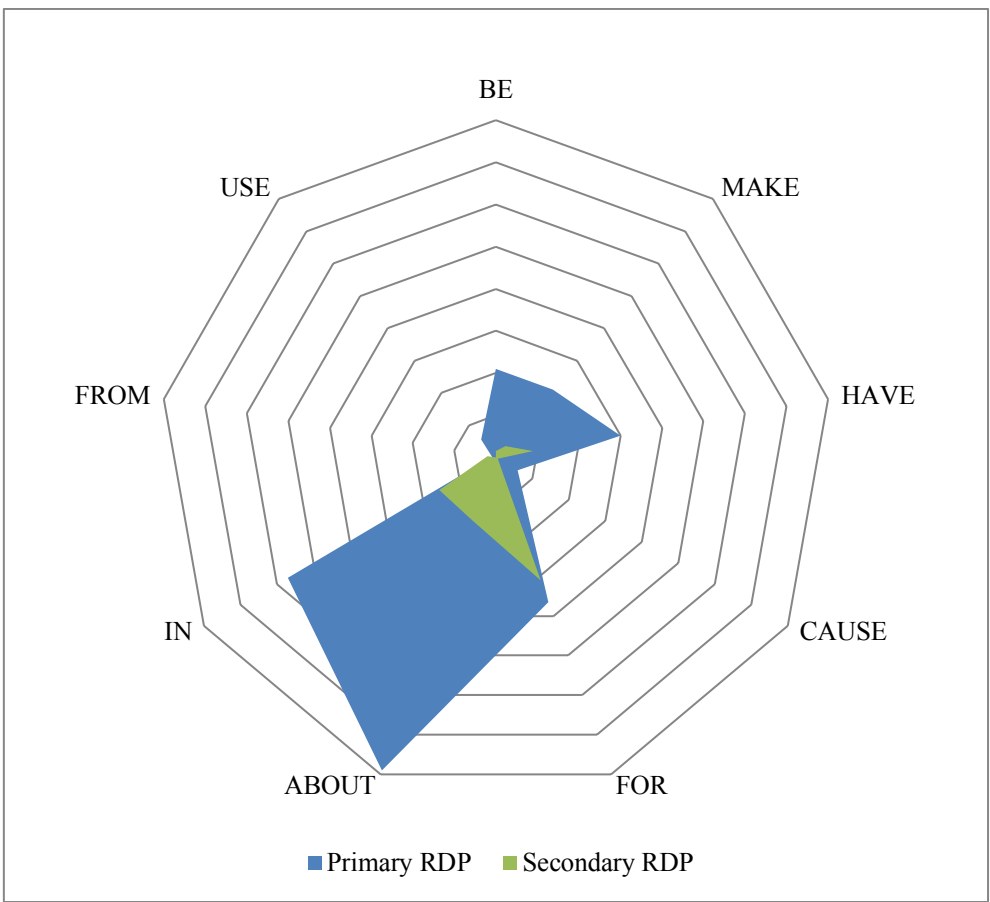


Figure 1. Meaning areas for Primary and Secondary RDPs

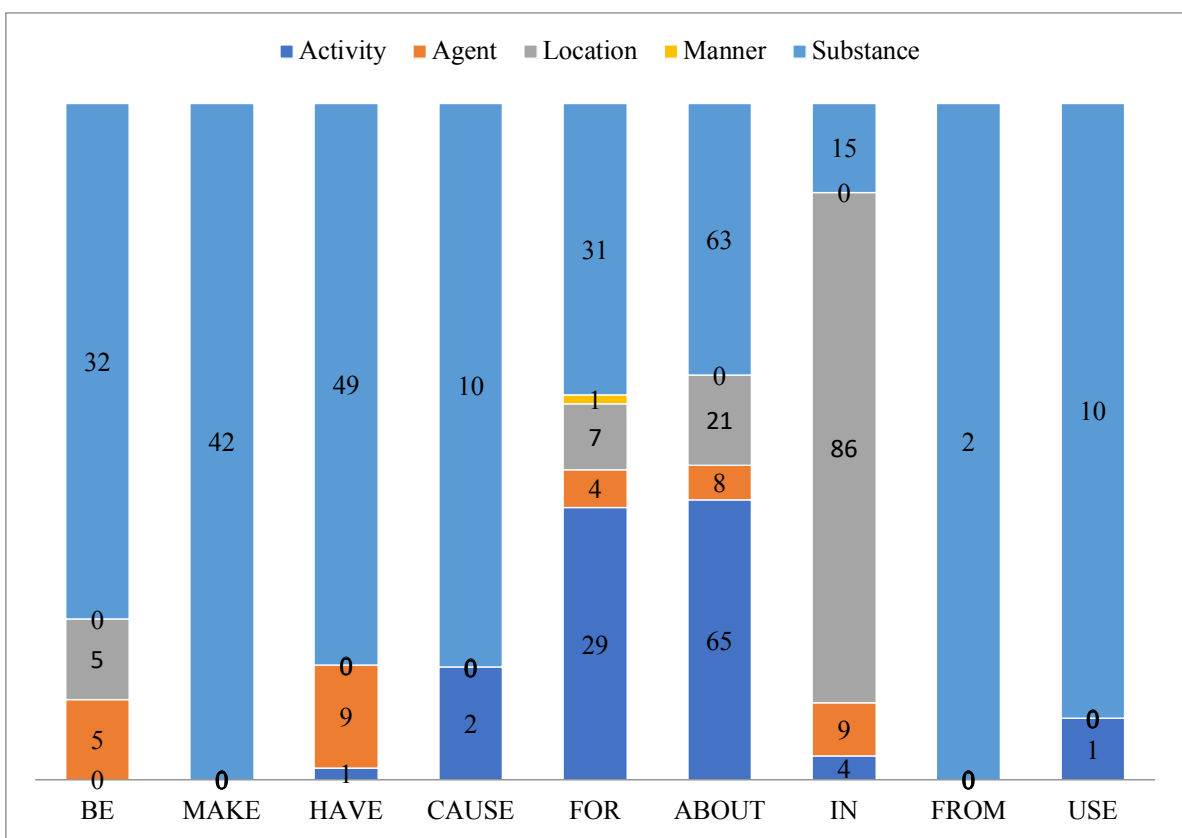


Figure 2. Semantic role of modifiers per RDP

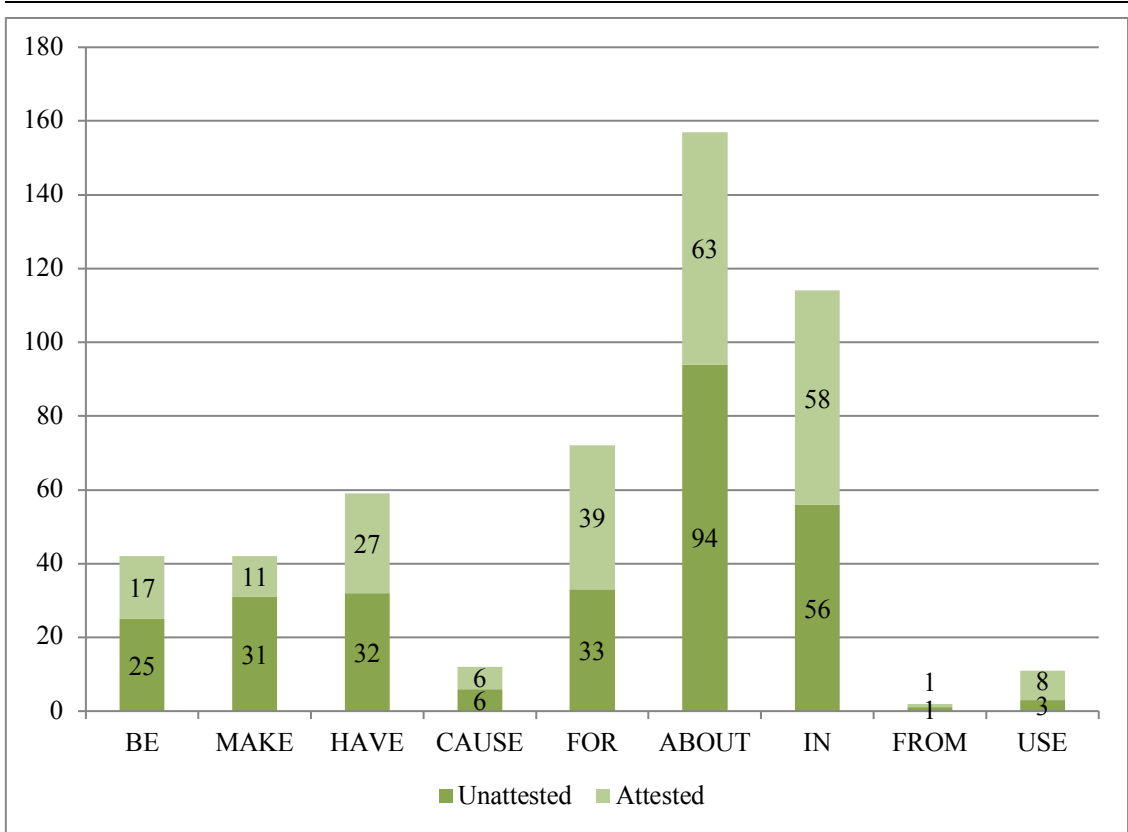


Figure 3. Attested and unattested units (OED) per RDP