The "have-it-that" construction: A corpus-based analysis[*]

Pedro Ureña Gómez Moreno

University of Granada

Speakers do not always attribute agency straightforwardly when they communicate. While complying with the maxims of explicitness and relevance, they may depict states of affairs headed by an identifiable source. More often than not, however, it seems they leave out this source for a number of reasons and through different mechanisms. This paper is a corpus-based study of one such non-identifying structures, namely the extrapositional *have-it-that* construction, in examples such as *Several hypotheses have it that land-use changes*. Drawing on data from the BNC, this paper investigates the use, distribution and functioning of the *have-it-that* construction. The paper also highlights the usefulness of simple collexeme analysis in revealing systematic co-selection relationships within the construction.

Keywords: extraposition, collexeme analysis, construction, evaluation

1. Introduction

'Extraposition' can be generally defined as the process by which a clausal element is allocated in final position instead of the position that would be expected from the standard rules of sentence structure and word order in a specific language (cf. Herriman 2000, Seppänen & Herriman 2002). In order to define this linguistic mechanism more, a distinction is often made between 'subject extraposition' (SUBEX), as shown in Example (1), and 'object extraposition' (OBEX), illustrated in Example (2) (Quirk et al. 1985, Biber et al. 1999, Huddleston & Pullum 2002, among others):

(1) It doesn't matter what you say. (Quirk et al. 1985: 89)
(2) I find it strange that no one noticed the error   . (Huddleston & Pullum 2002: 67)

Syntactically, Example (1) sentences are made up of two subjects: the introductory subject, realised by the expletive *it*, and the postponed subject, realised by a clause. The same type of element duplicity holds for cases like Example (2), that is, the object position is occupied by *it*, whilst the postponed or extraposed object appears in sentence final position − providing that no other clausal elements occur in end position.

This paper is concerned with cases of OBEX of the type shown in Example (2), and, in particular, with examples like (3), in which the main predicator is realised by the verb *have*. In the remainder, I will refer to this latter type of *have*-sentence as the *have-it-that* construction (HITC):

(3) Legend has it that Rhodes, largest of the Dodecanese islands set in the blue
Aegean Sea, was home to the sun god Helios. (AMW 1110)[1]

This paper presents new research insofar as it deals with the HITC, an interesting case of vacuous OBEX bearing specific collocational properties that no other similar extrapositional cases present. A central idea is that the HITC meets some of the main defining features of constructions as these are postulated in the constructionist approach to language (Lakoff 1987, Fillmore et al. 1988, Goldberg 1995, Gonzálvez-García & Butler 2006, Ruiz de Mendoza & Mairal 2008; see Hoffmann & Trousdale 2013 for a comprehensive state-of-the-art work on Construction Grammar) at least as far as the following two aspects are concerned. Firstly, the HITC is a non-compositional conventionalised pairing of form and meaning, i.e. its semantic and pragmatic function cannot be predicted from its constituent parts. In this regard, I posit that the meaning of the construction is related to attribution. Secondly, the HITC has an identifiable pattern made up of both variable and invariable elements. Following Wulff et al. (2007), I contend that the patterned co-selection of lexical units to the subject slot of the HITC constitutes a fundamental aspect in the definition of this structure as a prototypical construction.

This paper is structured as follows. Section 2 reviews current literature on extraposition. Section 3 deals with data collection and description, and the method followed to conduct the research. Section 4 describes the grammatical properties of the HITC and its constituent parts. Section 5 introduces distributional data and gives an overview of the occurrence of the construction in the corpus sample. Section 6 illustrates the coherent mechanism of subject co-selection within the HITC. The results

reported derive from the application of simple collexeme analysis. Finally, Section 7 discusses the main semantic and pragmatic aspects of the construction.

2. Review of previous research on extraposition

Extraposition has received close attention in prior literature, such that its study has diversified at least in the following intertwined subareas of interest. An important area of research revolves around the most problematic aspects of so-called "anticipatory" *it*, one of the main components of extrapositional structures. In this regard, Seppänen (2002) argues for the semantic emptiness of *it*, while Kaltenböck (2002, 2003) regards *it* as an element with semantic content. Concerning the study of the factors originating and constraining extrapositional processes, it has been claimed that elements take extraposed final position when they are syntactically very complex, in which position they are processed and understood more easily both by the speaker and the addressee (Gómez-González 1997, Hewings & Hewings 2002, Fitriati 2006). Additionally, extraposition is widely considered a discourse-controlled mechanism by which the speaker can assign informative focus or thematic saliency to either the subject or the object by placing any of these elements at sentence-final position (e.g. Quirk et al. 1985, Collins 1994). As far as frequency is concerned, extraposition can be viewed as the default choice, while non-extraposed sequences represent the less frequent choice of the two (Mair 1990, Esser 2002, Collins 1994, Kaltenböck 2004).

While SUBEX has been the subject of much analysis, OBEX has been less studied. Regarding its underlying syntactic structure, Seppänen et al. (1995:13) argue that OBEX amounts to basically two types of sentence, complex transitive (S + P + O $_{\text{Expletive}}$ + C + O)[2], as in Example (4), and sentences containing a post-object adverbial (S + P + O $_{\text{Expletive}}$ + A + O), as in Example (5):

(4) They found it very strange that Mary had refused to help. (Seppänen et al. 1995: 13)

(5)   He got it into his head that the car had to be sold. (Seppänen et al. 1995: 13)

Further work by Kim & Sag (2005) focuses on the study of the verbs allowing postponement of the object. These authors posit a triple distinction between obligatory,

optional and questionable OBEX, as shown in Examples (6), (7) and (8), respectively. The HITC is classified within the first of these:

(6)   I blame it on you that we can't go.

(7)   Nobody expected (it) of you that you could be so cruel.

(8)   John thought ?(it) to himself that we had betrayed him. (Kim & Sag 2005:194)

Within the scope of OBEX, a final distinction must be established between 'non-vacuous' OBEX, on the one hand, and 'vacuous' OBEX, on the other. As attested, in Kiparsky & Kiparsky (1971), and Bergh (1997), one of the earliest mentions of vacuous extraposition appeared in transformational work by Rosenbaum (1967). The notion of vacuity in all these authors accounts for sentences in which the anticipatory and the postponed objects co-occur in close proximity in the sentence, rather than separated, as in Example (9). In contrast, as assumed from standard non-vacuous cases, OBEX involves morphosyntactic separation between the two objects, as shown in Example (10):

(9)   I love it that you've asked me to go away. (CEF 739)

(10) He had never made it his business to enquire into the ramifications of his family. (HGV 4754)

One interesting feature of OBEX is that, while in non-vacuous cases like Example (10) extraposition is mainly explained as the result of the speaker trying to avoid a heavy object (e.g. *to enquire into the ramifications of his family*) displacing a light object complement (e.g. *his business*), in vacuous cases "end-weight" does not satisfactorily account for extraposition. In the latter case, there is no evident reason why the expletive is required or why the object does not appear immediately after the predicator, considering there is no clause element like *his business* in Example (10) that can be misplaced (see Miller 2001 for discussion).

Vacuous OBEX, of which the HITC as instantiated in Example (3) is a case in point, is particularly rare in English, with not many verbs either accepting or otherwise motivating a structure of this type. Indeed, Bergh (1997) provides a short list of all the verbs that can appear as the predicator of a vacuous OBEX. This list distinguishes between prepositional verbs (*rely upon*, *see to*, *depend upon*, *get over*), regular verbs

(*regret*, *dislike*, *resent*, *deplore*) and verbs, such as *have*, *take*, *put*, which change their core meaning when followed by expletive *it*. Examples (11), (12) and (13) are corresponding instances (Bergh 1997: 37):

(11) I rely upon it that John will resign.

(12) I regret it that John will resign.

(13) I take it that John will resign.

Section 3 presents the methodology followed in this study with a view to providing further insights into the grammatical and semantic facets of the HITC. The remainder of this paper addresses two main questions: (i) what is the form, distribution and meaning of the HITC as observed in the corpus? (ii) what are the main collostructional factors to consider the HITC as a construction?

3. Methodology

The following subsections present the strategies for data collection and analysis followed for the study of the HITC. Section 3.1 presents the corpus and describes the resources used for both the collection and the analysis of the sample. Section 3.2 defines the different stages of this analysis, including the initial retrieval of examples, the filtering process and the statistical study.

3.1 Data collection and description

The study of the HITC presented below is based on the analysis of 6,418 examples drawn from the British National Corpus (BNC). The BNC is made up of approximately 100 million words from both written (90%) and spoken materials (10%), and gathers language from a variety of sources, such as newspapers, academic documents or conversations.

To collect the data for the study, the corpus was accessed through the CQP edition of the BNCweb interface[3] because of the potential of this on-line application for loading, storing and easily handling corpus data (Hoffmann & Evert 2002). In addition to more general tools for corpus management, the BNCweb also has tools for headword and lemma search, as well as for part-of-speech search. Another feature of the interface is that it enables to obtain frequency distribution of queries based on the genre categories in the BNC as well as the so-called "derived text type" scheme (Hoffman et al. 2008). This scheme is a classification comprising six major categories under which specific sub-genres are grouped (cf. Lee 2001).

In order to retrieve all HITC cases from the BNC, two formal criteria were considered: (i) the invariable part of the construction, i.e. the expletive *it*; (ii) the variable part, which consists of the lemma *have* in the form of *have*, *has*, *had* or *having*, and the word *that*, which can be omitted (see Section 4). Thus, I ran a query of the sequence *{have} it*, where the curly brackets and the blank space at the end of the sequence allowed to retrieve all word types of the lexeme *have* and all HITC examples with elliptical *that*, respectively. In fact, the selected query criterion yielded the widest range of examples including the HITC from the BNC. Subsection 3.2 describes in more detail the procedure that followed the initial collection of data.

3.2 Procedure

The analysis of the corpus examples comprised four phases: (i) data filtering, (ii) data classification, (iii) distribution analysis, and (iv) collexeme analysis. First of all, the examples were manually disambiguated so as to filter out instances that were not relevant to the study. Four cases were discarded as a result of the filtering process. Firstly, cases where *have* showed an openly possessive meaning, as in Example (14):

(14) Did he have it that time? (KC0 5059)

Secondly, cases in which *that* functioned as a determiner within a noun phrase (NP), as in Example (15):

(15) You've never had it that way? (FPX 1695)

Thirdly, examples were also ruled out where *that* functioned as an intensifier of an adjective or an adverb as shown in Example (16):

(16) She could have it that loud at ten o'clock. (KPV 6525)

Finally, idiomatic expressions that were not related to the reporting meaning of *have* seen in Example (3). That is the case, for instance, in Example (17):

(17) If the ayes have it tomorrow, Dr Cole-King can expect to be a vicar by this time next year. (K25 2281)

After the filtering process, as many as 330 examples were pertinent to the analysis for having the HITC.

In the second step, two different classifications were made of the clean sample. Firstly, examples were categorised according to the realisation of the subject element, and secondly, according to the extraposed element. This was a necessary preliminary step for the statistical analysis described in step four. The classifications are described in greater detail in the analysis of the formal properties of the HITC in Section 4.

The third step consisted in obtaining basic descriptive statistics of the distribution of the HITC both in the spoken and written modes across different text types. This was done with a view to finding basic though valuable information on the use of the construction in British English. In fact, as shown in Section 5, the normalised figures show a significant incidence of the construction around the written component of the corpus. Statistical analysis at this stage was carried out by means of the BNCweb's built-in tools for distribution and dispersion.

Finally, to investigate the co-selection relationships within the HITC, step four involved using a collostructional test, which relies on the filtered dataset as the input. Collostructional analysis, as conceived and defined by Stefanowitsch & Gries (2003: 214), "always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction". The study of collostructional phenomena comprises three different types of analysis, namely (i) 'collexeme', (ii) 'distinctive-collexeme', and (iii) 'co-varying collexeme' analyses. The first inspects the degree of attraction or repulsion of a specific lexeme to a particular slot

in a construction; the second allows finding coincidences among different structures with respect to one or more collexemes; and the third investigates the relationship between collexemes within one specific construction. Section 6 presents results derived from the application of basic collexeme analysis to the HITC using a machine script compiled in Gries (2007) for use in the *R* environment (*R* Core Team 2011). In order to apply a collexeme analysis to study the HITC, four data were needed: firstly, the frequency of the collexemes, i.e. the lexical units that are attracted to the HITC; secondly, the frequency of the collexemes in all other constructions; thirdly, the frequency of the construction with lexemes other than the collexemes in the HITC, fourthly, the frequency of all other constructions with lexemes other than the collexemes studied. These frequencies were then submitted to a Fisher exact test using the aforementioned script (Stefanowitsch & Gries 2003: 217).

4. Formal and syntactic properties

The HITC shows some characteristics concerning its clausal components that call for a separate analysis from the other types of vacuous cases. As regards the subject, the corpus shows that the most frequent realisation of this element is an NP headed by a singular common noun, as in Example (18). This NP can optionally be modified, as in Example (19), totalling 297 instances in the corpus (90%):

(18) This theory has it that, as a breed, doctors like to be masters of every possible situation. (A92 134)

(19) Popular mythology among mothers has it that the first child is always the most 'difficult'. (CEE 165)

Rather more infrequently, the subject is an NP headed by a proper noun (21 cases, 6.4%) or a pronoun (12 cases, 3.6%), as in Examples (20) and (21), respectively:

(20) Bernie would always have it that he had known Malcolm since he was a mod in the sixties. (A6E 337)

(21) We have it that Nozick's account of knowledge succeeds in defusing one sceptical argument but not another. (F9K 778)

One relevant aspect of the morphology of the HITC is that it licences a rather infrequent type of subject consisting of a singular noun used as the head of an uncountable non-specific NP (cf. Allan 1980, Quirk et al. 1985, Bache & Davidsen-Nielsen 1997, Huddleston & Pullum 2002, Keizer 2007). This structure occurs frequently in the corpus, amounting to 182 cases, 55.2% of the entire sample. Subjects of this type are shown in the following Examples (22) to (26):

(22) Rumour has it that Stevo signed some of his first acts. (ACN 165)

(23) Legend has it that the high altar marks spot where Harold died. (BPC 1622)

(24) Report has it that they have now virtually committed racial suicide. (CCB 1256)

(25) Word has it that an unlikely combination of forces have come together in a bid to establish a new standard. (CSX 123)

(26) Story has it when Minton hung this portrait at Allen Street it so offended an art critic that he spat at it. (F9U 1194)

In all these instances the base form of the noun is selected where either an explicit determiner (e.g. *The legend/s has/have it that*) or a zero determiner (e.g. *Legends have it that*) is normally required. The zero determiner has been frequently related to the occurrence of both mass nouns and count nouns, in singular and plural contexts respectively (see Allan 1980). Bache & Davidsen-Nielsen (1997), for example, distinguish five different major uses of this determiner. Examples by these authors are *Time is still a mystery to science*, and *Traffic had to be diverted because roads were flooded*. These authors also record the use of *Word got round that he already resigned*, which is attributed the specific function of emphasising the conceptual nature of the subject rather than its referent. The corresponding periphrases with the plural form in Examples (27) and (28) rightly point to the differences between "singular generics" in contrast to "plural generics":

(27) Rumour/s has/have it that Stevo signed some of his first acts. (ACN 165)

(28) *Animal was/Animals were kept under light anaesthesia. (HWS 5317)

Besides its morphosyntactic realisation, the subject of the HITC becomes especially important for its collocational properties. This issue will be dealt with in more detail in Section 5.

Have is the only verb that can appear as the predicator of the HITC and its use responds to an infrequent sense of this verb[4]. In this use, *have* is not possessive but expresses support of ideas. This epistemic sense is an early development of the basic possessive meaning. The OED documents this new sense back to the early-mid 15th century, when the first instances appeared. Written material from the second part of the 16th century attests a more sustained occurrence at successive stages of English. Example (29) belongs to the earliest records and Example (30) to a later period:

(29) Thouȝ it mai be had by tho textis that God schal ȝeue and do. (OED)

(30) All the Town has it, that Miss Caper is to be married to Sir Peter. (OED)

The OED offers two entries for the stance use of *have*. The first of them describes *have* in the examples just explained, and to which the etymological data abovementioned refers to, whereas the second entry illustrates cases of the verb *have* plus the auxiliary *will (not)*. In both cases the meaning is claimed to be "to assert" and "maintain", with no further significant differentiation as regards meaning. Due to their close semantic proximity, no distinction would be established in the present study between both entries.

The possessive meaning underwent a process of semantic extension from which the new stance reading arose. Despite the apparent difference that may appear to lie between the original meaning and the acquired sense, it still seems possible, however, to trace back how such extension occurred from the source domain "possession" − either "physical" (e.g. *I have a car*) or "non-tangible" (e.g. *I have an idea*) − to the epistemic domain. In this view, the HITC ultimately conveys the subject's standpoint that what is stated in the extraposed element does not belong to the sphere of his/her personal viewpoint, or to the ideas and conceptions that he/she claims, holds or has.

This shift from the possessive meaning to the domain of stance did not occur exclusively within the limits of the verb *have*; rather, it is necessarily linked to another building element of the HITC, the expletive *it*. Both the verb and the expletive form a non-compositional semantic unit, such that the stance meaning is only expressed providing that the latter occurs. This strong relationship felicitously explains the compulsory nature of *it* within the HITC, as exemplified by Example (31):

(31) Legend has *(it) that the green of the water comes from the fairies washing their clothes in the lochan! (CME 697)

Bergh (1997) explains that the connection between the verb and the expletive is so close that the omission of the latter renders the sentence ungrammatical:

> […] It is possible that extraposition in this context was once optional and that frequent usage over time has modified the meaning of the predicates in such a way that it is now obligatory. (Bergh 1997: 39)

Finally, the last element of the HITC consists of a clause or phrase representing the propositional content that the speaker attributes to a source whose identity is unknown or uncertain. This content is most frequently expressed by means of a *that*-clause (63.6%), like in Example (32) or a zero *that*-clause (6.7%), like in Example (33):

(32) Legend has it that a real giant terrorised the locals. (CHK 1719)
(33) Rumour has it he's getting wilder and wilder as time goes on. (HHB 1565)

It must be noticed that there are a number of examples (27%) in which the HITC does not function at sentence but at clause level commonly proceeded by *as* and *so*, like in Examples (34) and (35), respectively:

(34) As a Chinese proverb has it, 'ivory does not come from a rat's mouth'. (FBA 214)
(35) Cornwall's own Bluebeard, so legend would have it, had his castle on the banks of the River Lynher, close to Saltash. (B0G 47)

In this case, the construction preserves both the stance meaning and the reporting function found in examples thus far, and therefore will be considered variations of the standard HITC at sentence level. The same also holds true of other minor variants showing a low rate of occurrence in the sample (2.7%), like in Examples (36) and (37):

(36) Of interest near the entrance to the old church is a floor-mounted grate that protects the Pietra degli Innocenti, the Stone of the Innocents, on which, legend has it, the Roman Emperor Valentian executed four innocent court officials framed by jealous colleagues. (ANB 1021)

(37) It is anyone's guess why Mr Mitterrand applauded, though he is said to get on better with Mrs T than Euro-legend has it. (AA5 611)

There is a noticeable difference between the HITC and other OBEX cases in that the latter can show a greater number of realisations of the extraposed element. Kim (2005: 150), for example, distinguishes the following realisations of the extraposed object: *to-*infinitive clause (e.g. *In the circumstances I do not think it unreasonable of me to ask for the return of my subscription*); *for* NP *to-*infinitive clause (e.g. *His face becomes dusky, because the muscle spasm makes it impossible for him to breathe*); *if-*clause (e.g. *You may find it convenient if he or she acts as a sort of chairman of any discussions you may have*) and *ing-*clause (e.g. *They're not finding it a stress being in the same office*).

Another important syntactic feature of the subordinator relates to its centrality within the HITC. Kaltenböck (2006) focuses specifically on the analysis of the factors which influence the choice of *that-*complements and zero *that-*clauses in extraposed subjects. His claim is that one of the main factors influencing omission in this case lies in the text type, so that informal texts favour the omission of the subordinator. In the case of the HITC, and as seen in the tight connection between the VP and the expletive, there is likewise optionality as regards its omission without yielding an ungrammatical or deviant sentence. For instance, consider Examples (38) and (39):

(38) Rumour had it the owner of the Gazi burned it down for the insurance money. (A6C 783)

(39) One legend has it they came from the Island of the Sun, now in Bolivian waters. (APC 1845)

In sum, a number of morphosyntactic features of the constituents making up the HITC show little variation if compared to other vacuous and non-vacuous cases of OBEX.

5. Distribution across modes and text types

One first relevant aspect about the distribution of the HITC concerns its occurrence in both written and spoken language. The data from the BNC, shown in Table 1, indicate that there is a clear tendency for HITC to appear in the written mode more frequently than in the spoken, though the total number in the whole corpus is in any case relatively low:

Table 1. Distribution of the HITC in the written and spoken modes in the BNC

Mode	Tokens	HITC	Dispersion	Tokens pmw[*]
Written	87,903,571	321	255/3,140	3.6
Spoken	10,409,858	9	9/908	0.9
Total	98,313,429	330	264/4,048	3.3

[*] "pmw" stands for "per million words"

As for the frequency across text types, Table 2 shows that the main focus of occurrence of the HITC clusters around the written component and, more specifically, the generic group "Other published written material", which subsumes the following types of written sources: advert, Hansard, institutional documents, instructional documents, letters, miscellanea and pop lore (adapted from Hoffmann et al. 2008: 265). A second related fact is that spoken occurrences of the construction feature low in the chart, in sharp contrast with other categories, such as journalese, which has a stronger presence in the corpus. The group "Other spoken material" comprises 23 spoken genres (e.g. debates, meetings, lectures) from the context-governed component of the BNC (Hoffmann et al. 2008: 265).

Table 2. Distribution of the HITC across categories in the BNCweb

Categories	Tokens	HITC	Dispersion	Tokens pmw[*]
Other published written material	17,924,109	94	69/710	5.24
Non-academic prose and biography	24,178,674	117	95/744	4.84
Fiction and verse	16,143,913	44	38/452	2.73
Academic prose	15,778,028	42	31/497	2.66
Newspapers	9,412,174	17	16/486	1.81
Unpublished written material	4,466,673	7	6/251	1.57
Other spoken material	6,175,896	7	7/755	1.13
Spoken conversation	4,233,962	2	2/153	0.47
Total	98,313,429	330	264/4,048	3.36

[*] "pmw" stands for "per million words"

In what follows, I report on the findings from a collostructional analysis drawn on the corpus sample described in Section 3. As I try to show, this statistical method is of great importance in the empirical definition of the HITC as a construction. The results show the strong co-selection of the construction relative to the subject slot and, more importantly, they provide empirical support for the argument that variable slots within a construction must be semantically compatible with the meaning of the construction (cf. Goldberg 1995, Wulff et al. 2007: 267).

6. Subject co-selection

The results obtained from the simple collexeme analysis reveal strong co-selection of the HITC upon the subject element. As Table 3 shows, most lexemes (lemmas) attracted to the subject slot have an inanimate referent and mostly refer to verbal processes, such as *rumour* and *legend*, as well as mental processes, such as *wisdom* and *theory* (see Appendix for details):

Table 3. Collostructional strength of the subject within the HITC. The ten most attracted lexemes

Lexical units	Observed frequency (subject)	Coll. strength	Relation
rumour	80	299.5	attraction
legend	63	233.3	attraction
tradition	16	40	attraction
proverb	7	26.3	attraction
myth	9	25.2	attraction
gossip	8	25	attraction
wisdom	6	16.6	attraction
theory	8	14.7	attraction
folklore	4	13.5	attraction
mythology	4	12.7	attraction

On the other hand, the results indicate that there are units that are clearly repulsed. As shown below, the lower part of Table 4 features the personal pronouns *you* and *he*, which are thus less likely candidates to become the subject of the HITC:

Table 4. Collostructional strength of the subject within the HITC. The ten least attracted lexemes

Lexical units	Observed frequency (subject)	Coll. strength	Relation
line	1	1	attraction
view	1	1	attraction
form	1	0.9	attraction
term	1	0.9	attraction
other	1	0.8	attraction
world	1	0.7	attraction
one	1	0.5	attraction
way	1	0.5	attraction
you	1	0.5	repulsion
he	1	0.4	repulsion

Interestingly, the pronoun *we* occurs as a moderately attracted unit (coll. strength 2.2), which can be partly explained by the fact that all the cases in the corpus containing *we* as the subject of the construction are scientific, and in all of them the pronoun serves the function of generalising among several authors rather than particularising. See Example (40):

(40) So far we have it that Nozick's account of knowledge succeeds in defusing one sceptical argument but not another. (F9K 778)

The data from the analysis suggests two possible classifications of the HITC as regards the semantic content of the subject. A first group of Subjects includes specific, animate entities which make reference to a fully identifiable entity − most frequently, a person who is known to both the speaker and the hearer (e.g. *Shakespeare*) − as well animate entities with generic reference (e.g. *traditionalists*). In both cases the subject embodies the role of 'sayer' and appears as the source of new information within a certain state of affairs[5]. In contrast, the second group holds the bulk of subjects consisting of entities with a non-specific, inanimate referent (e.g. *rumour*(*s*)) as well as inanimate entities with definite reference (e.g. *the myth*). In this case, the subject does not lead to an individual identifiable entity but to a more abstract and/or generic origin. Thus, what is of interest is that the collexematic approach unequivocally reveals that the HITC favours unidentifiable sayers, while it disprefers lexical units which can disclose the identity of a given source of information.

While this first coarse-grained classification according to parameters of specificity, definiteness and animateness contributes to the semantic profiling of the subject, the collexeme analysis can still serve as a means for additional semantic classification of

the HITC. Specifically, consideration of the list of attracted units allows distinguishing among distinct lexical domains. Table 5 summarises the main identifiable groups:

Table 5. Classification of the subject of the HITC according to the main semantic domains

Referent's lexical domain	Lexical realisation
(i) Anonymous voices	rumour, gossip, speculation, leaks
(ii) Scientific or professional sources	argument, hypothesis, theory, report, account
(iii) Traditional and folklore sources	legend, lore, myth, wisdom, story
(iv) Verbal sources	word, iteration
(v) Known sources	Shakespeare

This second classification identifies clear-cut, homogeneous groups of subjects organised according to a common hyper-ordinate semantic field. The most straightforward conclusion is that the HITC shows a strong semantic preference for subjects that refer to anonymous voices as well as more established and scientific sources. The first group involves voices which the speaker considers the approximate source of the information he/she is reporting. These can be evaluative, depending on the context, so that, for example, words such as *gossip* or *leaks* are almost invariably negative, since they always denote situations in which there is speculation and non-informed conversation normally aimed at revealing private details of the life and activities of other people. The second and third groups involve a rather different declaration of anonymity. More specifically, the units in the second group show that the source should be considered a serious and reliable basis, mainly the word of science or professionals (e.g. *hypothesis*, *theory*), while, in the third group, it is not scientific argumentation that is highlighted, but the seriousness of the source rests upon tradition and former experience (e.g. *legend*, *story*). In both the second and third groups the reporting person shows that the information that he/she gives is trustworthy. The fourth set is more heterogeneous, containing units that evoke any facet of human saying or thinking, and whose denotata do not permit in either case an approximation to the original source of information. The focus of the speaker is not reliability in this case but mainly a neutral recognition of an idea of unknown origin (e.g. *iteration*, *misconception*). Finally, the last semantic group includes mainly deictical elements and identifiable sayers (e.g. *we*, *Heidegger*), which help the speaker and the hearer recognise and rely on the referent who first produced a piece of information.

As has been shown above, the collostructional analysis has proven to be useful in providing a quantitative analysis of co-selective patterns within the HITC. The construction, in fact, shows a highly internal coherence since it is decidedly selective in the type of lexical units it attracts to itself. The classification according to lexical domains, moreover, has shown that the semantic co-selection is much richer than traditionally stated. Nonetheless, this classification cannot be interpreted tightly. Rather, the semantic domains mentioned above could require further refinement and subdivision. It must also be noted that a more comprehensive explanation of the HITC ultimately demands interpreting the context and the role of participants, which will be discussed in the following section.

7. Semantic and pragmatic functions

This section focuses on the semantic description as well as the analysis of the pragmatic factors involved in the selection of the HITC. As I discuss below, the construction is rooted in the speaker's attitude towards how he/she perceives states of affairs, and how he/she chooses to communicate them. In doing so, this section takes a more qualitative approach than the strictly statistical analysis in Section 6.

The HITC constitutes first and foremost a discourse strategy for reporting facts, speech, thoughts and opinions. Speakers thus use it mainly to inform his/her addressee(s) about what someone else said at a more or less remote point in time. An especially relevant notion in this context is that of evaluation of status. The main assumption behind this concept, as proposed in Hunston (2011), is that, any act of communication in which a speaker attributes propositional content to another speaker constitutes in itself and alignment of that content with the world which is being evoked. The evaluation of status therefore is not concerned with judging the truth value of a proposition but with the attributive functions derived from situating that content within a particular state of events. According to Hunston (2011: 28), this type of attribution can be expressed by a number of units, including, among others, verbs, nouns and adjectives governing *that*-clauses (e.g. *it has been suggested that*; *it is probable that*), adverbs (e.g. *probably*, *supposedly*) and other structures (e.g. *according to*; *as is well known*).

The notion of status is central to the description of the HITC, precisely because its inherent meaning is that of attribution. More importantly, attribution is essentially a

constructional and non-compositional meaning, and is therefore preserved in all the examples in the corpus, irrespective of the realisation of the two variable elements within the construction. Consider Examples (41) and (42):

(41) One view has it that they are the trusted preservers of law and order. (ASB 725)

(42) Business mythology has it that computer firms in Japan are no good at software. (B7B 453)

In the preceding examples, the HITC attributes the content in the *that*-clause to an external sayer − *one view* and *business mythology*, respectively. This sayer is presented as the ultimate source of information which is responsible for the truth of the proposition. The HITC therefore construes facts "as they are", enabling the speaker to take a more or less distant position in place and time with respect to the original source, while in turn inviting the hearer to evaluate the contents being reported.

The HITC, however, must be approached in a broader perspective because, as I intend to show below, the co-selective trends observed within the construction can have discourse functions that go beyond the merely attributive. A distinction is made here between 'identifying' and 'non-identifying' reporting. In the former, the speaker who is reporting makes overt reference to the source of information, either by citing a proper name or any data that clearly reveals its identity. In the latter, in contrast, what is being reported is brought into discourse without any explicit mention as to who originally uttered the message. This distinction is grounded in LaPolla's (1995) Role and Reference model of focus structure – this in turn attested to be based on Lambrecht (1994). This model foresees different activation states of participants in different communicative settings. According to this framework, whether the actual referent is accessible to the addressee depends on the level of specificity and contextual information accessible. Finally, concerning non-identifying structures, some scholars have coined the expression "defocused agents" pointing to the fact that the agent is presented as occupying a background position, both syntactically and pragmatically (Sansò 2006). In the remainder I discuss some of the main factors influencing the selection of non-identifying and non-specific HITC (as discussed in Section 6) and the implications of that choice. In this regard, it is important to note that there may not be a one-to-one relationship between the construction and a specific pragmatic function, and

that the context plays a fundamental role in explaining a given instance of the construction.

The most important factor for the occurrence of a non-identifying HITC, as suggested by examples in the corpus, is the speaker's lack of information about the origin of the reported state of affairs. That is, the speaker who is acting as a reporter, does not know or pretends to not know the source of specific information, and, thus, fills the informative gap by attributing the source to a generic entity. This function derives from virtually every example dealt with so far, as in Examples (43) and (44):

(43) Later sources had it that the king was murdered: Geoffrey le Baker, writing thirty years after the event, provides the vivid details […]. (E9V 132)

(44) Bazaar rumour has it that as many as 100 of these anti-American clerics have been locked away; but the government has taken no public stand against them. (ABH 2646)

Frequent subjects in this use are anonymous-voices, either informed, as in Example (43), or rumour-like, as in Example (44). In either case, there is an underlying implicature that the speaker is unaware of the actual sayer of an utterance. In this regard, the semantic conversion of *have* as a reporting verb may be now clear. The speaker who is reporting does not only blur the source of information but he/she also conceals the degree of involvement with the message being uttered.

The HITC can involve more complex communicative scenarios in which the speaker's lack of knowledge about the sayer is due to the fact that the original message is the result of a more or less wide number of different sayers that the reporter cannot enumerate, and, consequently, chooses to group them together under a non-specific subject. The resulting statement has the effect of "multiple-source attribution". That is the case, for instance, in Examples (45) and (46):

(45) Some utterly ignorant reports have had it that he was motivated by egotism or else arrogance. (AR3 731)

(46) He is 'an embarrassment to the numerous English historians' who would have it that modernism in poetry was a temporary, American-inspired distraction from a native tradition. (A2J 16)

However, speakers' lack of knowledge does not always explain the occurrence of the HITC. In sharp contrast to the situations presented above, the HITC can also account for cases in which the speaker *does* know the sayer of the original message but chooses to shroud it in secrecy so as to save the negative face of the latter (see Brown & Levinson's 1987 politeness theory). In Examples (47) and (48) the speaker is cautious about revealing sensitive information about a given state of affairs, which triggers the choice of the HITC:

(47) Word has it that Unix System Labs has structured its pricing schedule so that its OEMs could bring Destiny out for $350. (CTH 403)

(48) Rumour has it that the barbecue team is ready and willing to provide its services to Aberdeen on request. (HAR 499)

In certain cases, the context and the selection of the sayer may result in an act of evaluation thus gearing the attributive meaning towards a more attitudinal stance. In this view, in the preceding examples, the depersonalisation can be associated with a positive or negative evaluation on the part of the speaker about the information which he/she is reporting and/or the source.

Presenting information without a transparent sayer can also indicate a subtle way of eliciting information rather than preserving anonymity. Especially in conversation, the speaker may wish to determine whether the interlocutor(s) know/s (or ignore/s) the identity of the sayer in order to obtain that information or otherwise to ascertain it. In the exchange in Example (49) speaker A throws a rumour with the subsequent reaction by speaker B, who completes the informative gap left by the former:

(49) SPEAKER A: Oh rumour has it that her husband was in the forces.
SPEAKER B: Yeah now this is an old rumour and er, he was an officer, er I think in the RAF and he had erm an accident his plane crashed and he died and this is why she dresses in military uniform all the time. (KP1 3683)

Nevertheless, whether or not the speaker is actually concealing the sayer from the hearer in a premeditated way will depend on the situational context and the participants' involvement in the events being reported.

Finally, the examples in the corpus suggest that, in other contexts, the speaker takes a more authoritative perspective on the discourse, imbuing it with a more convincing and sound appearance. In this case, the speaker addresses the interlocutor by using a respectable source to present his/her argument. The reporting of events here is aimed at being perceived as an unrivalled source of knowledge. This is evident in cases in which the HITC consists of scientific and folklore-related sayers. Consider Examples (50) and (51):

(50) One acne theory has it that an enzyme deficiency permits testosterone. (CDR 1931)

(51) Tradition has it that at the cremation of Gautama Buddha in India in 543 BC the tooth survived the flames. (CK2 1074)

Both examples coincide in marking the facts reported as the stream of thought but the focus of authority differs. In this regard, in Example (50) – which is an extract from an academic written record – the speaker claims that the ideas that he/she is communicating are the result of contrasted study and patterned behaviour, rather than speculation. In contrast, in Example (51) the speaker appeals to the tradition based on beliefs and customs that have existed for a long time. In both examples it may be argued that the selection of an authoritative source may constitute a positive stance towards the proposition, and a subtle attempt on the part of the speaker to make the hearer agree on the nature of the facts being stated.

8. Conclusions

This paper has investigated the formal and functional aspects of the *have-it-that* construction. The study provides an insightful analysis of extraposition viewed and described as a construction. The research also contributes to our understanding of object extraposition, which has received less attention in the relevant literature, largely focused on subject extraposition and the semantics of the expletive *it*.

Drawing on the latest quantitative collostructional analysis and data from the British National Corpus, this paper has proved, in the first place, that there is a systematic selection as regards the lexemes that can realise the subject slot within the construction.

In particular, it is argued that the construction selects lexical units with non-specific reference on a coherent and regular basis. In this regard, the findings of this study confirm that object extraposition, as suggested by some authors, is a strategy for sayer depersonalisation and referential indeterminacy. Secondly, it is also shown that the HITC is used for the evaluation of status, i.e. it serves as a vehicular construction to report propositional content as well as to convey other speakers' stance and beliefs.

It must be noticed that the present paper has only examined examples in British English and that further research on other varieties of English would be required for a more detailed account of the *have-it-that* construction. Further research is also needed to investigate extraposition as a construction. From this perspective, we could understand how speakers use conventionalised form-meaning pairings to convey depersonalised information and how they use the *have-it-that* construction to emphasise certain facts over others within discourse.

Notes

1. Unless otherwise stated, examples in this paper have been retrieved from the British National Corpus. The reader is referred to the data section.

2. S(ubject), P(redicator), O(bject), C(omplement), A(dverbial).

3. BNCweb (version 4.2) (http://corpora.lancs.ac.uk/BNCweb/).

4. There are other verbs of stance that could eventually actualise the predicate of the HITC, namely *say* and *hold*. However, these are only marginally acceptable and rather infrequent. Thus, they will not be considered in this study.

5. 'Sayer' is used in this paper to refer to a participant who expresses a message in a state of affairs (see Halliday 1994). This sayer is considered the original source of information (e.g. *rumours*, *legend*, *myth*, etc.).

References

Allan, K. 1980. "Nouns and countability". *Language*, 56 (3), 541-567.

Bache, C. & Davidsen-Nielsen, N. 1997. *Mastering English: An Advanced Grammar for Non-native Speakers*. Berlin: Walter de Gruyter.

Bergh, G. 1997. "Vacuous extraposition from object in English". *Studia Neophilologica*, 69 (1), 37-41.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Brown, P. & Levinson, S. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Collins, P. 1994. "Extraposition in English". *Functions of Language*, 1 (1), 7-24.

Esser, J. 2002. "Sampling and categorizing fronted constructions in the BNC". In A. Fischer, G. Tottie & H. M. Lehmann (Eds.), *Text Types and Corpora: Studies in Honor of Udo Fries*. Tübingen: Gunter Narr, 131-138.

Fillmore, C. J., Kay, P. & O'Connor, M. 1988. "Regularity and idiomaticity in grammatical constructions: The case of *let alone*". *Language*, 64 (3), 501-538.

Fitriati, A. 2006. "An analysis of sentences with nonreferential *it*". *Journal of Language and Literature*, 10 (1), 22-37.

Gómez-González, M. A. 1997. "On subject it-extrapositions: Evidence from present-day English". *Revista Alicantina de Estudios Ingleses*, 10 (1), 95-107.

Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: Chicago University Press.

Gonzálvez-García, F. & Butler, B. 2006. "Mapping functional-cognitive space". *Annual Review of Cognitive Linguistics*, 4 (1), 39-96.

Gries, S. Th. 2007: online. *Coll.analysis 3.2: a program for R for Windows 2.x.* Available at http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/ coll.analysis.r(accessed August 2011).

Halliday, M. A. K. 1994. "Systemic theory". In R. Asher & J. Simpson (Eds.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 4505-4508.

Herriman, J. 2000. "Extraposition in English: A study of the interaction between the matrix predicate and the type of extraposed clause". *English Studies*, 81 (6), 582-99.

Hewings, M. & Hewings, A. 2002. "'It is interesting to note that…': A comparative study of anticipatory 'it' in student and published writing". *English for Specific Purposes*, 21 (4), 367-383.

Hoffmann, S. & Evert, S. 2002: online. *BNCweb (CQP edition). Version 4.2.* Available at: http://www.bncweb.info (accessed August 2011).

Hoffmann, S., Evert, S., Smith, N., Lee, D. Y. W. & Prytz, Y. B. 2008. *Corpus Linguistics with BNCweb: A Practical Guide*. Frankfurt am Main: Peter Lang.

Hoffmann, T. & Trousdale, G. (Eds.) 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Huddleston, R. & Pullum, G. 2002. *The Cambridge Grammar of English.* Cambridge: Cambridge University Press.

Hunston, S. 2011. *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. London/New York: Routledge.

Kaltenböck, G. 2002. "That's it? On the unanticipated 'controversy' over anticipatory *it*: A reply to Aimo Seppänen". *English Studies*, 83 (6), 541-550.

Kaltenböck, G. 2003. "On the syntactic and semantic status of anticipatory *it*". *English Language and Linguistics*, 7 (2), 235-55.

Kaltenböck, G. 2004. "Using non-extraposition in spoken and written texts: A functional perspective". In K. Aijmer & A. B. Sternström (Eds.), *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: John Benjamins, 219-242.

Kaltenböck, G. 2006. "'*That* is the question': complementizer omission in extraposed *that*-clauses". *English Language and Linguistics*, 10 (2), 371-396.

Keizer, E. 2007. *The English Noun Phrase: The Nature of Linguistic Categorization*. Cambridge: Cambridge University Press.

Kim, J. 2005. "Subject extraposition and object extraposition in English: Corpus findings and a constraint-based approach". *Studies in Generative Grammar*, 15 (2), 145-164.

Kim, J. & Sag, I. A. 2005. "English object extraposition: A constraint-based approach". In S. Mueller (Ed.), *Proceedings of the HPSG05 Conference, University of Lisbon*. Stanford: CSLI Publications, 192-212.

Kiparsky, P. & Kiparsky, C. 1971. "Fact". In D. D. Steinberg & L. Jakobovits (Eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge: Cambridge University Press, 345-369.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lambrecht, K. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents.* Cambridge: Cambridge University Press.

LaPolla, R. J. 1995. "Pragmatic relations and word order in Chinese". In P. Downing & M. Noonan (Eds.), *Word Order in Discourse*. Amsterdam: John Benjamins, 297-329.

Lee, D. Y. W. 2001. "Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle". *Language Learning & Technology*, 5 (3), 37-72.

Mair, C. 1990. *Infinitival Complement Clauses in English: A Study of Syntax in Discourse*. Cambridge: Cambridge University Press.

Miller, P. H. 2001. "Discourse constraints on (non)extraposition from subjects in English". *Linguistics*, 39 (4), 683-701.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

R Core Team. 2011: online. *R: A Language and Environment for Statistical Computing*. Available at: http://www.R-project.org (accessed August 2011).

Rosenbaum, P. S. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge, Mass.: The MIT Press.

Ruiz de Mendoza, F. & Mairal, R. 2008. "Levels of description and constraining factors in meaning construction: An introduction to the Lexical Constructional Model. *Folia Linguistica*, 42 (2), 355-400.

Sansò, A. 2006. "'Agent defocusing' revisited: Passive and impersonal constructions in some European languages". In W. Abraham & L. Leisiö (Eds.), *Passivization and Typology: Form and Function*. Amsterdam: John Benjamins, 229-270.

Seppänen, A. 2002. "On analysing the pronoun *IT*". *English Studies*, 83 (5), 442-462.

Seppänen, A. & Herriman, J. 2002. "Extraposed subjects vs. postverbal complements: On the so-called obligatory extraposition". *Studia Neophilologica*, 74 (1), 30-59.

Seppänen, A., Granath, S. & Herriman, J. 1995. "On so-called 'formal' subjects/objects and 'real' subjects/objects". *Studia Neophilologica*, 67 (1), 11-19.

Stefanowitsch, A. & Gries, S. Th. 2003. "Collostructions: Investigating the interaction between words and constructions". *International Journal of Corpus Linguistics*, 8 (2), 209-243.

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at: http://www.natcorp.ox.ac.uk/ (accessed August 2011).

The Oxford English Dictionary. 1989. A. J. Simpson & E. S. C. Weiner (Eds.). Oxford: Clarendon Press.

Wulff, S., Stefanowitsch, A. & Gries, S. Th. 2007. "Brutal Brits and persuasive Americans: Variety–specific meaning construction in the into-causative". In G. Radden, K. M. Köpcke, T. Berg & P. Siemund (Eds.), *Aspects of Meaning Construction*. Amsterdam: John Benjamins, 265-281.

Appendix

Collostructional strength of the Subject element within the HITC

Lexical units	Observed frequency (subject)	Coll. strength	Relation
rumour	80	299.5	attraction
legend	63	233.3	attraction
tradition	16	40	attraction
proverb	7	26.3	attraction
myth	9	25.2	attraction
gossip	8	25	attraction
wisdom	6	16.6	attraction
theory	8	14.7	attraction
folklore	4	13.5	attraction
mythology	4	12.7	attraction
story	7	12.3	attraction
jargon	3	9.1	attraction
convention	4	8.6	attraction
song	4	8	attraction
cliché	2	6.8	attraction
burler	1	5.5	attraction
slogan	2	5.4	attraction
doctrine	2	4.7	attraction
out	1	4.7	attraction
hypothesis	2	4.6	attraction

journalese	1	4.6	attraction
bürger	1	4.4	attraction
account	3	4.3	attraction
style council	1	4.3	attraction
tale	2	4.2	attraction
estimate	2	4.1	attraction
heidegger	1	4.1	attraction
look back in anger	1	4.1	attraction
phrase	2	4	attraction
report	3	3.9	attraction
nik	1	3.8	attraction
copernicus	1	3.6	attraction
parlance	1	3.5	attraction
euphemism	1	3.4	attraction
horowitz	1	3.4	attraction
lore	1	3.4	attraction
unreality	1	3.4	attraction
brooke-rose	1	3.3	attraction
paraphrase	1	3.3	attraction
talk	2	3.3	attraction
traditionalist	1	3.3	attraction
paul	2	3.2	attraction
chandler	1	3.1	attraction
guidebook	1	3.1	attraction
misconception	1	3.1	attraction
school	3	3.1	attraction
working party	1	3.1	attraction
bernie	1	3	attraction
dogma	1	3	attraction
hype	1	3	attraction
objector	1	3	attraction
proponent	1	3	attraction
the daily telegraph	1	3	attraction
antony	1	2.9	attraction
idiom	1	2.9	attraction
inscription	1	2.8	attraction
martyr	1	2.8	attraction
leak	1	2.7	attraction
mama	1	2.7	attraction
saying	1	2.7	attraction
nelson	1	2.5	attraction
wordsworth	1	2.5	attraction
conscience	1	2.3	attraction
consensus	1	2.2	attraction
fiction	1	2.2	attraction
pat	1	2.2	attraction
shakespeare	1	2.2	attraction
speculation	1	2.2	attraction
we	5	2.2	attraction
conception	1	2.1	attraction
historian	1	2	attraction
word	2	2	attraction
interpretation	1	1.8	attraction
some	3	1.7	attraction
comment	1	1.6	attraction
legislation	1	1.6	attraction

writer	1	1.6	attraction
survey	1	1.5	attraction
version	1	1.5	attraction
argument	1	1.4	attraction
title	1	1.4	attraction
analysis	1	1.3	attraction
doctor	1	1.3	attraction
source	1	1.3	attraction
model	1	1.2	attraction
record	1	1.2	attraction
rule	1	1.2	attraction
someone	1	1.2	attraction
research	1	1.1	attraction
line	1	1	attraction
view	1	1	attraction
form	1	0.9	attraction
term	1	0.9	attraction
other	1	0.8	attraction
world	1	0.7	attraction
one	1	0.5	attraction
way	1	0.5	attraction
you	1	0.5	repulsion
he	1	0.4	repulsion

Author's address

Pedro Ureña Gómez Moreno

Departamento de Filologías Inglesa y Alemana

Facultad de Filosofía y Letras

University of Granada

Campus Universitario de Cartuja s/n

18071 Granada

Spain

pedrou@ugr.es