

# ***If*-conditionals and Modality: Frequency patterns and theoretical explanations**

Costas Gabrielatos  
Edge Hill University

## **Abstract**

It has often been claimed that conditionals have a special relation to modality. This study empirically tests this claim by examining the frequency of modal marking in a number of conditional and non-conditional structures using a corpus-based approach. It then seeks to provide explanations for the emerging frequency patterns in light of the tenets of two linguistic theories: Lexical Grammar and Construction Grammar. This juxtaposition was motivated by the significant overlap in their tenets: both theories take into account meaning (semantic and pragmatic), as well as lexical and grammatical factors.

## **1. Motivation and Background**

The connection between conditionality and modality has long been asserted (Sweetser 1990:141). Comrie (1986:89) claims that a conditional “never expresses the factuality of either of its constituent propositions.” Turner (2003:135) presents the intuitive view that “conditionals are not part of fact-stating discourse: conditionals, instead, express uncertainties.” Similarly, Huddleston and Pullum (2002:741) state that “*If P (then) Q* is a weaker statement than *Q* on its own”, adding that “the conditional construction is conducive to the expression of modality” (ibid.:744).<sup>1</sup> Nuyts (2001:352) reports that “conditionals have an intimate link with the domain of epistemic qualification.” Palmer (1986:189) comments that “modality seems [...] to be doubly marked in conditionals.” More precisely, Dancygier (1998:72) states that “the presence of *if* in the construction marks the assumption in its scope as unassertable. As a result, the assumption in the apodosis, which belongs to the same mental space as the protasis, is not treated as asserted either” (see also Bybee et al. 1994:208). However, the relation between modality and conditionals has not been empirically investigated – apart from Gabrielatos (2007b), a critical summary of which is provided at the end of this section.

Indirect support for the above claims has come from the multi-dimensional analysis of texts carried out by Biber (1988), who statistically established seven “dimensions”, that is, “sets of linguistic features that typically occur together in texts” (Biber et al. 1998:146), because “they are used for a shared set of communicative functions in those texts” (Biber 1988:101). Two dimensions are pertinent to this study, as they are the only ones containing conditional subordination. The dimension of the “overt expression of persuasion” (Biber 1988:111) is closely related to the use of modality, in that it “marks the degree to which persuasion is marked overtly, whether overt marking of the speaker’s own point of view, or an assessment of the advisability or likelihood of an event presented to persuade the addressee” (Biber 1988:111). This dimension comprises seven features, five of which are overt modal markers: conditional subordination, prediction modals (e.g., *will*), necessity modals (e.g., *must*), possibility modals (e.g., *may*), and suasive verbs (lexical verbs with deontic or volitional senses, e.g., *demand*, *suggest*) (Biber 1988:102-115). The two remaining features (infinitives and split auxiliaries) are also linked to the use of modality: the presence of infinitives can be seen as a necessary corollary of the presence of modal and suasive verbs. Biber (1988:111) states that in this

dimension the adjective or verb acting as the head “frequently encodes the speaker’s attitude or stance towards the proposition encoded in the infinitival clause”, as in ‘hoped to see it’. The presence of split auxiliaries (i.e., auxiliaries separated from the main verb by adverbs) is likely due to the use of adverbs, which are “often modals” (Biber 1988: 111), as in (1).<sup>2</sup>

- (1) Conversely, if the definition changed, there would probably be a corresponding change in the pupil's actions in that context. [FB6 455]

The second dimension pertinent to this study is that of “involved versus informational production” (Biber 1988:107). It contrasts linguistic features “associated with an involved, non-informational focus” with those associated with “high informational focus and a careful integration of information in the text” (ibid.:104). Four “involved” features are relevant to the expression of modality: conditional subordination, possibility modals (e.g., *may*), general hedges (e.g., *maybe*), private verbs, which can be used with modal meaning (e.g., *feel, know*), as they “are used to flag uncertainty or lack of precision in the presentation of information” (Biber 1988:106). The co-occurrence of conditionals with modality in the above two dimensions is rendered more significant if we consider that the twelve linguistic features listed above are not included in any of the other five dimensions; that is, modality seems to be unique to these dimensions.

Empirical support for the extent of modal marking (henceforth ‘modal load’ or ‘ML’) in *if*-conditionals was provided in Gabrielatos (2007b), who used keyword analysis to compare a sample of 853 *if*-conditionals from the written BNC with two reference corpora: the written BNC Sampler and FLOB, which were used as representative samples of written language use in general.<sup>3</sup> Both comparisons showed that the sample of *if*-conditionals contained modal markers in much higher frequency than the reference corpora. However, limitations in the scope, data, and methodology suggest that the above results should be treated with caution. The manual analysis of the sample indicated that about 15 percent of *if*-clauses had non-conditional uses. Also, the sample included *even if* structures, which are not unanimously regarded as purely conditional (Huddleston & Pullum 2002:985-986; Quirk et al. 1985:1099). This did not allow for comparisons between *if*-conditionals and conditional-concessives with *even if*. What is more, the study did not examine conditionals with other subordinators (e.g., *assuming, in case*); therefore, it could not establish whether the higher frequency of modal markers in the sample of *if*-conditionals was unique to this structure, or a shared characteristic of all conditionals. For the same reasons, Gabrielatos (2007b:13) acknowledges that it was not possible to establish whether there were differences between particular conditional and non-conditional structures. Also, keyword analysis effectively treats the compared corpora as two lists of word-forms. As a result, comparisons included clauses that were extraneous to the *if*-conditionals (Gabrielatos 2007b:11; see also Gabrielatos 2005, 2010:17-21), which is likely to have resulted in the over- or under-estimation of the frequency of modal marking in *if*-conditionals (see Ball 1994:299-300). Equally importantly, frequency comparisons of word-forms cannot be seen as a reliable approach when the focus is on modal senses, which can be expressed by a variety of lexical, morphological, and syntactic means (see section 3.3 below). To address the above limitations, this study has expanded the scope of the analysis, examining a much larger number of relevant structures, and has used a more fine-grained methodology, which employs manual analysis of random samples and tailor-made metrics.

## 2. Lexical Grammar and Construction Grammar

This section will outline the two theoretical frameworks informing this study, namely Construction Grammar (CxG) and Lexical Grammar (LG). The fairly detailed outline and comparison is motivated by four interrelated reasons:

- a) Both LG and CxG acknowledge the interaction of lexis and grammar, and both account for semantic and pragmatic meaning. However, they differ considerably in the way these aspects are utilized and, more so, prioritized in linguistic description.
- b) The two theories would offer different explanations for the ML of conditionals.
- c) LG informed the selection of the theoretical constructs and methodological techniques employed in Gabrielatos (2007b). As this approach had important shortcomings in both respects (see section 1 above), this led to seeking explanations from the perspective of CxG.
- d) Their respective tenets and constructs have methodological implications (see section 3 below).

### 2.1. Lexical Grammar

Lexical Grammar (Sinclair, 2004:164) can be more usefully regarded as a family of closely related approaches sharing the tenet that grammar emerges from the interaction and patterning of lexis in discourse (Hoey 2005:1; Sinclair 1991:100). The impetus for these approaches, and their unifying element, is the notion of “collocation”, which was proposed by Firth (1957:195-196) as an approach to establishing meaning rather than to deriving grammatical description, distinguishing “meaning by collocation” from the “conceptual or idea approach to the meaning of words.” According to Sinclair (1991:115-116), examining the collocates of a word is akin to its “semantic analysis.” Collocation is a relation of “mutual expectancy” between words (Firth 1968:181) – that is, collocation is approached as a probabilistic relation (Hoey 1991:6-7; Partington 1998:16; Stubbs 2002:64). More importantly, LG sees collocation as “a purely lexical relation [...] which ignores any syntactic relation between the words” (Stubbs 2001:64). Collocations are usually established automatically, through the statistical calculation of the actual and expected frequencies of occurrence of the node and each collocate, within a span of 4-5 words on the right and left of the node (see Hoffmann et al. 2008:142-158; McEnery & Hardie 2012:122-133). A related construct, which takes account of grammar (while retaining the primacy of lexis posited in LG) is “colligation”: the co-occurrence of a) content and function words, and b) words and grammatical categories (Hoey 1997:8; Sinclair 2004:174; Stubbs 2002:238).

Regarding meaning, LG posits two constructs, both hinging on co-occurrence: “semantic preference” (Stubbs 2001:65) is the attraction “between a lemma or word form and a set of semantically related words” (Sinclair 2004:174), whereas “semantic prosody” is the pragmatic meaning expressed by the use of word co-occurrences in context (Sinclair 1996:87-88). LG recognizes that word-forms can be polysemous, and that different senses of a word can have their own collocation patterns (Hunston & Francis 2000:37, 83). However, Hunston and Francis (2000:85) add that not all words with a similar sense can be expected to have the same patterns. The above, and in particular the issue of polysemy, are directly relevant to the interpretation of results in the light of LG and CxG (see section 4 below).

Sinclair (1996) posited the “extended unit of meaning” (ibid.:75), or “lexical item” (ibid.:90), comprising the following components: a) The core (a word or phrase); b) its collocates; c) its

semantic preference; d) its semantic prosody; e) its colligations (see also Cheng et al. 2009:239, 248-249; Stubbs 2009:123-126). However, not all components are of equal status: apart from the core, only collocation (i.e., patterning of word-forms) and semantic prosody (i.e., pragmatic meaning) are obligatory, the rest being optional (Cheng et al. 2009:239; Sinclair 2004:174; Stubbs 2009:124). It is worth noting that the only LG constructs compatible with the notion of a conditional structure are those of colligation and semantic preference – both deemed optional in the make-up of extended units of meaning.

## 2.2. Constructions and Construction Grammar

Constructions are “conventionalized pairings of form and function” (Goldberg 2006:1), with particular elements pertaining to their form and meaning. The former specify morphological, phonological, lexical, and syntactic properties, whereas the latter specify semantic, pragmatic, and discourse attributes (Croft & Cruse 2004:258; Fillmore et al. 1988:501; Fried & Östman 2004:18-21, 30). Crucially, a core characteristic of CxG is that it does not make a distinction between lexis and grammar: words and even morphemes are regarded as constructions in their own right (Fried & Östman 2004:12, 18, 28; Goldberg 2006:5, 18; Goldberg & Jackendoff 2004:532-533). In light of the latter characteristic, Tomasello (2003:100) provides a particularly useful definition: “A construction is prototypically a unit of language that comprises multiple linguistic elements used together for a relatively coherent communicative function, with sub-functions performed by the elements as well.” Similarly, Fillmore (1986:163) sees constructions as being “sensitive to subtle structural patterns whose total effects cannot be seen as the *compositional product* of its parts but must be described in terms of separate *grammatical constructions*” (italics in the original). At the same time, the meaning of the construction is not divorced from the meanings of its parts, as “they strongly motivate the meaning of the whole construction” (Dancygier & Sweetser 2005:41). The implication of the above is that examining the meaning of a conditional construction can usefully take into account the meaning of its component parts: the subordinator, the protasis (the subordinate clause), and the apodosis (the main clause) – and the nature of the link holding between protasis and apodosis.

The characteristics discussed above give rise to the “principle of no synonymy” between constructions (Goldberg 1995:67-71), which entails that differences in the formal components of constructions can be expected to result in differences in their semantic/pragmatic function. Conversely, differences in their meaning components can be expected to be reflected in differences in their lexical and/or grammatical properties. The principle has received empirical support; for example, Wulff (2006) demonstrated that *go-V* and *go-and-V* constructions cannot be considered synonymous. The principle of no synonymy is pivotal for this study, as it hinges on the interaction between two or more of the following:

- a) the type of construction (e.g., conditional, indirect interrogative);
- b) the subordinator of the construction (e.g., *if*, *in case*, *whether*);
- c) (in bi-partite constructions) the subordinate and matrix parts;
- d) (in bi-partite constructions) the type of link holding between the two parts.

## 2.3. LG and CxG: Main differences and theoretical implications

LG and CxG converge in two main respects: both see lexis and grammar as interconnected, and both posit units which specify elements of form and semantic/pragmatic function (extended unit of meaning and construction, respectively). As far as the aims and scope of this study are concerned, the fundamental difference between the two approaches is the relative importance they assign to lexis and grammar. LG treats lexis as central, determining the grammar of the

patterns they form (hence the optionality of colligation), whereas CxG treats lexical, semantic/pragmatic, morphosyntactic, and phonological attributes as equally important components of a construction. A second important difference is that LG privileges syntagmatic relations (linear patterns) over paradigmatic relations, usually to the exclusion of the latter, whereas CxG includes both dimensions within a construction, while assigning equal status to both. Due to the above differences, each theory would provide different explanations of the ML of *if*-conditionals. LG would explain it as the result of the semantic preference of the subordinator (established via its collocation patterns), whereas CxG would explain it as being part of the meaning component of the conditional construction. However, conditionals can have quite complex and entangled surface forms, such as containing elements extraneous to the construction, or having elements of the protasis embedded in the apodosis (Gabrielatos 2005, 2010:17-25). In this light, an approach that favours linear co-occurrence relations, usually limited to a span of 4-5 words around the node, does not seem useful. It will be shown in the analysis that the collocation patterns of *if* can be regarded as reflecting not only its semantics, but also the semantics of the structures of which it is a component part.

### 3. Aims, Data, and Methodology

#### 3.1. Aims

The analysis has two related aims. The first aim is to establish the ML of the sample of *if*-conditionals in relation to the samples of the following structures:

- Written British English seen as a whole (*baseline*).
- Conditionals with other subordinators (*assuming, in case, on condition, provided, supposing, unless*).
- Conditional-concessives with *even if* and *whether*.
- Non-conditional structures taken collectively.
- Non-conditionals with subordinators containing *if*: indirect interrogatives with *if* and structures of comparison with *as if*, as well as the same type of structures introduced by *whether* and *as though*, respectively.
- Non-conditionals introduced by the conjunctions *when* and *whenever*, as they have been compared to unmodalized conditionals (e.g., Athanasiadou & Dirven 1996:617, 1997:62; Palmer 1990:174-175).

The second aim is to examine the extent to which the ML patterns emerging from the analysis can be better explained with recourse to the tenets of LG or CxG. More specifically, the study will examine if the ML patterns are due to the meaning of the subordinator (as LG would predict) or the nature of the structure (as CxG would predict).

The baseline and non-conditionals will provide initial reference points against which the ML of the structures in focus can be compared. Comparisons between *if*-conditionals and other conditionals will help investigate whether all conditionals have similar ML, and, if this is not the case, the extent to which the ML of a conditional could be attributed to its subordinator. Comparisons with indirect interrogatives with *if*, as well as with conditional-concessives with *even\_if* and structures of comparison with *as\_if*, will help investigate the extent to which the ML can be ascribed to the conditional construction and/or the presence of *if* in the subordinator. Comparisons with conditional-concessives with *whether*, indirect interrogatives with *whether*, and structures of comparison with *as\_though* will provide a reference point for conditional-concessives, indirect interrogatives, and structures of comparison that do not contain the word *if* in their subordinators – offering further opportunities to examine the influence of *if* on the

ML. The latter aim is of theoretical significance, as LG would predict that the ML is due to the semantic preference of *if* (i.e., its significantly frequent co-occurrence with modal markers), whereas CxG would predict that the ML results from the combination of the nature of the construction family (conditionals) and the subordinator *if*, which differentiates *if*-conditionals from other members of this construction family (e.g., conditionals with *assuming*).

### 3.2. Data and Extraction Techniques

The analysis was carried out manually on random samples from the written BNC (henceforth BNCw), using *BNCweb* (Hoffmann et al. 2008). In order to maximize the number of relevant structures in each sample, complex queries were used. More precisely, some queries specified the word strings to be returned, whereas others specified the word clusters to be excluded by means of regular expressions (see Appendix).<sup>4</sup> The word clusters to be included in, or excluded from, queries were established through the combination of at least two of the following techniques:

- a) Identifying relevant clusters in lists provided in descriptive grammars. For example, Quirk et al. (1985:1182) list introductory verbs used in interrogative structures with *if* (e.g., *wonder if*).
- b) Narrowing the query to a word tagged for a particular grammatical property in the BNC (e.g., *when* as a subordinator).
- c) Examining frequent collocates of relevant subordinators (e.g., *assuming*, *in case*), as well as sorted concordances of the collocations, in order to establish word clusters to include in, or exclude from, the query.

All the resulting samples were checked manually, and the structures not matching the intended content were removed. The samples initially comprised 1000 s-units (unless BNCw contained fewer instances).<sup>5</sup> Table 1 specifies the sample sizes after the manual cleaning, and the shorthand to be used when referring to the structures in each sample.

**Table 1.** Samples used in the study

Code	Content	N of s-units
<i>if-cnd</i>	Conditionals with <i>if</i>	959
<i>assuming-cnd</i>	Conditionals with <i>assuming</i>	727
<i>in_case-cnd</i>	Conditionals with <i>in case</i>	945
<i>provided-cnd</i>	Conditionals with <i>provided</i>	859
<i>supposing-cnd</i>	Conditionals with <i>supposing</i>	213
<i>on_condition-cnd</i>	Conditionals with <i>on condition</i>	205
<i>unless-cnd</i>	Conditionals with <i>unless</i>	989
<i>even_if-cc</i>	Conditional-concessives with <i>even if</i>	995
<i>whether-cc</i>	Conditional-concessives with <i>whether</i>	184
<i>if-q</i>	Indirect interrogatives with <i>if</i>	978
<i>whether-q</i>	Indirect interrogatives with <i>whether</i>	809
<i>as if-c</i>	Structures of comparison with <i>as if</i>	995
<i>as though-c</i>	Structures of comparison with <i>as though</i>	999
<i>when-t</i>	Structures expressing time with the conjunction <i>when</i>	902
<i>whenever-t</i>	Structures expressing time with the conjunction <i>whenever</i>	959
<i>baseline</i>	Sample from the whole BNCw	872
<i>non-cnd</i>	Non-conditional structures	856

Two observations can be made regarding the use of complex queries in this study (and, by extension, similar studies). In terms of methodological significance, they can be used to derive random samples with higher precision, that is, a higher proportion of relevant content (Baeza-Yates & Ribeiro-Neto 1999:75; Chowdhury 2004:170). For example, the random sample of 1,000 s-units returned by the query ‘if’ in Gabrielatos (2006) yielded 781 conditionals (a precision of 78.1 percent), whereas the complex query used in this study (see Appendix) had a precision of 95.9 percent. In terms of descriptive significance, complex queries can furnish insights regarding the lexical characteristics of particular structures. The manual examination of the sample of non-conditionals suggests that the vast majority of conditionals in BNCw are introduced by the subordinators *assuming*, *if*, *in case*, *supposing*, and *unless* – as the sample derived through the exclusion of these words contained a mere 0.5 percent of conditionals. More significantly, considering that the seven conditional structures examined here account for about 99.5 percent of all tokens of conditionals, *if*-conditionals appear to represent about 80 percent of conditionals in written British English. This supports Dancygier’s (1998:14) statement that “*if* is seen as the primary exponent of conditional meaning in English.” The above suggest that the conditionals selected for analysis here account for the overwhelming majority of all conditionals, and, thus, provide a comprehensive picture – with *if*-conditionals being an excellent candidate for a case study.

### 3.3. Modal Markers

This study adopts an inclusive view of modal markers, that is, it recognizes that modality in English can be marked via a large range of lexicogrammatical means (Halliday 2004:89, 354-356; Hoye 1997; Huddleston & Pullum 2002:117, 147-149, 172-175; Leech 2004:14-16, 36-40, 116; Lyons 1977:451-452, 769, 794, 805-806, 815, 820; Nuyts 2001:29; Palmer 1986:4-6, 97, 108-115, 126; Palmer 1987:44-46; Perkins 1983:106-108; Quirk et al. 1985:137, 147, 188, 236-239). More precisely, the following types of modal markers are recognized in this study (Quirk et al. 1985:137, 147, 188, 236-239):

1. Morphological marking of lexical verbs (not themselves modal): imperative, past tense, past/present subjunctive.
2. Auxiliary or lexical verbs with modal meaning:
  - central modals (*can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, *would*)
  - marginal modals (e.g., *dare*, *need*, *ought to*)
  - modal idioms (e.g., *had better*, *would rather*)
  - semi-auxiliaries (e.g., *be able to*, *be going to*, *be obliged to*)
  - catenative verbs (e.g., *appear to*, *seem to*)
  - mental state predicates (e.g., *think*, *believe*, *hope*, *wish*)
3. Structures involving nouns with modal sense (e.g., *The possibility exists that ...*).
4. Structures involving adjectives with modal sense (e.g., *probable*, *desirable*): BE + modal adjective + infinitive / *that*-clause
5. Adverbs with modal sense (e.g., *possibly*, *probably*)

As some of the above lexicogrammatical means (e.g., past tense) do not mark modality in all contexts, their modal nature was established via the manual examination of the sentences in the samples. When the sentence itself did not provide enough contextual clues, sentences preceding and/or following it were also examined.

### 3.4. Metrics

The quantitative examination of the ML of different structures for purposes of comparison needs to address the fact that some attested constructions will be more “grammatically intricate” than others (Halliday 2004:654), that is, they will have more constituent clauses. The number of clauses plays a significant role in a corpus-based quantitative analysis, in that the more embedded clauses a structure contains, the more opportunities for modal marking exist. Essentially, then, the analysis should examine the proportion of those modalization opportunities that are realized within each sample. In order for the comparisons of the ML among the different constructions examined in this study to be meaningful (and, hence, useful) the ML metric needs to be normalized. Following Ball (1994:297-299), the clause, rather than the word-form, is used as the unit of analysis.

One metric employed in the analysis is an adaptation of “lexical density” (Halliday 2004:654-655): the average number of content words per clause.<sup>6</sup> For the purposes of this study, this metric was adapted to the calculation of the average number of modal markings per clause, and was termed ‘modal density’ (henceforth, MD) (Gabrielatos 2010:50-52). Please note that MD should not be understood as the average number of modalized clauses. The utility of MD is demonstrated by examining (2) and (3) below.

- (2) If we could keep to a blue theme for leotards it would make a lovely contrast with the scarves. [KAF 72]
- (3) If you are worried or have questions about the illness, try to find someone you can trust to talk to about it. [CJ9 2271]

Although both (2) and (3) carry the same number of modal markings (two) – *could* and *would* in (2), and the imperative (*try*) and *can* in (3) – they comprise different number of clauses (two and four, respectively), so (2) has twice the MD of (3). Also note that, although (2) and (3) differ substantially in the number of their constituent clauses, they contain almost the same number of words (19 and 22, respectively). Therefore, if we approached the extent of their modal marking by simply taking into account the number of words in each (as in a keyword analysis), we would be misled in concluding that the two examples were modalized almost to the same extent. MD analysis can be seen as a form of manual collocation analysis, with the span being variable, yet clearly defined, in that the span is the clause.<sup>7</sup> Also, in some respects, the methodology used in this study shows similarities to “collostructional analysis” (Gries & Stefanowitsch 2004; Stefanowitsch & Gries 2003). However, as was mentioned above, this study is not concerned with the frequency of individual lexical modal markers, but with the collective frequency of lexicogrammatical means of modal marking. An equally important difference is that, in this study, the slot is not syntactically fixed, as both lexical and morphological modal marking can occur in different clausal constituents (e.g., different phrases). The similarity between the two approaches rests on their focus on a) elements within (constituent parts of) particular types of structures, and b) meaning, as “the meaning of a construction tends to harmonize with the meanings of the lexical elements that typically occur in it” (Hilpert 2014:392).

However, on its own, MD may not be a dependable indicator, as a high MD may be the result of a number of heavily modalized structures in the sample (see also Gries 2008:404). In such a case, a sample might show a high MD (relative to another sample) despite a large number (even



the majority) of structures in it being modally unmarked. For example, in (4), a single clause contains three modal markings (*perhaps, might, like to*).

- (4) If you live in the Wallingford area and have a railway interest perhaps you might like to join this enthusiastic group and give them a few hours of your time. [CJ7 109]

To take account of such an eventuality, the analysis will combine MD with a second metric, ‘modalization spread’ (MS), which approaches modal marking in an either/or manner, namely, the proportion of modally marked structures in a sample, irrespective of the number of modalizations (Gabrielatos 2010:51-52). Table 2 provides a summary of the characteristics and utility of the two metrics. The combination of the two metrics (MD and MS) provides the modal load (ML) of the target structures. It must be clarified that a single ML score cannot be calculated, as MD and ML examine the extent of modal marking from different perspectives. Rather, the ML is examined in two complementary ways: a) through its graphical representation, by plotting the MD and MS values of the target structures and b) through the clustering of the structures in focus based on the combination of the MD and MS values, using hierarchical cluster analysis (see also Gabrielatos et al. 2010:307-308).

Table 2. Summary of the two complementary ML metrics

	Modal Density	Modalisation Spread
Definition	Average number of modal markings per clause.	Proportion of constructions that carry at least one modal marking.
Expression	Number of modal markings per 100 clauses.	Proportion (percent) of modalized constructions.
Utility	Helps comparison by normalising the complexity of the constructions in the sample.	Corrects for heavily modalized constructions in the sample.

Hierarchical cluster analysis is a family of statistical techniques used in assigning objects (in our case, structures) to groups (clusters) according to their degree of similarity/dissimilarity in relation to a set of variables (in our case, MD and MS scores), with the objective of creating a classification (Everitt 1993:1, 6-7; Gan et al. 2007:3-5, Romesburg 1984:2). This study employs the “agglomerative” method, which begins by treating each structure as a separate cluster, and then assigns the structure to clusters according to the (dis)similarity of their MD and MS values; that is, it is a bottom-up technique (Everitt 1993:55-57; Gan et al. 2007:9).<sup>8</sup> As it makes no initial assumptions of similarity, the agglomerative method is consistent with the empirical nature of the study. The degree of (dis)similarity is measured using the “Euclidian distance”, which computes the square root of the sum of the squares of the pairwise differences in the MD and/or MS values (Gan et al. 2007:326). To account for the different nature of MD and MS, their values are standardized using the z-score (Everitt 1993:38-40, 47; Romesburg 1984:83, 211-213). The distance between clusters (or already established clusters and structures not yet assigned to a cluster) is calculated using “average group linkage”: the average of the distances between all the values in each cluster (Sneath & Sokal 1973:222). This determines the allocation of constructions to (initial or existing) clusters, as well as the conflation of existing clusters into more inclusive ones. This method has been shown to consistently produce clear and useful classifications (Adamson & Bawden 1981:208). The analysis will also examine pairwise differences in MD and MS of adjacent constructions (when ranked in reverse order of their MD or MS values). Differences above 25 percent will be considered sizeable; statistical

significance will be calculated via the Bayesian information criterion (BIC). BIC is calculated using the log-likelihood (LL) value of the frequency difference and the combined size of the compared corpora (N), as follows:  $BIC \approx LL - \log(N)$ . A BIC value of two or above is deemed to show dependable differences (for details, see Wilson 2013). If structures with the same subordinator show similar ML, regardless of the type of structure, this can be seen as support for LG. If different structures show different ML, regardless of their subordinator, this can be seen as support for CxG.

#### 4. Analysis and Discussion

It would be useful to preface the analysis with some quantitative observations, which put the observed ML patterns in perspective. Modal marking in written British English (*baseline*) is anything but infrequent: on average, we can expect two in five s-units to be modalized (MS=40.94), and one modal marking per four clauses (MD=27.72). Regarding the word *if*, the overwhelming majority (85 percent) of *if* tokens can be expected to be subordinators of conditional structures (Gabrielatos 2007b:3); the rest are subordinators of indirect interrogatives and, as part of multi-word units, subordinators of conditional-concessives (*even if*), and structures of comparison (*as if*) (Quirk et al. 1985:1110). This clearly indicates that, syntactically, *if* is not a free agent; its collocation patterns only exist within a small number of structures.<sup>9</sup>

##### 4.1. Examination of Structures as a Whole

Initial useful indications are derived when examining the MD and MS values separately (tables 3 and 4 show the structures in descending order of MD and MS values). *If*-cnd have the highest MD and twice the MD of the *baseline*, and are the only structure with an MD above 50 percent. Apart from *supposing*, which has similar MD with *if*-cnd (the difference is only 15 percent, with BIC=-3.19) all other structures show both sizeable and statistically significant MD differences (BIC>31.35). Differences are less clear when looking at MS. Although *if*-cnd have the highest MS, and, again, about twice the *baseline* MS, they have similar MS with the lower-ranking conditionals (*provided*-cnd, *assuming*-cnd, *unless*-cnd, *supposing*-cnd). In general, however, purely conditional structures (as opposed to conditional-concessives) have higher MD and MS than the *baseline*. Only two conditionals show similar MD values with the *baseline*, and all MS differences but one are sizeable and statistically significant.<sup>10</sup> At this point, it must be stressed that conditionals do not form a unified category in terms of their MD and MS values, as certain members have similar, or even lower, MD and MS when compared to some non-conditional structures. For example, *in\_case*-cnd and *on\_condition*-cnd have similar MD and MS with *if*-q, *whether*-q, *as\_if*-c, and *as\_though*-c. The similarity of the MD and MS values of the two interrogative structures and the two comparison structures would suggest that the ML is not merely due to the presence or absence of *if*, but can be attributed to the semantic nature of the entire structure. However, this is not the case with the two concessive-conditionals: *even\_if*-cc have higher MD and MS than *whether*-cc (66 percent, BIC=28.33 and 36 percent, BIC=1.55, respectively), which points towards an explanation in terms of the respective modal attraction of the subordinators, rather than the semantic nature of conditional-concessives. Still, *as\_if*-c has lower MD and MS than *as\_though*-c, which suggests that the ML does not depend solely on the semantic nature of the subordinator. This is also supported by the observation that constructions with the same subordinator do not consistently show similar MD and MS: *if*-cnd have significantly higher MD and MS than *if*-q, and *whether*-q have significantly higher MD than *whether*-cc (but similar MS).

Table 3. MD of structures

<b>Structure</b>	<b>No. of modalizations</b>	<b>No. of clauses</b>	<b>MD</b>
<i>if-cnd</i>	1449	2607	55.58
<i>supposing-cnd</i>	294	608	48.36
<i>unless-cnd</i>	1242	2852	43.55
<i>assuming-cnd</i>	987	2340	42.18
<i>even_if-cc</i>	1271	3038	41.84
<i>provided-cnd</i>	1214	2929	41.45
<i>whether-q</i>	781	2077	37.60
<i>if-q</i>	829	2312	35.86
<i>as_though-c</i>	845	2437	34.67
<i>in_case-cnd</i>	830	2467	33.64
<i>as_if-c</i>	744	2320	32.07
<i>on_condition-cnd</i>	167	532	31.39
<i>whenever-t</i>	775	2477	31.29
<i>baseline</i>	588	2121	27.72
<i>whether-cc</i>	137	544	25.18
<i>when-t</i>	530	2490	21.29
<i>non-cnd</i>	405	1902	21.19

Table 4. MS of structures

<b>Structure</b>	<b>No. of modalized structures</b>	<b>No. of structures</b>	<b>MS</b>
<i>if-cnd</i>	769	959	80.19
<i>provided-cnd</i>	652	859	75.90
<i>assuming-cnd</i>	546	727	75.10
<i>unless-cnd</i>	736	989	74.42
<i>even_if-cc</i>	731	995	73.47
<i>supposing-cnd</i>	152	213	71.36
<i>whether-q</i>	504	809	62.30
<i>if-q</i>	585	978	59.82
<i>in_case-cnd</i>	560	945	59.26
<i>as_though-c</i>	588	999	58.86
<i>on_condition-cnd</i>	115	205	56.10
<i>as_if-c</i>	551	995	55.38
<i>whether-cc</i>	100	184	54.35
<i>whenever-t</i>	519	959	54.12
<i>baseline</i>	357	872	40.94
<i>when-t</i>	350	902	38.80
<i>non-cnd</i>	295	856	34.46

The above results lead us back to the question of whether the ML of a structure can be ascribed to the semantic preference of the subordinator, irrespective of the type of structure (as would be predicted by LG), or the nature of the structure (e.g., conditional, indirect interrogative), perhaps with variations according to the particular subordinator (as would be predicted by CxG). Dancygier (1998:14) recognizes the importance of this question, stating that “[t]he most controversial aspect of conditional constructions is the contribution of *if* itself to the meaning of the utterance as a whole.” Further indications will be sought in considering the combination of MD and MS values in the cluster analysis, in conjunction with the examination of scatterplots.

Figure 1 shows the intersection of MD and MS values: the more to the top-right corner a value is depicted, the higher the ML (dotted lines indicate the *baseline* values). Here, it is shown more clearly that *if*-cnd have a distinctly higher ML than all other structures, and that most other conditionals share a similar ML. This observation is supported by the examination of ML clustering (Figure 2): the earlier ML values cluster, the more similar they are; the numbers on the top horizontal axis indicate the stage at which a (new) cluster is formed. Most conditionals (including *if*-cnd) are placed in the same penultimate cluster, but *if*-cnd is the last structure to join this cluster (stage 15), and can therefore be seen as being distinct from the other conditionals in terms of their ML. Also, *in\_case*-cnd and *on\_condition*-cnd belong to a different penultimate cluster from other conditionals, and they do not combine with each other until stage 10. Similarly, the two conditional concessives (*even\_if*-cc, *whether*-cc) belong to different penultimate clusters, and the two comparison structures (*as\_if*-c, *as\_though*-c) only combine in stage 6 – only the two indirect interrogatives (*if*-q, *whether*-q) cluster together fairly quickly (stage 4). These patterns would seem to challenge explanations in terms of the nature of the structure. At the same time, the two structures with *if* (*if*-cnd, *if*-q) belong to different penultimate clusters, while the two structures with *whether* (*whether*-cc, *whether*-q) only combine just before the penultimate cluster. In sum, the clustering of MD and MS values does not provide adequate support for a clear interpretation in terms of either the semantic preference of the subordinator, or the nature of the structure.

The ML analysis has also provided evidence that unmodalized conditionals cannot be usefully treated as synonymous to *when*-t or *whenever*-t, as is proposed in some classifications of conditionals, which also indirectly treat these structures as if they were typically unmodalized (e.g., Athanasiadou & Dirven 1996:617, 1997:62; see also Palmer 1990:174-175). In Figure 2, *when*-t, *whenever*-t and *if*-cnd belong to completely different clusters. Also, the analysis has shown that *when*-t and *whenever*-t are not routinely unmodalized, quite the contrary: more than half of *whenever*-t and more than one-third of *when*-t can be expected to be modalized (MS=54.12 and MS=38.80, respectively) – see Table 4 above.

Figure 1: ML of Structures

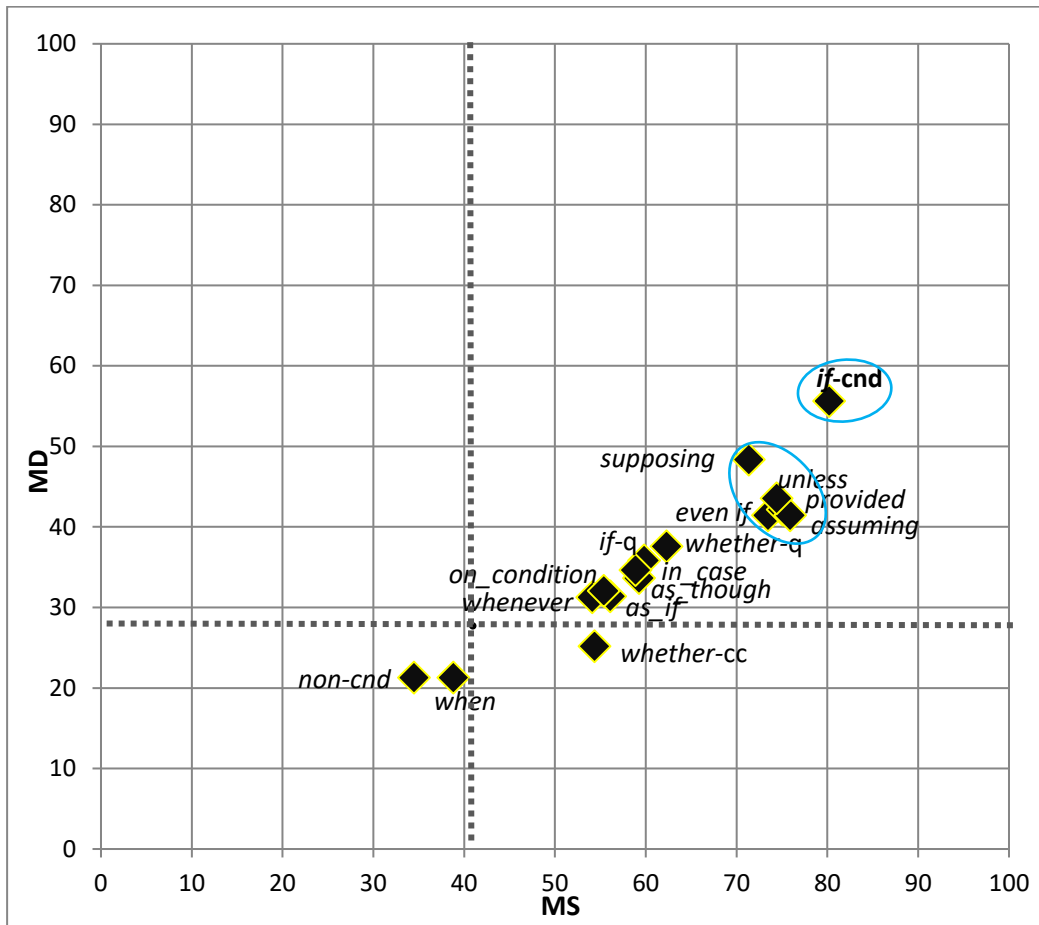
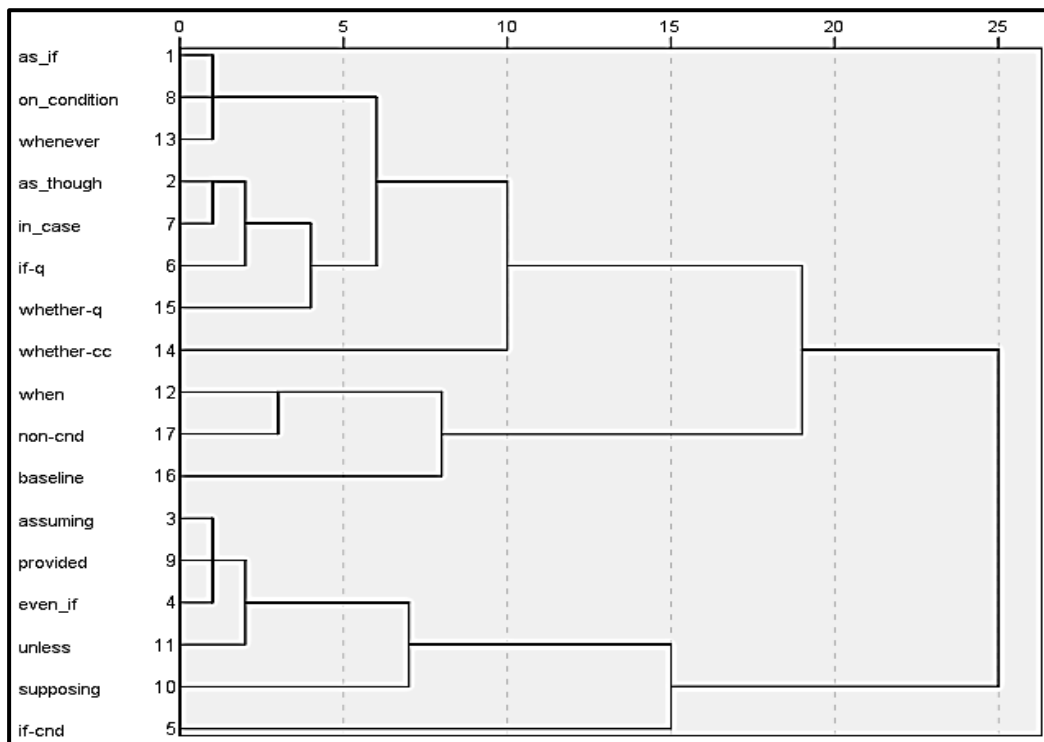


Figure 2. ML Clusters: structures



At this point we need to consider that the ML of bipartite structures may not reflect the semantic preference of their subordinator within the usual short collocation span of 4-5 words, and/or that any putative colligations of the subordinators are restricted to the subordinate part of the structures. One way towards resolving the issue of the influence of the subordinator is to reconsider the notion of semantic preference and, by extension, the notion of collocation as a determinant of meaning. LG posits that word-forms derive meaning through their collocational patterning (i.e., from the meaning of their collocates). However, the argument can be reversed: it can be equally argued that it is the particular meaning(s) of an item which determine its collocational patterns (see Leech 1981:17; Lyons 1977:613). What is more, this LG tenet is not devoid of circularity: if we start from a null position (i.e., that no word has meaning on its own, but derives it from its collocates) and apply the same principle to its collocates, we end up with frequent co-occurrence of meaningless forms. Lyons (1977:265) cautions against adopting either of the two extreme views:

We must not go from the one extreme of saying that the collocations of a lexeme are determined by its meaning or meanings (where meaning is defined independently of syntagmatic considerations) to the other extreme of defining the meaning of a lexeme to be no more than the set of its collocations.

However, even a moderate view of the relation between collocability and meaning, that is, that the meaning of a word affects, but does not fully determine, the collocates it attracts, does not help resolve the question of why, for instance, the ML of *if*-cnd differs significantly from that of *if*-q. An approach that seems more promising in resolving the circularity described above is the examination of collocation within “grammatical matrices” (Mitchell 1971:65). This approach is consistent with the tenets of CxG: instead of positing that the meaning of a word is reflected in its collocates (irrespective of grammatical context), we can posit that the meaning of the collocates a word attracts (its semantic preference) depends on the meaning this word has when it is part of a given construction (see also Goldberg 1995:12-13; Tomasello, 2003:160-161). That is, it could be argued that the ML differences of structures with the same subordinator, such as *if*-cnd and *if*-q, are due to polysemy. This approach allows us to hypothesize, for example, that the word-form *if* in conditionals has a different meaning from the word-form *if* in indirect interrogatives (henceforth *if*<sub>cnd</sub> and *if*<sub>q</sub>, respectively), with the latter being synonymous with *whether* in indirect interrogatives (*whether*<sub>q</sub>). As a result of their different senses, the hypothesis would go, *if*<sub>cnd</sub> and *if*<sub>q</sub> have different semantic preference patterns – in this case, higher or lower attraction to modal markers. This hypothesis will be examined in the next section, which looks at the ML of the subordinate and matrix parts of the structures in focus.

#### 4.2. ML in Subordinate and Matrix Parts

This section will examine the ML of the subordinator within the syntactic unit it introduces (i.e., the subordinate part of the structure). This is compatible with both LG and CxG: in LG, extended units of meaning are established by examining patterns within a short span around the core (Sinclair 1996:75-90); in CxG, constructions can be sub-parts of more complex constructions (Goldberg 1995:78). In addition, these values will also be compared to the ML of the matrix parts of these structures, which will determine the extent to which each part contributes to the ML of the whole structure (Tables 5 and 6).

Table 5. Subordinate parts: MD

<b>Construction</b>	<b>No. of modalizations</b>	<b>No. of clauses</b>	<b>MD</b>
<i>if-q_S</i>	529	1340	39.48
<i>if-cnd_S</i>	487	1235	39.43
<i>as_if-c_S</i>	462	1232	37.50
<i>as_though-c_S</i>	459	1294	35.47
<i>whether-q_S</i>	398	1139	34.94
<i>supposing-cnd_S</i>	117	352	33.24
<i>even_if-cc_S</i>	393	1313	29.93
baseline	588	2121	27.72
<i>on_condition-cnd_S</i>	76	286	26.57
<i>in_case-cnd_S</i>	298	1180	25.25
<i>provided-cnd_S</i>	324	1382	23.44
<i>whenever-t_S</i>	254	1112	22.84
<i>assuming-cnd_S</i>	209	960	21.77
<i>unless-cnd_S</i>	269	1351	19.91
<i>whether-cc_S</i>	23	230	10.00
<i>when-t_S</i>	96	1112	8.63

Table 6. Subordinate parts: MS

<b>Structure</b>	<b>No. of modalized structures</b>	<b>No. of structures</b>	<b>MS</b>
<i>if-q_S</i>	452	978	46.22
<i>if-cnd_S</i>	400	959	41.71
baseline	357	872	40.94
<i>as_if-c_S</i>	400	995	40.20
<i>whether-q_S</i>	325	809	40.17
<i>as_though-c_S</i>	387	999	38.74
<i>supposing-cnd_S</i>	81	213	38.03
<i>even_if-cc_S</i>	331	995	33.27
<i>on_condition-cnd_S</i>	65	205	31.71
<i>in_case-cnd_S</i>	253	945	26.77
<i>provided-cnd_S</i>	226	859	26.31
<i>whenever-t_S</i>	238	959	24.82
<i>unless-cnd_S</i>	221	989	22.35
<i>assuming-cnd_S</i>	153	727	21.05
<i>whether-cc_S</i>	22	184	11.96
<i>when-t_S</i>	74	902	8.20

The ML comparison of subordinate parts would seem to provide some support for the semantic influence of the subordinator, but far from conclusively (Table 5). The subordinate parts of *if*-cnd, *if*-q, and *as\_if*-c have very similar MD and MS, but this is not the case for *even\_if*-cc. It could be argued that this does not pose a problem for LG, as it could be posited that *as\_if* and *even\_if*, although containing the word *if*, are individual lexical items, and can be expected to have their own semantic preferences. This argument, however, is not supported by the ML of the two constructions with *whether*, which exhibit an extremely large MD difference: *whether*-q has more than three times the MD and MS of *whether*-cc (BIC 42.79 and 36.54, respectively). On the other hand, there are no consistent patterns in terms of construction family. For example, the subordinate parts of most conditionals (*assuming*-cnd, *in case*-cnd, *on condition*-cnd, *provided*-cnd) have MD values no higher than the *baseline*. The picture is equally unclear when looking at MS values (Table 6). Neither all structures of the same type, nor all subordinate parts with the same subordinator, have similar MS. For example, *whether*-q has almost four times the MD and MS of *whether*-cc (BIC 42.79 and 36.54, respectively). Finally, we need to briefly discuss an interesting pattern in the ML of subordinate parts: most values can be interpreted as showing degrees of repulsion (rather than attraction) to modality (Renouf & Banerjee 2007), as all but two MS values, and more than half of the MD values, are below the baseline. The lack of consistent indications supporting either the influence of the subordinator or the constructional nature is more clearly demonstrated in the scatterplot and clusters in terms of ML (Figures 3 and 4). Most structures of the same type (conditionals, conditional-concessives, indirect interrogatives) do not cluster together (with the exception of comparison structures, which meet at stage 3), but neither do structures with the same subordinator (*if*, *whether*). We will return to this pattern at the end of this section.

Figure 3. ML of subordinate parts

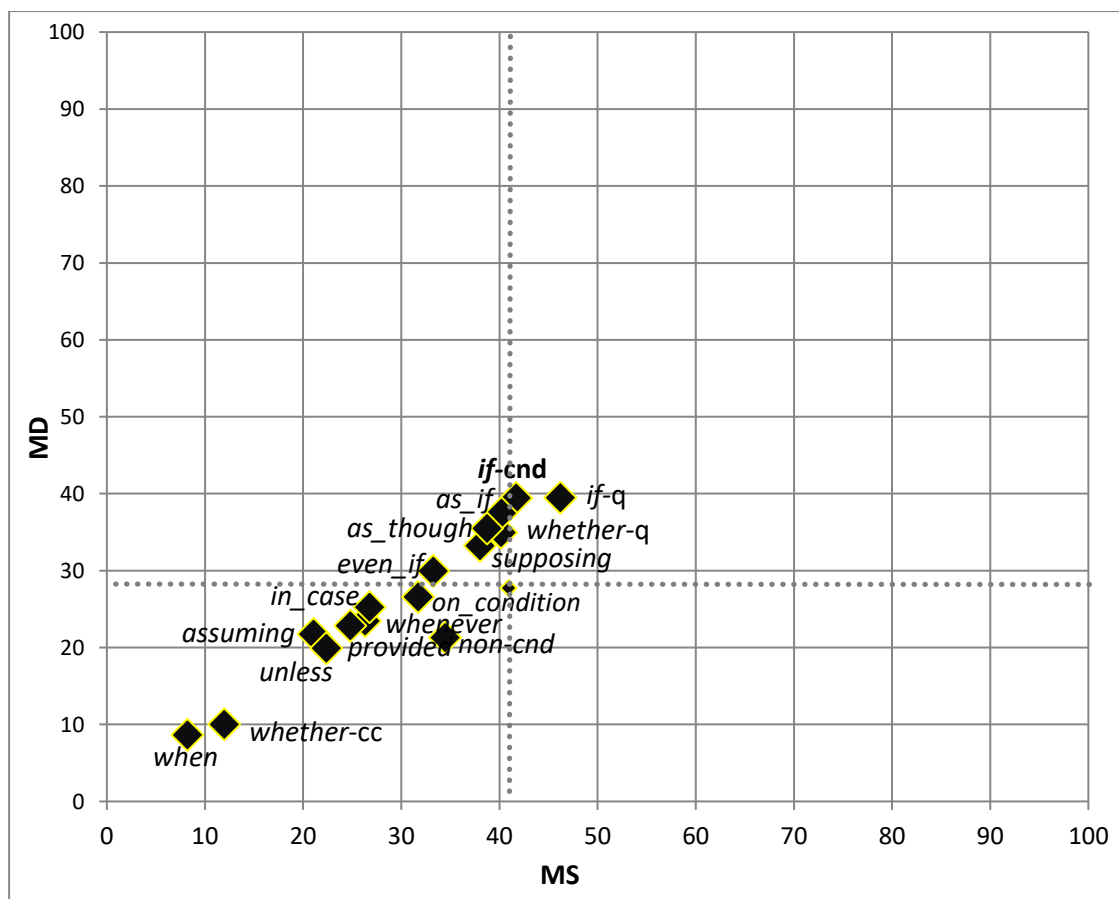
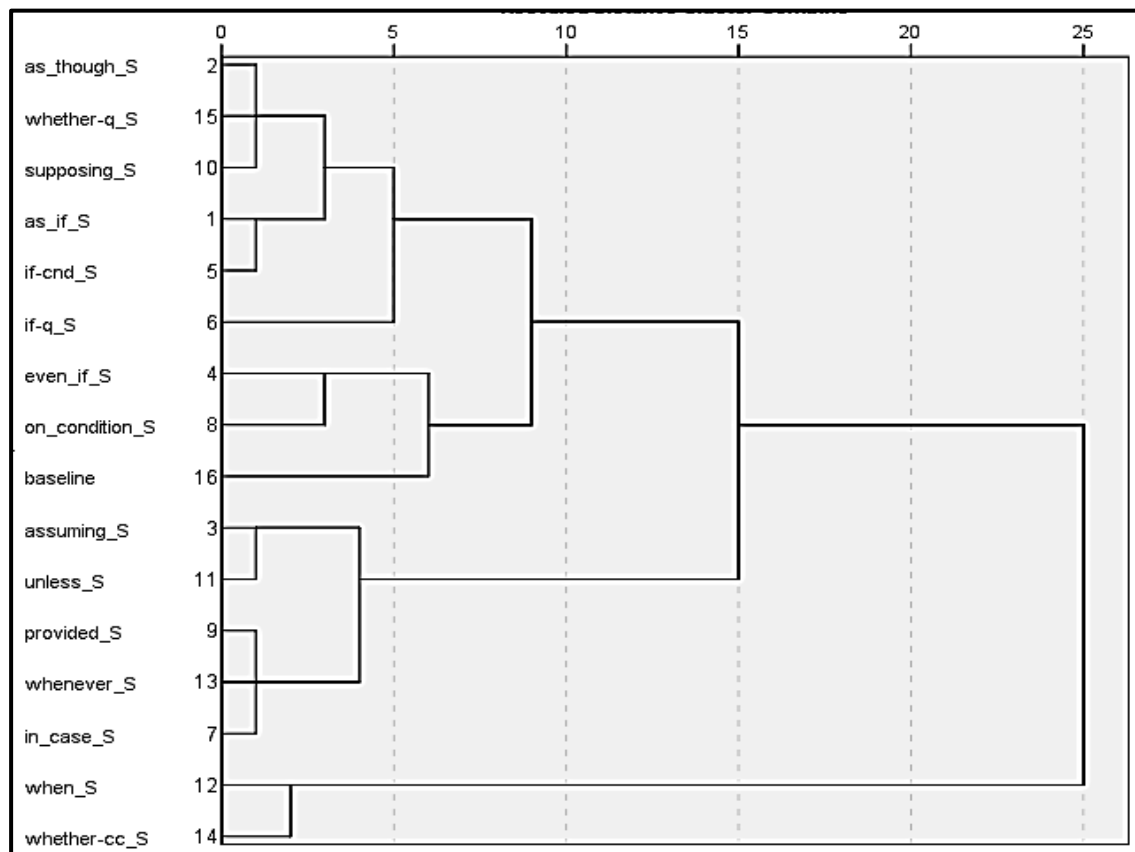




Figure 4. ML clusters: subordinate parts



At this point we will re-visit the hypothesis that polysemy can explain the ML patterns of subordinate parts. The subordinators *if* and *whether* will be used as a case study, as they are found in four different structures (*if-cnd*, *if-q*, *whether-cc*, *whether-q*). For the sake of the argument we will posit different meanings for *if* and *whether* according to the structure they are part of (*if<sub>cnd</sub>*, *if<sub>q</sub>*, *whether<sub>cc</sub>*, *whether<sub>q</sub>*), while also positing that the ML of the subordinate part of each structure is due to the semantic preference of each of the four different form-sense combinations. For the hypothesis to stand, the subordinate parts of *if-cnd* must have different ML from those of *if-q*, with the same holding between *whether-cc* and *whether-q*. Table 7, which summarizes the relevant results of the ML analysis of subordinate parts, shows that this is not consistently the case. The symbol ‘>’ indicates sizeable and statistically significant differences, whereas ‘=’ indicates similar MD and MS values.

Table 7. Clustering of ML of subordinate parts with *if* and *whether*

Metric	Clustering of subordinate parts		
MD	<i>if-cnd</i> = <i>if-q</i> = <i>whether-q</i>	>	<i>whether-cc</i>
MS	<i>if-cnd</i> = <i>if-q</i> = <i>whether-q</i>	>	<i>whether-cc</i>
MD+MS	<i>if-cnd</i> = <i>if-q</i> = <i>whether-q</i>	>	<i>whether-cc</i>

Although *whether-cc* and *whether-q* have significantly different ML (*whether-q* has almost four times the MD and MS of *whether-cc*), *if-cnd* and *if-q* have identical MD and similar MS (Tables 5 and 6). Translated into LG notions, Table 7 shows that, as regards modality, *whether<sub>cc</sub>* and *whether<sub>q</sub>* show different semantic preferences, but *if<sub>cnd</sub>* and *if<sub>q</sub>* have almost identical ones. The latter ML similarity is consistent with Dancygier's (1998:195) conclusion that "the *if* of embedded questions is clearly related to the *if* of conditional constructions." Therefore, the hypothesis that differences in ML patterns between structures sharing the same subordinator can be ascribed to these subordinators being polysemous does not seem to be supported by the analysis so far. The above results seem to support the interpretation that any ML similarities/differences are due to the respective meaning components of a structure, rather than solely the semantic preference of its subordinator. For example, the differences in the ML of the subordinate parts of *whether-q* and *whether-cc* (Table 7 above) can be ascribed to the different structures they are components of, rather than (only) the putative polysemy of *whether*. In this regard, the analysis does not support treating the word *if* as entering into a variety of linear patterns, each reflecting the various senses of *if*. What is more, given that about 85 percent of the instances of *if* are in conditionals, a grammar-independent collocation analysis of *if* would predominantly reveal the words and senses co-occurring with *if* within conditionals.

The remainder of this section will examine the ML of structures while also taking into account the ML of their two parts. More specifically, it will discuss the relative effect of the ML of subordinate and matrix parts of the structures in focus, by examining the ratio of MD and MS in subordinate and matrix parts, respectively. For example, the MD ratio will be derived by dividing the MD of the subordinate part with the MD of the matrix part. Figure 5 depicts the MD and MS ratios plotted against each other. The intersection of the dotted lines shows the position of a balanced ML, that is, when subordinate and matrix parts have the same MD and MS (ratio=1). Structures depicted in the top right-hand quarter show higher ML in the subordinate part, whereas those depicted in the bottom left-hand quarter show higher ML in the matrix part. Pertinent to the discussion is the marked difference in ML balance between the structures with *if* and *whether*. The cluster analysis (Figure 6) confirms the lack of consistent patterning of ratios in terms of subordinators: *if-cnd* belongs to a completely different cluster from *if-q*, as do *whether-cc* and *whether-q*. On the other hand, there is some consistent clustering in terms of structure types, but not in early clusters. Conditionals, indirect interrogatives and comparison structures only meet very late (in the penultimate cluster), with early clusters consisting of a mix of constructions. The above do not seem to support an interpretation in terms of the collocational attraction of the words *if* or *whether*. Rather, the clustering supports approaching the subordinate parts as constructions in their own right, while also accepting them as an integral part of the more complex construction they are a component of. More precisely, the ML of the subordinate part in a bi-partite construction can be understood to be in line with the semantic function of the construction as a whole, while the balance of the ML between its subordinate and matrix part can be seen to reflect the semantic link between them (itself a component of constructional meaning).

Figure 5. Ratios of MD and MS in Subordinate and Matrix Parts

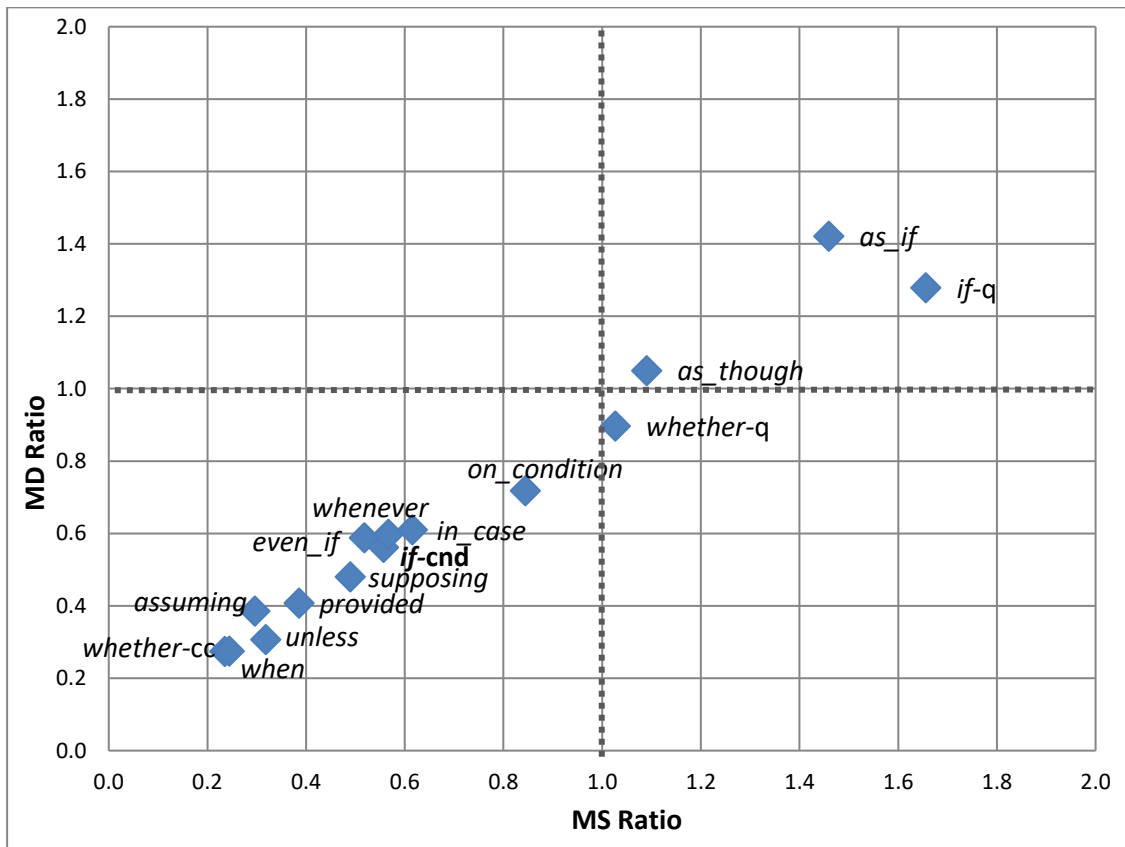
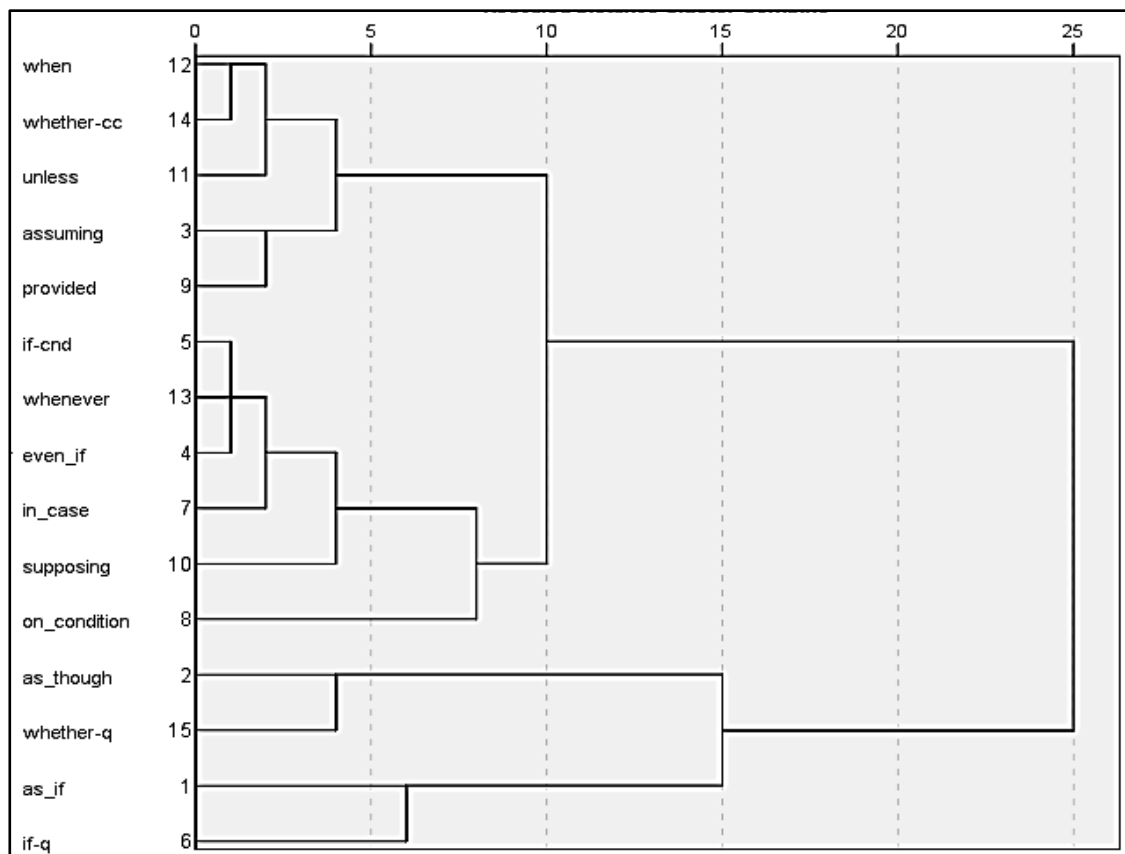


Figure 6. ML Clusters: subordinate/matrix ratio



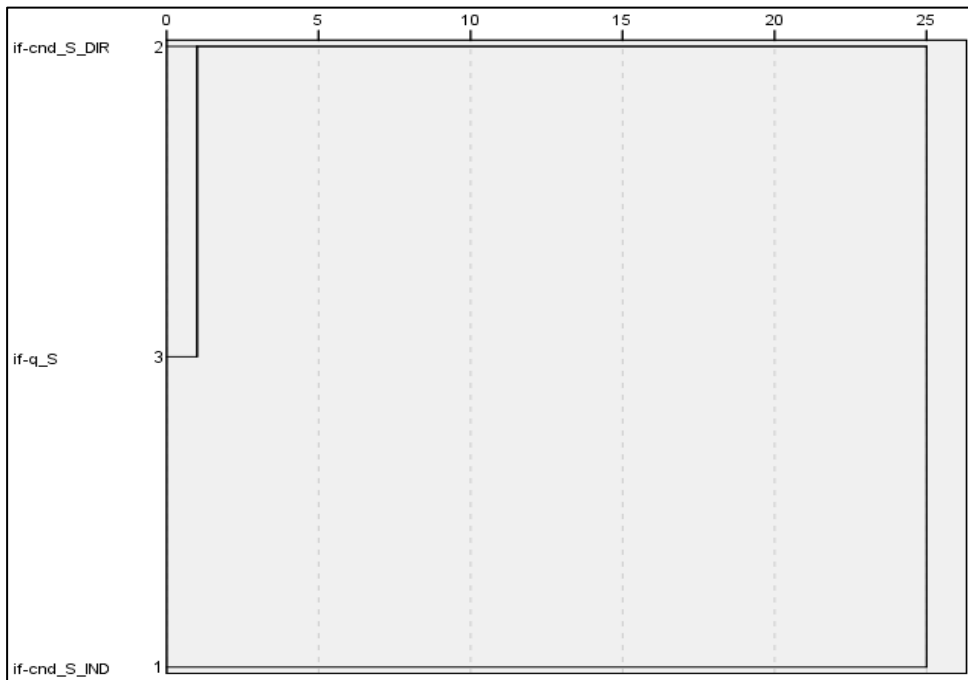
### 4.3. Focus on Types of *If*-conditionals

The final part of the ML analysis will focus on the two main sub-types of *if*-cnd: “direct” (DIR) and “indirect” (IND) (Quirk et al. 1985:1088-1097). In DIR, the realization of the content of the apodosis – that is, the action, situation, prediction, inference etc. expressed in it – depends on the realization, actuality or factuality of the content of the protasis. For example, in (5) the prediction expressed in the apodosis is directly dependent on the actualization of the premise in the protasis. In IND, it is not the content of the apodosis that is contingent on the protasis, but either the relevance of its uttering, or the wording/clarity of its content. This is illustrated in (6): the content of the evaluative comment in the apodosis is not contingent on the listener’s understanding or agreement; rather the protasis functions as a hedged introduction to the comment in the apodosis. The semantic difference of the protasis-apodosis link in DIR and IND is mirrored by their syntactic difference: the protasis is an adjunct in DIR, but a disjunct in IND (Quirk et al. 1985:612-631, 1072).

- (5) If we can assemble a package of cash, stock options, and newly issued shares as a good inducement, I think we’ll convince the key manager and he’ll persuade the others to sell. [FPB 108]
- (6) He’s not a bad sort for a brother if you know what I mean [AN7 3257]

When examining the whole structures, DIR and IND *if*-cnd have very similar ML, despite their parts having different syntactic links. An LG interpretation would posit that this is due to the semantic preference of *if*, albeit requiring us to accept a much wider collocation span than posited in LG. Even when looking only at the ML of subordinate parts (i.e., the more immediate vicinity of *if*) the MD and MS differences between the subordinate parts of DIR and IND are only marginally sizeable (about 25 percent) and not statistically significant (BIC  $-4.45$  and  $-5.00$ , respectively). Therefore, it does not seem warranted to posit more senses for the putative polysemous *if* (i.e., adding *if*<sub>cnd-DIR</sub> and *if*<sub>cnd-IND</sub>), particularly as the hypothesis of a polysemous *if* (i.e., *if*<sub>cnd</sub> and *if*<sub>q</sub>) was not supported by the results (see section 4.2 above). This conclusion is strengthened by the clustering of the ML of the subordinate parts of the three structures introduced by *if* on its own: *if*-cnd-DIR, *if*-cnd-IND, and *if*-q. As shown in Figure 7, *if*-cnd-DIR immediately cluster with *if*-q, while it takes twenty-five iterations for them to join *if*-cnd-IND.

Figure 7. *If*-structures: ML of subordinate parts



The very different ML of the subordinate parts of *if-cnd-DIR* and *if-cnd-IND* is explained when we examine the MD and ML ratios in their subordinate and matrix parts. As seen in Figure 8, IND have a higher ratio than DIR, which indicates that the similar ML of the whole structures hides differences in the distribution of ML in their subordinate and matrix parts: DIR have a much more uneven ML distribution in their respective parts than IND (Figure 9).

Figure 8. ML ratios in DIR and IND

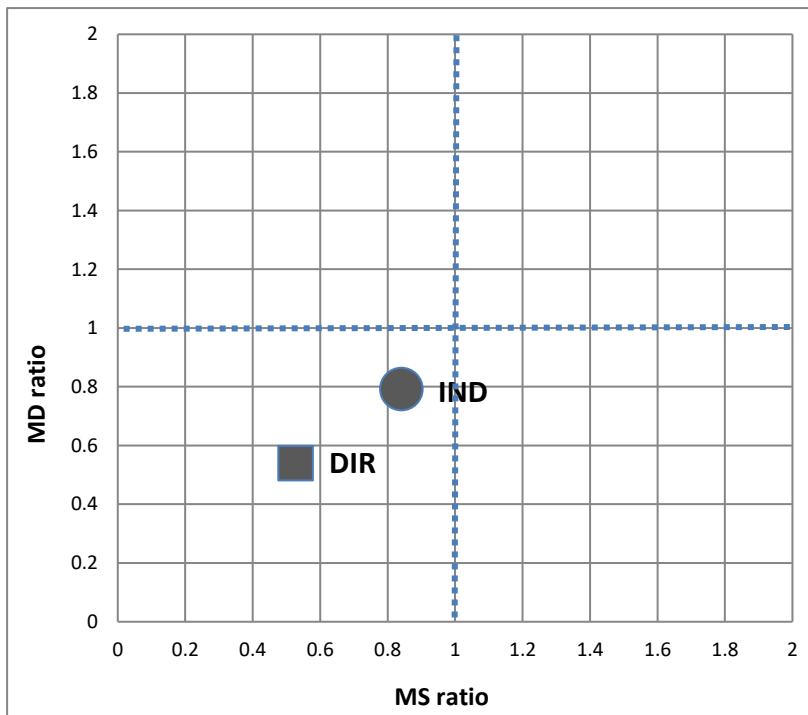
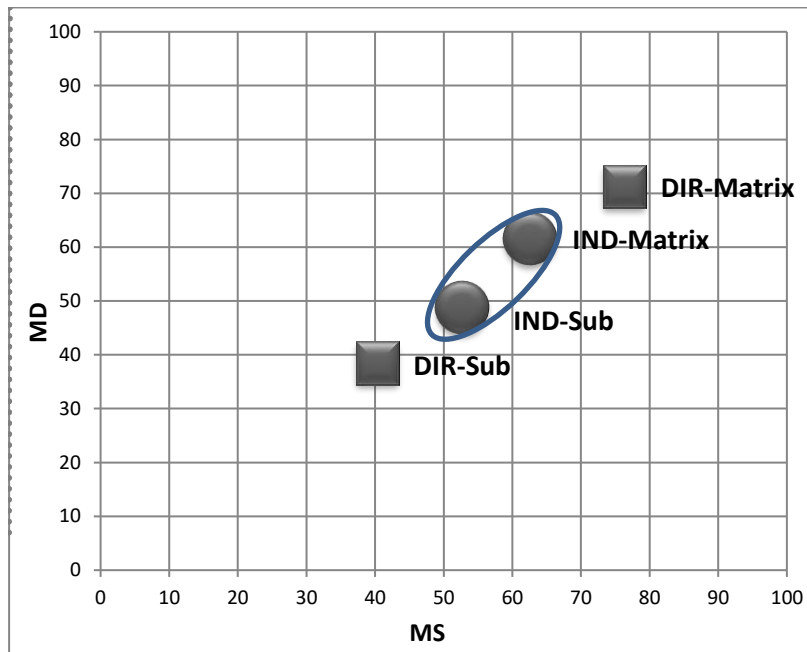


Figure 9. *If*-conditionals: ML of subordinate and matrix parts of DIR and IND



## 5. Conclusions

This study provided a quantitative measure of the connection between conditionals and modality. The analysis revealed that the ML of conditionals, and *if*-conditionals in particular, is significantly higher than the baseline and non-conditional constructions, as well as most, but not all, non-conditional constructions. More importantly, *if*-conditionals showed a distinctly higher ML than other conditional constructions. Further studies could usefully examine the ML of structures in which *if* and *as if* act as complements (as suggested by López-Couso & Méndez-Naya 2001, 2014, 2015), and any implications this may have for the classification of *if*-conditionals. It would also be useful to examine the frequency of individual modal markers in relation to the different types of *if*-conditionals.

The analysis also showed that the observed ML patterns cannot be fully explained by recourse to the type or semantic nature of either the subordinator or the structure, when each is considered on its own. An explanation in terms of the semantic preference of the subordinator (which emerges from its collocation patterns), that is, an explanation in terms of LG, would need to posit a polysemous *if* – a hypothesis that was not supported by the analysis. More importantly, the putative multiple senses of *if* would need to be defined with reference to particular grammatical structures – for example, *if*<sub>cond</sub> would need to be defined as ‘the sense of *if* when it is the subordinator in conditional structures’. In addition, the collocation span itself would need to be defined syntactically (i.e., co-occurrence within the subordinate part of a structure), rather than, as LG posits, lexically (i.e., co-occurrence within a particular number of word-forms around the node). This is because *if* is used in a small number of structures; therefore, a syntax-independent collocational analysis of *if* would not reveal useful patterns, as it would provide a homogenized picture of its semantic preference – a picture that would be further distorted by the influence of the most frequent structure that *if* is part of (*if*-conditionals). However, neither would an interpretation in terms of the nature of a structure provide a

satisfactory explanation, as it would fail to account for the different ML patterns of the same type of structure with by different subordinators.

The analysis seems to support the interpretation that the ML is the result of the combined effect of the subordinator and the type of structure. That is, the results support an explanation that is fully consistent with the tenets of CxG, as the subordinator can be a component of different constructions. More so, CxG accounts for the interaction between all components of a construction through the “principle of no synonymy” (Goldberg 1995:67); that is, morphosyntactic and/or lexical differences between constructions lead to semantic/pragmatic differences, and vice versa. In this light, CxG clearly demonstrates a stronger explanatory power, since a construction specifies morphosyntactic, lexical and semantic attributes (among others), which it also treats as having equal importance. In other words, the observed ML patterns cannot be seen to “belong” (Hunston & Francis 2000:1-2) to different senses of *if*. Rather, it seems more plausible to explain the ML patterns in terms of similarities and differences between different constructions, which are, in turn, influenced by the meaning of their respective components (which are themselves constructions). This is not to say that the constructs of collocation and its semantic extension, semantic preference, are not useful. On the contrary, they can be both incorporated into the lexical and semantic components of a construction. However, this would entail treating collocation as a lexicogrammatical, rather than a purely lexical, relation (Gabrielatos 2018:266; see also Hughes 2018). For example, in the case of *if*-constructions, the collocation pattern and resulting semantic preference of the subordinator *if* would be one of the constructional meaning components interacting with higher level components – i.e., those relating to the function of the construction as a whole, and each of its two constituent parts. It can be posited that the semantic component specifies the probability that particular (combinations of) modality types can be expected to be marked within particular elements of a construction – for example, within the protasis and apodosis of an *if*-conditional. The analysis has also provided indications that constructions belonging to different families may also share characteristics. For example, *if*-cnd and *if*-q have comparable ML in their subordinate parts – a feature which can be attributed to the semantic nature of *if* (seen as a construction in its own right). At the same time, not all conditional constructions exhibited similar ML, whereas sub-types of *if*-conditionals (DIR and IND) showed diverging MLs in their constituent parts. The above seem to suggest that constructional family members are further differentiated by a number of interacting constructional elements – in our case, the subordinator and the type of conditional.

In conclusion, the analysis has provided strong indications that CxG rather than LG can account for the ML patterns found in this study. It was also shown that ML patterns are sensitive to different combinations of constructional attributes, as would be predicted by the principle of no synonymy. This suggests that subordinators, rather than being the core of a lexical item (as posited by LG), are more usefully seen as one of many components defining a construction (as posited by CxG). Consequently, if a semantic attraction of the subordinator can be posited, this has to be understood as being influenced by the type of construction that the subordinator is used in. In this light, semantic preference could be more usefully treated as part of a construction’s semantic component.

## Notes

1. Although the term ‘construction’ is sometimes used in a general sense (e.g., Huddleston & Pullum 2002; Quirk et al., 1985), to avoid confusion, it will only be used here when referring to the units posited by Construction Grammar; in all other instances (barring direct quotations), the term ‘structure’ will be used.
2. All examples are from the samples of the written BNC used in this study.
3. The BNC sampler is a shorter version of the BNC, comprising two one-million-word sub-corpora of written and spoken British English (<http://www.natcorp.ox.ac.uk/getting/sampler.html>). FLOB (Hundt et al. 1999) is a representative corpus of written British English comprising texts published in the early 1990s (<http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html>).
4. I am indebted to Stefan Evert and Neil Millar for their invaluable help in constructing the regular expressions.
5. An s-unit is a sequence of words delimited by sentence-boundary markers (e.g., full-stop, question mark, exclamation mark) (Sperberg-McQueen & Burnard 2019).
6. Halliday’s use of the metric is, in turn, an adaptation of the initial definition of lexical density, that is, the percentage of the tokens in a text that are content words (Ure, 1971).
7. See also Kim & Choi (1999) and Gabrielatos (2007a), who use the text as the variable collocation span.
8. The cluster analysis was carried out using the SPSS statistical software.
9. It has also been suggested that, in some cases, *if* and *as if* can be analysed as complementizers (López-Couso & Méndez-Naya, 2001, 2014, 2015). However, this does not affect the present argument, as, even with this addition, the number of structures that *if* exists within is still small.
10. One conditional structure (*on\_condition*-cnd) shows a small and non-significant MD difference with *baseline* (13.2 percent, BIC=-5.92), while another (*in\_case*-cnd) shows a small but significant MD difference (21.3 percent, BIC=4.59). The only non-significant MS difference was between *on\_condition*-cnd and *baseline* (37 percent, BIC=-1.18).

## References

- Adamson, George W. & David Bawden. 1981. Comparison of hierarchical cluster analysis techniques for automatic classification of chemical structures. *Journal of Chemical Information and Computer Sciences* 21. 204-209.
- Athanasiadou, Angeliki & Rene Dirven. 1996. Typology of *if*-clauses. In Eugene H. Casad (ed.), *Cognitive linguistics in the Redwoods: The expansion of a new paradigm in linguistics*, 609-654. Berlin: Mouton de Gruyter.
- Athanasiadou, Angeliki & Rene Dirven. 1997. Conditionality, hypotheticality, counterfactuality. In Angeliki Athanasiadou & Rene Dirven (eds.), *On conditionals again*, 61–96. Amsterdam: John Benjamins.
- Baeza-Yates, Ricardo & Berthier Ribeiro-Neto. 1999. *Modern information retrieval*. London: Addison Wesley.
- Ball, Catherine N. 1994. Automated text analysis: Cautionary tales. *Literary and Linguistic Computing* 94. 265-302.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Cheng, Winnie, Chris Greaves, John Sinclair & Martin Warren. 2009. Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics* 30(2). 236-252.
- Chowdhury, Gobinda G. 2004. *Introduction to modern information retrieval*. 2nd edn. London: Facet Publishing.
- Comrie, Bernard. 1986. Conditionals: A typology. In Elizabeth Closs Traugott, Alice Ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson (eds.), *On conditionals*, 77-99. Cambridge: Cambridge University Press.
- Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dancygier, Barbara. 1998. *Conditionals and prediction: Time, knowledge and causation in conditional constructions*. Cambridge: Cambridge University Press.



- Dancygier, Barbara & Eve Sweetser. 2005. *Mental spaces in grammar: Conditional constructions*. Cambridge: Cambridge University Press.
- Everitt, Brian S. 1993. *Cluster analysis*. 3rd edn. London: Edward Arnold.
- Fillmore, Charles J. 1986. Varieties of conditional sentences. *Proceedings of the Third Eastern States Conference on Linguistics*, 163-182. Columbus, Ohio: Ohio State University Department of Linguistics.
- Fillmore, Charles J. 1998. The mechanisms of "Construction Grammar." In Shelley Axmaker, Annie Jaissner & Helen Singmaster (eds.), *General Session and Parasession on Grammaticalization*. Proceedings of the Fourteenth Annual Meeting of Berkeley Linguistics Society, February 13-15, 1998, 35-55. Berkeley: Berkeley Linguistics Society.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64(3). 501-538.
- Firth, John R. 1957. Modes of meaning. *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Firth, John R. 1968. A synopsis of linguistic theory. In Frank R. Palmer (ed.), *Selected papers of J.R. Firth 1952-59*, 168-205. London: Longmans.
- Fried, Mirjam & Jan-Ola Östman. 2004. Construction Grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction Grammar in a Cross-language Perspective*, 11-86. Amsterdam: John Benjamins.
- Gabrielatos, Costas. 2005. Elliptical and discontinuous *if*-conditionals: Co-text, context, inference, and intuitions. Paper given at Corpus Linguistics 2005, University of Birmingham, UK, 14-17 July 2005.
- Gabrielatos, Costas. 2006. Corpus-based evaluation of pedagogical materials: *If*-conditionals in ELT coursebooks and the BNC. Paper given at 7th Teaching and Language Corpora Conference, University Paris 7 Denis Diderot, Paris, France, 1-4 July 2006.
- Gabrielatos, Costas. 2007a. Selecting query terms to build a specialized corpus from a restricted-access database. *ICAME Journal* 31. 5-43.
- Gabrielatos, Costas. 2007b. *If*-conditionals as modal colligations: A corpus-based investigation. In Matthew Davies, Paul Rayson, Susan Hunston & Pernilla Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference: Corpus Linguistics 2007*. Birmingham: University of Birmingham.
- Gabrielatos, Costas. 2010. A corpus-based examination of English *if*-conditionals through the lens of modality: Nature and types. Lancaster, UK: Lancaster University PhD thesis.
- Gabrielatos, Costas. 2018. The lexicogrammar of *BE interested*: description and pedagogy. In Sebastian Hoffmann, Andrea Sand, Sabine Arndt-Lappe & Lisa Marie Dillmann (eds.), *Corpora and lexis*. Language and Computers, Vol. 81, 240-276. Leiden: Brill.
- Gabrielatos, Costas, Eivind Torgersen, Sebastian Hoffmann & Sue Fox. 2010. A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics* 38(4). 297-334.
- Gan, Guojun, Chaoqun Ma & Jianhong Wu. 2007. *Data clustering: Theory, algorithms and applications*. Philadelphia: ASA-SIAM.
- Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele & Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80(3). 532-568.
- Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403-437.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97-129.
- Halliday, Michael A.K. 2004. *An introduction to Functional Grammar*. 3rd edn. London: Arnold.
- Hilpert, Martin. 2014. Collocation analysis: Measuring associations between constructions and lexical elements. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 391-404. Amsterdam: John Benjamins.

- Hoey, Michael. 1991. *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hoey, Michael. 1997. From concordance to text structure: New uses for computer corpora. In Patrick James Melia & Barbara Lewandoska (eds.), *PALC '97: Practical Applications in Language Corpora*, 2-23. Łódź: Łódź University Press.
- Hoey, Michael. 2005. *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoffmann, Sebastian, Stefan Evert, Nick Smith, David Lee & Ylva Berglund-Prytz. 2008. *Corpus Linguistics with BNCweb - a practical guide*. Frankfurt am Main: Peter Lang.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hughes, Jennifer. 2018. Neural evidence for a single lexicogrammatical processing system. Paper given at Corpus Approaches to Lexicogrammar 2018, Edge Hill University, UK, 16 June 2018.
- Hundt, Marianne, Andrea Sand & Rainer Siemund. 1999. *Manual of information to accompany the Freiburg-LOB Corpus of British English 'FLOB'*. Freiburg: Englisch Seminar, Albert-Ludwigs-Universität Freiburg.
- Hunston, Susan & Gill Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Kim, Myoung-Cheol & Key-Sun Choi. 1999. A comparison of collocation-based similarity measures in query expansion. *Information Processing & Management* 35(1). 19-30.
- Leech, Geoffrey. 1981. *Semantics*. London: Penguin Books.
- Leech, Geoffrey. 2004. *Meaning and the English verb*. 3rd edn. London: Longman.
- López-Couso, María José & Belén Méndez-Naya. 2001. On the history of *if*- and *though*-links with declarative complement clauses. *English Language and Linguistics* 5(1). 93-107.
- López-Couso, María José & Belén Méndez-Naya. 2014. The use of *if* as a declarative complementizer in English: some theoretical and empirical considerations. In Alejandro Alcaraz-Sintes & Salvador Valera-Hernández (eds.), *Diachrony and synchrony in English Corpus Linguistics*, 85-107. Bern: Peter Lang.
- López-Couso, María José & Belén Méndez-Naya. 2015. Secondary grammaticalization in clause combining: From adverbial subordination to complementation in English. *Language Sciences* 47. 188-198.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Mitchell, T.F. 1971. Linguistic 'goings on'. Collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum* 2. 35-69.
- Nuyts, Jan. 2001. *Epistemic modality, language, and conceptualisation: A cognitive-pragmatic perspective*. Amsterdam: John Benjamins.
- Palmer, Frank R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- Palmer, Frank R. 1987. *The English verb*. 2nd edn. London: Longman.
- Palmer, Frank R. 1990. *Modality and the English modals*. 2nd edn. Cambridge: Cambridge University Press.
- Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins.
- Perkins, Michael R. 1983. *Modal expressions in English*. London: Frances Pinter.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive grammar of the English language*. London: Longman.
- Renouf, Antoinette & Jayeeta Banerjee. 2007. Lexical repulsion between sense-related pairs. *International Journal of Corpus Linguistics* 12(3). 415-444.
- Romesburg, H. Charles. 1984. *Cluster analysis for researchers*. Belmont, CA: Wadsworth.
- Sinclair, John. 1991. *Corpus concordance collocation*. Oxford: Oxford University Press.
- Sinclair, John. 1996. The search for units of meaning. *Textus* 9(1). 75-106.
- Sinclair, John. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.
- Sneath, Peter H.A & Robert R. Sokal. 1973. *Numerical taxonomy*. San Francisco: Freeman.
- Sperberg-McQueen, C.M & Lou Burnard. 2019. TEI P5: Guidelines for electronic text encoding and interchange. The Text Encoding Initiative Consortium. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (17 February, 2019).

- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Stubbs, Michael. 2001. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Stubbs, Michael. 2002. Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7(2). 215-244.
- Stubbs, Michael. 2009. The search for units of meaning: Sinclair on empirical Semantics. *Applied Linguistics* 30(1). 115–137.
- Sweetser, Eve E. 1990. *From etymology to pragmatics*. Cambridge: Cambridge University Press.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Harvard: Harvard University Press.
- Turner, Ken. 2003. On neo- and post-Gricean conditionals. *International Journal of Pragmatics* 14. 135-154.
- Ure, Jean. 1971. Lexical density and register differentiation. In George Ernest Perren & John Leslie Melville Trim (eds.), *Applications of linguistics. Selected papers of the second International Congress of Applied Linguistics, Cambridge 1969*, 443-452. London: Cambridge University Press.
- Wilson, Andrew. 2013. Embracing Bayes factors for key item analysis in corpus linguistics. In Markus Bieswanger & Amei Koll-Stobbe (eds.), *New approaches to the study of linguistic variability*, 3-11. Frankfurt: Peter Lang.
- Wulff, Stefanie. 2006. *Go-V* vs. *go-and-V* in English: A case of constructional synonymy? In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, 101-125. Berlin: Mouton de Gruyter.

### Appendix: extraction of samples

Code	Content	Query	S-units
<i>if-cnd</i>	Conditional structures with <i>if</i>	<p>Excluded clusters:  <i>as if, ascertain if, ask if, asked if, asking if, asks if, certain if, consider if, decide if, determine if, discover if, doubt if, doubted if, doubtful if, enquire if, even if, hear if, knowing if, known if, remember if, see if, seeing if, sure if, tell if, uncertain if, unsure if, wonder if, wondered if, wondering if, wonders if</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• concessive uses</li> <li>• interrogative uses</li> <li>• <i>when and if</i></li> <li>• metalinguistic uses</li> </ul>	959
<i>assuming-cnd</i>	Conditional structures with <i>assuming</i>	<p>Excluded clusters:  <i>am assuming, are assuming, be assuming, been assuming, is assuming, was assuming, were assuming, 's assuming, 're assuming, 'm assuming, about assuming, after assuming, again assuming, against assuming, almost assuming, already assuming, as assuming, automatically assuming, before assuming, besides assuming, by assuming, derived assuming, each assuming, even assuming, ever assuming, for assuming, from assuming, government assuming, him assuming, in assuming, involves assuming, just assuming, like assuming, not assuming, now assuming, of assuming, on assuming, since assuming, than assuming, thus assuming, to assuming, upon assuming, usually assuming, when assuming, while assuming, without assuming</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• non-conditional uses (e.g., making an assumption, assuming a role)</li> <li>• verbless fragments</li> <li>• duplicates</li> </ul>	727

Code	Content	Query	S-units
<i>in_case-cnd</i>	Conditional structures with <i>in case</i>	<p>Excluded clusters: <i>in case of, just in case</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• <i>case</i> = container</li> <li>• <i>case</i> = instance</li> <li>• <i>case</i> = legal term</li> <li>• <i>case</i> = typographical term</li> <li>• Spelling variations of <i>just in case</i> (e.g., <i>jus' in case</i>)</li> <li>• compounds (e.g., <i>case study, case law</i>)</li> <li>• duplicates</li> </ul>	945
<i>provided-cnd</i>	Conditional structures with <i>provided</i>	<p>Included clusters: <i>provided that, provided the, provided this, provided I, provided you, provided he, provided she, provided it, provided we, provided they</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• <i>provided</i> = <i>decided/decreed/ordered</i> etc</li> <li>• duplicates</li> </ul>	859
<i>supposing-cnd</i>	Conditional structures with <i>supposing</i>	<p>Included words: <i>supposing</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• <i>by/for/in/of supposing</i></li> <li>• <i>be supposing</i></li> <li>• verbless fragments</li> </ul>	213
<i>on_condition-cnd</i>	Conditional structures with <i>on condition</i>	<p>Included cluster: <i>on condition</i></p>	205
<i>unless-cnd</i>	Conditional structures with <i>unless</i>	<p>Included word: <i>unless</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• verbless fragments</li> <li>• metalinguistic uses</li> <li>• misuse (instead of <i>lest</i>)</li> </ul>	989

Code	Content	Query	S-units
<i>even_if-cc</i>	Conditional-concessive structures with <i>even if</i>	Included cluster: <i>even if</i>  Instances removed manually: <ul style="list-style-type: none"> <li>indirect interrogative uses (<i>even</i> is not part of <i>even if</i>)</li> <li>metalinguistic uses</li> <li>verbless fragments</li> </ul>	995
<i>whether-cc</i>	Conditional-concessive structures with <i>whether</i>	Included word: <i>whether</i>  Concessive uses established through manual examination.	184
<i>if-q</i>	Indirect interrogatives with <i>if</i>	Included word clusters: <i>ascertain if, ascertaining if, ask if, asked if, asking if, asks if, certain if, confirm if, confirms if, confirmed if, confirming if, consider if, decide if, determine if, discover if, doubt if, doubted if, doubtful if, enquire if, establish if, establishes if, established if, establishing if, hear if, knowing if, known if, prove if, proves if, proved if, proving if, remember if, see if, seeing if, sure if, tell if, uncertain if, unsure if, verify if, verifies if, verified if, verifying if, wonder if, wondered if, wondering if, wonders if</i>  Instances removed manually: <ul style="list-style-type: none"> <li>conditional uses</li> <li>duplicates</li> </ul>	978
<i>whether-q</i>	Indirect interrogatives with <i>whether</i>	Included word: <i>whether</i>  Indirect interrogative uses established through manual examination.	809
<i>as_if-c</i>	Structures of comparison with <i>as if</i>	Included cluster: <i>as if</i>  Instances removed manually: <ul style="list-style-type: none"> <li>metalinguistic uses of ‘as if’</li> <li>‘as’ used as a conjunction before an <i>if</i>-conditional</li> <li>non-recoverable ellipsis</li> </ul>	995
<i>as_though-c</i>	Structures of comparison with <i>as though</i>	Included cluster: <i>as though</i>  Instance removed manually: <ul style="list-style-type: none"> <li>duplicate, due to repetition of ‘as though’</li> </ul>	999

Code	Content	Query	S-units
<i>when-t</i>	Structures expressing time with the conjunction <i>when</i>	<p>Included word: <i>when</i>, tagged as a conjunction</p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• functions as relative (modifies NP), denoting time point/period or parallel/simultaneous actions – not consequent events.</li> <li>• introduces nominal clauses, with a function similar to indirect interrogatives with <i>if</i> (e.g., <i>I wonder when she was here</i>)</li> </ul>	902
<i>whenever-t</i>	Structures expressing time with the conjunction <i>whenever</i>	<p>Included word: <i>whenever</i>, tagged as a conjunction</p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• non-bi-partite constructions indicating unspecified time point/period</li> </ul>	959
<i>baseline</i>	Sample from the whole BNCw	<p>Regular expression: "&lt;s&gt; [] expand to s" (which returned all s-units in BNCw).</p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• verbless fragments, titles and headings</li> </ul>	872
<i>non-cnd</i>	Non-conditional structures	<p>Excluded words and clusters: <i>assuming, if, in case, supposing, unless</i></p> <p>Instances removed manually:</p> <ul style="list-style-type: none"> <li>• verbless fragments, list stems, list items headings, titles</li> <li>• conditionals (e.g., with <i>provided</i>)</li> </ul> <p>The word-form <i>provided</i> was not excluded, because an examination of a random sample of 200 instances indicated that its use as a verb is much more frequent than its use as a conditional subordinator – and its exclusion from the <i>noncnd</i> sample would potentially skew it. It was, therefore, decided to remove any conditionals with <i>provided</i> during the manual analysis.</p>	856