

***If*-conditionals: Corpus-based classification and frequency distribution**

Costas Gabrielatos (Edge Hill University)

1 Focus and motivation

This paper discusses the frequency distribution of the types of *if*-conditionals recognised in the corpus-based classification developed in Gabrielatos (2010: 230-265). It is pertinent to mention at the outset that *if*-conditionals have been estimated to account for about 80 per cent of all conditional constructions in written British English (Gabrielatos 2010: 49). The classification was partly adapted from Quirk et al. (1985: 1072-1097), and was based on two interrelated criteria: a) the nature of the link between the two parts of a conditional, (henceforth, *protasis* and *apodosis*) and b) the modal nature of the apodosis. The quantitative analysis discussed here provides insights into the nature of each type, and the ways that the interaction of the type of link between protasis and apodosis, and the type of modality expressed by the apodosis gives rise to their potential for use in communication.

The motivation for the development of a corpus-based classification of *if*-conditionals was the realisation that existing classifications have not been tested on representative samples of actual use, and, as a result, exhibit particular limitations (Gabrielatos 2010: 152-188). These limitations can be better understood when we consider the distinction between *introspection-informed*, *data-informed*, and *corpus-based* classifications (adapted from Gabrielatos 2010: 10-13). *Introspection-informed* classifications, and the examples used to support them, are derived merely from the analyst's introspections and informal observations. *Data-informed* classifications are supported by attested examples of use (e.g. taken from newspapers, novels, television, internet, overheard conversations, or corpora). However, these examples are selected ad hoc (even when the source is a corpus) to exemplify types that have been formulated on the basis of introspection or informal (i.e. unsystematic) observations, and can have no claim to being representative. *Corpus-based* classifications are based on an appropriate representative corpus, and adhere to the "principle of total accountability" to the data (Leech 1992: 112). That is, the analysis and resulting theoretical interpretations have to account for all relevant items in the corpus sample (in our case, *if*-conditionals) – no items are ignored or discounted, however inconvenient they may be for the proposed classification. In addition, corpus-based classifications can provide information on the frequency and distribution of particular types. Classifications that are not informed by the examination of representative samples of natural occurring language can be expected to reflect the analyst's introspections rather than actual language use; that is, even if they use attested examples, they leave open the possibility that types of *if*-conditionals may have been left out, because they are not accessible via the analyst's introspection, or have escaped the analyst's attention, or, worse still, because they are incompatible with the proposed classification. More specifically, the detailed examination of existing classifications of conditionals in Gabrielatos (2010: 10-13, 152-188) identified the following interrelated shortcomings:

- a. They are either data-informed or introspection-informed.
- b. They are not comprehensive; that is, they do not account for all *if*-conditionals.
- c. Some of the examples contradict the definitions of the types they are intended to illustrate.
- d. There are instances of conditional constructions which belong to two (or more) types, which the classification presents (explicitly or implicitly) as being mutually exclusive.
- e. They do not provide quantitative and distributional information.

The above shortcomings motivated the development of a classification of *if*-conditionals which met all of the following interrelated criteria (expanded from Gabrielatos 2010: 230-231):

- a. *Representativeness*: it is based on a representative sample of language use.
- b. *Comprehensiveness*: it caters for all items in the sample, and all examples (actual or introspective) provided in the studies examined in Gabrielatos (2010: 152-188).
- c. *Discrimination*: it avoids overlap between types or sub-types posited (explicitly or implicitly) as mutually exclusive.
- d. *Consistency*: definitions, examples and comments support one another, that is, they do not undermine or contradict one another.
- e. *Theoretical salience*: it contributes to the understanding of the nature of the linguistic item under classification.

2 Classification of *if*-conditionals

Following Fillmore (1986), Dancygier (1998), and Dancygier and Sweetser (2005), the classification in Gabrielatos (2010: 230-265) treats conditionals as bi-partite constructions, consisting of the *protasis* (henceforth, *P*), in which a condition is expressed, and the *apodosis* (henceforth, *A*), in which a comment related to the condition in the protasis is provided (see also Sweetser, 1990: 125). Therefore, the classification treats *if*-conditionals as whole entities (constructions), while also catering for their bi-partite nature. More specifically, the classification recognises that *A* defines the semantic function of the conditional, while *P* acts as a modal marker for *A*; that is, the content of *A* is presented as being contingent to the hypothesis in *P*.¹ Consequently, each conditional is classified according to the following two dimensions (see Sections 2.1 and 2.2 for examples and discussion):

- The nature of the link between *P* and *A*; that is, the relation between the content of *P* and the content of *A* (henceforth, *P-A link*).
- The semantic function of the construction, expressed via the modal marking of *A*.

2.1 Classification according to *P-A link*

Regarding the *P-A link*, the classification emerging from the analysis makes the core binary distinction between *direct* and *indirect* conditions proposed in Quirk et al. (1985: 1088-1097) – with adaptations in definitions and subtypes (for a detailed discussion, see Gabrielatos 2010: 263-265).

In *direct conditionals* (*DIR*), the realisation of the content of *A* – that is, the action, situation, or notion expressed in *A* – depends on the realisation, actuality or factuality of the content of *P*. For example, in (1), the result expressed in *A* is directly dependent on the action described in *P*. That is, if the garden path is designed as described in *P*, then the effect described in *A* is expected to be achieved; if not, the effect is not expected to be achieved.

- (1) Using a zig-zagging technique with your garden path will create a feeling of greater space and, if it is placed between the plantings, will enhance the "separate room" effect. [AOG 1476]²

In *indirect conditionals* (*IND*), what is contingent on *P* is not the content of *A*, but the relevance of its very uttering, or the wording of its content, or its clarity for the addressee and, consequently, the accurate indication of (contingent aspects of) the referent. Two subtypes of

IND are distinguished, *relevance* (IND-R) and *comment* (IND-C) conditionals – exemplified by (2) and (3) respectively:

- (2) If antibiotics are likely to clear up the infection, why are we having this long discussion?
[CH1 5292]
- (3) He’s not a bad sort, for a brother if you know what I mean. [AN7 3257]

In (2), a *relevance* conditional, the discussion referred to in A is taking place regardless of whether P holds (i.e. whether antibiotics are likely to clear up the infection). What is contingent on P is the relevance of the question in A; that is, the uttering of the question is relevant only if P holds. In (3), a *comment* conditional, the brother’s qualities mentioned in A remain the same irrespective of whether the listener understands what the speaker means by ‘not a bad sort’. What is contingent on P is not the personal qualities of the speaker’s brother (as the speaker may see them), but the wording used in the description of these qualities; that is, whether the words selected by the speaker manage to successfully communicate to the listener the speaker’s assessment of his/her brother’s personal qualities.

In *if*-conditionals, the difference in the semantic link of P and A in DIR and IND is reflected in their respective syntactic link: P is an adjunct in DIR, but a disjunct in IND (Quirk et al., 1985: 612-631, 1072).³ Similarly, the two sub-types of IND are also syntactically distinct. In IND-R, P is a content disjunct, that is, it comments on the actual content of the apodosis, whereas, in IND-C, P is a style disjunct, as it “convey[s] the speaker’s comment on the style and form” of the apodosis (Quirk et al., 1985: 615). IND-R can further be divided into two sub-types, termed *prerequisite* (*PREQ*) and *pretext* (*PTXT*), exemplified by (4) and (5) respectively:

- (4) If you have recently arrived in this country special rules apply and it is even more important to get good advice before claiming benefits. [CJ9 1287]
- (5) It was never like this, and my father was an Old Bastard if you must know. [EDJ 2007]

In (4), a *PREQ* conditional, the special rules for new arrivals mentioned in A are in force regardless of the status of the addressee (the content of P). What is contingent on P is that, in order for these rules to apply to the addressee and, therefore, for the information in A to be relevant to him/her, the addressee needs to be a new arrival. That is, the content of P holding is a prerequisite for the relevance of uttering A – rather than a prerequisite for A holding. By contrast, in (5), a *PTXT* conditional, the speaker’s comment on his/her father’s character is expressed regardless of whether the addressee must know. In this light, P functions as a pretext for uttering A: if P appears to provide the addressee with a choice (Dancygier 1998: 90), this amounts to the addressee being able to disregard the content of A as irrelevant if they deem that P does not hold – it does not negate the communication of the information it contains. In general, by being expressed within a *PTXT* conditional, the speech act in A is presented as cancellable (see Levinson, 1983: 118-120), in that the activation of the speech act is presented as depending on P holding. However, it must be stressed that, in the above types, P is anything but surplus to the interaction. It is exactly because of the existence of P that the strong negative opinion in the apodosis of (5) is not presented as voluntarily expressed by the speaker, but as a response to a hypothetical request by the hearer.

2.2 Classification according to semantic function

The semantic function of a conditional construction refers to what Leech (1983: 30) defines as the “sense ... of an utterance”, or its “face-value interpretation” (1983: 31), as opposed to its

“pragmatic force”, which is “represented as a set of implicatures” (1983: 30). For example, a conditional may superficially express the modal notion of volition, whereas its pragmatic force is that of an invitation or suggestion, as in (6) (see also Dancygier and Sweetser 2005: 10; Gabrielatos 2010: 253-263):

- (6) 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]

The classification of conditionals according to the semantic function of the construction is in line with the conception of conditional constructions as being closely related to modality (Gabrielatos 2007, 2010, 2011, 2013, 2019, 2020). In this dimension, four types of conditionals are recognised, as the apodoses of each expresses one of the modality types in Gabrielatos (2010: 134-147). This classification of modality types is a more fine-grained adaptation of the binary distinction posited in Quirk et al. (1985: 219-239), but it is also informed by all other classifications of modality (for a detailed discussion, see Gabrielatos 2010: 57-151). The terms describing each type are prefaced by ‘attitude to’, in order to stress the inherent subjectivity of modality, that is, in order to render explicit that the speakers’ expression of modal notions, or the degree or strength of the modality, may or may not reflect reality. This classification is compatible with all previous classifications, in that the types it posits can be combined to form types recognised in all other classifications (see Table 1 – adapted from Gabrielatos 2010: 142).

Table 1: Equivalencies of types in the different classifications of modality

Types				Source
Likelihood	Propensity	Non-Directed Desirability	Directed Desirability	Gabrielatos (2010)
Extrinsic		Intrinsic		Quirk et al. (1985)
Epistemic	Non-Epistemic (Root)			Coates (1983)
Epistemic	Agent/Speaker-Oriented			Bybee et al. (1994)
Modalization	Modularity			Halliday (2004)
Logical	Personal			Biber et al. (1999)
Epistemic	Dynamic		Deontic	Palmer (1986, 1990)

Attitude to Likelihood (LK). This type encompasses assessments of actuality, factuality, truth, possibility, likelihood, or probability. Seen from a different angle, it comprises the expression of knowledge, belief, inference, hypothesis, guess, prediction, or speculation. For example, (7) functions semantically as a conditional inference. However, these notions are not treated as discrete, but are seen as overlapping. For instance, a prediction can be made on the basis of observation (or inferences based on observation), or belief, or be a mere guess. In turn, belief and knowledge refer to a person’s attitude towards actuality, in that a person’s putative knowledge may not necessarily correspond to reality.

- (7) If physicists had tried to discover a way to release nuclear energy before 1939, they would have worked on anything else rather than the field which finally led to the discovery of fission, namely radiochemistry. [B78 1973]

Attitude to Propensity (PP). This type involves judgements about ability, capability, skills, aptitude, feasibility, potentiality, tendency, or propensity, as they relate to animate or inanimate entities, concepts, states, processes, or relations. For example, (8) expresses a conditional ability. This type is closely related to LK, in that assessments of likelihood may be

based on inherent properties (Palmer 1990: 38, Quirk et al. 1985: 221-222). However, PP is distinct from LK, in that the speaker stops at expressing his/her attitude to the existence of the above properties – any inferences regarding the likelihood of actualisation are the prerogative of the hearer.

- (8) If I can live with them, so can everyone else. [FS9 2538]

Attitude to Desirability: Directed (DD) and Non-Directed (DN). The final two types are also related, albeit in a different way. They both express desirability, that is, the way that the speaker would like states of affairs to have been in the past, or be in the present or future. However, attitude to desirability may manifest itself in two ways. On the one hand, speakers may actively seek to have their desires implemented, by attempting to directly manipulate the action of others (or even their own) through the use of language. The notions communicated in this way are those of obligation, duty, requirement, promise, advice, suggestion, invitation, prohibition, or permission. This type of modality is termed *directed desirability (DD)*. On the other hand, speakers may opt to use indirect ways in trying to have their desires implemented. They may, superficially, merely express what states of affairs they would like to see materialising, or how they would like an existing state of affairs to develop, without any explicit attempt to influence, through linguistic means, the thinking or behaviour of others (or themselves) to that direction. This involves the expression of such notions as volition, intention, willingness, wish, hope, desire or need. This type is termed *non-directed desirability (DN)*. For example, (9) offers conditional advice via explicit DD marking, whereas (10) expresses a conditional hope.

- (9) This is the best "bargain offer" pensioners have ever had, and any woman over 60 or man over 65 should take advantage of it if possible. [C8Y 946]

- (10) If anything can be salvaged from the tragedy it's hoped the publicity surrounding his death will help his work become more well known. [K21 3757]

2.3 Combining the two classification dimensions

Although the classification according to each dimension on its own is useful, taking account of both dimensions provides further insights into the effects of their interaction, and results in a more nuanced understanding of the types of conditionals. By way of demonstration, let us examine (11) and (12) below – both having an LK function, but being DIR and IND respectively.

- (11) There is a certain kudos in it, like the attitude towards Patrick Chauvel (who has been wounded several times covering wars), and there's Capa's: "If your photos aren't good enough you're not close enough." [APL 350]

- (12) If you have the stomach for it minced worms make a wonderful fly food! [C96 375]

In (11), a DIR, A is a confident inference which can be drawn (in the particular context) in the event that the content of P is a fact. Of course, (11) can also be pragmatically interpreted as advice, along the lines of, 'The closer you are to your subject, the better the photograph will be'. However, this implicature need not be drawn – more precisely, there is nothing in the form or meaning of (11) that necessarily points towards its interpretation as advice. In (12), an IND, A may be superficially presented as a conclusion contingent on the content of P being a fact, but the connection is not that straightforward. In (11), if P does not hold (i.e. if the photos are indeed good enough), then A does not hold either (i.e. you are indeed close enough). By

contrast, in (12), the factual status of A remains the same, whatever the factual status of P: minced worms make a wonderful fly food irrespective of whether one has the stomach to actually handle minced worms. It is this incongruity, stemming from the IND P-A link, which directs the reader to work out two interrelated implicatures: a) A is a suggestion, and b) P does not specify a condition, but is an indirect warning that some people may find handling minced worms gruesome. In short, (12) is intended as a hedged (i.e. modalised) suggestion. In sum, when examining conditionals, both classification dimensions must be considered.

2.4 Notation of types and sub-types of *if*-conditionals

At this point we need to clarify the notation that will be used to refer to a) types of modality and b) *if*-conditional constructions classified according to one or both of the dimensions discussed above. The four modality types will be referred to using the labels LK (likelihood), PP (propensity), DD (directed desirability) and DN (non-directed desirability). Types of *if*-conditional constructions according to one or both of the classification dimensions discussed above will be referred to by longer labels. Constructions classified only by their modal function (irrespective of whether they are DIR or IND) will be referred to as ‘*if*-cnd’ followed by the type of modality they express. For example, *if*-cnd-LK will indicate an *if*-conditional with an LK function. Similarly, direct *if*-conditionals will be indicated by ‘DIR’ followed by the modal function of the construction. For example, DIR-DD will indicate a DIR *if*-conditional with a DD function. Finally, indirect conditionals will be referred to using ‘IND’ followed by the sub-type they belong to. For example, IND-R-PREQ will indicate a relevance prerequisite indirect *if*-conditional.

3 Data and methodology

The study employed two random samples from the written BNC (BNCw), accessed via BNCweb (Hoffmann et al. 2008). The first random sample contained *if*-conditionals; the second contained sentences from the whole BNCw, and was used to represent written British English as a whole (henceforth, *baseline*). The *baseline* sample was used for comparison purposes (see discussion on modal load below). The two random samples were derived as follows. For each sample, the respective complex queries in Table 2 (Gabrielatos 2010: 46)⁴ were used to derive two sets of 1000 random s-units.⁵ The 1000 s-units in the sample of *if*-conditionals were then examined manually to remove non-conditional uses of *if* (concessive, interrogative, metalinguistic) – resulting in a random sample of 959 *if*-conditionals. Similarly, the 1000 s-units in the *baseline* sample were manually examined to remove verbless fragments, titles, and headings – resulting in a random sample of 872 s-units.

Table 2: Queries used to return the samples

Sample	Query used to derive sample	Interpretation
<i>if</i> -conditionals	MU(meet "if"%c [word != "as ascertain ask asked asking asks certain consider decide determine discover doubt doubted doubtful enquire even hear knowing known remember see seeing sure tell uncertain unsure wonder wondered wondering wonders"%c] -1 -1)	Return s-units with <i>if</i> , but excludes those containing the following 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>
baseline	"<s> [] expand to s"	All s-units in BNCw

The baseline sample was further manually examined to identify the proportion of each of the four modality types discussed in Section 2.2 above. The sample of *if*-conditionals was also examined manually to identify modal markers in P and A (and the modality type they expressed) and determine the type of each *if*-conditional. Apart from examining the proportion of each type of *if*-conditionals, the study also examined the *modal load* (ML) of each type, that is, the extent of its modal marking, using two complementary metrics: *modal density* (MD) and *modalisation spread* (MS), as summarised in Table 3 (for details, see Gabrielatos 2010: 50-52, 2019: 309-311).

Table 3: The two complementary metrics used to establish modal load

	Modal Density (MD)	Modalization Spread (MS)
Definition	The average number of modal markings per clause.	The proportion of constructions that carry at least one modal marking.
Expression	The number of modal markings per 100 clauses.	The percentage of modalized constructions.
Utility	It helps comparison by normalizing the complexity of the constructions in the sample.	It corrects for heavily modalized constructions in the sample.

Comparisons of the proportions of different types of conditionals and their respective ML were carried out using two complementary approaches. The size of pairwise differences were measured using %DIFF: the percent difference of the normalised frequencies of two types, or their MD and MS values (Gabrielatos 2018). The %DIFF scores were also tested for statistical significance using the Bayesian Information Criterion (BIC). A BIC value of 2 or above is deemed to show dependable differences (for details, see Wilson 2013). Patterns in the ML of different types were examined via a) the graphical presentation of MD and MS scores in scatterplots, and b) hierarchical cluster analysis (for details, see Gabrielatos 2010: 52-54, 2019).⁶

4 Types of *if*-conditionals: Quantitative analysis and discussion

This section will first examine the frequency distribution of the *if*-conditionals in the sample according to the two classification dimensions: a) the type of P-A link, and b) the semantic function of the construction. The discussion will be complemented with the comparison of the distribution of the four modal functions in DIR and IND. The focus will then turn to the distribution of types and sub-types of indirect *if*-conditionals. The section will conclude with a discussion of interrogative apodoses. Please note that the analysis excludes 16 constructions lacking apodoses – which made it impossible to ascertain either classification dimension (for details, see Gabrielatos 2010: 233-236). Therefore, the quantitative analysis is based on 943 *if*-conditionals.

4.1 Distribution of DIR and IND in the sample

The *if*-conditionals in the sample were predominantly DIR, with IND representing just under 10 per cent of the tokens (Table 4).

Table 4: Distribution of types by syntactic link

Type (n=943)	Raw Frequency	Proportion (%)
DIR	852	90.3
IND	91	9.7

The low proportion of IND may be due to the fact that they tend to be more frequent in speech than in writing, as IND lend themselves to the expression of politeness, or the performance of speech acts – both being more characteristic of spoken interaction (Gabrielatos 2010: 268). Mato-Míguez (2016: 165) found that, in the Diachronic Corpus of Present-Day Spoken English (Aarts and Wallis 2006), IND accounted for 32.6 per cent. However, comparisons with the quantitative results of other corpus-based studies are not always straightforward, as the classifications they employ may differ. Lastres-López (2020: 69-72), examining a sample of 445 *if*-conditionals from ICE-GB, found a fairly balanced distribution: ideational conditionals, which are DIR (Lastres-López 2020: 71) accounted for about 45 per cent of the conditionals in the sample, with interpersonal conditionals (which are mainly IND) accounting for about 54 per cent. However, the ideational-interpersonal distinction does not fully correlate with the DIR-IND one, as interpersonal conditionals also include ‘epistemic’ conditionals, in which P expresses a premise and A expresses a related conclusion (Sweetser 1990: 116-117), which can be DIR or IND. Also, the speech-writing distinction may not be useful on its own, as the level of formality also needs to be considered. Examining formal spoken registers (courtroom and parliamentary discourse), Lastres-Lopez (2019: 61-68) found that interpersonal conditionals accounted for no more than 25 per cent of the corpus samples. Finally, Gabrielatos (2010: 268-270) examined 130 *if*-conditionals in simulated speech (dialogue in fiction and direct quotations of spoken statements in written texts) and found that the proportion of IND was much higher than in written language, but still low (18.5 per cent). The above seem to indicate that IND tend to be more frequent in informal natural speech.

We now turn to the distribution according to the semantic function of the construction, irrespective of the type of P-A link. As Table 5 shows, LK is clearly the most frequent semantic function, followed by DD – collectively accounting for the vast majority (88 per cent) of the *if*-conditionals in the sample.

Table 5: Distribution of types by semantic function in the sample of *if*-conditionals

Type of <i>if</i> -conditional (n=943)	Raw Frequency	Proportion (%)
<i>if</i> -cnd-LK	610	64.7
<i>if</i> -cnd-DD	220	23.3
<i>if</i> -cnd-PP	79	8.4
<i>if</i> -cnd-DN	34	3.6

The results in Table 5 indicate that (at least in written British English) *if*-conditionals are predominantly used to express assessments of likelihood (e.g. prediction, hypothesis, inference), with *if*-cnd-LK accounting for almost two-thirds of the *if*-conditionals in the sample. However, *if*-conditionals expressing DD notions (e.g. obligation, promise, permission) are also quite frequent, with *if*-cnd-DD accounting for about one in four *if*-conditionals in the sample. The frequencies of *if*-conditionals expressing one of the four modality types reported in Table 5 can also be usefully contrasted with the overall proportions of these modality types in the baseline sample (Table 6).

Table 6: Distribution of modality types in the baseline sample

Modality type (n=588)	Raw Frequency	Proportion (%)
LK	316	53.7
DD	141	24.0
PP	85	14.5
DN	46	7.8

The comparison of the proportion of the modal functions of *if*-conditionals with the overall distribution of modality types reveals some interesting patterns. The first observation is that the frequency ranking of the types of *if*-conditionals in terms of their semantic function is the same as that of the corresponding modality types in the baseline sample. However, there are distinct differences in the actual proportions, as only the proportion of *if*-cnd-DD corresponds with the proportion of DD modality (23.3 per cent and 24 per cent, respectively). The proportions of *if*-cnd-PP (8.4 per cent) and *if*-cnd-DN (3.6 per cent) are almost half of the proportions of the corresponding modality types: PP (14.5 per cent) and DN (7.8 per cent). That is, PP and DN modal notions tend to be more frequent outside *if*-conditionals. In contrast, *if*-cnd-LK are the only type of conditional with a larger proportion than LK modality overall (64.7 per cent and 53.7 per cent, respectively). The latter observation can be explained with recourse to the notion of “modal harmony” (Huddleston and Pullum 2002: 179-188, 767-770), that is, the combination of two or more modal markers expressing the same modality type and, thus, reinforcing it. For example, in ‘He must surely have made a mistake’ (2002: 767), the modal markers *must* and *surely* both express LK modality. Modal harmony in *if*-conditionals can be approached in two complementary ways. As a construction, P sets up a possible world (Bybee et al., 1994: 208) and, thus, marks “the assumption in its scope as unassertable” (Dancygier (1998: 72). In this light, P acts as an LK modal marker for A, irrespective of any additional modal marking in P (Gabrielatos 2010, 2019, 2020). In the case of *if*-cnd-LK, the LK function of P is modally harmonious with the LK function of A, as well as the LK sense of the modal marker *if*. In view of the above, it can be tentatively hypothesised that language users tend to prefer *if*-conditionals in which P and A are modally harmonious.

4.2 The interaction of P-A link and semantic function

The semantic function of the construction is employed differently in DIR and IND, due to the different type of link holding between P and A. More precisely, in IND, the semantic function is employed to elicit the working out of implicatures on the part of listeners (Gabrielatos 2010: 236-264; see also Elder and Jaszczolt 2016: 39). Therefore, it is necessary to also compare the distribution of the four semantic functions within DIR and IND (Table 7).

Table 7: Distribution of types in terms of semantic function in DIR and IND

Type (n=943)	DIR (n=852)	IND (n=91)	<i>Proportion (%) in DIR</i>	<i>Proportion (%) in IND</i>	%DIFF	BIC
<i>if-cnd-LK</i>	562	48	66.0	52.7	25.1	-4.59
<i>if-cnd-PP</i>	52	27	6.1	30.0	386.1	28.50
<i>if-cnd-DD</i>	209	11	24.5	12.2	102.9	-0.34
<i>if-cnd-DN</i>	29	5	3.4	5.6	61.4	-5.98

As Table 7 shows, there is only one semantic type (*if-cnd-LK*) in which the proportion of DIR and IND is fairly balanced. In the other three cases, there are sizeable differences: *if-cnd-PP* and *if-cnd-DN* are more frequent in IND, whereas *if-cnd-DD* are more frequent in DIR. However, due to the overall low frequency of IND, only the very large difference in the case of *if-cnd-PP* is statistically significant. Seen from a complementary angle, the picture that seems to be emerging is that, in both DIR and IND, the LK function is by far the most frequent, representing the semantic function of more than half of the constructions in either type – with its frequency being slightly more pronounced in DIR. The difference between DIR and IND, therefore, lies in their second most frequent semantic function: for DIR, it is clearly DD, whereas for IND, it is clearly PP (Table 6 above). One way that these two differences in the distribution of *if-cnd-DD* and *if-cnd-PP* between DIR and IND can be explained is by reference to the type of link holding between P and A in each type, and, as a result, the communicative functions that each is better suited to perform. This is in line with Leech's (1983: 27) comment that "[i]n so far as grammar is motivated, it is motivated at least in part by pragmatic considerations". More precisely, DIR and IND lend themselves to the direct or indirect (respectively) expression of DD functions (e.g. giving instructions or advice). In DIR, DD functions are predominantly expressed directly by explicit DD marking in A, as in (13); although DD marking is sometimes tempered by LK marking, as in (14).⁷

- (13) For example, if the number of tower cranes on site is critical, all tower crane movements must be separately identified in the computer model. [HSE 1424]
- (14) Well, if you didn't put it under my door yourself, maybe you'd better just check with Reception downstairs. [HGM 461]

In IND, however, DD functions are frequently expressed pragmatically, via conversational implicatures (Grice, 1975; Levinson, 1983: 97-166). That is, instead of explicitly seeking to influence the behaviour of the listener by using overt DD modal marking in A, the speaker presents an action/event as feasible/potential (via overt PP marking), expecting the listener to infer the DD function. What prompts the working out of the implicature on the part of the listener is the type of P-A link in IND, and relevant co-textual and contextual clues (Gabrielatos 2010: 185-188). For example, in (15), the potentiality of drilling a hole exists irrespective of the reader's preference.

- (15) It is no good pretending that you are ever going to do fine woodwork with this All Purpose saw, but for rough cutting and shaping it has many uses around the house. One useful technique is plunge cutting into a surface, such as floorboards when these need to be lifted. You hold the machine on the surface and gradually tilt the blade downwards to saw through. It takes a bit of practice, but this is a useful technique to master. Of course, if you prefer, you can drill a hole through which the tip of the blade can be inserted. [A16 983]

In (15), the oddity of presenting the speaker's preference as a condition for the potentiality of drilling a hole triggers the working out of the implicature that the speaker is presenting it as an indirect suggestion (as the hole would make the insertion of the blade easier). In a direct and concise fashion, this suggestion would be probably expressed using the imperative (e.g. 'To insert the tip of the blade [more easily], drill a hole'). Of course, the same inference could be prompted without the use of a conditional (e.g. 'You can also drill a hole through which the tip of the blade can be inserted'). However, its presentation as the A of an IND, and the use of 'prefer' in P, renders the suggestion more polite, as it is presented as contingent on the listener's preference – rather than simply on the utility of drilling a hole. The same communicative effect can also be achieved through overt DN marking in the apodoses of IND: the speaker superficially expresses his/her wish or hope that an action/event will come about, while also indirectly prompting the listener to infer the DD function – as in (16):

- (16) Nevertheless if there is anything I can do for you in the way of friendship, now or in the future, I hope you will not hesitate to call upon me. [APR 2258]

The explanation put forward above seems to be further supported by two other findings. First, there is a significant difference in the combined frequency of PP and DN functions between DIR and IND: they are performed in IND almost four times more frequently than in DIR, and the difference is highly significant (BIC=24.55). Second, the sample did not contain any cases of DIR conditionals performing DD functions indirectly; that is, in DIR, all such functions were performed through explicit DD marking. However, the above should not be taken to imply that DD functions are not performed explicitly in IND conditionals (as (17) demonstrates), but that they are expected to be rare.

- (17) If you wish, you may authorise someone to be your agent to draw your pension or allowance and pay your charge for you. [GXJ 4398]

To sum up, there seem to be distinctly different frequency patterns for the explicit marking of DD functions in DIR and IND. When examining each type (Table 8), we see that, in DIR, DD functions are only expressed explicitly (BIC=283.70) – whereas, in IND, explicit and implicit DD marking seem equally probable overall (54.5 per cent and 45.5 per cent respectively; BIC=-3). Seen from a complementary angle (Table 9), when a DD function is expressed explicitly via an *if*-conditional, this conditional is almost certain to be DIR rather than IND (97.2 per cent and 2.8 per cent, respectively; BIC=237.21) – whereas, if the DD function is expressed implicitly, then it is almost certain that the *if*-conditional is IND rather than DIR (0 per cent and 100 per cent, respectively; BIC=4.63).

Table 8: Explicit/implicit marking of DD function in DIR and IND

Type of <i>if</i> -conditional	Explicit DD function	Implicit DD function	<i>Explicit</i> %	<i>Implicit</i> %
DIR (n=209)	209	0	100.0	0
IND (n=11)	6	5	54.5	45.5

Table 9: Explicit/implicit marking of DD function in DIR and IND

DD function	DIR	IND	<i>DIR</i> %	<i>IND</i> %
Explicit (n=215)	209	6	97.2	2.8
Implicit (n=5)	0	5	0	100

It seems, therefore, that the nature of IND conditionals is harmonious with the indirect performance of DD functions, through the explicit marking of the apodosis for PP modality or, to a lesser degree, DN modality, rather than DD. Consequently, IND lend themselves to the polite expression of directives, suggestions, etc. much more readily than DIR. Conversely, there are strong indications that DIR lend themselves to the explicit performance of DD functions – although IND are an alternative, albeit a much less frequent one. Seen from the user’s perspective, once a user decides to use of conditional construction to perform a DD function implicitly, then an IND would seem to be the appropriate choice (see also Quirk et al. 1985: 1097). In addition, it does not seem unwarranted to argue that when the choice is to explicitly perform a DD function via a conditional, a DIR could be expected to be deemed by users more appropriate than an IND.

So far, IND have been treated as a unified group. However, as IND subtypes perform different communicative functions, their distribution in the sample must be examined, in order to establish whether their different functions are reflected in their relative frequency, and their patterning in terms of ML and type of modal marking.

4.3 Distribution of types and sub-types of IND

As Table 10 shows, the differences between comment (IND-C) and relevance (IND-R) *if*-conditionals are reflected in their uneven distribution, with IND-C accounting for less than 5 per cent of IND tokens. That is, the IND in the sample are predominantly used to give advice or suggestions, or communicate hedged opinions, particularly strong or strongly-worded ones. A tentative explanation for the large discrepancy in the frequency of the two subtypes of IND is that speakers may have cause to express tentative views or perform indirect speech acts (i.e. use IND-R) much more often than they may have cause to express uncertainty regarding the accuracy of their linguistic choices or the attributes of real-life referents in their utterances (i.e. use IND-C).

Table 10: Frequency distribution of IND subtypes

Type (n=91)	Raw Frequency	<i>Proportion</i> (%)
IND-C	4	4.4
IND-R	87	95.6

IND-C are too infrequent in the sample for the examination of their distribution to be useful (i.e. dependable); therefore we will only examine the distribution of IND-R (Table 11). The two subtypes of IND-R, PREQ and PTXT, are fairly equally distributed. IND-R-PREQ are one-third more frequent than IND-R-PTXT (%DIFF=35.1), but the difference is not statistically significant (BIC=-3.21).

Table 11: Frequency distribution of IND-R

Type (n=87)	Raw Frequency	Proportion (%)
IND-R-PREQ	50	57.5
IND-R-PTXT	37	42.5

About one in five IND-R-PREQ (18 per cent) are bridging conditionals (BRDG), that is, they require additional bridging inferences for their interpretation (see Section 2.1 above; for details, see Gabrielatos 2010: 247-252). Conversely, there are no IND-R-PTXT-BRDG in the sample. Of course, this should not be taken to mean that we can confidently discount their existence; rather, we are only warranted to conclude that they are much less likely to occur than IND-R-PREQ-PTXT. The not inconsiderable proportion of BRDG among PREQ conditionals in the sample, in conjunction with the absence of BRDG-PTXT tokens, and the nature of PTXT conditionals, do suggest that PTXT conditionals may not be readily compatible with requiring the addressee to draw bridging inferences. PTXT conditionals are essentially strong, or strongly-worded, statements or questions. Their communicative force is mitigated by superficially, and strategically, presenting the statement/question in A as conditional on the addressee's volition or propensity, the user's obligation, or, less frequently, a hypothetical action or event (Gabrielatos 2010: 236-265). For example, in (18), the opinion that 'it probably didn't cover the petrol' is expressed regardless of whether 'she sat down to work it out' – although it is superficially presented as the condition on which it depends. In fact, (18) is meaningful only if the condition is understood to refer to her awareness of the probability, rather than to the probability itself.

- (18) It probably didn't cover the petrol, if she sat down to work it out, which she didn't. [FB0 1592]

Simply put, the fact that the statement in the apodosis of PTXT is presented regardless of the condition in their protases renders PTXT conditionals thinly veiled statements. As such, they are less indirect than PREQ, and, therefore, they do not necessarily require bridging inferences. Granted, PREQ conditionals are quite similar to PTXT ones, as their apodoses, too, provide statements or pose questions, regardless of the superficial condition in the protasis. However, PREQ protases refer to real-world aspects (via LK marking) or the ability of the speaker (via PP marking), rather than the listener's volition (via DN marking) or the speaker's obligation (via DD marking). This renders the condition in PREQ less superficial or redundant, and the dependence of A on P less tenuous than in PTXT. Consequently, it is arguably more likely that the user of a PREQ conditional may choose to increase tentativeness by eliciting bridging inferences. The argument is supported by the comparison of the ML in the protases of PREQ and PTXT, as their respective ML seem to reflect the different functions that P performs in PREQ and PTXT (see Section 5.2 below).

4.4 Interrogative apodoses in DIR and IND

Due to the small number of tokens with interrogative apodoses (37 – a mere 3.9 per cent of the constructions in the sample), this section will only examine their distribution in terms of their P-A link. The analysis revealed that IND show a clear preference for interrogative A (Table 12). Almost one in five IND (18.7 per cent) have an interrogative A, in contrast to a mere 2.3 per cent of DIR tokens, and the difference is highly significant (BIC=25.66).

Table 12: Distribution of interrogative-A in DIR and IND

Type	Raw Frequency	Proportion (%)
Interrogative-A in DIR (n=852)	20	2.3
Interrogative-A in IND (n=91)	17	18.7

The extremely low proportion of interrogative A in DIR should perhaps not be surprising, as they are predominantly used to express assessments of likelihood and, to a lesser extent, provide hedged directives, advice, suggestions etc. (see Table 4 above). In order to interpret the comparatively high proportion of interrogative A in IND we need first to take into account that they are all found in IND-R-PREQ, almost half (40 per cent) of which have interrogative-A. This seems to be harmonious with their nature, as, in IND-R-PREQ, P sets out the conditions rendering the comment or question in A relevant to the interaction – while also helping clarify the nature of that relevance. For example, in (19), the speaker uses P to clarify the background against which his/her question in A can be best addressed:

- (19) I have a pair of juvenile Oscars in a four foot tank. Recently the larger of the two has developed a white nodule on its chin. This has grown rapidly and is now 2-3mm in diameter. The Oscar is untroubled by it and the other fish are not affected. Could it be a nodular disease? If so, do you know of any effective treatment? [CGH 707]

The lack of interrogative A in IND-R-PTXT in the sample can only be seen as suggesting their lower frequency compared to IND-R-PREQ, rather than their non-existence. This is because questions in A are compatible with the nature of IND-R-PTXT conditionals, and the communicative functions that they are suited to perform. Simply put, the protases of IND-R-PTXT are used to cushion the effect of the content of A, which, in the sample, is mainly strong (or strongly-worded) statements and directives. For example, potentially embarrassing questions, as in (20) (not in the sample), fit the profile of IND-R-PTXT perfectly.

- (20) (*Abberley watches, solicitous. He nods. Gets anxious about Max.*)
 ABBERLEY: Would you mind turning your cardiograph around a little so that I can see it?
 (*Max swivels set. Dead screen visible to us. Abberley watches. Max plugs himself in again. Normal graph. Abberley is relieved.*)
 ABBERLEY: How's sex, if I may ask? [FRH 1037]

As will be seen in Section 5.2, the type of P-A link in IND-R-PREQ conditionals, and the attendant function of their protases, is also reflected in the ML of their protases.

5 Modal load according to type

5.1 ML in DIR and IND

The analysis revealed that DIR and IND have similar ML: their MD is almost identical, and the MS difference is very small and not statistically significant (Table 13 and Figure 1).

Table 13: Comparison of ML in DIR and IND

Type	Modalisations	Clauses	Modalised	Constructions	MD	MS
DIR	1306	2341	695	852	55.79	81.57
IND	136	246	67	91	55.28	73.63
					%DIFF=0.9	%DIFF=10.8
					BIC=-7.85	BIC=-6.19

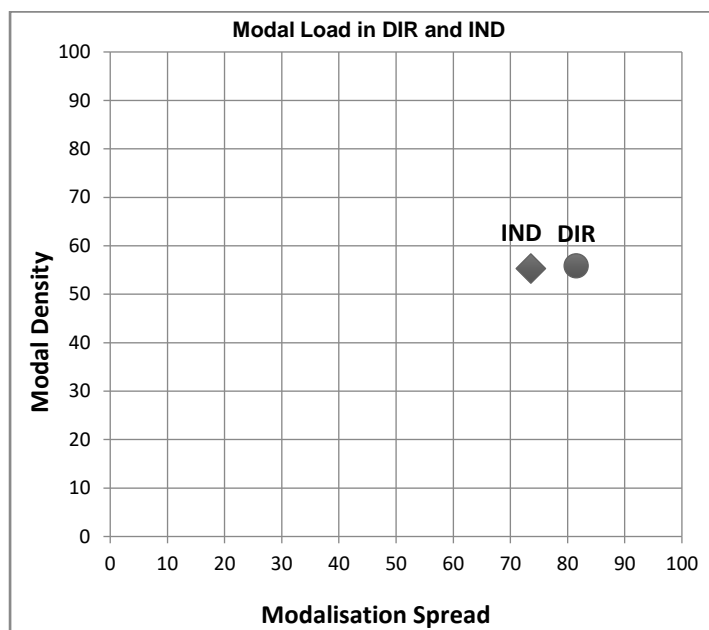


Figure 1: Comparison of ML in DIR and IND

At first glance, it would seem that the ML reflects neither of the two core elements of the proposed classification; that is, a) the distinction between DIR and IND, and b) the centrality of modal marking in distinguishing between subtypes. This can be seen as problematic for the classification, particularly in view of *if*-conditionals being treated as constructions (see Section 2). This is because construction grammar posits that differences in function are expected to be reflected in differences in form – e.g. in terms of lexis and/or syntax (Goldberg 2006: 1, Fillmore, Kay & O’Connor 1988: 501, Croft & Cruse 2004: 258, Fried & Östman 2004:18-21). In this light, classification differences would be expected to correlate with differences in meaning and form.

However, such differences are revealed when we compare the ML of the P and A of DIR and IND, respectively (Tables 14 and 15). Although the differences are neither large, nor statistically significant, two clear interrelated patterns can be observed. First, DIR have higher ML than IND in their apodoses, whereas IND have higher ML than DIR in their protases (Figure 2).

Table 14: ML of P in DIR and IND

Type	Modalisations	Clauses in P	Modalised P	Number of P	MD	MS
DIR	422	1100	344	852	38.36	40.38
IND	56	115	48	91	48.70	52.75
					%DIFF=26.9	%DIFF=30.6
					BIC=-4.46	BIC=-4.04

Table 15: ML of A in DIR and IND

Type	Modalisations	Clauses in A	Modalised A	Number of A	MD	MS
DIR	884	1241	651	852	71.23	76.41
IND	80	131	57	91	61.54	62.64
					%DIFF=16.6	%DIFF=22.0
					BIC=-5.41	BIC=-4.66

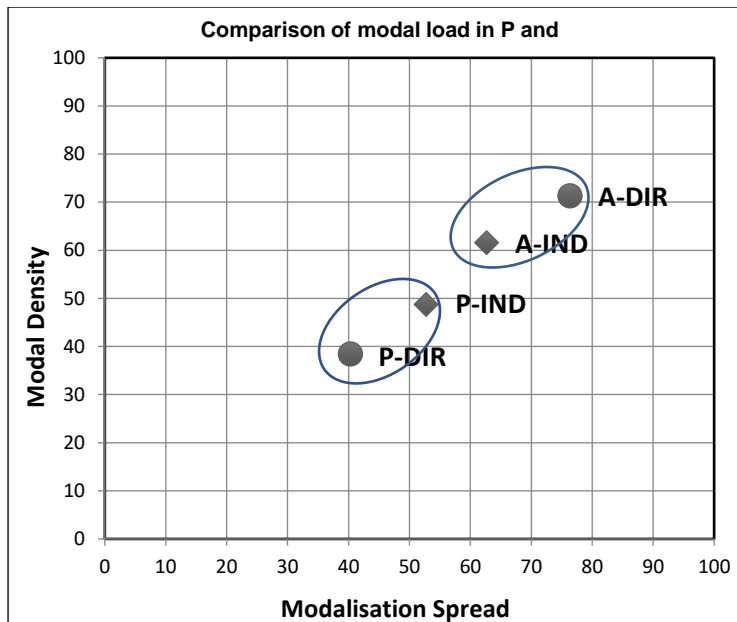


Figure 2: DIR and IND: Comparison of ML in P and A

Second, the ratio of the ML in P to the ML in A (henceforth, P/A ML ratio) is much higher in IND than in DIR (Figure 3). As was shown in Gabrielatos (2010, 2011, 2013, 2019), the P/A ML ratio seems to be useful in differentiating between the semantic components of bi-partite constructions. The comparison of P/A ML ratios revealed that the distribution of modal load in DIR is clearly uneven: apodoses have almost twice the MD and MS of protases (%DIFF values are 85.7 per cent and 89.2 per cent, respectively), with both differences being highly significant (BIC values are 108.29 and 88.85, respectively). In comparison, IND conditionals have a fairly balanced ML in P and A, with MD and MS differences being fairly small (%DIFF values are 25.4 per cent and 18.8 per cent, respectively), and not statistically significant (BIC values are -3.80 and -4.43, respectively). The differences in the P/A ratios of DIR and IND are visually represented in Figure 3: the more towards the upper right-hand corner the ratio is depicted, the more balanced it is. IND are quite close to perfect balance, that is, they are very close to having equal ML in P and A. On the other hand, DIR are depicted almost at the intersection of MD and MS ratio values of 0.5 – indicating that, in DIR, P have about half the ML of A.

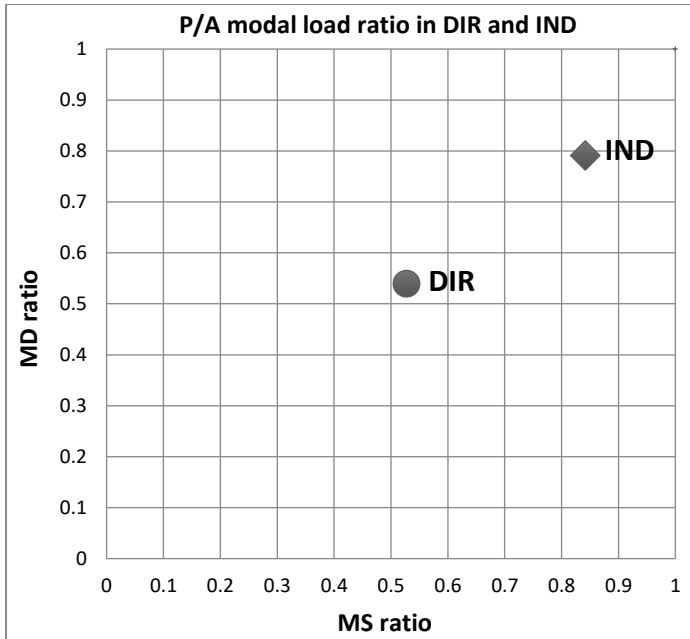


Figure 3: P/A ML ratio in DIR and IND

Both patterns can be explained with recourse to the nature of DIR and IND conditionals. The difference in the P/A ML ratios is essentially the result of the following characteristic: although A have a higher ML than P in both DIR and IND, this is much more pronounced in DIR (Figures 4 and 5, respectively).

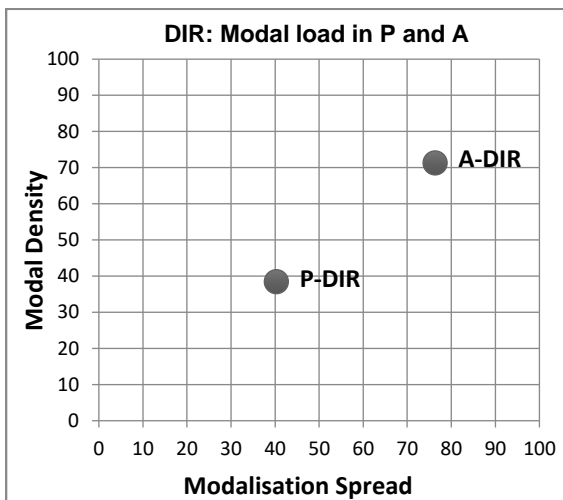


Figure 4: ML in P and A: DIR

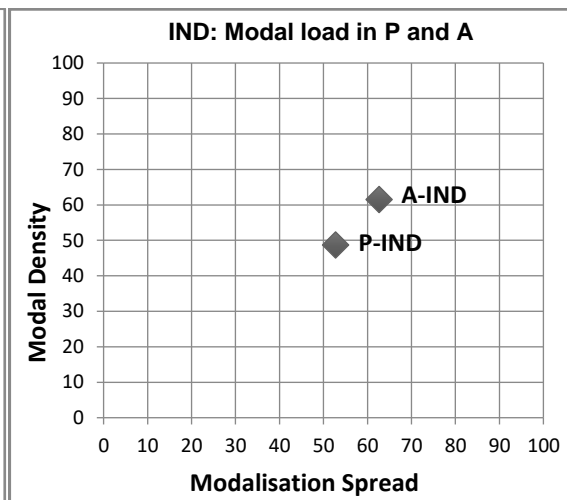


Figure 5: ML in P and A: IND

The higher ML of A (as compared to P) in both DIR and IND is arguably because the function of both types relies on overt modal marking in A, albeit for different reasons (see Gabrielatos 2010: 236-265). In DIR, the semantic function of A hinges directly on the overt modal marking; in IND, the overt modal marking provides the basis for eliciting implicatures on the part of the addressee. Therefore, the interrelated patterns described above seem to hinge mostly on differences in the ML of P. In IND, the condition is often expressed in terms of the listener's permission or volition and, consequently, their protases have cause to be modalised more frequently than in DIR. Conversely, the protases of DIR do not necessarily need additional modalisation (i.e. apart from the modal marking by *if*). The combination of these two trends

can be seen to result in protases having a much higher ML in IND than in DIR conditionals (Figure 4 above).

The results reported so far seem to support the proposed classification distinctions within the framework of construction grammar, as classification differences are reflected in differences in form and function. The analysis will conclude with the examination of the distribution of *if*-conditionals with no additional modal marking (Section 5.2), and a more detailed discussion of modal marking patterns in the protases of different types and sub-types of *if*-conditionals (Section 5.3).

5.2 Modal load in protases

Aspects of the ML of protases have already entered the discussion, as they were an integral part of the ML comparison between DIR and IND. This section will complete the examination, by comparing the ML in the protases of DIR and IND-R subtypes.⁸ In order to provide a point of reference, comparisons will also involve the ML of the baseline.

Looking at DIR, a first observation (Figure 6, Table 16) is that all sub-types have protases with an MD higher than the baseline (indicated by the horizontal dotted line in Figure 6), and the difference is highly significant. On the other hand, their MS values are, overall, comparable to that of the baseline (vertical dotted line), as differences are very small.

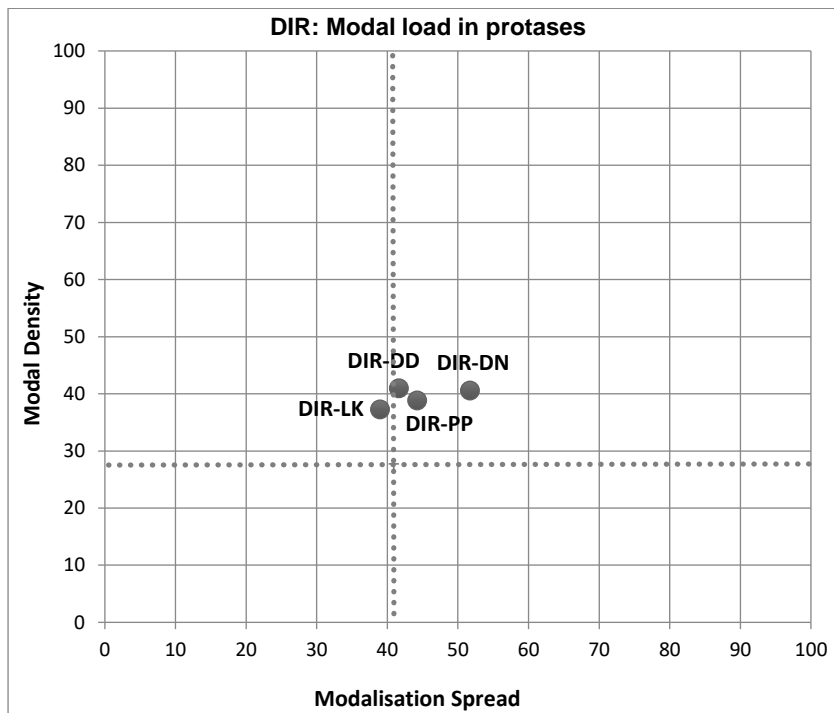


Figure 2: Difference in the ML of P and A in IND

Table 16: ML of P in DIR sub-types

Type	Modalisations in P	Clauses in P	Modalised P	Number of P	MD of P	MS of P
DIR-LK	272	730	219	562	37.26	38.97
DIR-PP	26	67	23	52	38.81	44.23
DIR-DD	109	266	87	209	40.98	41.63
DIR-DN	15	37	15	29	40.54	51.72
Baseline	588	2121	357	872	27.72	40.94

The MS of protases in all DIR subtypes indicates that they are modalised at least as often as the baseline. In turn, their significantly higher MD indicates that protases in DIR have a higher than average proportion of modal marking despite being already modally (LK) marked by *if* – see (21)-(24) below (DIR-LK, DIR-PP, DIR-DD, DIR-DN, respectively).

- (21) If there was to be a common external policy, economic and strategic, which appeared more and more desirable, as between the parliamentarily self-governing populations around the world that were deemed all to be parts of one empire, the logical but crazy conclusion must be to defy the impracticability that had been so clear in the eighteenth century and to envisage an Imperial parliament. [A69 1478]
- (22) Mistakes of law made by judges of the High Court acting in their capacity as such can be corrected only by means of appeal to an appellate court; and if, as in the instant case, the statute provides that the judge's decision shall not be appealable, they cannot be corrected at all. [FDW 104]
- (23) If, rather than face unemployment, you would at least be prepared to consider staying on with a reduced salary or a lower level of seniority, you should make that clear. [B08 1103]
- (24) If you should decide to concentrate on one particular nursing specialty then you will probably want to undertake a clinical nursing studies course. [CHT 248]

The above findings suggest that, in addition to sharing the same type of P-A link, DIR subtypes also exhibit similar ML patterns. However, the examination of ML patterns in terms of LK marking (Table 17 below) suggests that DIR should not be treated as a fully homogeneous group.

Table 17: LK modal load of P in DIR sub-types

Type	LK markings	Clauses in P	P with LK marking	Protases	MD of LK	MS of LK
DIR-LK	201	730	165	562	27.53	29.36
DIR-PP	10	67	9	52	14.93	17.31
DIR-DD	47	266	35	209	17.67	16.75
DIR-DN	11	37	11	29	29.73	37.93
Baseline	314	2121	212	872	14.80	24.31

Table 17 shows clearly that there are distinct differences among DIR sub-types as regards the frequency of LK modality in their protases. However, due to the small numbers involved, the differences are not statistically significant. Issues of statistical significance notwithstanding, it seems wise to examine these differences, as some are quite pronounced. DIR-LK protases have much higher MD in LK marking than both DIR-PP (%DIFF=84.5) and DIR-DD (%DIFF=55.8), and DIR-DN protases have twice the MD of the baseline and DIR-PP. Also, DIR-DN protases have more than twice the MS in LK marking of both DIR-DD (%DIFF=126.5) and DIR-PP (%DIF=119.2). The distinct patterns in LK marking in the protases of different DIR sub-types are shown more clearly through the cluster analysis of ML values (Figure 7). What is more, these patterns are different from those shown in the clustering of the ML values of all modality types (taken collectively) in the same protases (Figure 8 – data from Table 17). The above analysis can be seen to support the proposed classification: the overall ML of the protases of DIR reflects their similarities as a group, while the ML of their LK marking in protases reflects the differences between DIR subtypes.

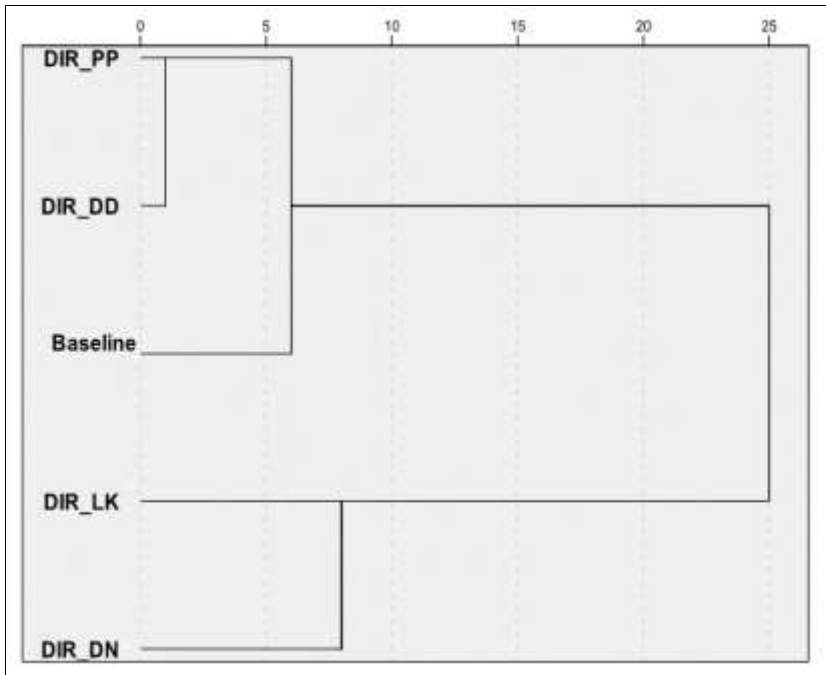


Figure 7: Clustering of DIR protases by ML (LK marking only)

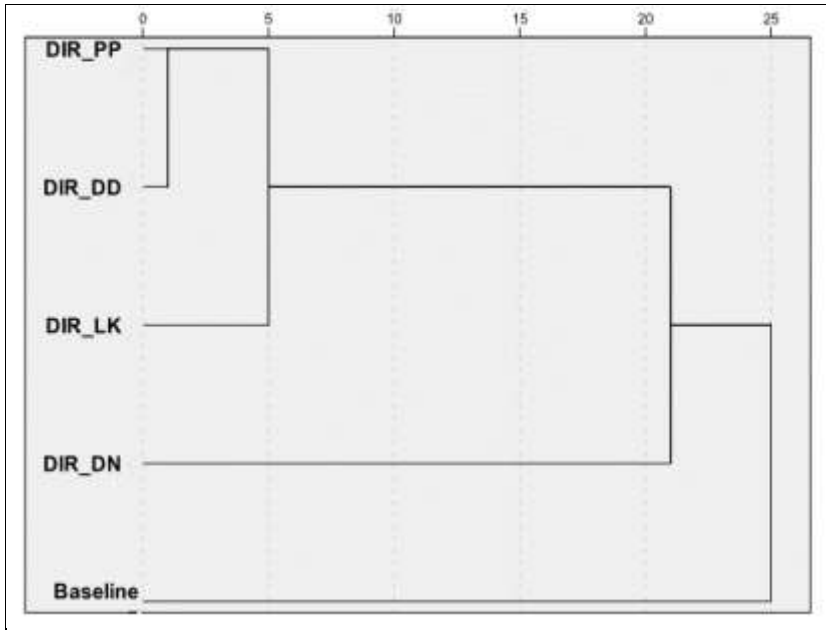


Figure 8: Clustering of DIR protases by ML (all modality types)

The examination of the correlation of ML patterns with particular types and sub-types of *if*-conditionals is concluded with the comparison of the ML in the protases of IND-R-PREQ and IND-R-PTXT. In Section 3.2 above, it was shown that IND-R-PREQ and IND-R-PTXT differ substantially in their respective proportions of unmodalised tokens, and this was ascribed to the different functions of their protases. Further evidence supporting this explanation is provided by the comparison of the modal load in their protases (Table 18, Figure 9).

Table 18: ML of P in IND-R-PREQ and IND-R-PTXT

Type	Modal markings in P	Clauses in P	Modalised P	Number of P	MD of P	MS of P
IND-R-PREQ	21	65	17	50	32.31	34.00
IND-R-PTXT	33	44	27	37	75.00	72.97
Baseline	588	2121	357	872	27.72	40.94

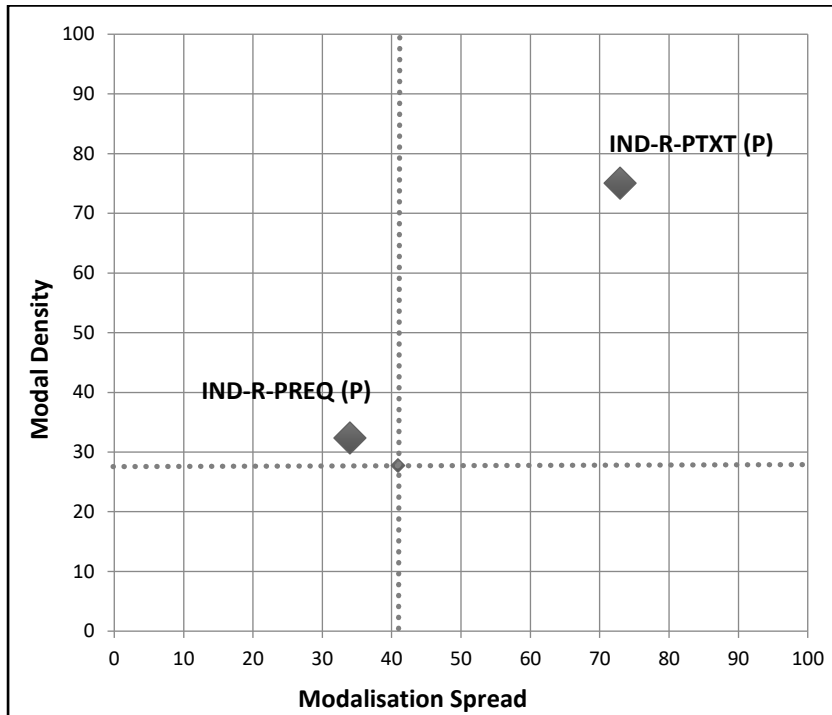


Figure 9: ML in the protases of IND-R subtypes

IND-R-PTXT protases have more than twice the MD and MS of IND-R-PREQ protases (%DIFF values are 132.1 and 114.6, respectively), with both differences being statistically significant, although only marginally in the case of MS (BIC values are 4.72 and 1.83, respectively). In fact, the MD and MS values of IND-R-PREQ are comparable to those of the baseline (the differences are very small and not statistically significant). This is because, in IND-R-PTXT, the statement in the apodosis is usually presented as being contingent on the addressee's permission or volition (Gabrielatos 2010: 247-252), which requires the protasis to be modally marked – as in (25). In contrast, IND-R-PREQ protases mostly refer to factual matters, and are, therefore, less in need of additional modal marking – as in (26).

- (25) Palma Nova's many lively bars and nightspots are all within reach of these well-maintained apartments and if you're wanting to eat out, a variety of good restaurants and cafés are close at hand. [AM0 740]
- (26) A bending spring can be used for bending 15mm and, if you are feeling strong, 22mm copper pipe. [HH6 473]

5.3 *if*-conditionals with no additional modal marking

In some *if*-conditionals, A is only modalised by P, while P itself is only modalised by the subordinator (*if*) (see Section 2, and Gabrielatos 2010: 233-235, 317-324). Such conditionals

will be referred to as ‘conditionals with no additional modalisation’ (CNAM), and can be either DIR or IND – as in (27) and (28), respectively.

(27) If the alarm gets no response, the timer goes ahead and switches off in the interest of safety and economy. [B77 1124]

(28) Replacing the motherboard or fitting one of these accelerator cards is a fairly simple job, if you're confident to do it. [HAC 5815]

The sample contains 181 CNAM (19.2 per cent of the sample). The proportion of CNAM is higher in IND than in DIR (26.4 per cent and 18.4 per cent, respectively), but the difference is not statistically significant (BIC=-4.40). The higher proportion of CNAM among IND is consistent with their nature, as they rely less often than DIR on overt modal marking to communicate a modal notion, relying more on the type of P-A link to lead addressees to infer the modal notion – as in (28) above. However, in terms of the semantic function of the construction (irrespective of the P-A link), the distribution is extremely uneven: almost all (97.8 per cent) of CNAM conditionals are *if-cnd-LK*, the rest being *if-cnd-DD*. This is not unexpected, as an LK semantic function does not necessarily need overt modalisation: P can be modalised only by *if*, while the content of A can be presented as a fact or truth (through absence of modal marking). The above two patterns are supported by the distribution of the CNAM *if-cnd-LK* tokens in DIR and IND. In the sample, more than a quarter (27.4 per cent) of DIR-LK and more than a third (39.6 per cent) of IND-LK are CNAM. In IND-LK, the proportion seems to be higher not only due to the reliance on the P-A link (as was mentioned above), but also because of the high frequency of interrogative A in IND (see Section 4.4 and Table 11 above), which tend to be unmodalised.

6. Conclusions and further work

The analysis has revealed that conditionals sharing the same subordinator (*if*) exhibit a multitude of patterns as regards the type of connection between their protases and apodoses, and their modal marking. More precisely, the findings indicate that the nature of types and sub-types of *if*-conditionals, and the communicative functions they are best suited to perform, are reflected in their frequency distribution in the sample. In turn, different types and sub-types of *if*-conditionals exhibit different ML patterns, including patterns relating to the presence/absence of additional modalisation. Again, these correlations between types and modal load seem to be explained with recourse to the nature of the P-A link and the modal function of the construction, which, in turn, determine the respective roles of P and A in each type. That is, different types and sub-types, as well as the two parts of the *if*-conditional construction, show different ML patterns, both in regard to modal marking collectively and the marking of particular modality types. It is also pertinent to add that conditionals with different subordinators (e.g. *in case*, *unless*) – regardless of (sub-)types – also demonstrate different ML patterns, and that conditionals, seen as a whole, have higher ML than non-conditional constructions (Gabrielatos 2010: 189-229, 2019).

The above suggest that it would be useful to approach conditionals as a “family of constructions” (Goldberg and Jackendoff, 2004: 535-536). All members of the family of conditional constructions share three core characteristics (Gabrielatos 2020: 166): a) they are bi-partite, consisting of a protasis and an apodosis (with each part consisting of one or more clauses), b) P modally marks A, and c) the factuality, actuality realisation, activation, or relevance of A depends on P. However, members of this family would also be expected to demonstrate differences (as summarised in the previous paragraph) stemming from the

interaction of the following elements: a) the syntactic type of P-A link (subordination or co-ordination), b) in the case of subordination, the particular subordinator, and c) the semantic function of the construction (see Gabrielatos 2020).

Further studies could usefully examine the distribution of types in conditionals with different subordinators, as well as conditionals in which protasis and apodosis are linked by co-ordination. It has also been suggested that, in some cases, *if* can be seen to function syntactically as a complementizer rather than a subordinator (López-Couso & Méndez-Naya 2014, 2015), as in ‘It would be a good idea if you hired a bodyguard’ (2015: 189). It would be useful to examine this claim and, if this is the case, to establish how such conditional constructions would relate to the classification proposed here. Finally, the results of this study would need to be supplemented with the examination of the distribution of the types of all conditionals and their ML patterns in speech, as well as in formal and informal contexts.

Acknowledgements

I am grateful to the reviewers for their useful comments and suggestions.

Notes

1. See also Bybee et al. (1994: 208), Dancygier (1998: 72), Fillmore (1990: 140), Halliday (2004: 89, 354-356), Høye (1997), Huddleston and 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; 1987:44-46), Perkins (1983: 106-108), Sweetser (1990: 127).
2. All examples are from the random sample of *if*-conditionals from the BNC (for details on the sample, see Section 3). Square brackets after examples provide the code of the sentence in the BNC: the first part indicates the text, the second part indicates the sentence.
3. Gabrielatos (2020: 165) hypothesises that, in IND conditionals, P and A are always linked by subordination, as no instances of IND with co-ordinated P and A were found in the corpus sample. For a discussion of conditionals in which P and A are linked by co-ordination, see Gabrielatos (2020: 163-167).
4. I am grateful to Stefan Evert and Neil Millar for their invaluable help in constructing the complex queries.
5. An s-unit is a sequence of words delimited by sentence-boundary markers—e.g., full-stop, question mark, exclamation mark (Sperberg-McQueen and Burnard 2019).
6. %DIFF and BIC values were calculated using Paul Rayson’s Excel spreadsheet (<http://ucrel.lancs.ac.uk/people/paul/SigEff.xlsx>); the cluster analysis was carried out using SPSS.
7. This LK marking is in addition to that of the protasis, within whose semantic scope the apodosis lies. Therefore, this modal marking is internal to the apodosis, rather than external to it – as is its modalisation by the protasis.
8. A comparison in terms of the top-level distinction in IND is not useful, due to the very small number of IND-C tokens (4).

References

- Aarts, Bas and Sean Wallis. 2006. The Diachronic Corpus of Present-Day Spoken English (DCPSE). Survey of English Usage, UCL, London. <https://www.ucl.ac.uk/english-usage/projects/dcpse> (28 February 2021).
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow, Essex: Longman.
- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm
- Croft, William D. and 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 and Eve Sweetser. 2005. *Mental spaces in grammar: Conditional constructions*. Cambridge: Cambridge University Press.
- Elder, Chi-Hé and Kasia M. Jaszczolt. 2016. Towards a pragmatic category of conditionals. *Journal of Pragmatics*, 98: 36-53.
- Fillmore, Charles J. 1990. Epistemic stance and grammatical form in English conditional sentences. *Papers from the twenty-sixth regional meeting of the Chicago Linguistic Society*, Volume 1, The Main Session: 137-162.
- 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.
- Fried, Mirjam and Jan-Ola Östman. 2004. Construction Grammar: A thumbnail sketch. In Mirjam Fried and Jan-Ola Östman (eds.). *Construction Grammar in a cross-language perspective*, 11-86. Amsterdam: John Benjamins.
- Gabrielatos, Costas. 2007. *If-conditionals as modal colligations: A corpus-based investigation*. In Matthew Davies, Paul Rayson, Susan Hunston and Pernilla Danielsson (eds.). *Proceedings of the Corpus Linguistics Conference: Corpus Linguistics 2007*, 1-20. University of Birmingham.
- Gabrielatos, Costas. 2010. A corpus-based examination of English *if*-conditionals through the lens of modality: Nature and types. PhD thesis, Lancaster University.
- Gabrielatos, Costas. 2011. *If-conditionals and modality: A corpus-based investigation*. *Corpus Linguistics 2011*, University of Birmingham, UK, 20-22 July 2011.
- Gabrielatos, Costas. 2013. Using corpus analysis to compare the explanatory power of linguistic theories: A case study of the modal load in *if*-conditionals. *Corpus Linguistics 2013*, Lancaster University, 23-26 July 2013.
- Gabrielatos, Costas. 2018. Keyness analysis: Nature, metrics and techniques. In Charlotte Taylor and Anna Marchi (eds.). *Corpus Approaches to Discourse: A critical review*, 225-258. Oxford: Routledge.
- Gabrielatos, Costas. 2019. *If*-conditionals and modality: Frequency patterns and theoretical explanations. *Journal of English Linguistics*, 47(4): 301-334.
- Gabrielatos, Costas. 2020. Conditionals, modality, and Schrödinger's cat: Conditionals as a family of linguistic qubits. In Pascal Hohaus and Reiner Schulze (eds.). *Re-assessing modalising expressions: categories, co-text, and context*, 141-172. Amsterdam: John Benjamins.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele and Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80(3): 532-568.

- Grice, Paul. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.). *Syntax and Semantics, 3: Speech Acts*, 41–58. New York: Academic Press.
- Halliday, Michael A.K. 2004. *An introduction to Functional Grammar*. 3rd edn. London: Arnold.
- Hoffmann, Sebastian, Stefan Evert, Nick Smith, David Lee and Ylva Berglund-Prytz. 2008. *Corpus linguistics with BNCweb - A practical guide*. Frankfurt am Main: Peter Lang.
- Hoye, Leo. 1997. *Adverbs and Modality in English*. London: Longman.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Lastres-López, Cristina. 2019. Conditionals in spoken courtroom and parliamentary discourse in English, French and Spanish: A contrastive analysis. In Teresa Fanego and Paula Rodríguez-Puente (eds.). *Corpus-based research on variation in English legal discourse*, 51-78. Amsterdam: John Benjamins.
- Lastres-López, Cristina. 2020. Beyond conditionality: On the pragmaticalization of interpersonal if-constructions in English conversation. *Journal of Pragmatics*, 157: 68-83.
- Leech, Geoffrey. 1983. *Principles of Pragmatics*. Cambridge: Cambridge University Press.
- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. In Jan Svartvik (ed.). *Directions in corpus linguistics: Proceedings of the Nobel symposium 82, Stockholm, 4-8 August 1991*, 105-122. Berlin: Mouton de Gruyter.
- Leech, Geoffrey. 2004. *Meaning and the English verb*. 3rd edn. London: Longman.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- López-Couso, María José and Belén Méndez-Naya. 2014. The use of *if* as a declarative complementizer in English: Some theoretical and empirical considerations. In Alejandro Alcaraz-Sintes and Salvador Valera-Hernández (eds.). *Diachrony and synchrony in English Corpus Linguistics*, 85-107. Bern: Peter Lang.
- López-Couso, María José and 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
- Mato-Míguez, Beatriz. 2016. The expression of directive meaning: A corpus-based study on the variation between imperatives, conditionals, and in subordinate if-clauses in spoken British English. PhD Thesis, University of Santiago de Compostela.
- 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.
- Perkins, Michael R. 1983. *Modal expressions in English*. London: Frances Pinter.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Sperberg-McQueen, C. Michael and 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).
- Sweetser, Eve E. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Wilson, Andrew. 2013. Embracing Bayes factors for key item analysis in corpus linguistics. In Markus Bieswanger and Amei Koll-Stobbe (eds.). *New approaches to the study of linguistic variability*, 3-11. Frankfurt am Main: Peter Lang.