

Exploring *if/whether* variation in subordinate interrogatives¹

Cristina Lastres-López, University of Santiago de Compostela

Abstract

This paper investigates the variation between *if* and *whether* (e.g. *I don't know if/whether you are right*), on the grounds that prior studies on complementation in English have not sufficiently addressed the variation between these two complementizers. Based on data extracted from the British component of the *International Corpus of English*, the results from the study show that although *if* and *whether* are mostly interchangeable complementizers, there are grammatical, semantic and stylistic constraints governing the choice of conjunction. The results suggest that a number of factors can influence the choice of complementizer, such as the matrix verb or the presence of an explicit alternative in clauses introduced by the conjunction *or*. In addition, extralinguistic factors such as age and sex are also revealed as playing a role in the choice of complementizer.

Keywords: *if*; *whether*; variation; complementizers

1. Introduction

This paper intends to tackle the issue of *if/whether* variation in subordinate interrogatives, as illustrated in (1) and (2) below, since this type of clausal complementation has received very little attention in the study of the grammar of English.

(1) I don't know if they are any good though <ICE-GB: S1B-005 #173: 1: A>

¹ For generous financial support, I am grateful to the following institutions: the Spanish Ministry of Education, Culture and Sports (grant FPU15/02519), the Spanish Ministry of Economy, Industry and Competitiveness (grants FFI2014-52188-P and FFI2017-86884-P), the European Regional Development Fund and the Regional Government of Galicia (Directorate General for Scientific and Technological Promotion, grants ED431D 2017/09 and ED431B 2017/12). Thanks are also due to Bas Aarts, Teresa Fanego, and Sean Wallis; as well as to two anonymous reviewers, for valuable comments and suggestions on an earlier version of this paper.

(2) I don't know whether he was giving it or taking it <ICE-GB: S1A-005 #253>

If and *whether* have sometimes been assumed to be interchangeable complementizers² in contexts such as the one presented in examples (1) and (2) above (Huddleston and Pullum *et al.* 2002: 973), although Eckardt (2007: 462) suggests that there exist, in fact, slight semantic differences between the two. Other researchers even argue that it cannot be demonstrated that *if* and *whether* occur in free distribution (Gawlik 2013: 131). Therefore, the aim of this paper is to examine their distribution by means of a corpus-based study, analysing data extracted from the British component of the *International Corpus of English* (henceforth ICE-GB) (Nelson, Wallis and Aarts 2002).

After this brief introduction, the paper is structured as follows. Section 2 contains a review of the literature on the topic. Section 3 explains the methodology, corpus and data extraction process. Section 4 presents the results and discusses the variation between both complementizers from a corpus-based perspective, analysing a number of factors, namely: (i) frequency, (ii) verb in the matrix clause, (iii) polarity of the matrix clause, (iv) explicit choice with *or*, (v) text type, (vi) sex of the speaker, and (vii) age of the speaker. Finally, Section 5 closes the paper with some concluding remarks.

2. Literature review

If-clauses have been widely studied over the years (Traugott, Meulen, Reilly and Ferguson 1986; Athanasiadou and Dirven 1997; Couper-Kuhlen and Kortmann 2000; Declerck and Reed 2001; among many others). However, research has mainly centred around adverbial clauses, more specifically, conditional clauses such as the one presented in (3).

(3) If it rains, I will stay at home (Wierzbicka 1997: 19)

² Throughout the paper, I employ the terms *complementizer* and *conjunction*, mostly interchangeably, when I refer to *if* and *whether*. It should be noted however, that *complementizer* is a label that refers to the syntactic function that *if* and *whether* have, whereas the label *conjunction* refers to the part of speech. Although both concepts are interrelated, they are, obviously, different.

In sentences such as (3) “the situation in the matrix clause is contingent on that in the subordinate clause” (Quirk, Greenbaum, Leech, and Svartvik 1985: 1088), that is to say, if fulfilled, the condition expressed in the *if*-clause—raining—will ensure that the proposition in the main clause—staying at home—is true. These structures, apparently similar in form to example (1) illustrated above, are in fact very different. While (3) is an adverbial clause, (1) and (2) are instances of complementation, in which the subordinate complement clause functions as direct object of the verb in the main clause and is therefore one of its arguments; more specifically, such *if/whether* clauses can be classified as indirect interrogative clauses. An *if*-clause such as the one presented in (1) cannot be understood as an adverbial clause, as shown by the fact that the clause needs the presence of the main verb (Huddleston 1984: 371), which functions as the licensing element of the subordinate clause; as well as by the fact that *if* cannot be replaced by any phrase with a similar adverbial meaning such as *on the condition that* (López-Couso and Méndez-Naya 2015: 190).

Clausal complementation, as in (1) and (2), has been studied from many different perspectives, both diachronically and synchronically, and depending on whether the complement clause is finite or non-finite, for instance. Despite the fact that *if/whether* variation occurs with finite complement clauses, studies on finiteness have primarily concentrated on the variation between clauses introduced by *that* or by a *zero* complementizer (Elsness 1984; Finegan and Biber 1995; Kaltenböck 2006; Kearns 2007; Shank, Plevoets and Cuyckens 2014, to name but a few). Less studied, however, is interrogative complementation, introduced by *if* and *whether*, as shown in (1) and (2) above. In particular, constructions in which these two conjunctions can easily alternate have been overlooked in the literature on the topic, with only a few exceptions such as Stuurman (1990), Eckardt (2007) or Gawlik (2013). This gap in the literature on complementation can be explained on the grounds that these are minor complementizers in terms of frequency, as well as the fact that it is sometimes difficult to distinguish “between declarative, interrogative and conditional uses in the case of *if*”, as pointed out by López-Couso and Méndez-Naya (2001: 94). Yet since, as already mentioned, previous studies show that it cannot be demonstrated that *if* and *whether* occur in free distribution in structures of this type (Gawlik 2013: 131), it would seem worthwhile to try to

elucidate some factor or factors that may govern the choice by speakers of one complementizer over the other.

2.1 Defining subordinate interrogatives

As noted by Huddleston (1984: 371), subordinate interrogatives or interrogative subordinate clauses, as opposed to their direct *yes/no* question counterparts, require the presence of a verb in the main clause which serves as a licensing element for the complement clause, as is the case with *wonder* in (4a) below. Another difference with respect to direct questions is that subject-operator inversion does not apply to indirect interrogatives, as shown in (4a), which may be compared to (4b), which illustrate an indirect (subordinate) interrogative and a direct one respectively.

(4a) I wonder if they've gone out <,> <ICE-GB:S1A-083 #237:1:A>

(4b) Have they gone out?

The controlling verbs that license *if*-complement clauses constitute a closed set. They are usually factual verbs (Quirk *et al.* 1985: 1180-1182), such as *know* or *wonder*, as illustrated in (1), (2) and (4a) above. Aarts, Chalker, and Weiner (2014: 154) distinguish factual verbs of two types: *public verbs of speaking* and *private verbs of thinking* (see Section 4.2). The examples presented in (1), (2) and (4a) fall under the latter category. Examples of public verbs of speaking include lexical verbs such as *affirm*, *confirm* and *declare*, among others. Example (5) shows a subordinate interrogative with a public verb of speaking.

(5) Will the minister in that case confirm to the House whether it is the case that the HSC's plan of action has been returned to the HSC by the Secretary of State? <ICE-GB:S1B-055 #016:1:D>³

³ The original utterance in <ICE-GB:S1B-055 #016:1:D> is “Will will the minister i in that case confirm to the House whe whether it is the the the case that the the HSC’s plan of action has been returned to the HSC by the Secretary of State”. However, here and henceforth, repetitions and hesitations, which are very frequent in spoken discourse have not been included in the examples for

Similarly, as far as *whether* is concerned, Huddleston and Pullum *et al.* (2002: 975) also argue that these clauses typically occur with particular lexical verbs, such as *explain*, *investigate*, *judge*, *ponder* or *study*, among others, which favour the use of *whether* instead of *if*.

As suggested in specific studies on the topic (López-Couso and Méndez-Naya 2001; Eckardt 2007; among others) as well as grammars (Huddleston and Pullum *et al.* 2002: 975), some *if*-clauses may be ambiguous or difficult to categorise. Example (6) below illustrates a case of an utterance which may have two readings, as a complement or as an adverbial clause, that is, as a subordinate interrogative or as a conditional. It should be noted, however, that this ambiguity only occurs in spoken discourse; in writing, a comma would be necessary before the *if*-clause in the conditional reading.

(6) Sue will tell us if she is hungry (Eckardt 2007: 463)

Eckardt (2007: 463) proposes two variants to disambiguate the utterance shown in (6), as illustrated in examples (7a) and (7b) below. (7a) displays a conditional reading of (6). By contrast, (7b), disambiguated by means of a coordinate, is a subordinate interrogative, an instance of indirect reported speech whose direct counterpart would be *Is she hungry?*

(7a) If she is hungry, Sue will tell us (so).

(7b) Sue will tell us if she is hungry (and what she wants to eat).

2.2 *If and whether: Semantic similarities and differences*

Overall *if* and *whether* are similar in semantic terms and interchangeable in most contexts. According to Ransom (1988: 370) both complementizers are used to indicate “indeterminate truth and action”, that is to say, *if* and *whether* show that the speaker is uncertain about the truth value of the proposition expressed by the complement clause. Eckardt (2007: 455) claims that “there seems to be no officially acknowledged difference between the two [complementizers]” but also

purposes of clarity. In addition, for an easier reading of the examples, punctuation marks have also been added where necessary.

suggests slight semantic and pragmatic nuances between *if* and *whether* that may lead speakers to choose between one complementizer over the other. Eckardt (2007: 455-457, 461) argues that one of such differences is that *if* carries a bias which makes the positive answer to be perceived by the speaker as the relevant one. Such premise does not hold, however, in the case of *whether*.

In addition, the choice between these complementizers also seems to be an issue of scope: while *if* indicates a wider range of possibilities, *whether* is more restricted in this respect, usually pointing to mutually exclusive alternatives. Even when *if* is restricted to only two possibilities, it seems that the two alternatives are viewed “as two different, but equally relevant possibilities” (Eckardt 2007: 462). This semantic difference that points to *whether* as a more restrictive complementizer can probably be traced back to the fact that, historically, *whether* derives from an interrogative pronoun, Old English *hwæþer*, which had the meaning ‘which of the two’ (Ransom 1988: 371; see also OED s.v. *whether* pron., adj., and conj. I), while *if* does not have this semantic restriction and has been employed as a conjunction throughout its recorded history (OED s.v. *if* conj.).

2.3 If and whether: stylistic differences

In addition to the semantic differences between *if* and *whether* discussed in section 2.2, the use of both complementizers has also been connected to stylistic differences (Biber, Johansson, Leech, Conrad and Finegan 1999; Gawlik 2013). In this respect, *if* is said to be more informal and colloquial than *whether*.⁴ This explains the high frequency of indirect interrogative *if* in conversations and fiction in corpus-based studies such as Biber *et al.* (1999: 691-693). Clauses introduced by *whether*, on the other hand, have been reported to be “more neutral in their stylistic range” (Biber *et al.* 1999: 691), although not necessarily more formal.

⁴ It is worth noting that for some scholars “there is a **clear** stylistic difference” [emphasis mine] (Biber *et al.* 1999: 691) between *if* and *whether*, while for others the difference is much less evident: “*if* [is] **slightly** more informal than *whether*” [emphasis mine] (Huddleston and Pullum *et al.* 2002: 974). This may be explained on the grounds that Biber *et al.*’s claim is based on corpus data, while Huddleston and Pullum *et al.*’s is not.

2.4 Grammatical constraints on the variation between *if* and *whether*

So far, the discussion has focused on the semantic and stylistic differences that may lead to the choice of complementizer. However, it should be noted that this variation is restricted grammatically in some respects, with some contexts in which no variation between the two complementizers is possible.

A frequent context in which no variation between *if* and *whether* is grammatically permitted is when the conjunction is immediately followed by the “elliptical negative clause *or not*” (Biber *et al.* 1999: 690), as in (8a) and (8b).

(8a) And you couldn't remember now whether or not you, other people, had used the steps [...] <ICE-GB:S1B-066 #064:1:A>

(8b) *And you couldn't remember now if or not [...] ⁵

In addition to this, which is in keeping with the historical origins of *whether*, as detailed in Section 2.2, Huddleston and Pullum *et al.* (2002: 973-974) also identify other contexts where *if* is ungrammatical and only *whether* is permitted, namely:

a) In the exhaustive conditional construction, as in (9) below.

(9) I'm going to see her whether/*if you like it or not. ⁶

b) When the interrogative clause is infinitival, as in (10) below ⁷.

(10) She can't make up her mind whether/*if to accept.

c) When the interrogative clause precedes the superordinate predicator, as in (11) below.

(11) Whether/*if it will work we shall soon find out.

d) As complement to *be* or as supplement to an NP, as in (12) and (13) below.

⁵ As is conventional, ungrammatical examples are preceded by an asterisk.

⁶ Examples (9) to (14) are extracted from Huddleston and Pullum *et al.* (2002: 973-974).

⁷ The restriction for *if* when the clause contains an infinitive was also earlier mentioned by Dixon (1991: 234-235).

(12) The question you have to decide is whether/*if guilt has been established beyond reasonable doubt.

(13) The question, whether/*if the commissioner exceeded the terms of reference, will need to be carefully investigated.

e) When the interrogative clause is complement of a preposition, as in (14) below.

(14) It depends on whether/*if we have enough time left.

It seems necessary to highlight these contexts in which the choice between *if* and *whether* is not possible and only one of the complementizers can occur. From what has been mentioned on the grammatical constraints of both complementizers, we can conclude that *if* imposes more restrictions than *whether* on the syntactic environment in which it is used. Therefore, these contexts in which the alternation between the two complementizers is not possible have been excluded from the corpus-based analysis that follows (see Section 3).

3. Corpora and methodology

As mentioned in the introduction, this paper addresses the topic of *if/whether* variation from a corpus-based perspective. The data analysed in the present study have been extracted from ICE-GB, a one-million word corpus that contains written and spoken English from the 1990-1992 period.⁸ The fact that ICE-GB is a fully tagged and parsed corpus allows researchers to carry out complex grammatical searches by using the *International Corpus of English Corpus Utility Programme* (ICECUP). The data used for this study have been automatically retrieved with the help of a *Fuzzy Tree Fragment* (henceforth FTF). FTFs allow researchers to carry out complex grammatical queries by means of completing a tree diagram to look for that specific query. Each node in the FTF contains three slots—for function, form or category, and features respectively—which can be completed or not depending on the type of search needed. Thus, FTFs can be as general or as specific as required for the study. Figure 1 below shows the FTF constructed for subordinate

⁸ See the Appendix for an overview of the general structure and text types in ICE-GB.

interrogatives introduced by *if*. As observed in Figure 1, it has been specified that the clauses introduced by *if* need to perform the function of direct object. The same FTF has been constructed for the conjunction *whether*.

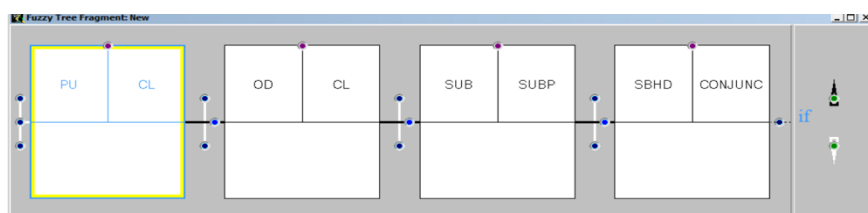


Figure 1. FTF used to extract cases of subordinate interrogatives introduced by the conjunction *if*.

The number of tokens obtained from the automatic extraction with the FTFs was 112 for *if* and 74 for *whether*. Although the FTFs used in this study allow a more direct extraction—in a way which would be impossible in a non-annotated corpus—the tokens have been manually checked after the automatic retrieval, since some of the cases in which the variation between *if* and *whether* is not possible (see Section 2.4) could be retrieved with the FTF query. After this manual analysis, a total of 7 tokens were discarded (4 in the case of *if* and 3 in the case of *whether*) since they corresponded to cases in which no alternation between *if* and *whether* is possible, as discussed in section 2.4 and illustrated in example (15); or to utterances that were incomplete or there was a change of topic or false start that left the message unfinished.

(15) She would hesitate whether to <unclear-word> him <ICE-GB:S1A-020 #124:1:C>

Therefore, the final number of tokens analysed after the discarding was 108 for *if*, and 71 in the case of *whether*. After the extraction and discarding, a database was created to analyse manually the instances of subordinate interrogatives introduced by *if* and *whether*, and to annotate each token according to a number of variables, namely, (i) verb in the matrix clause, (ii) polarity of the matrix clause, (iii) explicit choice with *or*, (iv) text type, (v) sex of the speaker, and (vi) age of the speaker.

4. Results and discussion

4.1 Frequency

The automatic extraction with the FTFs to examine the distribution of *if* and *whether* in subordinate interrogatives in ICE-GB revealed a preference for the former complementizer, as shown in Table 1 below. Taking the total number of *if*- and *whether*- subordinate interrogative clauses as a baseline, the results reveal that in roughly 60% of the cases there is a preference for *if*. This confirms the findings previously attested in larger corpora, such as the 40 million-word *Longman Spoken and Written English Corpus*, used as the basis for Biber *et al.*'s (1999) grammar.

Table 1. Absolute frequency of *if* and *whether* as complementizers in subordinate interrogatives

Complementizer	Frequency
<i>If</i>	108 (60.34%)
<i>Whether</i>	71 (39.66%)

4.2 Verb in the matrix clause

One of the variables considered to try to determine the choice for speakers between one complementizer and the other was the verb used in the main clause, since it licenses the embedded complement clause. Table 2 displays the absolute frequencies and percentages of lexical verbs controlling *if*- and *whether*-clauses.

Table 2. Absolute frequency and percentage of matrix verbs in subordinate interrogatives introduced by *if* and *whether*

<i>If</i>		<i>Whether</i>	
<i>know</i>	42 (38.89%)	<i>know</i>	20 (28.17%)
<i>wonder</i>	28 (25.92%)	<i>ask</i>	12 (16.90%)
<i>see</i>	14 (12.96%)	<i>wonder</i>	12 (16.90%)
<i>ask</i>	8 (7.41%)	<i>doubt</i>	5 (7.05%)
<i>say</i>	4 (3.70%)	<i>remember</i>	3 (4.23%)
<i>doubt</i>	3 (2.78%)	<i>say</i>	3 (4.23%)
<i>decide</i>	2 (1.85%)	<i>consider</i>	2 (2.81%)
<i>find out</i>	2 (1.85%)	<i>decide</i>	2 (2.81%)
<i>tell</i>	2 (1.85%)	<i>see</i>	2 (2.81%)

<i>think</i>	2 (1.85%)	<i>tell</i>	2 (2.81%)
<i>remember</i>	1 (0.92%)	<i>check</i>	1 (1.41%)
		<i>confirm</i>	1 (1.41%)
		<i>debate</i>	1 (1.41%)
		<i>describe</i>	1 (1.41%)
		<i>determine</i>	1 (1.41%)
		<i>indicate</i>	1 (1.41%)
		<i>leave aside</i>	1 (1.41%)
		<i>recall</i>	1 (1.41%)
Total	108 (100%)	Total	71 (100%)

As shown in Table 2, *whether*-clauses are controlled by a larger set of lexical verbs than *if*. In addition to this, and in line with other previous corpus-based studies (Biber *et al.* 1999: 692–3), verbs such as *know* and *wonder* are highly frequent in both constructions. *See* is considerably more frequent in *if*-clauses than in their *whether* counterparts, often in the construction *let's see + complement clause*, with a meaning close to *find out*, as in (16).

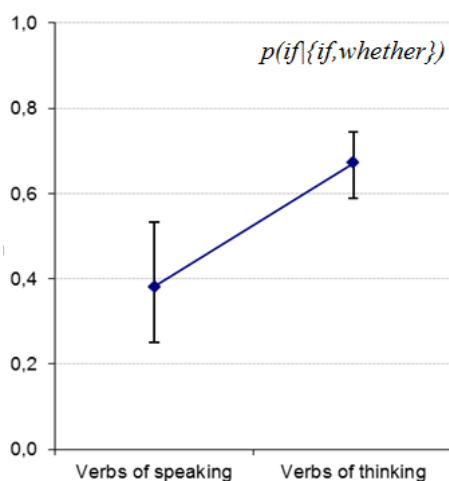
(16) Let's see if it fits <,> <ICE-GB:S1A-077 #171:1:B>

If we group these factual verbs that control the complement clause into the categories of *public verbs of speaking* and *private verbs of thinking* (Aarts *et al.* 2014: 154; see also Quirk *et al.* 1985: 1180–2), as discussed in section 2.1, we can observe a clear preference for verbs of thinking in both *if*- and *whether*-clauses, as shown in Table 3 below. The category of public for factual verbs includes “speech act verbs introducing indirect statements” (Quirk *et al.* 1985: 1180), whereas private verbs are those that express “intellectual states such as belief and intellectual acts such as discovery” (Quirk *et al.* 1985: 1181). Verbs classified under the label of private verbs of thinking include *know*, *wonder*, *see*, *doubt*, *decide*, *think*, *remember* and *consider*; while the rest of the verbs listed in Table 2 have been assigned to the category of public verbs of speaking. For more examples of verbs included under each category others than those listed in this paper, see Quirk *et al.* (1985: 1180–1182) and Aarts *et al.* (2014: 154).

Table 3. Absolute frequency and percentage of *if* and *whether* with public verbs of speaking and private verbs of thinking

	<i>If</i>	<i>Whether</i>
Public verb of speaking in the matrix clause	16 (14.81%)	25 (35.21%)
Private verb of thinking in the matrix clause	92 (85.19%)	46 (64.79%)
Total	108 (100%)	71 (100%)

The results of a chi-square test carried out show that the difference in the probability of using *if* with verbs of speaking and thinking is statistically significant at the 0.05 error level, as displayed in Figure 2 below, which plots 95% Wilson confidence intervals.

Figure 2. Probability of using *if* with verbs of speaking and thinking

4.3 Polarity of the matrix clause

Another variable examined in the corpus-based study was the polarity of the matrix clause. Seventy-three of the *if*-clauses display affirmative polarity, while 35 show negative polarity; in the case of clauses introduced by *whether*, again, affirmative polarity is more prominent—

49 clauses—whereas only 22 clauses show negative polarity. This is graphically depicted in Figure 3 below. A chi-square test shows that this difference, however, is not statistically significant at 0.05. Therefore, the polarity of the matrix clause does not influence the choice of complementizer by speakers.

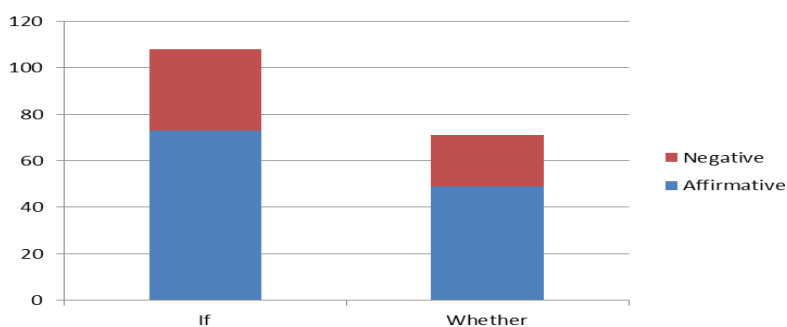


Figure 3. Affirmative and negative polarity of the matrix clause for *if* and *whether*

Dixon (2005) argues that certain lexical verbs show restrictions with respect to the polarity of the clause. This is the case, for example, of the verb *doubt*, which “commonly takes a *whether* complement in a positive clause” (Dixon 2005: 239). Preferences for certain lexical verbs for affirmative or negative clauses have not been explored in this paper, but this could explain some of the differences in frequencies encountered for *if* and *whether* here.

4.4 *Explicit choice with or*

In subordinate interrogatives introduced by *if* and *whether*, the speaker considers two possible scenarios; since these constructions are the indirect counterparts of *yes/no* questions (see Section 2.1). Both situations can either be explicitly mentioned in the clause, as it happens in (17); or not, but inferred from the context, as in (18).

Syntactically, as discussed in section 2.4, when *whether* is immediately followed by *or not*, no variation is possible with *if*. However, when *or not* occurs sentence-finally there is no restriction on the use of *if*, as can be observed in (17).

(17) I mean, I don't know if the guy added it on or not <,> <ICE-GB: S1A-039 #285:1:B>

(18) I just wondered if I can <,> if these are to take away <ICE-GB: S1A-077 #292:1:C>

When the choice between the alternatives is explicitly signalled in discourse, this can be marked by *or not*, as in (17). However, the choice can also be between two completely different situations, as in (19), in which the speaker does not propose making amendments or not, but rather, making amendments or proposing a separate motion.

(19) And we have to decide if we are going to propose any amendments to the Executive motions <,> or a separate motion from UCAUT <ICE-GB: S1B-077 #119:1:A>

Thus, another variable taken into account for the study was whether the explicit presence of a choice for the speaker in the complement clause marked by *or*, as in (17) or (19), as opposed to (18), triggered the use of one complementizer over the other. The results, as illustrated in Figure 4, reveal a clear preference for *whether* when the choice between possibilities is made explicit; while *if* is more likely when such a choice is not explicitly signalled in the clause. The clear preference for *whether* when the choice between two possible scenarios are explicitly mentioned in the clause is ultimately also connected to the etymology of this complementizer, which refers to mutually exclusive alternatives, as discussed in Section 2.2.

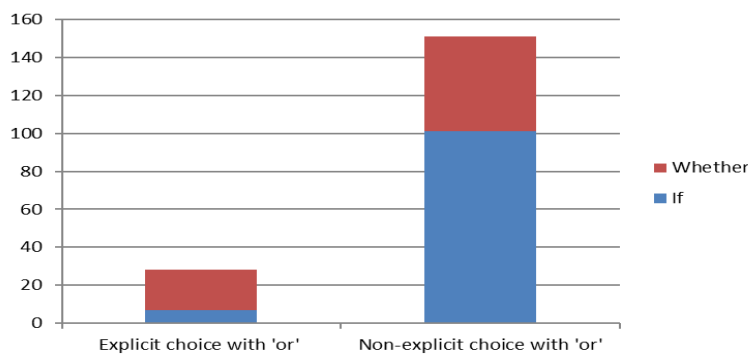


Figure 4. Presence or absence of an explicit choice with *or* in *if*- and *whether*-clauses

Figure 5 below displays the probability of *if* occurring when there is an explicit choice with the disjunctive conjunction *or* in the complement clause, or when it is not the case. The graph plots 95% Wilson confidence intervals, showing that the difference between *if* and *whether* when a choice marked by *or* appears explicitly in the clause is statistically significant at the 0.05 error level.

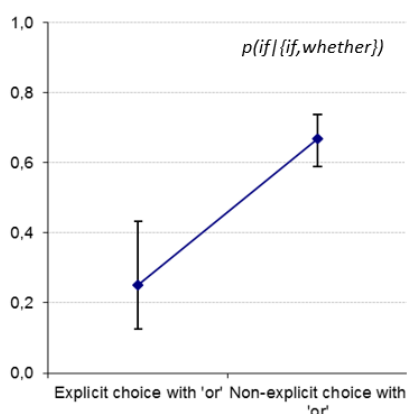


Figure 5. Probability of using *if* with the presence or absence of explicit choices with *or* in the complement clause

4.5 Text type

Since the variation between *if* and *whether* can be influenced by stylistic factors, it would not be surprising that there are differences in the variation of these complementizers across different genres. However, contrary to expectation, there is no clear difference between *if* and *whether* in spoken and written English in the sample analysed. It is not the case that one of the complementizers is used more frequently in speech and the other in writing, but rather, both complementizers occur much more frequently (>80%) in speech than writing, as shown in Table 4 below.

Table 4. Absolute frequency and percentage of *if* and *whether* in speech and writing

	<i>If</i>	<i>Whether</i>
Spoken	88 (81.48%)	59 (83.10%)
Written	20 (18.52%)	12 (16.90%)
Total	108 (100%)	71 (100%)

Nevertheless, a more detailed analysis between different subgenres in spoken and written discourse reveals clear preferences in certain genres for one complementizer over the other. Table 5 below displays the distribution of *if* and *whether* across the different spoken subgenres contained in ICE-GB. Since the different text types contained in ICE-GB do not have the same number of words, frequencies have been normalised per million words (pmw) to allow the comparison across genres.

Table 5. Distribution of *if* and *whether* across spoken genres (absolute and normalised frequencies per million words)

Text type	Number of words	<i>If</i>		<i>Whether</i>	
		N	pmw	N	pmw
Direct conversations	185,193	51	275.3	23	124.2
Telephone calls	20,415	8	391.9	1	49.0
Classroom lessons	42,208	1	23.7	1	23.7
Broadcast discussions	43,921	4	91.1	6	136.6
Broadcast interviews	22,147	3	135.5	2	90.3
Parliamentary debates	21,060	3	142.5	5	237.4
Legal cross-examinations	21,179	3	141.6	6	283.3
Business transactions	20,544	8	389.4	2	97.4
Spontaneous commentaries	42,472	2	47.1	1	23.5

Unscripted speeches	66,065	2	30.3	3	45.4
Demonstrations	22,563	2	88.6	1	44.3
Legal presentations	21,735	0	0	3	138.0
Broadcast news	42,962	0	0	2	46.6
Broadcast talks	43,506	1	23.0	2	46.0
Non-broadcast talks	21,592	0	0	1	46.31
Total	637,562	88	138.0	59	92.54

The results for speech report that *if* most frequently occurs in telephone calls, closely followed by business transactions and direct conversations; whereas *whether* is especially salient in legal cross-examinations and parliamentary debates. Overall, as pointed out by Biber *et al.* (1999: 691-693), genres in which language is more colloquial favour the use of *if*, as is the case of direct conversations and telephone calls. On the other hand, correlating with stylistic variation, the alternative with *whether* is more recurrent in more formal contexts, such as legal presentations; broadcast discussions, news, and talks; parliamentary debates; legal cross-examinations and legal presentations.

Turning now to written subgenres, as illustrated in Table 6, it is more difficult to extract conclusions from the very few data found, a total of 20 *if*-clauses and 12 *whether*-clauses. The results show that *if*- and *whether*-clauses appear more frequently in social letters than in other types of written texts. Letters, both of the social and business type, show a preference for *if*, although *whether* is also found. In contrast, in other subgenres the preference for one complementizer is much clearer, as is the case of fiction, in which all the instances correspond to *if*-clauses. This can be explained on the grounds of stylistic factors, since fiction, in general, is considered to be a more informal register.

Table 6. Distribution of *if* and *whether* across written genres (absolute and normalized frequencies per million words)

Text type	Number of words	<i>If</i>		<i>Whether</i>	
		N	pmw	N	pmw
Students' untimed essays	21,304	0	0	0	0
Students' examination scripts	21,225	0	0	0	0
Social letters	31,085	8	257.4	5	160.8
Business letters	30,491	4	131.2	2	65.6
Academic writing	85,586	0	0	1	11.7
Popular writing	86,645	1	11.5	2	23.1
Press reports	41,539	1	24.1	0	0
Administrative/regulatory writing	21,142	1	47.3	1	47.3
Skills and hobbies	21,199	0	0	0	0
Press editorials	20,719	0	0	1	48.3
Fiction	42,646	5	117.2	0	0
Total	423,581	20	47.21	12	28.3

Preference for certain patterns of matrix verbs and complementizers in specific genres should be further explored in future research. For example, the findings reported by Gawlik (2013: 137) for spoken academic American English show that the verb *know* occurs much more frequently preceding *if* than *whether* in this genre, whereas in this paper, where the focus is on general discourse and not on a specific genre, similar frequencies were encountered for this verb with both complementizers.

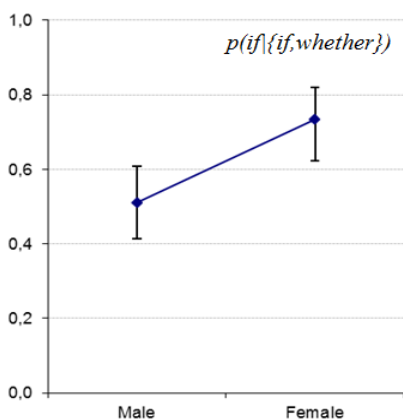
4.6 Sex of the speaker

Sociolinguistic aspects such as sex and age (see section 4.7 for age) have also been taken into account in the analysis. Table 7 below displays the distribution of *if* and *whether* by sex. The results reveal that percentage-wise more instances of *if* were uttered by women, whereas *whether* shows the reverse pattern, with more than 70% of the instances of this complementizer being used by males.

Table 7. Absolute frequency and percentages of *if* and *whether* according to sex⁹

	<i>If</i>	<i>Whether</i>
Male	50 (48%)	48 (71%)
Female	55 (52%)	20 (29%)
Total	105 (100%)	68 (100%)

The results of a chi-square test show that the probability that the choice of one complementizer over the other is motivated by the sex of the speaker is statistically significant at the 0.05 error level, as shown in Figure 6 below, which plots 95% Wilson confidence intervals.

Figure 6. Probability of using *if* by sex

4.7 Age of the speaker

Moving on to another sociolinguistic variable, the age of the speaker, the results indicate a preference for young speakers to use *if*. There is a clear correlation between this fact and the stylistics of *if*, since the language of younger speakers is more informal and colloquial. As shown in Figure 7 below, more than 70% of the participants aged 18-25 in the corpus choose the complementizer *if*. This percentage decreases slightly if we

⁹ For the variable of sex, a total of 6 tokens (3 instances of *if* and 3 instances of *whether*) have been excluded from the analysis since the corpus does not provide information about the sex of the speaker in certain texts.

consider the 26-45 age group, with 68% preferring *if*. However, the 46-65 age group shows the reversed preference, opting for *whether* over *if*, with only 45% of the speakers using the latter complementizer.

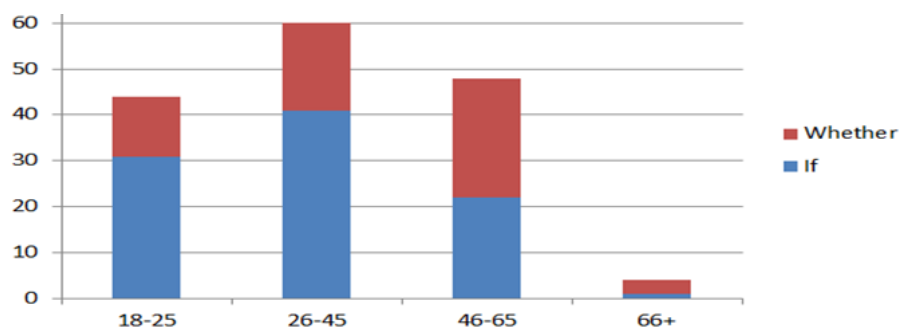


Figure 7. Use of *if* and *whether* according to age of the speaker¹⁰

Thus, we can conclude from this that there is a tendency for younger speakers to use *if* and that the tendency shifts towards *whether* in older users of the language. Such preference for one of the complementizers correlating with the age of the speakers seems to be progressive and in order to test the significance of these results, a chi-square test was carried out considering two age groups only, one between 18 and 25 years old and the other for 46 years old or more. Figure 8 below shows the probability of *if* in the 18-25 and 46+ age groups, which is statistically significant at the 0.05 error level. The graph displays 95% Wilson confidence intervals.

¹⁰ For the variable of age, a total of 23 tokens (13 instances of *if* and 10 instances of *whether*) have been excluded from the analysis since the corpus does not provide information about the age of the speaker in certain texts.

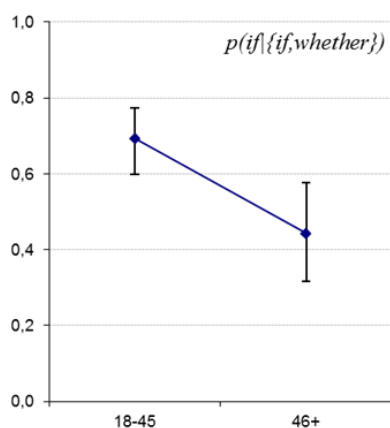


Figure 8. Probability of using *if* by age of the speaker.

5. Conclusions

This study has aimed to elucidate some of the factors that trigger the choice of subordinate interrogative complementizers, *if* and *whether*, by speakers; a topic which has received comparatively little attention in the literature on complementation in English. Although *if* and *whether* are interchangeable in most contexts, in certain grammatical environments some syntactic restrictions for *if* must be taken into account, making *if/whether* variation not possible under certain conditions, as discussed in section 2.4. The variation between *if* and *whether* is also conditioned by the semantics of both conjunctions, *if* being wider in scope than *whether*, with the latter being (etymologically) restricted to only two possibilities. In addition to this, there are stylistic differences in the use of these complementizers. Corpus-based studies, such as Biber *et al.* (1999: 691-693) have demonstrated that *if* is more characteristic of informal registers and this has been confirmed by the data retrieved from my corpus.

The corpus-based analysis on ICE-GB has revealed a number of factors that may influence the choice of complementizer by the speakers. First, the type of lexical verb may affect this choice. The results reveal that private verbs of thinking trigger the use of *if* more frequently than *whether*. If we analyse lexical verbs individually, however, we will observe that verbs such as *know* and *wonder* are highly frequent as matrix verbs licensing both *if*- and *whether*-clauses. Second, having an

explicit choice between alternatives in the complement clause may lead to the employment of *whether* instead of *if*. Third, differences in the use of *if* and *whether* across different subgenres of speech and writing have also been observed. Some genres show a clear preference for one complementizer over the other; a fact that correlates with stylistic variation (informal registers such as direct and telephone conversations choosing *if*, and formal registers such as legal presentations or broadcast discussions opting for *whether*). Fourth, sociolinguistic factors have also been proved to be significant in the choice of *if* over *whether*. Regarding sex, men have been shown to have a preference for *whether*, while women opt for *if*. The age of the speaker has also been found to be a determinant factor for the choice of complementizer. The data have revealed that younger speakers show a tendency to use *if*, whereas older speakers choose *whether* instead. Again, this difference can be explained in relation to stylistic variation, with younger speakers using more informal and colloquial language. Other factors such as the polarity of the main clause or the choice of spoken versus written discourse have been found to be not significant in playing a role for choosing *if* or *whether*.

The results obtained suggest that it might be worthwhile to carry out a study focusing on *if* and *whether* usage diachronically, to check the extent to which the variables examined also played a role in earlier English.

Appendix: Structure of ICE-GB¹¹

The table below displays the general structure of the spoken and written components of ICE-GB, with the number of texts of each type indicated in brackets.

Spoken Texts (300)	Dialogues (180)	Private (100)	face-to-face conversations (90) phonecalls (10)
		Public (80)	classroom lessons (20) broadcast discussions (20) broadcast interviews (10) parliamentary debates (10) legal cross-examinations (10) business transactions (10)
	Monologues (100)	Unscripted (70)	spontaneous commentaries (20) unscripted speeches (30) demonstrations (10) legal presentations (10)
		Scripted (30)	broadcast talks (20) non-broadcast speeches (10)
	Mixed (20)		broadcast news (20)
Written Texts (200)	Non-printed (50)	Non-professional writing (20)	untimed student essays (10) student examination scripts (10)
		Correspondence (30)	social letters (15) business letters (15)
	Printed (150)	Academic writing (40)	humanities (10) social sciences (10) natural sciences (10) technology (10)
		Non-academic writing (40)	humanities (10) social sciences (10) natural sciences (10) technology (10)
		Reportage (20)	press news reports (20)
		Instructional writing (20)	administrative/regulatory (10) skills/hobbies (10)
		Persuasive writing (10)	press editorials (10)
		Creative writing (20)	novels/stories (20)

¹¹ Extracted from <http://www.ucl.ac.uk/english-usage/projects/ice-gb/design.htm>, based on Nelson *et al.* (2002: 307-308).

References

- Aarts, Bas, Sylvia Chalker and Edmund Weiner. 2014. *Oxford Dictionary of English Grammar*. Oxford: Oxford University Press.
- Athanasiadou, Angeliki and René Dirven. (eds.). 1997. *On Conditionals Again*. Amsterdam: John Benjamins.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Couper-Kuhlen, Elizabeth and Bernd Kortmann. (eds.). 2000. *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter.
- Declerck, Renaat and Susan Reed. 2001. *Conditionals: A comprehensive empirical analysis*. Berlin: Mouton de Gruyter.
- Dixon, Robert M. W. 1991. *A New Approach to English Grammar, On Semantic Principles*. Oxford: Clarendon Press.
- Dixon, Robert M. W. 2005. *A Semantic Approach to English Grammar*. Oxford: Oxford University Press.
- Eckardt, Regine. 2007. "The syntax and pragmatics of embedded yes/no questions". *On Information Structure, Meaning and Form. Generalizations across languages*. Eds. Kerstin Schwabe and Susanne Winkler. Amsterdam: John Benjamins. 447-466.
- Elsness, Johan. 1984. "That or zero? A look at the choice of object clause connective in a corpus of American English". *English Studies* 65(6): 519-533.
- Finegan, Edward and Douglas Biber. 1995. "That and zero complementisers in late modern English: Exploring ARCHER from 1650-1990". *The Verb in Contemporary English: Theory and Description*. Eds. Bas Aarts and Charles F. Meyer. Cambridge: Cambridge University Press. 241-257.
- Gawlik, Oskar. 2013. "On 'if' and 'whether' complement clauses of 'see', 'wonder', and 'know' in contemporary spoken academic American English: A corpus-based study". *Respectus Philologicus* 24(29): 131-41.
- Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Huddleston, Rodney and Geoffrey K. Pullum *et al.* 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

- ICE-GB = International Corpus of English, British component. Compiled at the Survey of English Usage, University College London. <http://www.ucl.ac.uk/english-usage/projects/ice-gb/>
- Kaltenböck, Gunther. 2006. "...That is the question': Complementizer omission in extraposed that-clauses". *English Language and Linguistics* 10(2): 371–396.
- Kearns, Kate. 2007. "Epistemic verbs and zero complementizer". *English Language and Linguistics* 11(3): 475-505.
- López-Couso, María José and 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 and Belén Méndez-Naya. 2015. "Secondary grammaticalization in clause combining: From adverbial subordination to complementation in English". *Language Sciences* 47: 188-98.
- Nelson, Gerald, Sean Wallis and Bas Aarts. 2002. *Exploring Natural Language: The British Component of the International Corpus of English*. Amsterdam: John Benjamins.
- OED = Oxford English Dictionary Online. <http://www.oed.com/> (accessed 12 December 2017).
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Ransom, Evelyn N. 1988. "The grammaticalization of complementizers". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 364-74.
- Shank, Christopher, Koen Plevoets and Hubert Cuyckens. 2014. "A diachronic corpus-based multivariate analysis of 'I think that' vs. 'I think zero'". *Corpus Methods for Semantics. Quantitative studies in polysemy and synonymy*. Eds. Dylan Glynn and Justyna A. Robinson. Amsterdam: John Benjamins. 279-303.
- Stuurman, Frits. 1991. "If and whether: Questions and conditions". *Lingua* 83: 1-41.

- Traugott, Elizabeth Closs, Alice Ter Meulen, Judy Snitzer Reilly and Charles A. Ferguson. (eds.). 1986. *On Conditionals*. Cambridge: Cambridge University Press.
- Wierzbicka, Anna. 1997. "Conditionals and counterfactuals: Conceptual primitives and linguistic universals". *On Conditionals Again*. Eds. Angeliki Athanasiadou and René Dirven. Amsterdam: John Benjamins.15-59.