RiCL Research in Corpus Linguistics

Masked by annotation: Minor declarative complementizers in parsed corpora of historical English

María José López-Couso – Belén Méndez-Naya
University of Santiago de Compostela / Spain

Abstract – This article discusses some of the potential problems derived from the syntactic annotation of historical corpora, especially in connection with low-frequency phenomena. By way of illustration, we examine the parsing scheme used in the *Penn Parsed Corpora of Historical English* (PPCHE) for clauses introduced by so-called 'minor declarative complementizers', originally adverbial links which come to be occasionally used in complementizer function. We show that the functional similarities between canonical declarative complement clauses introduced by the major declarative links *that* and zero and those headed by minor declarative complementizers are not captured by the PPCHE parsing, where the latter constructions are not tagged as complement clauses, but rather as adverbial clauses. The examples discussed reveal that, despite the obvious advantages of parsed corpora, annotation may sometimes mask interesting linguistic facts.

Keywords – annotation; parsing; English historical corpora; minor declarative complementizers; indeterminacy

1. INTRODUCTION[1]

Since the advent of the first computerized corpora in the 1960s (e.g. the compilation of the pioneering *Brown Corpus* by W. Nelson Francis and Henry Kučera), corpus linguistics has experienced exponential growth. In just half a century we have witnessed the creation of an impressive range of corpora, written, spoken and multimodal, small collections of data alongside mega-corpora and reference corpora side by side with different types of specialized corpora.

One of the milestones of modern corpus linguistics has undoubtedly been the development of various types of annotation systems, such as tagging and parsing, which

add significantly to the potentials of earlier collections of raw data, allowing the analyst to run quicker and more effective searches. It must be acknowledged, however, that linguistic annotation also has certain disadvantages, mostly because any annotation system implies the acceptance of particular theoretical premises, no matter how inclusive the annotators claim to be. Even the identification and tagging of a basic grammatical category such as 'preposition' can be controversial (see Huddleston and Pullum *et al*. 2002: 598–601, who include traditional subordinating conjunctions, such as *since* or *because*, under the category 'preposition', *contra* Quirk *et al*. 1985: 658–661). Linguistic annotation of historical material may be even more problematic, since parsing has to account for language change, especially if the annotation system is supposed to hold for successive stages in the history of a given language. Consider, for instance, the intrinsic difficulties in the annotation of so-called 'bridging contexts' in grammaticalization (Heine 2002), where parsing may mask cases of potential syntactic and semantic indeterminacy or ambiguity, which are central to our understanding of linguistic change.

Further problems with annotation may arise in the treatment of low-frequency phenomena, which tend to be overlooked. One of these low-frequency features is so-called 'minor declarative complementizers', a category which we have analyzed in depth for English from both a theoretical and a diachronic point of view (for an overview, see López-Couso and Méndez-Naya 2015). Minor declarative complementizers are connectives whose main function is that of marking various kinds of adverbial relations (e.g. condition, concession, purpose, comparison, etc.), but which also serve a secondary function to introduce finite complement clauses, as equivalents (or near equivalents) of the major declarative complementizers *that* and zero. Examples of such minor complementizers in English include *if*, *though*, *as if*, *as though*, *like*, *lest* and *but* (see examples (3a)–(3e) in Section 3 below).

In this article we draw attention to some of these minor declarative complementizers by examining the way in which they are annotated in the *Penn Parsed Corpora of Historical English* (Kroch and Taylor 2000; Kroch *et al*. 2004, 2016), with the aim of checking whether parsing overlooks diachronic facts and/or masks diachronic developments. The outline of the discussion is as follows. In Section 2 we introduce the *Penn Parsed Corpora of Historical English* and their common annotation scheme. Section 3, in turn, summarizes the most relevant information about minor declarative complementizers from our earlier research, focusing on the various structural and

semantic criteria which lead us to consider the clauses introduced by these connectives as complements rather than as adjuncts. Then, in Section 4 we show how such clauses are annotated in the corpora, as compared to canonical cases of finite complement clauses introduced by the major declarative complement-clause links *that* and zero. Finally, Section 5 offers some concluding remarks.

2. THE *PENN PARSED CORPORA OF HISTORICAL ENGLISH* AND THEIR ANNOTATION SYSTEM

A major landmark in the history of English historical corpora was the release in 1991 of the *Helsinki Corpus of English Texts* (HC), a project launched by Matti Rissanen and his collaborators at the Department of English of the University of Helsinki back in 1984. The HC is a 1.5 million-word corpus which contains text material from the time of the earliest written records of English in the eighth century up to the first decade of the eighteenth century, representing a wide range of genres, both formal (e.g. philosophical treatises) and informal (e.g. comedies).[2] Almost three decades after its publication and despite its small size according to modern standards, the HC still remains an excellent resource for corpus-based research on the long diachrony of English and is still successfully used world-wide as a "diagnostic" corpus (Rissanen 2008: 59) for the analysis of processes of language change taking place in the Old, Middle and Early Modern English periods.

In order to expand the potentialities of the original 'raw' version of the HC, a number of complementary corpora have appeared over the last couple of decades or so. Of special relevance for our purposes in this article are the annotated (tagged or parsed) editions for various historical sub-periods developed by a team of scholars based at the University of Pennsylvania and at the University of York, a project aimed at producing syntactically annotated corpora for the different stages in the history of the English language. Though based on the raw version of the HC, these corpora contain considerably larger text samples than those in the HC, together with some new material not available in the original corpus. The advantages of these complementary annotated corpora over their raw counterparts are more than evident: in addition to searches for simple words or word strings, they allow searching for syntactic constructions, including empty or covert

---

[2] For full details about the HC, see the third edition of the manual by Kytö (1996) and the corresponding entry for the corpus in the *Corpus Resource Database* (CoRD) at http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.

categories, such as empty subjects or zero complementizers, thus conveniently facilitating the analyst's task, especially when high-frequency phenomena are in focus.

At present, the *Penn Parsed Corpora of Historical English* (PPCHE) comprise the following datasets:

- *York-Helsinki Parsed Corpus of Old English Poetry* (YCOEP)

- *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (YCOE)

- *Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English*

- *Penn-Helsinki Parsed Corpus of Middle English, second edition* (PPCME2)

- *Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME)

- *York-Helsinki Parsed Corpus of Early English Correspondence* (PCEEC)

- *Penn Parsed Corpus of Modern British English, second edition* (PPCMBE2)

The PPCHE are presented in three different formats: simple or raw text, part-of-speech (POS) tagged and parsed text, which combines both POS and syntactic annotation (treebanks). For the annotated versions of the PPCHE, the compilers adopted a simplified version of the Principles and Parameters theory. The annotation scheme used is ultimately based on the system developed for the *Penn Treebank*, a corpus of over 4.5 million words of American English (Marcus *et al*. 1993). This system was adapted to historical material by Ann Taylor and Anthony Kroch for the second edition of the *Penn-Helsinki Parsed Corpus of Middle English* (Kroch and Taylor 2000) and was then revised for the 2016 update of the PPCHE.[3]

In the annotation manual (Santorini 2016) it is made clear that the primary goal of the Penn-Helsinki annotation system is to facilitate automated searches, "not to give the correct linguistic analysis of each sentence." In other words, practical purposes are clearly privileged over grammatical ones. In addition, Santorini mentions that in the annotation process subjective judgements have been avoided "since they are extremely error-prone." This explains why distinctions such as, for instance, adjectival vs. verbal passive participles are disregarded. Practical issues also prevail, for instance, in the different

---

[3] The annotation scheme used for the PPCHE has also been applied to various other historical datasets, both for English (e.g. the *Parsed Corpus of Middle English Poetry*; https://pcmep.net/index.php) and for other languages, including Portuguese, French, Icelandic and Japanese, among others. For a list of corpora sharing the same annotation scheme, see https://www.ling.upenn.edu/ppche/ppche-release-2016/other-corpora.html.

parsing given for the clauses introduced by *because* and variants, depending on whether the form is written together or apart. In the first case the clause following *because* is analyzed as an adverbial clause (CP-ADV), whereas in the second case the clause is treated as a complement clause (CP-THT):

> BECAUSE is treated as a fused form. When it is written together, the clause following it is treated as the CP-ADV complement of the compound head P+N. When it is written apart, the clause following it is treated as a THAT complement of the noun CAUSE.

```
( (IP-MAT (CONJ but)
          (NP-SBJ (NPR$ Balynes) (N oste))
          (MD myght)
          (NEG $not)
          (BE be)
          (VAN lette)
          (RP in)
          (PP (P+N because)
              (CP-ADV (C 0)
                      (IP-SUB (NP-SBJ (PRO he))
                              (HVD had)
                              (NP-OB1 (Q no) (N lady)))))
          (. .))
          (ID CMMALORY,63.2096))


( (IP-MAT (CONJ but)
          (PP (P by)
              (NP (N cause)                          ← by cause that
                  (CP-THT (C 0)
                          (IP-SUB (NP-SBJ (PRO he))
                                  (VBD knewe)
                                  (NEG not)
                                  (NP-OB1 (PRO$ his) (N sheld))))))
          (NP-SBJ (PRO he))
          (VBD demed)
          (CP-THT (C 0)
                  (IP-SUB (NP-SBJ (PRO it))
                          (BED was)
                          (NEG not)
                          (NP-OB1 (PRO he))))
          (. .))
          (ID CMMALORY,68.2300))
```

This parsing, which clearly favors automation, relies exclusively on the expressions' surface structure: *because* is analyzed as a one-word item that governs an adverbial CP, while in the variant *by cause*, the first word is marked as governing the noun phrase headed by the noun *cause*.

As mentioned in Section 1, in contrast to the annotation of contemporary data, the tagging and parsing of historical material pose special challenges to both annotators and corpus users, especially when it comes to the interpretation of items and constructions

undergoing processes of language change such as grammaticalization and lexicalization, which imply alterations in the status of a given item or construction over time (e.g. from lexical item to grammatical item; from syntactic construction to lexical item). The parsing system of the PPCHE conveniently tries to accommodate such changing diachronic facts. This applies in particular to differences between the Middle English corpus (PPCME2) and the later corpora. Thus, for example, the annotation reflects the emergence of conjunctions and adverbs out of phrases and clauses, providing a different parsing for the source constructions and for the grammaticalized elements. By way of illustration, consider the convincing explanation provided in the manual for the development of the subordinator *albeit*:

ALL BE IT (THAT), ALBEIT

In the PPCME2, ALL BE IT (THAT) clauses, like SO BE IT (THAT) clauses, are treated similarly to V1 conditionals. ALL is POS-tagged Q, surrounded by ADVP brackets, and treated as a daughter of CP-ADV. This is not intended as the correct analysis of the construction, but rather to fit in with the annotation of V1 conditionals.

```
( (IP-MAT (CONJ and)
          (PP (P atte)
              (NP (N risyng)
                  (PP (P of)
                      (NP (D the)  (N sonne)))))
          (NP-SBJ (PRO I))
          (VBD fond)
          (NP-OB1 (D the) (ADJ secunde) (N degre)
                  (PP (P of)
                      (NP (NPR Aries))))
          (IP-PPL (VAG sittyng)
                  (PP (P upon)
                      (NP (PRO$ myn) (N est) (N orisonte))))
          (, ,)
          (CP-ADV (ADVP (Q all))                          ← ALL BE IT
                  (IP-SUB (BEP be)
                          (NP-SBJ-1 (PRO it))
                          (CP-THT-1 (C that)
                                    (IP-SUB (NP-SBJ (PRO it))
                                            (BEP was)
                                            (ADJP (FP but) (ADJ litel))))))
          (. .))
  (ID CMASTRO,673.C1.364))
```

In the later corpora, ALBEIT (like HOWBEIT) is treated as a unitary adverb (when used absolutely) or as a unitary preposition (when introducing a subordinate clause).

```
(NODE (CP-CAR (WNP-1 (WPRO Which))
           (C 0)
           (IP-SUB (PP (P in)
                       (NP (NP-POS (D the) (N$ kinges))
                           (NS daies)))
                   (, ,)
                   (PP-LFD (P albeit)              ← ALBEIT
                           (CP-ADV (C 0)
                                   (IP-SUB (NP-SBJ (PRO he))
                                           (BED was)
                                           (ADVP (ADV sore))
                                           (VAN ennamored)
                                           (PP (P vpon)
                                               (NP (PRO her))))))
                   (, ,)
                   (ADVP-RSP (ADV yet))
                   (NP-SBJ-RSP=1 (PRO he))
                   (VBD forbare)
                   (NP-OB1 (PRO her))
       (ID MORERIC,55.118))
```

Table 1 extracted from the manual[4] summarizes the differences in the treatment of *albeit* in the PPCME2 and in the later corpora.

Item	PPCME2	Later corpora

ALL BE IT, ALBEIT (see Concessive clauses)	Always phrasal.
	`(Q all) (BEP be) (PRO it)`
		Unitary adverb or preposition.
		`(ADV albeit)`
		`(ADV (ADV31 al) (ADV32 be) (ADV33 it))`
		`(P albeit)`
		`(P (P31 al) (P32 be) (P33 it))`

Table 1: The annotation of *albeit* in the PPCME2 and in the later PPCHE corpora

Moreover, the annotators of the PPCHE also acknowledge the existence of ambiguity by explaining alternative analyses, even though the annotation finally opts for a default interpretation, as shown below in the case of the verb *do*:

In Middle English, DO can be ambiguous between a causative (ECM) main verb and a periphrastic auxiliary. The default in the PPCME2 is to treat ambiguous cases as causative except when a causative reading is impossible. Causative DO dies out in the course of Middle English, and so instances of DO in the later corpora that could in principle be treated as ambiguous and hence causative by default are instead uniformly treated as periphrastic.

Though recognizing ambiguity, practical purposes finally prevail in the annotation used in the PPCHE system. In addition to facilitating the retrieval of examples, this solution has the obvious advantage of enabling the automation of the annotation process.

---

[4] https://www.ling.upenn.edu/hist-corpora/annotation/index.html

In the remainder of this article we examine another controversial area of syntactic interpretation, namely so-called 'minor declarative complementizers', and discuss the way(s) in which such subordinators and the clauses introduced by them are treated in the PPCHE annotation scheme.

3. INTRODUCING MINOR DECLARATIVE COMPLEMENTIZERS

Most Present-day English reference grammars (see Quirk *et al*. 1985: 1047ff; Biber *et al*. 1999: 192ff) distinguish three main classes of subordinate clauses on the basis of their potential functions in the complex sentence: complement clauses, which realize functions that approximate to those of noun phrases; relative clauses, which resemble adjectives in function; and adverbial clauses, which are found in functions more closely associated with adverbial and prepositional phrases, expressing satellite relations and acting as adjuncts or modifiers. These different functional categories of subordinate clauses are introduced by various kinds of markers indicating the type of relating function which exists between the subordinate clause and its corresponding superordinate clause. The two types of subordinate-clause links which are relevant for the present discussion are those introducing complement clauses (i.e. complementizers) and subordinators which mark adverbial clauses of various kinds.

Finite declarative complement clauses are typically introduced by the complementizers *that* or zero, as in (1a)–(1b).

(1a) I noticed **that** *he spoke English with an Australian accent*. (from Quirk *et al*. 1985: 1049)

(1b) I know **Ø** *it's late*. (from Quirk *et al*. 1985: 1049)

In turn, adverbial clauses are normally marked by the presence of different subordinators in clause-initial position. These markers signal the various kinds of semantic relations which may hold between the main clause and the sub-clause, among them time (2a), reason (2b), condition (2c), concession (2d), exception (2e), (negative) purpose (2f), comparison (2g), etc. (see Quirk *et al*. 1985: 1077ff; Kortmann 1997: 79ff; Biber *et al*. 1999: 818ff).

(2a) **Since I saw her last**, she has dyed her hair. (from Quirk *et al*. 1985: 1078)

(2b) The flowers are growing so well ***because*** *I sprayed them*. (from Quirk *et al*. 1985: 1103)

(2c) ***If*** *Mary had visited her parents yesterday*, she would have known about their problems. (from Kortmann 1997: 85)

(2d) He can walk faster than I can, ***though*** *he is well over eighty*. (adapted from Quirk *et al*. 1985: 1097)

(2e) Rumsfeld: "There is no question ***but that*** *the invasion would be welcomed*." (COCA, 2006, MAG)

(2f) I sent the children to bed ***lest*** *they (should) hear their parents quarrel*. (from Kortmann 1997: 86)

(2g) He treats me ***as if*** *I am a stranger*. (from Quirk *et al*. 1985: 1110)

Interestingly, some of the adverbial connectives illustrated in (2a)–(2g) may also show a secondary or subsidiary function beyond the domain of adverbial subordination and are also used (or have been used at different stages in the history of English) to introduce finite declarative complements, as equivalents or near-equivalents of the major declarative complementizers *that* and zero. Illustrative examples are given in (3).

(3a) It would be a real comfort to me ***if*** *you would make me feel we belonged to each other*. (ARCHER, 1893pine.d6b; from López-Couso and Méndez-Naya 2014: 98)

(3b) and therfore, ***though*** *I be wrooth and inpaciente*, it is no merveille (c. 1390, Chaucer, Tale of Melibee 232.C2; from López-Couso and Méndez-Naya 2001: 99)

(3c) I don't doubt ***but that*** *she meant it*. (from Huddleston and Pullum *et al*. 2002: 971)

(3d) He suggested offering half to Sir Edward, fearing ***lest*** *"he shall thinke it to good for us and procure it for himselfe, as he served us the last time"*. (Brown G64 S25)

(3e) It seems ***as if*** (***as though***) *we're in a bad situation*, no matter your point of view. (adapted from COCA, 2018, SPOK)

However, while the variation between *that* and zero has been widely examined in the extensive literature on clausal complementation both from a synchronic and from a diachronic perspective (see, among many others, Elsness 1982, 1984; Warner 1982; Fanego 1990; Rissanen 1991; Finegan and Biber 1995; López-Couso 1996; Tagliamonte and Smith 2005; Kaltenböck 2006; Kearns 2007; Torres Cacoullos and Walker 2008), the complementizer function of these subordinators has been largely overlooked.[5] More importantly for the purposes of the present article, this neglect has had serious implications for the way in which these connectives and the clauses introduced by them have been annotated in parsed corpora.

In various publications we have drawn attention to the complementizer use of these originally adverbial links, which we have labelled 'minor declarative complementizers', and have dealt with their origin, development and present-day use. In particular, we have examined *but (that)* (López-Couso and Méndez-Naya 1998), *if* and *though* (López-Couso and Méndez-Naya 2001, 2014), *lest* (López-Couso 2007) and *as if*, *as though* and *like* (López-Couso and Méndez-Naya 2012a; 2012b). In these articles we have argued that even though clauses such as those italicized in (3a)–(3e) above resemble adverbial clauses at first sight, they nevertheless meet a number of criteria which favor a complement analysis. What follows summarizes the discussion of the criteria for complementhood that we proposed in López-Couso and Méndez-Naya (2015).

(a) Licensing. The most central criterion is licensing, inasmuch as complements depend on the presence of a predicate "that licenses them" (Huddleston and Pullum *et al.* 2002: 219). In contrast to the adverbial clauses in (2), the occurrence of the subordinate clauses in (3) requires the presence of a particular kind of predicate. In example (3d), for instance, the clause introduced by *lest* is perfectly grammatical with a predicate of fearing such as the verb *fear*, but it would be ungrammatical with the utterance predicate *say* (see (4)).[6]

(3d) He suggested offering half to Sir Edward, <u>fearing</u> ***lest*** *"he shall thinke it to good for us and procure it for himselfe, as he served us the last time"*.

---

[5] Minor declarative complementizers are discussed in passing in Lakoff (1968: 69, note 7); Huddleston (1971: 177−178); Warner (1982: 180−185, 221−224); Mitchell (1985: §§1960−1961); Noonan (1985: 104); Quirk *et al.* (1985: 1175, note a); McCawley (1988: 143); Dirven (1989: 134); Fanego (1990: 19−20); Huddleston and Pullum *et al.* (2002: 962, 1151−1152); Dancygier and Sweetser (2005: 229−230) and Taylor and Pang (2008: 130). The comparative complementizers *as if* and *as though* are discussed in greater detail in Bender and Flickinger (1999), Rooryck (2000) and Brook (2014, 2018).

[6] We follow here the classification of semantic predicates proposed by Noonan (1985).

(4) *He suggested offering half to Sir Edward, <u>saying</u> ***lest*** *"he shall thinke it to good for us and procure it for himselfe, as he served us the last time"*.

(b) Obligatoriness. As a direct consequence of licensing, complements are obligatory constituents in clause structure; in other words, their omission would compromise the grammaticality of the sequence. As can be seen, the italicized clauses in (3) are not omissible (see (5)), which clearly supports a complementation analysis for them. By contrast, adverbial clauses, like the concessive *though*-clause in (2d), can easily be left out, as shown in (6). As complements, the sub-clauses in (3) show therefore a higher degree of integration into the corresponding matrices than the adverbial clauses in (2).

(3e) It seems ***as if*** *we're in a bad situation*.

(5) *It seems.

(2d) He can walk faster than I can, ***though*** *he is well over eighty*.

(6)　He can walk faster than I can.

(c) Replacement by unambiguous declarative complement clauses. Further evidence in favor of the complement status of the subordinate clauses in (3) is provided by their ability to be replaced by prototypical declarative complement clauses, either finite or non-finite, "without any perceptible change of meaning" (Huddleston and Pullum *et al.* 2002: 962). Consider, for instance, the alternatives provided in (7) and (8) for the sequences in (3a) and (3e), respectively.

(3a) It would be a real comfort to me ***if*** *you would make me feel we belonged to each other*.

(7) It would be a real comfort to me *for you **to** make me feel we belonged to each other*.

(3e) It seems ***as if*** *we're in a bad situation*.

(8) It seems ***that*** *we're in a bad situation*.

(d) Impossibility of replacement by equivalent adverbial links. Another clear indication that the subordinators in bold in (3) realize a complementizer function is their ability to be replaced by prototypical declarative complementizers (see (3e) and

(8) above), while they are not interchangeable with other adverbial links belonging to their original semantic domains. Note, for instance, that the conditional subordinator *on condition that* cannot substitute for *if* in (3a) and that *but that* cannot be replaced by the marker of exception *except that* in (3c).

(3a) It would be a real comfort to me ***if*** *you would make me feel we belonged to each other.*

(9) *It would be a real comfort to me ***on condition that*** *you would make me feel we belonged to each other.*

(3c) I don't doubt ***but that/ that*** *she meant it.*

(10) *I don't doubt ***except that*** *she meant it.*

(e) Coordination with prototypical complements. Furthermore, the subordinate clauses in (3) can be coordinated with prototypical complement clauses, thus testifying to their complement status. In (11), for instance, an *if*-clause is coordinated with a *that*-clause, both clauses functioning as complements of the complement-taking predicate *feel*.

(11) Now, driving the horse and sulky borrowed from Mynheer Schuyler, he felt ***as if*** *every bone was topped by burning oil* <u>*and*</u> ***that*** *every muscle was ready to dissolve into jelly and leave his big body helpless and unable to move*. (Brown K14)

(f) Pronominalization. Sub-clauses like those in (3) also meet another criterion for complementhood, namely pronominalization (see McCawley 1988: 143): as shown in (12) and (13), the clauses in (3) can be recovered by the anaphoric elements *that* and *so*, respectively, just like complements do.

(3a) It would be a real comfort to me ***if*** *you would make me feel we belonged to each other*.

(12) **That** would be a real comfort to me.

(3e) It seems ***as if*** *we're in a bad situation*.

(13) It seems **so**.

(g) Pseudo-clefting. A final piece of evidence comes from pseudo-clefting: clauses introduced by minor declarative complementizers are co-referential with *what* in a pseudo-cleft construction, as shown by the comparison of (3a) and (14).

(3a) It would be a real comfort to me *if you would make me feel we belonged to each other*.

(14) What would be a real comfort to me would be *if you would make me feel we belonged to each other*.

The application of the aforementioned criteria clearly shows that the subordinate clauses introduced by *if*, *though*, *but*, *lest*, *as if* and *as though* in the examples in (3) should be analyzed as complements rather than as adjuncts, despite the fact that they are introduced by subordinators which typically function as adverbial connectives.

In our previous research, and taking as a starting point several historical and Present-day English corpora, we have shown that the adverbial function has been the original historical function for these links, while their complementizer use is a derived function. This is shown in Figure 1 (adapted from López-Couso and Méndez-Naya 2015: 191), which provides the time-depth of both the adverbial function (blue lines) and the complementizer use (red lines) of *if*, *though*, *but*, *lest*, *as if* and *as though*.
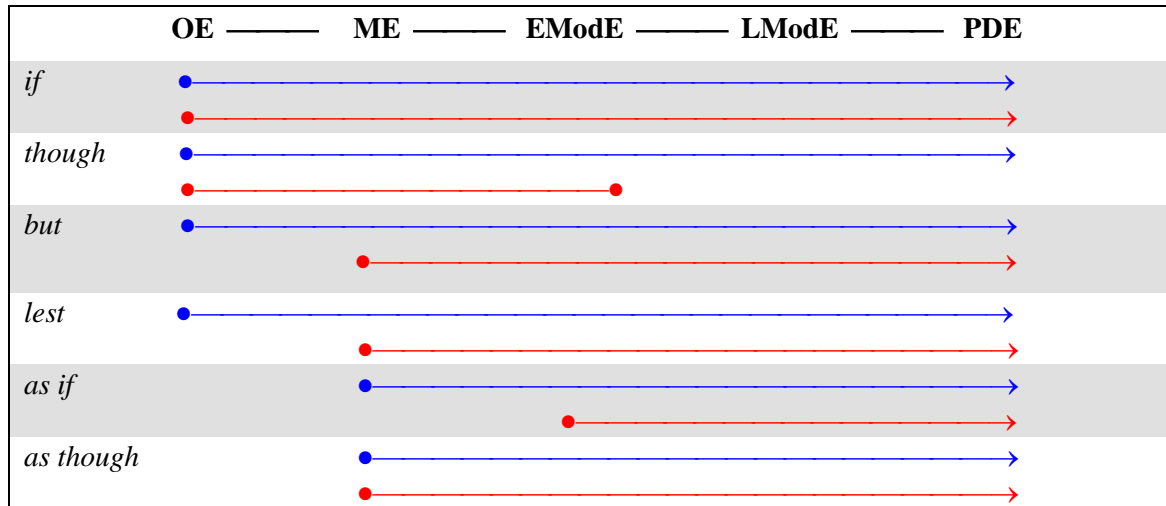

Figure 1: Timeline of *if*, *though*, *but*, *lest*, *as if* and *as though* in their functions as adverbial subordinators (blue lines) and declarative complementizers (red lines) (adapted from López-Couso and Méndez-Naya 2015: 191)

The adverbial function is not only the original use of these subordinators, but has also been the most frequent one all through their recorded history. Thus, for instance, the adverbial function of *lest* in the HC and ARCHER material represents almost 90% of all

occurrences of the subordinator (see López-Couso 2007). Similarly, adverbial *as if* and *as though* show a ratio of 3:1 in these corpora with respect to their complementizing function (López-Couso and Méndez-Naya 2012b).

The low frequency of the complementizer use of these subordinators is one of the reasons why minor declarative complementizers have been overlooked in the literature, where the default interpretation for the italicized clauses in (3) is the adverbial one. In view of this, the way in which clauses introduced by these minor complement-clause links are parsed in corpora is worth examining.

4. THE PARSING OF CLAUSES INTRODUCED BY MINOR DECLARATIVE COMPLEMENTIZERS IN THE *PENN PARSED CORPORA OF HISTORICAL ENGLISH*

As opposed to more traditional grammars like those by Quirk *et al*. (1985) and Biber *et al*. (1999), and in line with Huddleston and Pullum *et al*. (2002) and the Principles and Parameters framework (see Section 2 above), the annotators of the PPCHE establish a distinction between 'complementizers' and 'prepositions'. In the first group they include *that*,[7] zero, *as* and Middle English *þe*, while the second group comprises both traditional prepositions and traditional subordinating conjunctions (other than 'complementizers').[8] In other words, traditional subordinating conjunctions are treated as prepositions taking a clausal complement (CP = 'complementizer phrase', that is, a clause headed by a complementizer) and are consequently tagged P.[9] As to the types of finite subordinate clauses (tagged CP), the parsing system distinguishes adverbial, *that*-clauses, degree complements, questions, exclamations and relative clauses.[10] Examples (15) and (16) below illustrate the tagging of an adverbial and a *that*-clause, respectively.

---

[7] Note that *that* is regarded as a clause subordinator, which can introduce complement clauses and also non-*wh-* relative clauses; see Huddleston and Pullum *et al*. (2002: 1034, 1056–1057).

[8] Huddleston and Pullum *et al*. (2002: 600) also include traditional adverbs within prepositions. They regard them as prepositions without a complement.

[9] Similarly, Huddleston and Pullum *et al*. (2002: 604) regard traditional subordinating conjunctions as prepositions taking "non-expandable [i.e. zero] content clauses" as complements.

[10] All subordinate clauses that are headed by a complementizer are labelled CP. In addition, there are four types of subordinate clauses whose primary label is IP (infinitives, small clauses, adjunct participials and absolutes). All subordinate clauses, both CP and IP, have a dash tag indicating their type. Finally, there are reduced relative clauses, which are labelled RRC. For further details, see https://www.ling.upenn.edu/hist-corpora/annotation/index.html.

(15) and *when the Ink was made Cleer again by the Oyl of Vitriol*, the affusion of dissolv'd <font> Sal Tartari <$$font> seem'd but to Praecipitate, (BOYLECOL-E3-P1,137.30)

```
( (IP-MAT (CONJ and)
        (PP (P when)
            (CP-ADV (C 0)
                    (IP-SUB (NP-SBJ-1 (D the) (N Ink))
                            (BED was)
                            (VAN made)
                            (IP-SMC (NP-SBJ *-1)
                                    (ADJP (ADJ Cleer))
                                    (ADVP (ADV again)))
                            (PP (P by)
                                (NP (D the)
                                    (N Oyl)
                                    (PP (P of)
                                        (NP (N Vitriol))))))))
        (, ,)
        (NP-SBJ (D the)
                (N affusion)
                (PP (P of)
                    (NP (VAN dissolv'd)
                        (LATIN (CODE <font>) (FW Sal) (FW Tartari) (CODE
                        <$$font>)))))))
```

(16) I have already said, *the old general was kill'd by the shot of an arrow* (BEHN-E3-P1,155.118)

```
( (IP-MAT (NP-SBJ (PRO I))
        (HVP have)
        (ADVP-TMP (ADV already))
        (VBN said)
        (, ,)
        (CP-THT (CP-THT (C 0)
                        (IP-SUB (NP-SBJ (D the) (ADJ old) (ADJ general))
                                (BED was)
                                (VAN kill'd)
                                (PP (P by)
                                    (NP (D the)
                                        (N shot)
                                        (PP (P of)
                                            (NP (D an) (N arrow)))))
```

Let us now examine how structures involving minor declarative complementizers are parsed in the PPCHE. For practical purposes, we will focus on the Early Modern English period, when, as shown in Figure 1 above, the complementizer function is attested for all the items under study. By way of illustration, consider (17a), an example of a complement clause introduced by the minor declarative complementizer *if*, a complementizer which is typically associated with expressions meaning 'wonder' (López-Couso and Méndez-Naya 2001). This example will be compared with (17b), which features a parallel structure with the major complementizer zero.

(17a) Therefore <u>it</u> was no wonder ***if*** *we could not understand the Divine Essence* (BURNETROC-E3-P2,103.103)

(17b) […], <u>t</u>is noe wonder **Ø** *they should not like it*. (LOCKE-E3-P2,66.34)

According to our analysis, based on the criteria for complementhood outlined in Section 3 above, the two instances show an extraposed complement clause in subject function which is anticipated in the matrix clause by the dummy pronoun *it* (underlined in the examples). The only structural difference between the two instances lies in the choice of subordinator: the minor complementizer *if* in (17a) and the major complement-clause connective zero in (17b). Note that although the two clauses are structurally similar, complementizer selection clearly signals a semantic difference as regards the speaker's degree of commitment to the truth of the proposition expressed in the matrix clause, which seems to be higher in the case of (17b) than in (17a).

Here follows the parsing for these two sentences in the PPCEME:

```
(17a')  ( (IP-MAT (PP (ADV+P Therefore))
          (NP-SBJ (PRO it))
          (BED was)
          (NP-OB1 (Q no) (N wonder))
          (PP (P if)
             (CP-ADV (C 0)
                     (IP-SUB (NP-SBJ (PRO we))
                             (MD could)
                             (NEG not)
                             (VB understand)
                             (NP-OB1 (D the) (ADJ Divine) (N Essence)))))
          (. :))

(17b')  (NP-SBJ-1 (PRO $'t))
          (BEP $is)
          (CODE {TEXT:tis})
          (NP-OB1 (Q noe) (N wonder))
          (CP-THT-1 (C 0)
                  (IP-SUB (NP-SBJ (PRO they))
                          (MD should)
                          (NEG not)
                          (VB like)
                          (NP-OB1 (PRO it))))
          (. .))
```

As can be seen, even though the two examples contain parallel structures which only differ as regards the connective introducing the sub-clause (*if* in (17a) vs. zero in (17b)), the parsing fails to capture the obvious structural similarities which, in our view, exist between the two sequences. While in (17a) *if* is analyzed as a preposition P, taking a subordinate adverbial clause as complement (CP-ADV), in (17b) the sub-clause is tagged as a complement (CP-THT), attached immediately below the predicate (*be no wonder*).

(a) `(PP (P if)`
     `    (CP-ADV (C 0)`

(b) `(CP-THT-1 (C 0)`

We also find here a different treatment of the pronoun *it* in the two parsings: while in (17b) the zero clause is co-indexed with the anticipatory subject *it* (NP-SBJ-1), no co-indexing is present in the clause introduced by *if* in (17a) (NP-SBJ).

It should be noted, however, that the PPCEME is not completely consistent in the treatment of anticipatory *it* in such cases. In (18) below, an example of an *if*-clause complementing the matrix *it is meruayle*, very similar to the sequence in (17a), the *if*-clause is indeed co-indexed with the pronoun *it* in the matrix, just as the zero clause in our earlier instance (17b).

(18) and it is meruayle, *if thou scape with thy lyfe*, (FITZH-E1-H,101.376)

(18')
```
( (IP-MAT (CONJ and)
    (NP-SBJ-1 (PRO it))
    (BEP is)
    (NP-OB1 (N meruayle))
    (, ,)
    (PP-1 (P if)
        (CP-ADV (C 0)
                (IP-SUB (NP-SBJ (PRO thou))
                        (VBP scape)
                        (PP (P with)
                            (NP (PRO$ thy) (N lyfe))))))
    (. ,))
```

Similar analyses to the one provided for the *if*-clause in (17a) above are given in the PPCEME for clauses introduced by other minor declarative complementizers and for those featuring their *that*/zero counterparts. Examples (19a)–(19b) illustrate the use of *though*/*that* after a predicate meaning 'wonder' (López-Couso and Méndez-Naya 2001):

(19a) [...] meruayle it shall not be, ***thoughe*** *he be greued with pouertee*. (FITZH-E1-H,99.339)

(19a')
```
  (NP-OB1 (N meruayle))
    (NP-SBJ (PRO it))
    (MD shall)
    (NEG not)
    (BE be)
    (, ,)
    (PP (P thoughe)
        (CP-ADV (C 0)
                (IP-SUB (NP-SBJ-RSP (PRO he))
                        (BEP be)
                        (VAN greued)
                        (PP (P with)
                            (NP (N pouertee)))))
    (. .))
```

(19b) for it is maruell ***that*** *a sinner can without shame beholde this blessed Image?*
(FISHER-E1-H,1,399.198)

(19b')
```
( (IP-MAT (CONJ for)
         (NP-SBJ-1 (PRO it))
         (BEP is)
         (NP-OB1 (N maruell))
         (CP-THT-1 (C that)
                   (IP-SUB (NP-SBJ (D a) (N sinner))
                           (MD can)
                           (PP (P without)
                             (NP (N shame)))
                           (VB beholde)
                           (NP-OB1 (D this) (VAN blessed) (N Image))))
         (. ?))
```

The same holds for *lest*, which is associated with predicates denoting fear (López-Couso 2007), as in (20a) vs. (20b):[11]

(20a) Did Cobham fear ***lest*** *you would betray him in Jersey?* (RALEIGH-E2-P1,1,218.132)

(20a')
```
( (CP-QUE (IP-SUB (DOD Did)
                  (NP-SBJ (CODE <font>) (NPR Cobham) (CODE <$$font>))
                  (VB fear)
                  (PP (P lest)
                      (CP-ADV (C 0)
                              (IP-SUB (NP-SBJ (PRO you))
                                      (MD would)
                                      (VB betray)
                                      (NP-OB1 (PRO him))
                                      (PP (P in)
                                         (NP (CODE <font>) (NPR Jersey) (. ?)
```

(20b) for he feared ***that*** *should he continew at Court,* (PERROTT-E2-H,33.12)

(20b')
```
( (IP-MAT (CONJ for)
          (NP-SBJ (PRO he))
          (VBD feared)
          (CP-THT (C that)
                  (IP-SUB (CP-ADV (IP-SUB (MD should)
                                          (NP-SBJ (PRO he))
                                          (VB continew)
                                          (PP (P at)
                                             (NP (N Court)))))
```

A similar parsing is given for *as if*-clauses dependent on propositional attitude predicates such as *seem* (López-Couso and Méndez-Naya 2012a, 2012b), as shown in (21a)–(21b).

---

[11] As noted by an anonymous reviewer, the variant *lest that*, as in (i) below, occurs in uncontroversial cases of complement clauses in the PCEEC, though not in the PPCEME.
   (i) He ferythe ***lesse that*** *he schall neuer come home* (PASTON, I.656.9503)

Note that *as if* is not parsed as a unit in the PPCEME, but rather as two recursive prepositions.

(21a) so like our first parents before the fall, it seems ***as if*** *they had no wishes*, (BEHN-E3-P1,149.32)

(21a')

```
(NP-SBJ (PRO it))
(VBP seems)
(PP (P as)
   (PP (P if)
      (CP-ADV (C 0)
             (IP-SUB (NP-SBJ (PRO they))
                    (HVD had)
                    (NP-OB1 (Q no) (NS wishes)))
```

(21b) and it seemes **Ø** *Sir Robert Bevell thinks our demaunds very unreasonable*. (MASHAM-E2-P1,103.59)

(21b')

```
( (IP-MAT (CONJ and)
        (NP-SBJ-1 (PRO it))
        (VBP seemes)
        (CP-THT-1 (C 0)
               (IP-SUB (NP-SBJ (NPR Sir) (NPR Robert) (NPR Bevell))
                      (VBP thinks)
                      (IP-SMC (NP-SBJ (PRO$ our) (NS demaunds))
                             (ADJP (ADV very) (ADJ unreasonable)))))
        (. .))
```

The evidence provided for the (a) and the (b) sequences in (17) – (21) above so far indicates that parsing clearly masks crucial syntactic similarities between functionally parallel structures.

Another case in which syntactic structure seems to be masked by the parsing of the PPCEME concerns sequences such as (22), which involve insubordination, i.e. "the conventionalized main clause use of what, on *prima facie* grounds, appear to be formally subordinate clauses" (Evans 2007: 367). Note that in (22) the italicized clause is introduced by *as though*, but there is no clause in the context that can be claimed to be a matrix.[12]

(22) In that I am giltles? ***As though*** *they were gilty*. (MORERIC-E1-P1,37.179)

The parsing of (22) shows that the *as though*-clause is coded in the PPCEME as an adverbial subordinate clause:[13]

---

[12] We regard clauses of this kind as exclamatory clauses (López-Couso and Méndez-Naya 2012b: 324). See also Brinton (2014) on *as if*-exclamatory clauses.

[13] Note that the insubordinated clause in (22') is parsed as Fragment (FRAG): "FRAG should be thought of as a last resort for annotating material consisting of at least two constituents, for which there is not enough material to construct an IP" (https://www.ling.upenn.edu/hist-corpora/annotation).

(22')

```
( (FRAG (PP (P In)
           (CP-ADV (C that)
                   (IP-SUB (NP-SBJ (PRO I))
                           (BEP am)
                           (ADJP (ADJ giltles)))))
        (, ?)
        (PP (P As)
            (PP (P though)
                (CP-ADV (C 0)
                        (IP-SUB (NP-SBJ (PRO they))
                                (BED were)
                                (ADJP (ADJ gilty))))))
        (. .))
```

The foregoing discussion has shown that clauses introduced by the minor declarative complementizers *if*, *though*, *lest*, *as if* and *as though* are invariably parsed as adverbial clauses in the PPCEME. There is, however, one case in which clauses headed by one of our 'minor' links are taken in the PPCHE, though not consistently, to be complements rather than adjuncts. This involves the complementizer *but* (*that*). In López-Couso and Méndez-Naya (1998), we identified two different sub-types of the connective: on the one hand, *but*₁, which is equivalent to 'that not', and is therefore a negative complementizer marking the sub-clause as negative, as shown in (23); on the other, *but*₂, as in (24), which means 'that', and typically occurs after negated predicates which are themselves inherently negative, such as *not doubt* or *not deny*. *But*₁, by contrast, is excluded from such contexts and occurs either with negated predicates (e.g. *not know*) or with inherently negative ones (e.g. *be a shame*).

(23) It is impossible ***but that*** *offences will come* (1582 Rhem; from López-Couso and Méndez-Naya 1998: 162) [i.e. 'it is impossible that offences will not come']

(24) Nor will any Man deny ***but that*** *every thing which is just, is good*; (1695, R. Preston, Cons. of Ph. 180; from López-Couso and Méndez-Naya 1998: 167) [i.e. 'nor will any man deny that everything which is just, is good']

In the annotation manual, *but* is one of those words which are treated individually (see Section 2 above), precisely due to its multifunctional nature. In the PPCHE parsing scheme, it is analyzed as a coordinating conjunction and tagged CONJ (e.g. *Jill laughed but Mary cried*), as a focus particle (FP; e.g. *It cannot be but a great folly*) and as a preposition (P; e.g. *Nobody but you*). The criterion used by the corpus annotators to distinguish between FP and P is whether *but* can be naturally replaced by *except* or *than*, in which case it is tagged as P:

The distinction between the conjunction use of BUT on the one hand and the prepositional and focus particle uses on the other is generally clear, but the distinction between the latter two can be difficult. BUT is tagged as P if it can be replaced naturally by EXCEPT or THAN.

The manual also refers explicitly to cases of complementation with *but*, as follows:

Inherently negative or questioning verbs (DENY, DOUBT, FEAR, HINDER, LET, MISTRUST, PREVENT, QUESTION) as well as other verbs or degree words when negated sometimes take finite clausal complements preceded by BUT. As in the NEG ... BUT construction, BUT is tagged FP and attached low (that is, as part of the complement clause).

An example of the complementizer use of *but*, more specifically *but₂*, meaning 'that', is given as (25a), where the *but*-clause is coded as CP-THT, just in the same way as the zero-clause in (25b).

(25a) However, I doubt not **but** *he is well*; (NHADD-1710-E3-P2,54.17)

(25a')

```
( (IP-MAT (ADVP (WADV+ADV However))
        (, ,)
        (NP-SBJ (PRO I))
        (VBP doubt)
        (NEG not)
        (CP-THT (FP but)
                (C 0)
                (IP-SUB (NP-SBJ (PRO he))
                        (BEP is)
                        (ADJP (ADJ well))))
```

(25b) for I doubte not **Ø** *Sir <font> Thomas Wyat <$$font> hath bin examin'd of me, and hathe sayde what he could directly or indirectly*. (THROCKM-E1-H,I,68.C1.281)

(25b')

```
( (IP-MAT-SPE (CONJ for)
            (NP-SBJ (PRO I))
            (VBP doubte)
            (NEG not)
            (CP-THT-SPE (C 0)
                    (IP-SUB-SPE (NP-SBJ (NPR Sir) (CODE <font>) (NP
                     Thomas) (NPR Wyat) (CODE <$$font>))
                            (HVP hath)
                            (BEN bin)
                            (VAN examin'd)
                            (PP (P of)
                               (NP (PRO me))))
```

Interestingly, not all cases of *but₂* are parsed in this way. Consider (26), a very similar example to (25a) above, but in which *but* is tagged as P taking a CP-ADV, rather than as CP-THT, even though the context makes it clear that *but* cannot be replaced by *except*.

(26) This Bishop is a temporall Lord, notwithstanding his sprituall title; and no
doubt ***but*** *the flesh preuailes aboue the Spirit with him*; (JOTAYLOR-E2-
P1,3,85.C2.281-282)

(26')
```
 ( (CONJP (CONJ and)
        (NP (Q no)
            (N doubt)
            (PP (P but)
              (CP-ADV (C 0)
                    (IP-SUB (NP-SBJ (D the) (N flesh))
                          (VBP preuailes)
                          (PP (P aboue)
                              (NP (D the) (N Spirit)))
                          (PP (P with)
                              (NP (PRO him)))))))))
        (. ;))
```

The same parsing is provided for (27), an example of *but*₁ 'that not', with the inherently

negative predicate *be a great shame*.

(27) 'Iff thys be trew,' seyde Arthure, 'hit were grete shame unto myne astate ***but***
that *he were myghtyly withstonde*.' (CMMALORY-M4,45.1470)

(27')
```
 (NP-SBJ (PRO hit))
            (BED were)
            (NP-OB1 (ADJ grete)
                  (N shame)
                  (PP (P unto)
                      (NP (PRO$ myne) (N astate))))
            (PP (P but)
              (CP-ADV (C that)
                    (IP-SUB (NP-SBJ (PRO he))
                          (BED were)
                          (ADVP (ADV myghtyly))
                          (VAN withstonde))))
```

As seen, then, the parsing of the PPCHE recognizes that *but*-clauses can be complements,

but this annotation is not always consistent throughout the corpora and does not account

for the two uses of the complementizer *but*.

5. CONCLUDING REMARKS

In this article we have discussed some of the issues that may derive from the syntactic

annotation of corpora. In particular, we have examined the problems posed by the parsing

of structures containing minor declarative complementizers in the PPCHE *vis-à-vis*

canonical finite declarative complement clauses. We have shown that low-frequency

phenomena such as the one considered in this article may go unnoticed, masked by

annotation.

In our view, parsed corpora are undoubtedly useful for the analysis of highly-frequent and/or uncontroversial categories, but may overlook interesting features especially when dealing with low-frequency constructions, as shown here in connection with the annotation of minor declarative complementizers. While the parsing of the default finite complementation patterns with *that* and zero seems to be straightforward and therefore can be easily retrieved by means of the search engines, the annotation of complement clauses introduced by the minor complementizers discussed in this article is not completely devoid of problems. On the one hand, the parsing does not capture the obvious functional similarities between *that* or zero complement clauses and clauses headed by minor complementizers. On the other, the parsing of *if*, *though*, *lest*, *as if*, *as though* and *but* complement clauses is not always consistent, as shown in particular in the case of but_1 and but_2 and in the treatment of some anticipatory pronouns. Nevertheless, we believe that such minor weaknesses in the annotation of the Penn family of historical corpora do not at all diminish their value as indispensable tools for the study of the history of the English language.

REFERENCES

ARCHER 3.1 = *A Representative Corpus of Historical English Registers* version 3.1. 1990–1993/2002/2007/2010/2013/2016. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Bamberg, Freiburg, Heidelberg, Helsinki, Lancaster, Leicester, Manchester, Michigan, Northern Arizona, Santiago de Compostela, Southern California, Trier, Uppsala, Zurich.

Bender, Emily and Dan Flickinger. 1999. Diachronic evidence for extended argument structure. In Gosse Bouma, Erhard Hinrichs, Geert-Jan M. Kruijff and Richard Oehrle eds. *Constraints and Resources in Natural Language Syntax and Semantics*. Stanford, CA: CSLI, 1–19.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brinton, Laurel. 2014. The extremes of insubordination: Exclamatory *as if! Journal of English Linguistics* 42/2: 93–113.

Brook, Marisa. 2014. Comparative complementizers in Canadian English: Insights from early fiction. *University of Pennsylvania Working Papers in Linguistics* 20/2, Article 2.

Brook, Marisa. 2018. Taking it up a level: Copy-raising and cascaded tiers of morphosyntactic change. *Language Variation and Change* 30/2: 231–260.

Brown Corpus = *A Standard Corpus of Present-day Edited American English*, for use with Digital Computers (Brown). 1964, 1971, 1979. Compiled by W. N. Francis

and H. Kučera. Brown University. Providence, Rhode Island. http://icame.uib.no/brown/bcm.html

Dancygier, Barbara and Eve Sweetser. 2005. *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press.

Davies, Mark. 2008–. *The Corpus of Contemporary American English* (COCA): One billion words, 1990–2019. https://www.english-corpora.org/coca/.

Dirven, René. 1989. A cognitive perspective on complementation. In Dany Jaspers, Yvan Putseys, Wim Klooster and Pieter Seuren eds. *Sentential Complementation and the Lexicon: Studies in Honour of Wim de Geest*. Dordrecht: Foris, 113–139.

Elsness, Johan. 1982. *That* v. zero connective in English nominal clauses. *ICAME News* 6: 1–45.

Elsness, Johan. 1984. *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65: 519–533.

Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva ed. *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press, 366–431.

Fanego, Teresa. 1990. Finite complement clauses in Shakespeare's English I & II. *Studia Neophilologica* 62: 3–21; 129–149.

Finegan, Edward and Douglas Biber. 1995. *That* and zero complementisers in Late Modern English: Exploring ARCHER from 1650–1990. In Bas Aarts and Charles F. Meyer eds. *The Verb in Contemporary English: Theory and Description*. Cambridge: Cambridge University Press, 241–257.

HC = *Helsinki Corpus of English Texts*. http://www.ota.ox.ac.uk/desc/1477

Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer and Gabriele Diewald eds. *New Reflections on Grammaticalization.* Amsterdam: John Benjamins, 83–101.

Huddleston, Rodney. 1971. *The Sentence in Written English: A Syntactic Study Based on an Analysis of Scientific Texts*. Cambridge: Cambridge University Press.

Huddleston, Rodney and Geoffrey K. Pullum *et al*. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

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.

Kortmann, Bernd. 1997. *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin: Mouton de Gruyter.

Kroch, Anthony and Ann Taylor. 2000. *The Penn-Helsinki Parsed Corpus of Middle English* (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 4. http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4

Kroch, Anthony, Beatrice Santorini and Lauren Delfs. 2004. *The Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, release 3. http://www.ling.upenn.edu/ppche-release-2016/PPCEME-RELEASE-3

Kroch, Anthony, Beatrice Santorini and Ariel Diertani. 2016. *The Penn Parsed Corpus of Modern British English* (PPCMBE2). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 1. http://www.ling.upenn.edu/ppche-release-2016/PPCMBE2-RELEASE-1

Kytö, Merja. 1996. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts* (third edition). Helsinki: Department of English, University of Helsinki.

Lakoff, Robin. 1968. *Abstract Syntax and Latin Complementation*. Cambridge, MA: The MIT Press.

López-Couso, María José. 1996. A look at *that*/zero variation in Restoration English. In Derek Britton, ed. *English Historical Linguistics 1994*. Amsterdam: John Benjamins, 271–286.

López-Couso, María José. 2007. Adverbial connectives within and beyond adverbial subordination: The history of *lest*. In Ursula Lenker and Anneli Meurman-Solin eds. *Connectives in the History of English*. Amsterdam: John Benjamins, 11–29.

López-Couso, María José and Belén Méndez-Naya. 1998. On minor declarative complementizers in the history of English: The case of *but*. In Jacek Fisiak and Marcin Krygier eds. *Advances in English Historical Linguistics*. Berlin: Mouton de Gruyter, 161–171.

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. 2012a. On the use of *as if*, *as though* and *like* in Present-day English complementation structures. *Journal of English Linguistics* 40/2: 172–195.

López-Couso, María José and Belén Méndez-Naya. 2012b. On the origin and development of comparative complementizers in English: Evidence from historical corpora. In Nila Vázquez ed. *Creation and Use of Historical English Corpora in Spain*. Newcastle upon Tyne: Cambridge Scholars Publishing, 311–333.

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*. Bern: Peter Lang, 85–107.

López-Couso, María José and Belén Méndez-Naya. 2015. Secondary grammaticalization in clause combining: From adverbial subordination to complementation in English. *Language Sciences* 47: 188–198.

Marcus, Mitchell, Beatrice Santorini and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The *Penn Treebank*. *Computational Linguistics* 19: 313–330. Reprinted in Susan Armstrong ed. *Using Large Corpora*. Cambridge, MA: The MIT Press, 273–290.

McCawley, James D. 1988. *The Syntactic Phenomena of English*. Chicago: The University of Chicago Press.

Mitchell, Bruce. 1985. *Old English Syntax*. 2 vols. Oxford: Clarendon Press.

Noonan, Michael. 1985. Complementation. In Timothy Shopen ed. *Language Typology and Syntactic Description*, Vol. 2. Cambridge: Cambridge University Press, 42–140.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rissanen, Matti. 1991. On the history of *that*/zero as object clause links in English. In Karin Aijmer and Bengt Altenberg eds. *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, 272–289.

Rissanen, Matti. 2008. Corpus linguistics and historical linguistics. In Anke Lüdeling and Merja Kytö eds. *Corpus Linguistics: An International Handbook*, Vol. 1. Berlin: Mouton de Gruyter, 53–68.

Rooryck, Johan. 2000. *Configurations of Sentential Complementation: Perspectives from Romance Languages*. London: Routledge.

Santorini, Beatrice. 2016. *Annotation manual for the Penn Historical Corpora and the York-Helsinki Corpus of Early English Correspondence*. https://www.ling.upenn.edu/hist-corpora/annotation/index.html

Tagliamonte, Sali and Jennifer Smith. 2005. No momentary fancy! The zero 'complementizer' in English dialects. *English Language and Linguistics* 9/2: 289–309.

Taylor, John R. and Kam-Yiu Pang. 2008. Seeing as though. *English Language and Linguistics* 12/1: 103–139.

Torres Cacoullos, Rena and James A. Walker. 2008. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47: 1–43.

Warner, Anthony. 1982. *Complementation and the Methodology of Historical Syntax*. London: Croom Helm.

Corresponding author
María José López-Couso
Facultade de Filoloxía
Universidade de Santiago de Compostela
Av. de Castelao s/n
15782 Santiago de Compostela
Spain
e-mail: mljopez.couso@usc.es