

# Diachronic changes to the [(if the) truth BE told] construction – a corpus study

Research Article

John Potter

Lancaster University

Received 23 July 2024; accepted 2 December 2024

**Abstract:** [(If the) truth BE told] is an idiomatic construction in English with a number of pragmatic purposes. It can suggest that a proposition is generally known but rarely admitted, or that a proposition is a previously unknown personal admission. It can also act as a pragmatically weaker discourse marker. This study initially looks for early uses in the Early English Books Online (EEBO) corpus, finding examples from the late 16th and early 17th centuries. Diachronic changes in the use and form of the construction in American English are then examined using the Corpus of Historical American (Davies 2010), which contains texts written between 1810 and 2020. In COHA the [(if the) truth BE told] construction becomes considerably more frequent after 1980. Coinciding with this increase in frequency the construction becomes more lexically fixed and reduced in length. It also becomes more likely to appear at the left periphery of a clause. In addition, it appears to be moving towards 'extended intersubjectivity' (Tantucci 2017). As such, it is increasingly losing its pragmatic purposes of marking a rarely admitted truth or a personal admission, and behaving more like a simple discourse marker, connecting clauses.

**Keywords:** discourse markers • subjectivity/intersubjectivity • grammaticalization • lexicalization • frequency effects

© Sciendo

## 1 Introduction

This case study will present a diachronic investigation of the [(if the) truth BE told] construction in English. This idiomatic construction functions as a sentence adjunct and is used to mark new information and make certain pragmatic implications about it. It is idiomatic in the sense that it is non-compositional, its meaning cannot be determined from the words and grammatical structure (Evans 2019) but has to be learned holistically. In terms of the typology of idiomatic expressions set out by Fillmore et al. (1988) it is a decoding idiom, in contrast to an encoding idiom like *wide awake*, the meaning of which could be guessed by a language user unfamiliar with it. It is a substantive, or lexically filled, idiom rather than a formal (or schematic) one which is open to various lexical items (Fillmore et al. 1988). However, it does show slight variation in its lexical content. While *truth* and *told* are always present, any third person form of the verb *to be* can fill the verb slot, in addition to the bare infinitive. The construction can be used with or without *if* and *the*.

Fillmore et al. (1988) make the distinction between idioms that serve a particular pragmatic or rhetorical purpose, and those that do not. In (1) the main proposition, that the Italian government was reluctant to draft Sicilians, is presented as fact. The inclusion of *if the truth be told* implies that the Italian government would be unlikely or reluctant to admit to this.

- (1) And then, *if the truth be told*, the Italian government itself was not so anxious to draft Sicilians. (from 'Sicilian' by Puzo, M. 1984 (COHA))

\* Corresponding author: John Potter, E-mail: j.k.potter1@lancaster.ac.uk

Open Access. © 2025 John Potter, published by Sciendo.

 This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

In (2) the proposition is a personal admission by the speaker/writer (SP/W). Again, the use of *truth be told* implies some reluctance in presenting the information, this time on the part of the SP/W. As the information has been presented, despite the reluctance, this has further implications about the relationship between the SP/W and the addressee/reader (AD/R).

- (2) *Truth be told*, I'm a little superstitious about names. (from 'Love Walked In' by De los Santos, M. 2005 (COHA))

In (3) *if the truth be told* indicates that the information in the proposition is expanding upon previously given information. In this respect, despite its position at the right peripheral of the sentence, it is acting as a discourse marker connecting the sentences. It implies that the later information is less obvious, but it does not have the same sense of reluctance by the SP/W (or anybody else) to impart it.

- (3) Gabriel Stubbs feels neither remorse nor strong satisfaction. He feels very little *if the truth be told*. (from 'Bartholomew Fair Murders' by Turney, L. 1986 (COHA))

From these examples it is clear that the [(if the) truth BE told] construction does have a number of pragmatic purposes.

Fillmore et al. (1988) also make the distinction between grammatical and extragrammatical idioms. An expression like *kick the bucket*, while completely non-compositional, has a V-NP structure predicted by the familiar grammatical rules of non-idiomatic language. In contrast, an expression like *by and large* (a preposition conjoined with an adjective) has an unfamiliar structure not normally encountered, and can be considered extragrammatical. As noted, there is some variety in the form that [(if the) truth BE told] can take. Its fullest version (*if the truth were told*) has the form of a subjunctive *if*-clause. One would expect this to be subordinate to a main clause stating the consequences of the hypothetical act of truth-telling (*if the truth were told then X would happen*). When used idiomatically the fullest version of the [(if the) truth BE told] construction is an example of an insubordinate *if*-clause. These retain the structure of a non-finite subordinate clause, but they either have no syntactic link to an associated main clause or are entirely stand-alone (see Kaltenböck & Keizernot 2022 for an overview). In this respect it cannot be considered fully grammatical. The construction's most reduced version (*truth be told*) has the plain verb form of the present subjunctive, and is again a non-finite clause which is not subordinate to a main clause. These forms of the construction are extragrammatical to a speaker of Present Day English.

Insubordination refers to the phenomenon in which a subordinate construction is used as an autonomous main clause. Evans (2007) suggests that this phenomenon is more widespread than might be supposed. While they can be interpreted as subordinate clauses with a main clause that has been elided (but can be restored from context) (e.g. Haegeman 2003), this is often not unambiguously the case. For example, while (4) *may* suggest an elided main clause (e.g. *if you'll just come next door then we can start*) its main illocutionary purpose is interpersonal – a hedge intended to present an imperative more politely – a function often fulfilled by *if*-clauses (Vlasyan 2018).

- (4) "If you'll just come next door." (Kaltenböck & Keizernot 2022: 675)

Similarly, in (5) the most likely restored main clause (*if you say so then it must be true*) is in opposition to the illocutionary intent of the utterance, which is to express doubt on the part of the SP/W.

- (5) "She'll be here by ten." – "If you say so." (Kaltenböck & Keizernot 2022: 675)

Insubordinate clauses can perform a number of pragmatic purposes. They can be used for interpersonal purposes, such as mitigating 'face-threatening acts' (Brown & Levinson 1987) (as in 4); they can be used to express a modal meaning, such as a SP/Ws evaluation of a proposition (as in 5); or they can serve a discourse function, such as expressing interclausal relations (Evans 2007).

Evans (2007) outlines the diachronic process which results in a subordinate clause being reanalysed as an autonomous main clause with its own pragmatic purpose. Firstly, the subordinate clause is used with an elided main clause, and any main clause can be reconstructed by the listener if suggested by the context. Secondly, a restricted set of reconstructed main clauses become conventionally attached to a given subordinate clause. Thirdly, the subordinate clause becomes insubordinate and autonomous with its own conventionalised meaning, which is independent of any reconstructed main clause. Indeed, at this point it may not be possible to reconstruct a 'main' clause at all. It is reasonable to assume that insubordinate *if*-clauses have followed this path of development from conditional constructions with subordinate *if*-clauses (Mato-Míguez 2016). This kind of change can be interpreted as resulting from structural reanalysis, as defined by Langacker (1977). This requires a surface manifestation that is structurally ambiguous and can be interpreted in different ways e.g. as a subordinate clause with an elided main clause or as a standalone insubordinate clause. The structural reanalysis is covert until it results in surface manifestations that reflect the new interpretation (Hopper & Traugott 2003), such as the use of an insubordinate clause in a context where it is not possible to unambiguously restore a main clause. More recently, however, it has been suggested that such changes are driven by semantic reanalysis rather than structural reanalysis (Detges et al. 2021). The *if*-clause loses the semantic purpose of defining the condition under which a main clause is true. Instead, it develops its own conventionalised pragmatic purpose and, as this no longer relates to a main clause, it is able to function autonomously. Similarly, the diachronic change of *in fact* from VP adverbial to sentence adverb and then to syntactically independent discourse marker (discussed below) is not possible until a semantic change has permitted the change in syntactic position (Traugott 1999).

The semantic development of *in fact* is examined by Schwenter & Traugott (2000) in terms of the Invited Inferencing Theory of Semantic Change (see also Traugott & Dasher 2002). The use of a construction in a particular context invites inferences from that context, an 'utterance-token meaning'. If the construction is encountered in that context frequently enough to become salient in a community the inferences become part of its 'utterance-type meaning', conventionally attached to that construction. The inferences can then become part of the semantics, or 'coded meaning', of the construction. This does not involve the replacement of one meaning by another, but rather the expansion of a construction's semantics (Traugott & Dasher 2002). Older meanings may fall out of use or remain part of the polysemy of the construction. Using historical examples, Schwenter & Traugott (2000) trace the semantic development of *in fact* from the 17th century. Its earliest meaning (*in fact*<sub>1</sub>) suggests that an action has occurred and can be observed or attested. It is a VP adverbial, and *fact* (as the NP in a prepositional phrase) can be co-ordinated, e.g. *in fact and theory*. In the early 18th century, it appears as an epistemic sentence adverb (*in fact*<sub>2</sub>) suggesting that a proposition is true. This shift in meaning is the result of an 'invited inference' – that what is observable/attested must be true. However, explicitly stating that a proposition is true implies that there is reason to doubt it. As a result, later in the 18th century *in fact*<sub>2</sub> tends to be restricted to adversative contexts, suggesting that a proposition is true even though it contradicts a previous utterance. The use of *in fact*<sub>2</sub> to link utterances allows the development of *in fact*<sub>3</sub> – a syntactically independent discourse marker linking clauses. *In fact*<sub>3</sub> has lost its adversative element, instead suggesting that the current proposition is 'rhetorically stronger' than the previous one (but does not contradict it). In Present Day English *in fact* is polysemous, with all three meanings in current use (Schwenter & Traugott 2000).

In an earlier paper, Traugott (1999) discusses the evolution of *in fact* in terms of 'subjectification'. This is a process by which a construction with an 'objective' meaning, referring to reality, develops polysemies that instead reflect the SP/W's attitude or perspective. These become part of the coded meaning of that construction. 'Subjectification' tends to be unidirectional (Traugott 2003), as it is unlikely that a construction will originally have a subjective meaning and later develop a non-subjective meaning. Constructions can then develop further polysemy and become 'intersubjective', carrying implications about the SP/W's awareness of the AD/R. Again, these become part of the coded meaning of the construction (Traugott 2003). This can be epistemic, showing awareness of what the AD/R may know or feel about something, or social, paying attention to the AD/R's 'face'. This process is 'intersubjectification' and also tends to be unidirectional. Tantucci (2017) extends the semasiological process further by making the distinction

between ‘immediate’ intersubjectivity and ‘extended’ intersubjectivity. ‘Immediate’ intersubjectivity involves the SP/W’s awareness of a single interlocutor, while ‘extended’ intersubjectivity involves the SP/W’s awareness of further, third-party, AD/R’s, either specific or generic. This move from immediate to extended intersubjectivity is also a unidirectional process, and adds to the polysemy of the construction.

The *[(if the) truth BE told]* construction is first attested around the start of the 17th century (see Section 3). The earliest attested examples already demonstrate polysemy, making it difficult to identify which pragmatic purpose the construction served first. Considering the process of insubordination (Evans 2007), the pragmatic development of *in fact* (Schwenter & Traugott 2000; Traugott 1999), and the process of subjectification and intersubjectification (Traugott 2003; Traugott 1999), a reasonable hypothesis for the pragmatic development of *[(if the) truth BE told]* could be:

- i. a subordinate *if*-clause with a main clause stating the consequences of telling the truth (objective) >
- ii. marking a truth that (in the SP/W’s view) is known but people are reluctant to admit (subjective) >
- iii. marking an unknown truth that the SP/W is reluctant to tell (but will tell the AD/R) (intersubjective) >
- iv. a discourse marker connecting the current utterance to the previous one.

The use of *in fact*3 (Schwenter & Traugott 2000) is very similar to the discourse marker use of *[(if the) truth BE told]* (as in example (3)) – both link to a previous proposition, supporting and expanding upon it.

‘Discourse markers’ (DMs) are linguistic items with a pragmatic purpose, the term being popularised by Schiffrin (1987). They have subsequently been the subject of many studies and much discussion (Bernd et al. 2021). Some authors employ different terms to refer to DMs, such as ‘pragmatic markers’ (Brinton 2017) or ‘pragmatized units’ (Bolly 2014). A linguistic item can be considered a discourse marker if it conforms to the following criteria (based on Bernd et al. 2021):

- i. DMs are invariable expressions. There is no schematicity or variation in the form they take.
- ii. DMs are semantically independent. They do not contribute semantically to the main proposition, and can be omitted without changing the basic meaning of an utterance.
- iii. DMs are syntactically independent. They are outside the structure of the sentence.
- iv. DMs are prosodically independent. They are separated from the prosody of the main utterance with distinct intonation and stress (Traugott 1995).
- v. DMs have a metatextual function. They are non-truth-conditional (Brinton & Traugott 2005) and comment on the text itself rather than the referent of the text. They serve text organisation and reflect SP/W attitude and the relationship between SP/W and AD/R.

Additionally, by some definitions (e.g. Schiffrin 1987; Brinton & Traugott 2005; 2018) a construction can only be considered a DM if it signals a relationship between utterances. This is the definition used in this study. It can be seen that the examples of the *[(if the) truth BE told]* construction discussed so far conform to criteria ii, iii and v. However, only example (3) performs the function of linking clauses. As the examples are from written texts it is not possible to comment on criterion iv. The examples do not appear to conform to criterion i due to the variations in form.

As with insubordinate clauses, discourse markers have their origins in more prototypical linguistic constructions such as main clauses, subordinate clauses and prepositional phrases (Brinton & Traugott 2005). A number of theories have been proposed to account for the diachronic process by which these items become DMs. One explanation of the process is that it is an example of grammaticalization (e.g. Traugott 1995; Brinton & Traugott 2005; Brinton 2017). When a linguistic item undergoes grammaticalization it moves towards a grammatical rather than lexical function, loses autonomy and becomes more subject to the constraints of the language system (Lehmann 2004). The difficulty with considering DMs to be the result of grammaticalization is that they do not develop in the same way as typical grammatical items, and must be considered less ‘canonical’ examples (Traugott 1995). DMs differ from more canonical examples of grammaticalization in that they do not develop into clitics or inflections, they do not phonetically reduce or fuse with host words, they do not become a fixed and obligatory part of a syntactic structure and they do not reduce in scope (Brinton 2017). Their syntactic independence means they have greater freedom and can appear within a clause or at the left or right periphery. Bernd et al. (2021) argue that their independence means they no longer have syntactic scope, while their

metatextual function means their semantic scope has increased to the level of whole utterances and beyond. Indeed, their independence from sentence structure has led some (e.g. Waltereit 2002) to question whether DMs can be considered grammatical items at all.

'Pragmaticalization' has been suggested as an appropriate term to describe the process by which a lexical or grammatical item becomes a pragmatic item, independent from the syntactic structure and semantics of an utterance and fulfilling a metatextual or interpersonal role (Dostie 2009). This clearly references the particular properties of DMs that are problematic from a grammaticalization standpoint. However, proponents of grammaticalization (e.g. Traugott 2007; Brinton 2017) do not consider pragmaticalization to be a distinct process, and argue that many uncontroversial grammatical items carry pragmatic meanings, such as the polite use of the past tense in *what was your name?* (Brinton & Traugott 2005). The development of DMs has also been described as an example of lexicalization. Lexicalization, in this sense, is a process in which a common collocation or multiword unit enters the lexicon as a single unit and develops a conventionalised meaning that differs from, but may be related to, the original compositional meaning (Wischer 2000). While DMs do have pragmatic meanings that cannot be determined from their component parts, it is not obvious that they are typically lexical any more than they are typically grammatical. In fact, they seem to be "elusive to the distinction between lexical and grammatical expressions or categories" (Bernd et al. 2021: 12). Lexicalization and grammaticalization are complementary processes (Brinton & Traugott 2005) and share many similarities (Lehmann 2004). These similarities include:

- i. Gradualness – changes occur in small steps, with older and more innovative uses existing side-by-side in the polysemy of a construction.
- ii. Fusion – both lexicalization and grammaticalization involve the fixing and freezing of collocations, reducing and eliminating variation.
- iii. Coalescence – subsequent to fusion, constructions become reduced. This can lead to reduced forms and eventually to zero forms during grammaticalization, and the loss of phonological segments during lexicalization.
- iv. Demotivation – the loss of semantic compositionality of a construction. As the conventionalised meaning drifts further from the original compositional meaning speakers no longer analyse a construction as consisting of component parts.
- v. Unidirectionality – fusion, coalescence and demotivation obscure the origins of a construction leading to unidirectionality.

(Brinton & Traugott 2005)

Bolly (2014) suggests the over-arching term 'functionalization' to describe the process by which high frequency correlates with increasing (inter)subjectivity and lexical fixedness leading to the recategorization of linguistic units as either grammatical, pragmatic or lexical. Frequency is a driver for both lexicalization, in the sense of a multi-word construction becoming entrenched as a single unit, and grammaticalization (Bybee 2006; 2010). In turn, these processes affect the frequency of the recategorized linguistic units. There are a number of recognised effects of high token frequency (Bybee 2006). The first effect, 'reduction', is related to what Brinton & Traugott (2005) call 'coalescence'. As 'canonical' grammatical items are very high frequency (Brinton & Traugott 2005; Bybee 2010) they undergo extreme phonetic reduction. Less frequent items undergo less reduction. The second effect, 'conservation' (Bybee 2006), is related to 'fusion' (Brinton & Traugott 2005), high frequency items conserve their structure and become resistant to change. The third effect, 'autonomy' (Bybee 2006), is related to 'demotivation' (Brinton & Traugott 2005), high frequency items become semantically and syntactically opaque as they are processed as a single unit, and less clearly related to the constructions that they are etymologically derived from.

This study will first look at some early examples of the idiomatic use of the *[(if the) truth BE told]* construction. Then diachronic changes in the use of the construction in American English will be examined using the Corpus of Historical American (Davies 2010), which contains texts written between 1810 and 2020. Particularly this study will look for any changes in frequency, and how they correspond to changes in lexical fixedness, construction size, clause position and subjectivity/intersubjectivity.



## 2 Factors to be analysed

### 2.1 Lexical fixedness

'Lexical fixedness' refers to several processes related to 'functionalization' and token frequency. As a multi-word construction becomes entrenched it will resist any changes to its structure (Bybee 2006; Brinton & Traugott 2005). As the conventionalised meanings of a construction expand beyond its original compositional meaning, its structure becomes opaque to the speaker (Brinton & Traugott 2005). As a result, the construction is not analysed as being composed of parts, but instead processed as a single chunk (Bybee 2010). The construction loses any schematicity. This invariability, or lack of schematicity, is one of the criteria for defining a discourse marker (Bernd et al. 2021).

The *[(if the) truth BE told]* construction shows some schematicity, in that any third person form of the verb *to be* can fill the verb slot. This will be analysed in the COHA results to determine whether the schematicity is decreasing over time, and how this relates to frequency.

### 2.2 Construction size

Another process that is symptomatic of 'functionalization' and increasing token frequency is 'reduction' (Bybee 2006) or 'coalescence' (Brinton & Traugott 2005). For very high frequency items this can take the form of phonetic reduction (Bybee 2006; Brinton & Traugott 2005). While very frequent discourse markers may become phonetically reduced, this is not a criterion of discourse markers *per se* (Traugott 1995; Brinton 2017). They can undergo syntactic reduction (Thompson 2007) and lose 'phonological segments' (Brinton & Traugott 2005), such as the elision of elements from a construction. As this study is examining written examples from COHA it is not possible to examine phonetic reduction. However, the *[(if the) truth BE told]* construction can be used without *if* or *the*, therefore it is possible to analyse the COHA results to see whether there is an increasing tendency to use the shortest version, and how this relates to frequency.

### 2.3 Clause position

The position of a construction in relation to a main clause, whether it is internal or at the periphery, has been thought to determine the function of that construction. The left periphery has been associated with discourse markers (Traugott 1995) and expressing the SP/W's (subjective) attitude (Traugott 2012), while the right periphery has been associated with expressing the SP/W's (intersubjective) awareness of the AD/R (Beeching & Detges 2014). However, Traugott (2012; 2022) suggests that the links between these functions and position are not general, and instead the functions a particular construction performs in different clause positions are specific to that construction. Notwithstanding, the COHA results will be analysed to see if there is a tendency for the *[(if the) truth BE told]* construction to appear in a particular position.

### 2.4 Subjectivity/intersubjectivity

The COHA results will be examined to determine whether examples of the *[(if the) truth BE told]* construction demonstrate subjectivity, intersubjectivity (Traugott & Dasher 2002) or extended intersubjectivity (Tantucci 2017), whether there is a tendency for usage to move towards extended intersubjectivity, and how this relates to frequency.

## 3 Early examples

The Early English Books Online (EEBO) corpus consists of Early Modern English texts from the 1470s to the 1690s. A search of the EEBO (V3) corpus on CQPweb (Hardie 2012) found 33 examples of *truth BE told*, including some repetitions. Of these, the majority are subordinate clauses linked to a main

clause. Only three can be considered early examples of idiomatic use. The earliest reliably attested example is (6).

- (6) Truth gets hatred, and even brethren to love together (*if truth be told*) is a hard thing. (from 'Comfortable Notes Upon Genesis', Babbington, G. 1592 (EEBO A00730))

However, this sentence is ambiguous. The literal interpretation is entirely compositional. It is a conditional sentence with a subordinate *if*-clause (*if truth be told*) and a main clause stating the consequence (*even brethren to love together is a hard thing*) – it is hard for brethren to love each other if they tell the truth. The *if*-clause is part of the proposition and cannot be omitted without changing the meaning (despite the brackets which are in the original text). To arrive at the more idiomatic interpretation – to tell the truth, it is (generally) hard for brethren to love each other – requires that the link between the action and the consequence be broken. *If truth be told* is interpreted as an autonomous insubordinate clause and as such is “available for pragmatic interpretation” (Evans 2007: 374). It is not part of the main proposition, but instead is an evaluation of it – marking new information and implying that it is rarely stated. The second example (7) appears a few years later.

- (7) But yet I fear, *if that the truth were told*, That thou haste robbed, and fliest unto this hold. (from 'The Barrons' Wars in the Raigne of Edward the Second', Drayton, M. 1603 (EEBO A20811))

Here, *if that the truth were told* cannot be interpreted as part of a conditional proposition. What follows is not a consequence, but new information and the implication that it is being withheld. Furthermore, the construction is ‘face-threatening’ (in the sense used by Brown & Levinson (1987)) because it implies that the AD/R is the one withholding the truth. *If that the truth were told* can be removed without changing the meaning of the utterance. In the third example (8) we find a usage remarkably consistent with the way that the construction is used as a discourse marker in Present Day English. This example comes from a translation, by Thomas Shelton, of ‘Don Quixote’, published in 1652.

- (8) [...] Rozinante began to neigh, and Dapple to sigh, [...] though *if the truth were told*, Dapple's sighs and brayings were more then the Horse's neighing. (from 'The History of the Valorous and Witty Knight Errant, Don Quixote', Cervantes, M. / Shelton, F. 1652 (EEBO A31538))

*If the truth were told* is used to mark new information again, but here, as in (3), it expands upon previously given information. It does not have a sense of reluctance to impart the information, nor is it face-threatening. As this is part of a narrative rather than an ongoing speech act, the AD/R is not known and singular but generic and plural, the reader. In these respects, it is displaying extended intersubjectivity (Tantucci 2017).

## 4 Method and data frequency in COHA

A search was carried out in COHA for *truth BE told*, with any forms of the verb lemma *be*. The tokens retrieved were checked to determine whether they were examples of the idiomatic use of the construction, i.e. if the construction was not a subordinate *if*-clause linked to a main clause, and could be removed from utterance without changing the meaning, it was considered idiomatic. Table 1 and Figure 1 show the normalised frequencies (per million words) of the idiomatic and non-idiomatic tokens.

**Table 1.** Normalised frequencies (per million words) of idiomatic and non-idiomatic uses of [(if the) truth BE told] in COHA by decade.

| Decade                                                  | 1820s | 1830s | 1840s | 1850s | 1860s | 1870s | 1880s | 1890s | 1900s | 1910s |
|---------------------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Frequency of idiomatic examples (per million words)     | 0.000 | 0.292 | 0.127 | 0.181 | 0.100 | 0.076 | 0.249 | 0.196 | 0.182 | 0.260 |
| Frequency of non-idiomatic examples (per million words) | 0.286 | 0.146 | 0.127 | 0.000 | 0.000 | 0.038 | 0.149 | 0.098 | 0.182 | 0.043 |
| Decade                                                  | 1920s | 1930s | 1940s | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s |
| Frequency of idiomatic examples (per million words)     | 0.311 | 0.072 | 0.000 | 0.105 | 0.103 | 0.277 | 0.703 | 1.056 | 2.441 | 2.200 |
| Frequency of non-idiomatic examples (per million words) | 0.039 | 0.000 | 0.146 | 0.000 | 0.000 | 0.035 | 0.067 | 0.030 | 0.029 | 0.113 |

The results from COHA suggest that the idiomatic use of [(if the) truth BE told] in American English has been increasing considerably since 1980 after being at a consistently low level since the 1820s. The remainder of this study will compare the pre-1980 and post-1980 tokens of the idiomatic construction to determine whether the form or usage has changed between the two periods as the frequency has increased.

## 5 Lexical fixedness

Earlier examples of [(if the) truth BE told] in COHA appear to be partially schematic in that they have an open slot that can take any form of the verb *to be*, as in (9) and (10).

- (9) [...] who, *truth being told*, it must be admitted, did indeed know a great deal about the way it is with men. (from 'Ramona' by Hunt, H. 1884 (COHA))
- (10) I do dislike untidy gardens; flowers, and vines, and trees, all crowded together, and weeds too, *if the truth's told*. (from 'John Ward, Preacher' by Deland, M. 1888 (COHA))

After 1980, as the frequency increases, the forms of *be* become much more restricted.

Additionally, five of the earliest examples from the COHA results occurred with adjectives internal to the construction, as in (11).

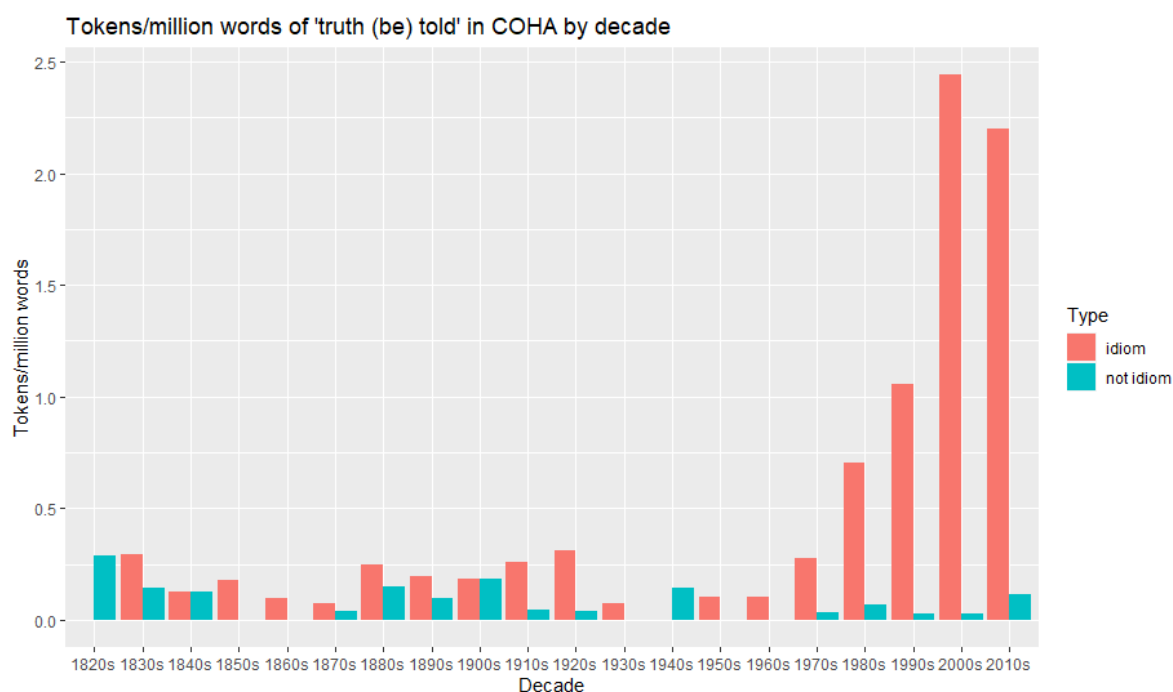
- (11) [...] and, *if the whole truth be told*, each of us fell asleep at once. (from 'If, Yes and Perhaps' by Hale, E.E. 1868 (COHA))

These adjectives come from a limited set (either *whole*, *plain* or *simple*) and the most recent example dates from 1907. While there are few examples of this in the earlier results, the fact that there are none in later examples may be further evidence of increasing lexical fixedness.

Table 2 shows all the forms of the verb *to be* from the COHA results and the number of tokens and percentages for each pre- and post-1980. Figure 2 shows the number of tokens for each pre-1980, while Figure 3 shows the number of tokens for each post-1980.

It can be seen that there is a greater variety of forms of *be* in the pre-1980 results, with examples spread between them. *Were* is the most frequent form, while *be* is the next most frequent. After 1980 the construction predominately contains *be*, with very few exceptions. The five examples of *truth is told*





**Figure 1.** Normalised frequencies (per million words) of idiomatic and non-idiomatic uses of [(if the) truth BE told] in COHA by decade.

**Table 2.** Forms of the verb 'to be' with number of tokens and percentages for each pre- and post-1980.

| Form of <i>be</i>         | 's    | be     | being | is    | was   | were   |
|---------------------------|-------|--------|-------|-------|-------|--------|
| <b>Tokens (pre-1980)</b>  | 2     | 19     | 1     | 0     | 4     | 34     |
| <b>% (pre-1980)</b>       | 3.33% | 31.67% | 1.67% | 0.00% | 6.67% | 56.67% |
| <b>Tokens (post-1980)</b> | 0     | 210    | 0     | 5     | 0     | 4      |
| <b>% (post-1980)</b>      | 0.00% | 95.89% | 0.00% | 2.28% | 0.00% | 1.83%  |

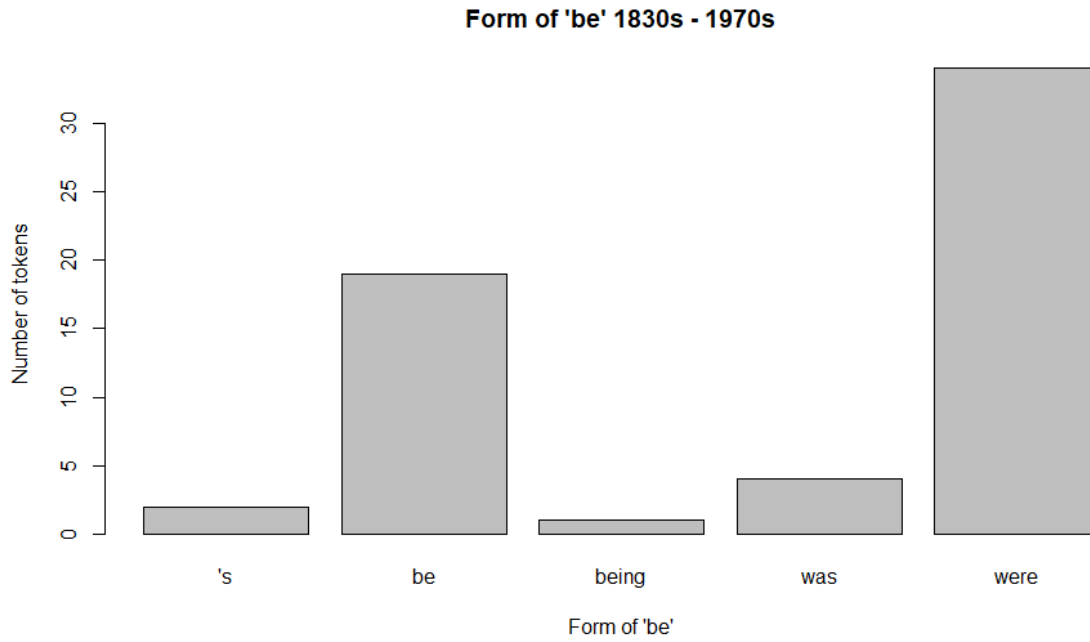
in the post-1980 data come from a single text and can probably be put down to an idiosyncrasy of a particular author. *Truth is told* does not appear in the pre-1980 results. A 2-sample test for equality of proportions with continuity correction shows that the reduction in the use of *were* between the pre- and post-1980 results is highly significant ( $X^2 = 115.78$ ,  $p < 2.2e-16$ ), as is the increase in the use of *be* ( $X^2 = 127.73$ ,  $p < 2.2e-16$ ).

## 6 Reduction

In the COHA results the [(if the) *truth* BE *told*] construction appears in its fullest form (12), with *if* elided (13), with *the* elided (14), or with both elided (15).

(12) *If the truth be told*, they wrote to each other at least twice a week. (from 'The Talismans of Shannara' by Brooks, T. 1993 (COHA))

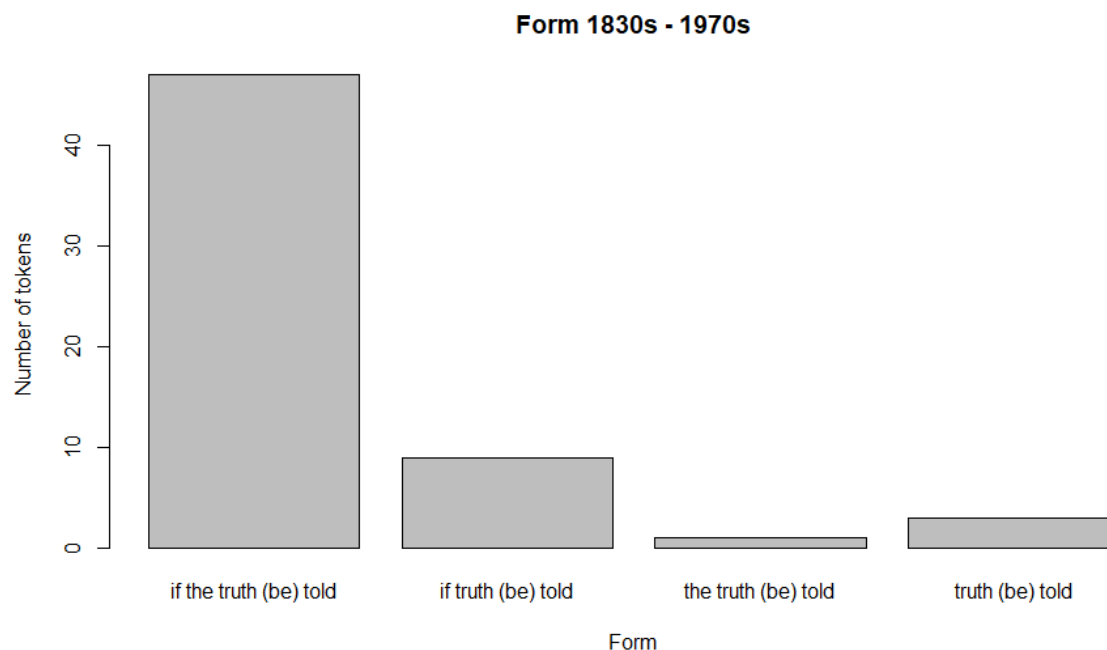
(13) And *the truth be told*, Watkins, I'd like that better myself. (from 'True Love' by Gold, H. 1982 (COHA))



**Figure 2.** Forms of the verb *to be* with number of tokens for each pre-1980.



**Figure 3.** Forms of the verb *to be* with number of tokens for each post-1980.



**Figure 4.** Forms with number of tokens and percentages for each pre-1980.

**Table 3.** Forms with number of tokens and percentages for each pre- and post-1980.

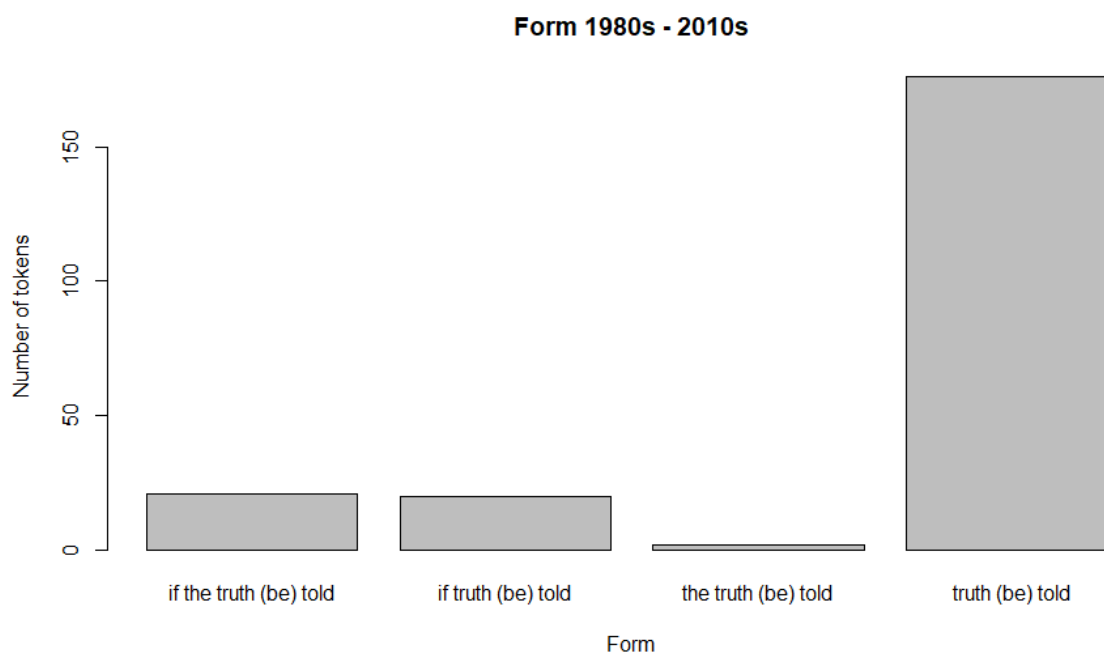
| Form                      | <i>if the truth BE told</i> | <i>if truth BE told</i> | <i>the truth BE told</i> | <i>truth BE told</i> |
|---------------------------|-----------------------------|-------------------------|--------------------------|----------------------|
| <b>Tokens (pre-1980)</b>  | 47                          | 9                       | 1                        | 3                    |
| <b>% (pre-1980)</b>       | 78.33%                      | 15.00%                  | 1.67%                    | 5.00%                |
| <b>Tokens (post-1980)</b> | 21                          | 20                      | 2                        | 176                  |
| <b>% (pre-1980)</b>       | 9.59%                       | 9.13%                   | 0.91%                    | 80.37%               |

(14) [...] a trick, too, that *if truth were told*, Lanier should have known about. (from 'Lanier of the Cavalry' by King, C. 1909 (COHA))

(15) *Truth be told*, I'd had just about enough of her wisdom. (from 'The Butcher's Wife', 1991 (COHA))

Table 3 shows the forms of the construction in the COHA results and the number of tokens and percentages for each pre- and post-1980 (including all the forms of *be*). Figure 4 shows the number of tokens for each form pre-1980, while Figure 5 shows the number of tokens for each form post-1980.

It can be seen that, prior to 1980, by far the most frequent form of the construction was the fullest version (*if the truth BE told*), while the most reduced version (*truth BE told*) is infrequent. After 1980 this changes entirely, with the majority of tokens being of the reduced version. A 2-sample test for equality of proportions with continuity correction shows that the reduction in the use of *if the truth BE told* between the pre- and post-1980 results is highly significant ( $X^2 = 117.05$ ,  $p < 2.2\text{e-}16$ ), as is the increase in the use of *truth BE told* ( $X^2 = 113.08$ ,  $p < 2.2\text{e-}16$ ).



**Figure 5.** Forms with number of tokens and percentages for each post-1980.

**Table 4.** Position of the construction relative to the main clause with number of tokens and percentages for each pre- and post-1980.

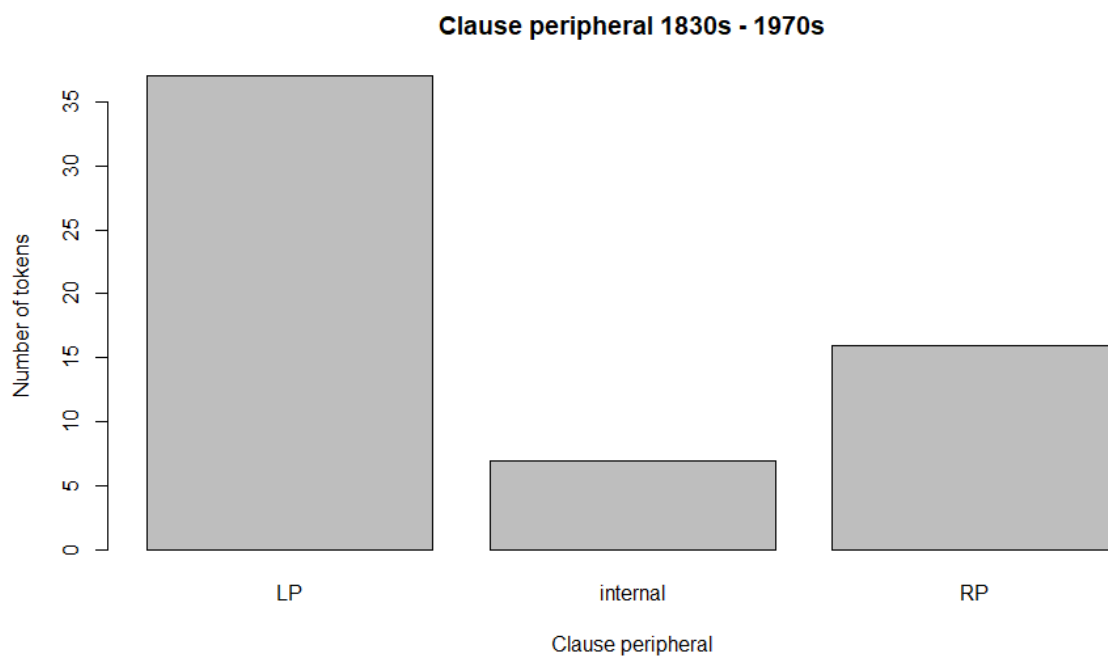
|                           | LP     | internal | RP     |
|---------------------------|--------|----------|--------|
| <b>Tokens (pre-1980)</b>  | 37     | 7        | 16     |
| <b>% (pre-1980)</b>       | 61.67% | 11.67%   | 26.67% |
| <b>Tokens (post-1980)</b> | 181    | 12       | 26     |
| <b>% (pre-1980)</b>       | 82.65% | 5.48%    | 11.87% |

## 7 Clause position

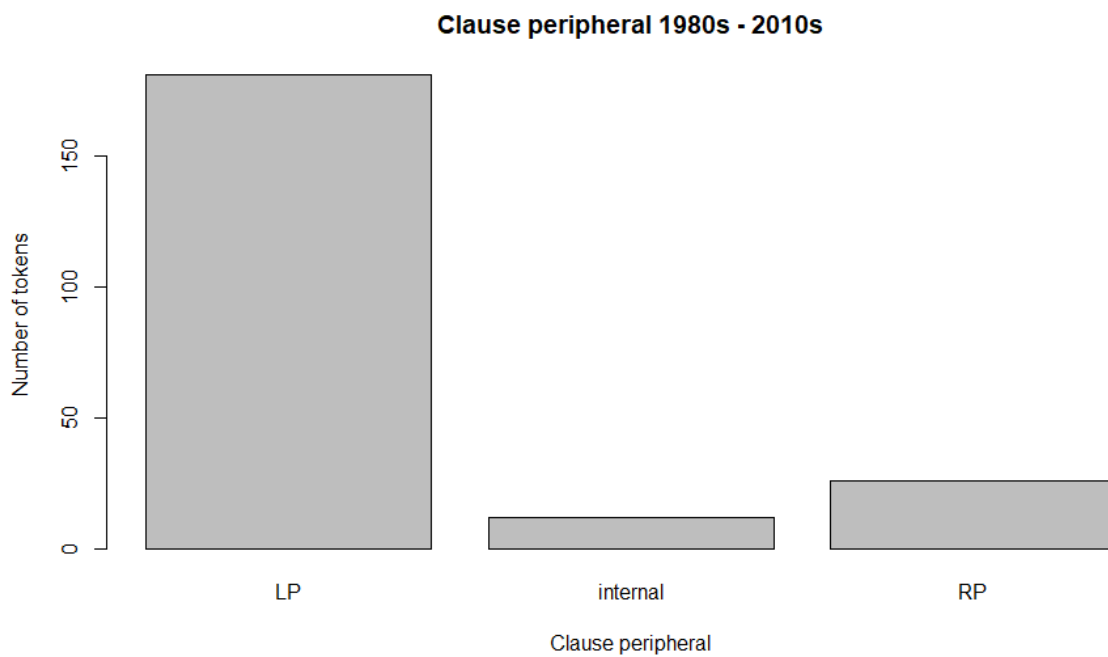
Table 4 shows the position of the *[(if the) truth BE told]* construction relative to the main clause in the COHA results and the number of tokens and percentages for each position pre- and post-1980. Figure 6 shows the number of tokens for each position pre-1980, while Figure 7 shows the number of tokens for each position post-1980. ‘LP’ indicates that the construction is external to the main clause on the left periphery, ‘RP’ indicates

that it is external to the clause on the right periphery, while ‘internal’ indicates that it is internal to the main clause.

It can be seen that there is an increased proportion of tokens external to the clause on the left periphery in the post-1980 results, while the proportion of tokens on the right periphery decreases. There are very few examples of the construction internal to the clause in either period, but the relative frequency for these also decreases post-1980. A 2-sample test for equality of proportions with continuity correction shows that the increase in tokens at the left periphery of the clause between the pre- and post-1980 results is significant ( $X^2 = 10.939$ ,  $p = 0.0009414$ ), as is the relative decrease in tokens at the right periphery ( $X^2 = 6.9458$ ,  $p = 0.008401$ ), the relative decrease in clause internal tokens is not significant ( $X^2 = 1.9497$ ,  $p = 0.1626$ ).



**Figure 6.** Positions of the construction relative to the main clause with number of tokens for each pre-1980.



**Figure 7.** Positions of the construction relative to the main clause with number of tokens for each post-1980.



**Table 5.** Number of tokens and percentages for subjective (S), immediate intersubjective (II) and extended intersubjective (EI) uses of [(if the) truth BE told] pre- and post-1980.

|                           | S      | II     | EI     |
|---------------------------|--------|--------|--------|
| <b>Tokens (pre-1980)</b>  | 25     | 11     | 24     |
| <b>% (pre-1980)</b>       | 41.67% | 18.33% | 40.00% |
| <b>Tokens (post-1980)</b> | 49     | 52     | 118    |
| <b>% (pre-1980)</b>       | 22.37% | 23.74% | 53.88% |

## 8 Inter-subjectification

The literal interpretation of (4) is a non-subjective example of [(if the) truth BE told], as is (16) which simply states the consequences of a hypothetical revelation. Sentences containing non-subjective uses of [(if the) truth BE told] were not included in this analysis as they had already been removed as being non-idiomatic (see above).

(16) I would be humiliated *if the truth were told* [...] (from 'A Decent Birth, A Happy Funeral' by Saroyan, W. 1941 (COHA))

In contrast, (17) does not refer to a potential revelation and its consequences. Instead, it is subjective and introduces information that, in the SP/W's view, is a self-evident truth that is unlikely to be admitted.

(17) I dare say there is faults enough on both sides, *if the truth was told* [...] (from 'Horse Shoe Robinson: A Tale of the Tory Ascendency' by Kennedy, J.P. 1835 (COHA))

In (18) [(if the) truth BE told] is intersubjective. It is used to introduce an admission that the SP/W predicts the AD/R to be unaware of. Also, by introducing information that could leave the SP/W vulnerable it indicates trust, and therefore could be considered supportive of the AD/R's 'face'.

(18) And *truth be told*, I don't like that arrogant kid. (from 'Good Luck, Miss Wyckoff' by Platt, P. 1979 (COHA))

(19), being part of a narrative rather than an ongoing speech act, is not addressed to a specific AD/R. [(if the) truth BE told] is used to introduce information that the SP/W judges any AD/R would be unaware of. Furthermore, not being addressed to a specific interlocutor, it can no longer be considered 'face'-supporting.

(19) *Truth be told*, I was psyched about trying the case for reasons of my own. (from 'In Firm Pursuit' by Samuels-Young, P. 2007 (COHA))

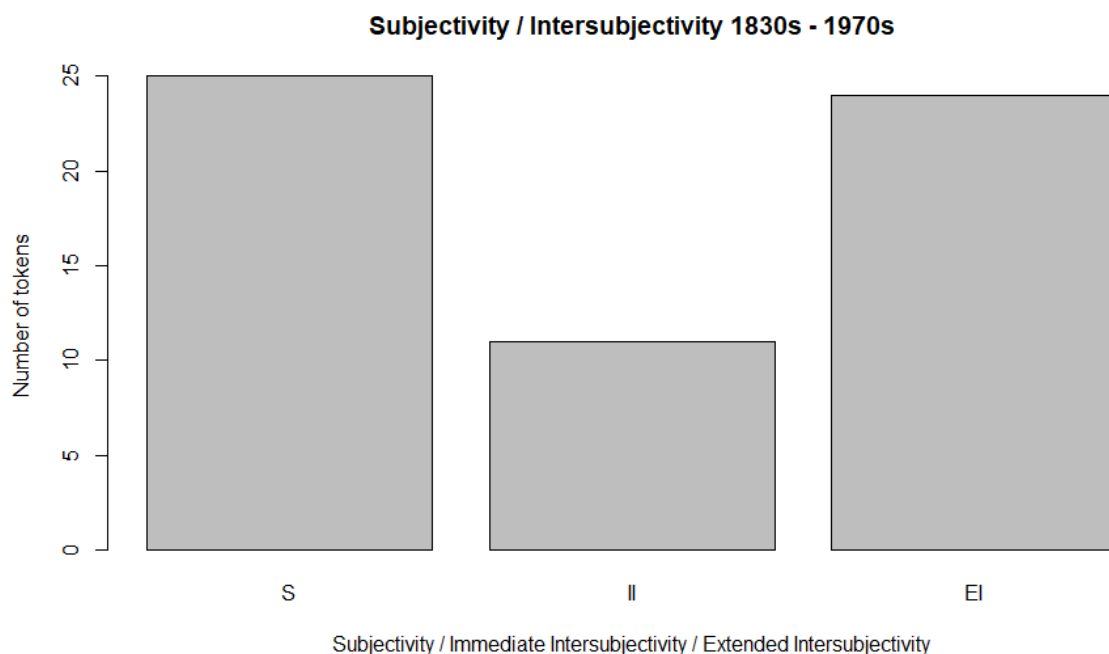
It is worth noting that in (3), from the 1652 translation of Don Quixote, [(if the) truth BE told] is part of a narrative, rather than discourse, and used to introduce new information to a generic AD/R. This suggests that the construction had been through this process of semasiological change and acquired an extended intersubjective meaning by the middle of the 17th century, while still retaining immediate intersubjective and subjective semantics.

In order to operationalise the stages of the process from subjectification to extended intersubjectification results from the COHA data were reviewed against the following definitions:

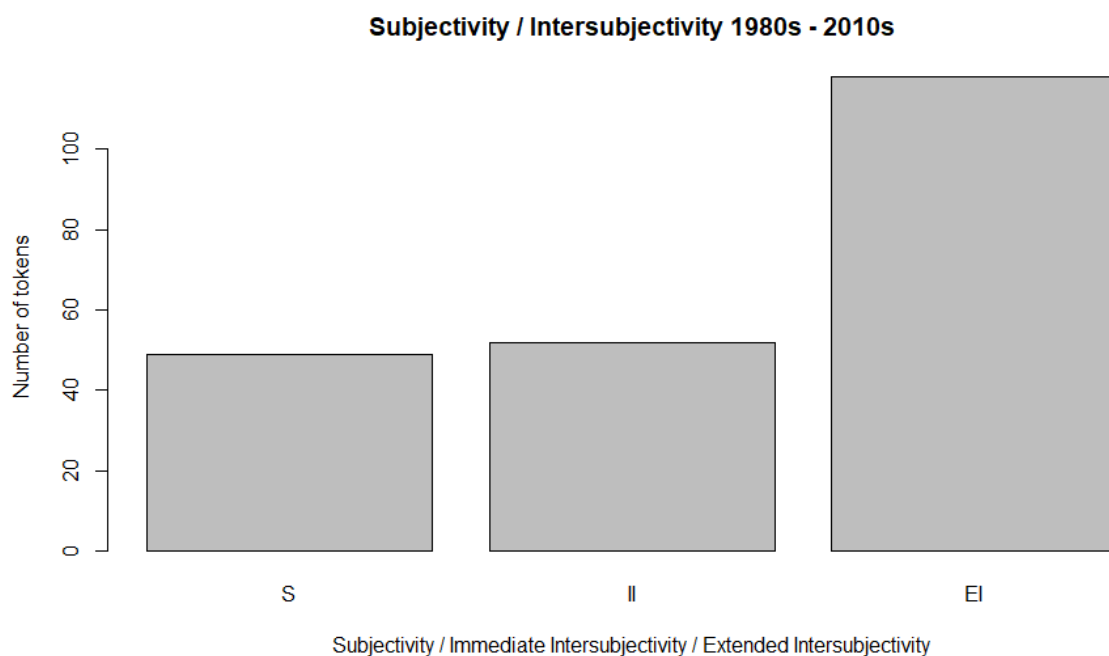
### Subjective

- [(if the) truth BE told] introduces information that the SP/W considers generally understood, but rarely expressed.
- SP/W does not know if the AD/R is already aware of the information.
- Can be replaced with *if people would admit it*, as demonstrated below for (17).

(20) I dare say there is faults enough on both sides, *if people would admit it* [...]



**Figure 8.** Number of tokens for subjective (S), immediate intersubjective (II) and extended intersubjective (EI) uses of [(if the) *truth* BE *told*] pre-1980.



**Figure 9.** Number of tokens for subjective (S), immediate intersubjective (II) and extended intersubjective (EI) uses of [(if the) *truth* BE *told*] post-1980.

## (Immediate) Intersubjective

- [(if the) truth BE told] introduces information that is an admission, or known only to SP/W (or at least judged as not known to AD/R).
- Occurs in an ongoing speech act.
- Suggests that the AD/R is trusted by the SP/W.
- Can be replaced with *between you and me*, as demonstrated below for (18). Cannot be replaced with *if people would admit it*.

(21) And *between you and me*, I don't like that arrogant kid.

## (Extended) Intersubjective

- [(if the) truth BE told] introduces information that is judged as unlikely to be known by a generic AD/R).
- Occurs in a monologue or narrative.
- Can be replaced with *in fact* (specifically the discourse marker *in fact*<sup>3</sup> (Traugott & Schwenker 2000)), as demonstrated below for (17). Cannot be replaced with *if people would admit it* or felicitously replaced with *between you and me*.

(22) *In fact*, I was psyched about trying the case for reasons of my own.

Table 5 shows the number of tokens and percentages for subjective, immediate intersubjective and extended intersubjective uses of [(if the) truth BE told] pre- and post-1980. Figure 8 shows the number of tokens for each pre-1980, while Figure 9 shows the number of tokens for each post-1980.

It can be seen that the proportion of tokens showing a subjective use of [(if the) truth BE told] is considerably lower post-1980, while the proportion of tokens showing immediate and extended intersubjective uses has risen. Carrying out a 2-sample test for equality of proportions with continuity correction indicates that the reduction in the proportion of subjective uses between the two periods is significant ( $X^2 = 8.0319$ ,  $p = 0.004596$ ), the increase in II uses is not significant ( $X^2 = 0.50962$ ,  $p = 0.4753$ ), while the increase in EI uses approaches significance ( $X^2 = 3.097$ ,  $p = 0.07844$ ). This suggests that, while the construction has had extended intersubjective semantics since the 17th century, a reduction in subjective uses and an increase in EI uses have coincided with the increasing frequency post-1980.

## 9 Discussion and conclusion

The [(if the) truth BE told] construction has a long history. Its fuller form (*if the truth BE told*) indicates its origins as a subordinate *if*-clause with a main clause in a conditional *if...then...* construction. Ambiguous sentences such as example (6) allowed the *if*-clause to be reanalysed as insubordinate and autonomous, making it available for pragmatic interpretation. As a result, it has developed several conventional pragmatic meanings. The COHA results indicate that the [(if the) truth BE told] construction has become more frequent since 1980. This increase in frequency has coincided with notable changes in the form most commonly used. Variation in the form of the verb *to be* used in the construction has practically disappeared, and the shortest form, without *if* or *the*, is used in the overriding majority of examples. Considered together, the results show a change in the prototypical form of the construction from *if the truth were told*, with some variation, to *truth be told*, with hardly any variation. As such, the construction has demonstrated 'coalescence' (Brinton and Traugott 2005) or 'reduction' (Bybee 2006) by the elision of elements. It has also demonstrated 'fusion' (Brinton & Traugott 2005), or what Bybee (2006) calls 'conservation', by losing its schematicity. It is notable, though, that the most frequent form of the verb *to be* in the pre-1980 examples (*were*) is not the form that appears to have become entrenched in the post-1980 results. There are only three examples of the shortest form of the construction in the pre-1980 results – one with *be*, one with *was* and one with *being* – so while *truth be told* is attested

it certainly does not seem to have been common. The increase in frequency has resulted in a loss of schematicity but, far from an older form of the construction being ‘conserved’, it is a previously rare form that has become fixed.

The increasing dominance of *truth be told* as the construction becomes more frequent has some possible consequences. By eliding *if* and *the* the origins of the construction are obscured. While *if the truth were told* implies that the truth is not told, *truth be told* does not. There is no longer a link between the form of the construction and the pragmatic implication that it is marking a truth that people are reluctant to tell. It has become less analysable, more opaque and more likely to be processed as a single chunk. At the same time, and perhaps as a result, the semantics of the construction have drifted further in the direction of extended intersubjectivity and a pragmatically weaker discourse marker. The EEBO corpus has an example of the [(*if the*) *truth* BE *told*] construction from 1652 (8) that behaves as a discourse marker, linking two clauses without implying reluctance to disclose something, so this has been part of the construction’s polysemy for a long time. Following the increase in frequency post-1980 the construction has nearly become an invariable expression, one of the criteria for considering a linguistic item to be a discourse marker (Bernd et al. 2021). The same period has seen a significant reduction in subjective uses of the construction and an increase in extended intersubjective uses that approaches significance. EI uses of the construction, that are addressed to multiple generic AD/Rs and do not pay attention to the ‘face’ of the AD/R, account for the majority of the examples post-1980. The significant increase in the proportion of tokens at left periphery, and a significant decrease in tokens at the right periphery, also imply that the usage has changed, even though Traugott (2012; 2022) suggests that links between function and position are construction specific. After centuries of polysemy, a period of increased frequency seems to have pushed the [(*if the*) *truth* BE *told*] construction further in the direction of discourse marker, and away from pragmatically richer meanings.

## References

- Beeching, Kate & Ulrich Detges. 2014. *Discourse functions at the left and right periphery: Crosslinguistic investigations of language use and language change*. Leiden: Brill.
- Bolly, Catherine. 2014. Gradience and gradualness of parentheticals. *Yearbook of Phraseology* 2014 (5). 25–56.
- Brinton, Laurel. 2017. *The evolution of pragmatic markers in English: Pathways of change*. Cambridge: Cambridge University Press.
- Brinton, Laurel J. & Elizabeth Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2006. From usage to grammar: The mind’s response to repetition. *Language* 82 (4). 711–733.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Davies, Mark. 2010. *The Corpus of Historical American English (COHA)*. Available online at <https://www.english-corpora.org/coha/>.
- Detges, Ulrich, Richard WALTERIT, Esme Winter-Froemel & Anne C. Wolfgruber (eds.). 2021. Whither reanalysis. Special issue of *Journal of Historical Syntax* 5 (32–39).
- Dostie, Gaétane. 2009. Discourse markers and regional variation in French. In Kate Beeching, Nigel Armstrong & Françoise Gadet (eds.), *Sociolinguistic variation in contemporary French*, 201–214. Amsterdam: John Benjamins Publishing Company.
- Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 366–431. Oxford: Oxford University Press.
- Evans, Vyvyan. 2019. *Cognitive linguistics: A complete guide*. Edinburgh: Edinburgh University Press.
- Fillmore, Charles, Paul Kay & Mary O’Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of ‘let alone’. *Language* 64 (3). 501–538.
- Haegeman, Liliane. 2003. Conditional clauses: External and internal syntax. *Mind & Language* 18 (4). 317–339.

- Hardie, Andrew. 2012. CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17 (3). 380–409.
- Heine Bernd, Gunther Kaltenböck, Tania Kuteva & Haiping Long. 2021. *The rise of discourse markers*. Cambridge: Cambridge University Press.
- Hopper, Paul & Elizabeth Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kaltenböck, Gunther & Evelien Keizer. 2022. Insubordinate *if*-clauses in FDG: Degrees of independence. *Open Linguistics* 8 (1). 675–698.
- Langacker, Ronald. 1977. Syntactic reanalysis. In Charles Li (ed.), *Mechanisms of syntactic change*, 57–140. Austin: University of Texas Press.
- Lehmann, Christian. 2004. Theory and method in grammaticalization: Grammatikalisierung. *Zeitschrift für Germanistische Linguistik* 32 (2). 152–187.
- Mato-Miguez, Beatriz. 2016. The expression of directive meaning: A corpus-based study on the variation between imperatives, conditionals, and insubordinate *if*-clauses in spoken British English. In María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo & Ignacio M. Palacios-Martínez (eds.), *Corpus linguistics on the move: Exploring and understanding English through corpora*, 291–312. Boston: BRILL.
- Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- Schwentner, Scott & Elizabeth Traugott. 2000. Invoking scalarity: The development of *in fact*. *Journal of Historical Pragmatics* 1 (1). 7–25.
- Tantucci, Vittorio. 2017. From immediate to extended intersubjectification: A gradient approach to intersubjective awareness and semasiological change. *Language and Cognition* 9. 88–120.
- Thompson, Sandra. 2007. Three frequency effects in syntax. In Joan Bybee (ed.), *Frequency of use and the organisation of language*, 269–278. Oxford: Oxford University Press.
- Traugott, Elizabeth. 1995. The role of the development of discourse markers in a theory of grammaticalization. [Paper presentation]. International Conference on Historical Linguistics (ICHL) XII, August 13–18, 1995, Manchester University, Manchester, United Kingdom.
- Traugott, Elizabeth. 1999. The rhetoric of counter-expectation in semantic change: A study in subjectification. In Andreas Blank & Peter Koch (eds.), *Historical semantics and cognition*, 177–196. Berlin: De Gruyter.
- Traugott, Elizabeth. 2003. From subjectification to intersubjectification. In Raymond Hickey (ed.), *Motives for language change*, 124–139. Cambridge: Cambridge University Press.
- Traugott, Elizabeth. 2007. Discussion article: Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic. *Catalan Journal of Linguistics* 6. 139–157.
- Traugott, Elizabeth. 2012. Intersubjectification and clause periphery. In Lieselotte Brems, Lobke Ghesquiere & Freek Van de Velde (eds.), *English Text Construction* 5 (1), special issue: *Intersections of intersubjectivity*, 7–28. Amsterdam: John Benjamins Publishing Company.
- Traugott, Elizabeth. 2018. Modelling language change with constructional networks. In Salvador Pons Bordería & Óscar Loureda Lamas (eds.), *Beyond grammaticalization and discourse markers: New issues in the study of language change* (Studies in Pragmatics 18), 17–50. Leiden: Brill.
- Traugott, Elizabeth. 2022. *Discourse structuring markers in English: A historical constructionalist perspective on pragmatics*. Amsterdam: John Benjamins Publishing Company.
- Traugott, Elizabeth & Richard Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Vlasyan, Gayane. 2018. Linguistic hedging in the light of politeness theory. [Paper presentation]. IX International Conference “Word, utterance, text: Cognitive, pragmatic and cultural aspects”, April 18–20, 2018, Chelyabinsk State University.
- Waltereit, Richard. 2002. Imperatives, interruption in conversation and the rise of discourse markers: A study of Italian *guarda*. *Linguistics* 40 (5). 987–1010.
- Wischer, Ilse. 2000. Grammaticalization versus lexicalization: ‘Methinks’ there is some confusion. In Olga Fischer, Annette Rosenbach & Dieter Stein (eds.), *Pathways of change: Grammaticalization in English*, 355–370. Amsterdam: John Benjamins Publishing Company.