

Using the BNC and the Spoken BNC2014 to Study the Syntactic Development of *I Think* and *I'm Sure*

Henrik Kaatari & Tove Larsson

To cite this article: Henrik Kaatari & Tove Larsson (2019): Using the BNC and the Spoken BNC2014 to Study the Syntactic Development of *I Think* and *I'm Sure*, English Studies, DOI: [10.1080/0013838X.2018.1558702](https://doi.org/10.1080/0013838X.2018.1558702)

To link to this article: <https://doi.org/10.1080/0013838X.2018.1558702>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 18 Apr 2019.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)

Using the BNC and the Spoken BNC2014 to Study the Syntactic Development of *I Think* and *I'm Sure*

Henrik Kaatari^a and Tove Larsson ^b

^aDepartment of Humanities, University of Gävle, Gävle, Sweden; ^bCentre for English Corpus Linguistics, Université catholique de Louvain, Louvain-la-Neuve, Belgium

ABSTRACT

This study investigates whether *I'm sure* seems to be on the same grammaticalisation trajectory as *I think*. It does so by tracking the frequency of these two constructions over time to explore (i) their distribution across clausal positions (syntagmatic variability) and (ii) the extent to which the complementiser *that* is omitted (paradigmatic variability). The study uses spoken data from the BNC and the newly compiled Spoken BNC2014. The results show that the two constructions exhibit remarkable similarity, not only in terms of their proportional distribution across clausal positions, but also in terms of their propensity for *that*-omission. For example, both constructions show adverb-like behaviour with regard to clausal positions. Furthermore, even though the time span covered is relatively short, a clear increase in *that*-omission was noted for *I'm sure*, mirroring the frequencies for *I think* very closely. It thus seems that *I'm sure* is on the same path as *I think*, despite differences in frequency entrenchment.

ARTICLE HISTORY

Received 14 April 2018
Accepted 27 August 2018

1. Introduction

The grammaticalisation of the subject-verb combination *I think* has received considerable attention in recent years.¹ Following Traugott and Heine, we view grammaticalisation as both a diachronic phenomenon and a synchronic phenomenon that can be studied “at a synchronically segmented moment in time”.² The question that has been addressed in many previous studies is whether *I think* followed by a *that*-clause should be reanalysed as an epistemic phrase; that is, whether or not *I think* is parenthetical in relation to the *that*-clause (with the parenthetical use then being associated with grammaticalisation).

(1) Uhm <,> I think I was <,> probably possessive and jealous of my mother³

CONTACT Henrik Kaatari  henrik.kaatari@hig.se

¹E.g., Thompson and Mulac, “A Quantitative Perspective”; Aijmer; Van Bogaert, “A Constructional Taxonomy”; Van Bogaert, “*I Think*”; Kaltenböck, “Explaining Diverging Evidence”; Kaltenböck, “Development of Comment Clauses.”

²Traugott and Heine, 1.

³From Kaltenböck, “Development of Comment Clauses,” 286.

Kaltenböck views constructions, as exemplified in (1), as exhibiting a discrepancy between form and function; he notes that “structurally they represent clauses, but functionally they are like disjunct adverbials conveying secondary information”.⁴ Although other, similar subject-verb combinations, such as *I guess*, *I suppose* and *I believe*, also have been described as exhibiting such grammaticalisation tendencies, *I think* stands out in being by far the most frequent subject-verb combination that can be followed by a *that*-clause.⁵

Far less attention, however, has been paid to the grammaticalisation of constructions in which an adjective is followed by a *that*-clause. One adjectival construction of special interest in this regard is *I'm sure*, since it has been shown to be similar to *I think* in two important respects.⁶ Firstly, the adjective *sure*, and in particular the construction *I'm sure*, have been shown to trigger *that*-omission.⁷ *I'm sure* is thus similar to *I think*, which strongly favours *that*-omission.⁸ Secondly, a recent empirical study has indicated that *I'm sure* readily occurs outside of its canonical clause-initial position, and that this construction thus exhibits syntactic flexibility similar to that of *I think*.⁹ There is therefore reason to believe that *I'm sure* might be on the same trajectory syntactically as its more frequent relative, *I think*, and, therefore, that it could be a candidate for grammaticalisation.

However, there are also differences between the constructions, pertaining for example to frequency entrenchment and the fact that the predicates belong to two different word classes. The question thus remains whether it is indeed possible to account for the development of the two constructions in the same way, which is explored in the present article. Specifically, the study uses comparable subsets of the spoken component of the BNC and the newly compiled Spoken BNC2014 to investigate the degree to which *I'm sure* seems to be on the same grammaticalisation trajectory as *I think*. The following research questions have been used:

- What are the frequency distributions across the clausal positions for *I think* and *I'm sure* respectively?
- To what extent is the complementiser *that* omitted in the two constructions?
- Does the development of *I'm sure* mirror that of *I think*, and if so, to what extent can such mirror development be evidence of grammaticalisation?

2. Previous Research

Several studies have investigated the syntactic status of *I think*, of which one of the most central is Thompson and Mulac.¹⁰ In that article, the authors put forth arguments to support the claim that there is a correlation between (i) the overall frequency with which *I think* is followed by a *that*-clause and the frequency with which *that* is omitted,

⁴Kaltenböck, “Development of Comment Clauses,” 287.

⁵See Thompson and Mulac, “A Quantitative Perspective”; Torres Cacoullos and Walker; Jaeger.

⁶See Kaatari, “On the Syntactic Status.”

⁷See Kaatari, “Variation across Two Dimensions”; Kaatari, “On the Syntactic Status.”

⁸See Thompson and Mulac, “A Quantitative Perspective,” 323; Thompson and Mulac, “The Discourse Conditions,” 244; Van Bogaert, “*I think*,” 306–7; Jaeger, 29; Torres Cacoullos and Walker, 20.

⁹Kaatari, “On the Syntactic Status”; see also Van Bogaert, “A Constructional Taxonomy” and Van Bogaert, “*I think*” on *I think*.

¹⁰Thompson and Mulac, “A Quantitative Perspective.”

and also a correlation between (ii) the frequency with which *that* is omitted and the ability of *I think* to occur in positions other than its canonical clause-initial position. With regard to the extent to which *I think* has grammaticalised, the authors see certain uses of *I think* as having been reanalysed into epistemic phrases:¹¹

the most frequent subjects and verbs occurring with what syntacticians have considered to be ‘*that*-less’ ‘complements’ [...] have in fact been reanalyzed by speakers as epistemic phrases, which have a degree of freedom not possible for subject-verb combinations; in particular they are ‘free’ to occur in other positions, just as epistemic phrases, such as epistemic adverbs, do in English.

Thompson and Mulac term instances in which *that* is retained *main clause uses* and instances in which *that* is omitted *parenthetical uses* of *I think*. Their proposed path of grammaticalisation follows the following steps: (i) main clause use of *I think* with *that* retained, illustrated in (2), (ii) parenthetical use of *I think* with *that*-omission, illustrated in (3), and (iii) parenthetical clause-final use of *I think*, illustrated in (4).¹²

- (2) I think that we are definitely moving towards being more technological.
- (3) I think exercise is really beneficial, to anybody.
- (4) It’s just your point of view you know what you like to do in your spare time I think.

Following this view, the parenthetical use of *I think* emerges when *that*-omission has become sufficiently prevalent. With the emergence of this parenthetical use, *I think* can be reanalysed as an epistemic phrase, which is now free to occur in other positions. This means that, through the process of grammaticalisation, *I think* has developed a new grammatical function, namely that of epistemic adverbs. The question remains, however, to what extent *I’m sure* can be considered to be on the same path.

For reasons of comparability with previous studies,¹³ we will, in this article, focus on two parameters of grammaticalisation that are relevant to Thompson and Mulac’s proposed path of grammaticalisation. These two parameters are investigation of syntagmatic and paradigmatic variability.¹⁴ Syntagmatic variability is concerned with the possibility of shifting a sign around within its construction, whereas paradigmatic variability refers to the possibility of a sign being omitted.¹⁵ In the present context, syntagmatic variability is thus concerned with the ability of *I think* and *I’m sure* to occur in positions other than their canonical clause-initial position, whereas paradigmatic variability is concerned with the extent to which *that* is omitted in clause-initial position.

There is far less research focusing on the syntactic development of *I’m sure*; nonetheless, there is reason to believe that *I’m sure* is similar enough to *I think* to merit similar treatment, as discussed below. First, the frequency of each of the predicates in relation to other verbs and adjectives, respectively, is very similar. In the class of verbs, THINK has been found to be the most frequent verb to be followed by a *that*-clause;¹⁶ similarly,

¹¹Thompson and Mulac, “A Quantitative Perspective,” 316.

¹²(2)–(4) from Thompson and Mulac, “A Quantitative Perspective,” 313.

¹³E.g., Thompson and Mulac, “A Quantitative Perspective”; Van Bogaert, “A Constructional Taxonomy”; Van Bogaert, “*I think*.”

¹⁴See Lehmann, 303.

¹⁵See Lehmann, 305–7.

¹⁶Thompson and Mulac, “A Quantitative Perspective,” 319; Jaeger, 29; Torres Cacoullos and Walker, 20.

in the class of adjectives, *SURE* has been found to be the most frequent adjective to be followed by a *that*-clause.¹⁷ The reason why this is important is that the rate of *that*-omission for predicates has been attributed to the frequency (i.e., the “expectedness”) with which these are followed by a *that*-clause (cf. the correlations proposed by Thompson and Mulac above). This also means that although grammaticalisation is strongly associated with semantic factors, frequency plays a vital role in grammaticalisation processes. For example, grammaticalised items tend to become phonetically reduced through a process of increase in frequency. This process is typically referred to as a *frequency effect*.¹⁸ The importance of frequency in grammaticalisation processes is linked to the probability, or the expectedness, of a given expression to be used in a given context.¹⁹

Second, another similarity between *I think* and *I’m sure* pertains to their semantics. Several previous studies have found a link between the semantics of the predicate and its propensity to be followed by *that*-omission. Both *THINK* and *SURE* belong to the class of predicates that most frequently are followed by a *that*-clause, namely those that express epistemic modality; that is, they comment on the truth value of the complement clause.²⁰ Dor shows that *that*-omission is determined by the property “truth-claim” of predicates, meaning that only predicates that present a truth-claim, or allow their meaning to be pragmatically extended to approximate a truth-claim, allow for *that*-omission. Moreover, Boye and Harder note that epistemic meaning is the common denominator for predicates that allow a parenthetical reading, which is why epistemic predicates are the focus of the present investigation of grammaticalisation.²¹ It is also worth noting that while there are several verbs that can be followed by a *that*-clause and that can readily be classified as epistemic predicates (e.g., *I guess*, *I believe*, *I reckon*), *SURE* is special in that most other adjectives which express epistemic modality are found in extraposed constructions (e.g., *CLEAR* and *LIKELY*, as in *it is clear that*, *it is likely that*); *SURE* is almost exclusively found to occur with a referring noun phrase as the matrix subject (as in *I’m sure*).²²

Even though the meanings expressed by *THINK* and *SURE* are very similar, the predicates are not semantically identical. Hooper makes a distinction between strong and weak assertive predicates, in which she classifies *SURE* as strong assertive and *THINK* as weak assertive. Hooper distinguishes between these two types by noting that weak assertive predicates serve “to weaken the claim to truth made by the complement”,²³ whereas strong assertive predicates “represent a rather strong commitment to the truth of the complement”.²⁴ Interestingly, Hooper extends the difference between weak and strong assertive predicates to say that only weak assertive predicates have the semantic and syntactic properties associated with grammaticalisation. The reason for this is that only weak predicates can be used parenthetically, as it is only the proposition in the *that*-clause that is being asserted for predicates of this kind. Strong assertive predicates, however, do not share these

¹⁷Kaatari, “Variation across Two Dimensions”; Torres Cacoullos and Walker, 20; Mindt, 16.

¹⁸See Bybee, “Phonology and Language Use”; Bybee, “Frequency of Use”; Hopper.

¹⁹Cf. Jurafsky et al.’s Probabilistic Reduction Hypothesis and Jaeger’s Uniform Information Density Principle.

²⁰Herriman; Mair; on epistemic modality, see Perkins, 10.

²¹Boye and Harder, 583.

²²Huddleston and Pullum, 964, state that *sure* is occasionally complemented by extraposed *that*-clauses (as in their cited example *it now seems sure there’ll be an election before the end of the year*). In corpus data, however, such examples are very rare. Examples of epistemic adjectives that are found to be moderately frequent with referring noun phrases as matrix subjects include *confident* and *convinced*.

²³Hooper, 101.

²⁴Hooper, 101.

properties, as the propositions made in both the main clause and the *that*-clause are being asserted.²⁵ Hooper further notes that certain members of the class of strong assertive predicates may develop into weak assertive predicates “through a reduction in semantic content that takes place through usage over a period of time”.²⁶ In this article, we will investigate whether this change seems to have taken place and, thus, whether *I’m sure* can now be used parenthetically.

In addition to *that*-omission being of interest in studies dealing with grammaticalisation, the topic of *that*-omission vs. *that*-retention has also been discussed in relation to other factors that are of relevance here. First of all, we should mention that the reason we are solely concerned with the first person singular pronoun form is that the type of the matrix subject has been found to be a significant factor in triggering *that*-omission, with *I* and *you* being the forms most associated with *that*-omission.²⁷ Furthermore, in terms of absolute frequency, Thompson and Mulac note that markers of epistemicity are skewed towards first person pronouns, which is why they focus on *I think* in particular.²⁸ In addition to the importance of the type of matrix subject, other factors of importance for *that*-omission/retention include the type of complement clause subject,²⁹ the length of the complement clause subject,³⁰ the length of the complement clause,³¹ and co-referentiality between the matrix subject and the complement clause subject.³² The *Complexity Principle* by Rohdenburg offers an explanation for the underlying mechanisms of *that*-omission; it states that “[i]n the case of more or less explicit grammatical options the more explicit one(s) will tend to be favored in cognitively more complex environments”.³³ In discussing the implications of the Complexity Principle on the variation between retaining or omitting *that*, Rohdenburg is mainly concerned with the type and length of the complement clause subject, and with separation between the matrix clause and the complement clause. According to Rohdenburg, “any elements capable of delaying the processing of the object clause and thus of the overall sentence structure favor the use of an explicit signal of subordination”.³⁴

All in all, it would seem that clausal position and *that*-omission vs. *that*-retention are two factors worthy of further attention in the context of grammaticalisation. Based on previous research and on the similarities between the constructions, we hypothesise that *I’m sure* is on the same grammaticalisation trajectory as *I think*; the extent to which this is the case will be investigated by studying the above-mentioned factors for both constructions.

²⁵Hooper, 101, notes that either of the assertions “may be semantically dominant in the discourse context” with strong assertive predicates. This is illustrated by the fact that strong assertive predicates are ambiguous as to which clause a question refers to. The response to the question *Are you sure you don’t want to have the artichokes tonight?* (from Hooper, 98) cannot be in the form of a simple yes or no without causing ambiguity as to which clause is being referred to in the answer. Instead, four different responses would be appropriate in the given context: *No, I’m not sure. Yes, I’m sure. No, I don’t. Yes, I do.*

²⁶Hooper, 112.

²⁷Jaeger; Rohdenburg; Thompson and Mulac, “A Quantitative Perspective.”

²⁸Thompson and Mulac, “A Quantitative Perspective,” 322.

²⁹Jaeger; Rohdenburg.

³⁰Jaeger; Rohdenburg.

³¹Jaeger.

³²Elsness; cf. also Shank et al.

³³Rohdenburg, 151.

³⁴Rohdenburg, 161.

3. Material and Method

3.1. Corpora

The corpus data used in the present study come from the *British National Corpus* (BNC)³⁵ and the newly compiled *Spoken BNC2014*.³⁶ The spoken component of the BNC was compiled between 1991 and 1994;³⁷ to ensure comparability to as great a degree as possible, only the demographically sampled (DS) part of the BNC is included in the present study.³⁸ The Spoken BNC2014 was compiled between 2012 and 2015.³⁹ Both corpora contain British English data compiled from both male and female speakers of all age groups (0–99 years old), from different socio-economic and geographical backgrounds.⁴⁰ The size of the subsets used in the present study is detailed in Table 1.

For reasons of simplicity, the demographically sampled component of the BNC will henceforth be referred to as the *BNC1994*, and the Spoken BNC2014 will henceforth be referred to as the *BNC2014*.

3.2. Method

In order to retrieve the relevant tokens for the present study, the following two strings were searched for in the BNC1994 and the BNC2014 using the corpus query processor CQPweb:⁴¹ <*I'm sure*> and <*I think*>. The study does thus not cover any tokens with other subjects or with intervening adverbials.⁴² With regard to *I'm sure*, the searches yielded 2,476 hits in total, all of which were gone through manually to exclude invalid tokens. To ensure manageability of the much more frequent construction *I think*, a random sample of 2,000 tokens was culled from each of the two corpora; these tokens were manually examined, and invalid tokens were excluded.

Examples of excluded tokens include those that are not complemented by a *that*-clause (5) and false starts (6). Only instances where the complement clause included a subject and a verb were included, which means that instances that included false starts where the same construction was repeated more than once (7) were only counted once.

(5) [S0454:] oh right mm so what's that? the drug just making her feel down? [S0455:] yes
I think so [S0454:] mm (BNC2014_S8PW)

(6) [S0543:] it's **I'm sure** it you expected it though didn't you? (BNC2014_S5MM)

(7) [S0331:] I'm sure he **I'm sure** he wrote it (BNC2014_SKHW)

All the valid tokens were coded manually for presence or absence of the complementiser *that* (+/– *that*); two examples are provided in (8) and (9). Instances where *that* did not function as a complementiser, such as referring *that* (10), were not included in the category + *that*.

³⁵Burnard; Lee.

³⁶Love et al.

³⁷Love et al., 321.

³⁸Cf. Love et al., 324.

³⁹McEnery et al., 312.

⁴⁰Love et al.

⁴¹Hardie.

⁴²See Section 2 and Kaatari, "On the Syntactic Status" for a more detailed discussion.

Table 1. Size of the subcorpora included for analysis.

Subset	Number of words
Spoken BNC1994_DS	4,233,962
Spoken BNC2014	11,422,617

- (8) Well **I think that** he's got some educational interest (BNC1994_PS1E4_KC4)
 (9) I **I think** any resemblance is purely superficial (BNC1994_PS0PP_KDM)
 (10) Oh yeah, wacko **I think** that's insulting [...] (BNC1994_PS1K5_KSV)

The valid tokens were further coded manually for clausal position: clause-initial (11), clause-medial (12) or clause-final (13). These categories are described in Quirk et al.,⁴³ where the clause-initial position is “that preceding any other clause element” and the clause-final position (*end position* in Quirk et al.) is defined as “the position in the clause following all obligatory elements”; the clause-medial position covers all other positions.⁴⁴

- (11) **I'm sure** she'll settle in though (BNC2014_SR8V)
 (12) granny **I'm sure** would have loved to have met you (BNC2014_SKDX)
 (13) it would be very different if you had to pay for it **I'm sure** [S0144:] hm yeah (BNC2014_S27D)

The fact that the BNC2014 more readily gives information about the speaker turns than the BNC1994 greatly facilitated the classification of clause position in the former. An example of a token for which the speaker-turn information helped disambiguate the classification is found in (14), which could otherwise have been erroneously categorised as clause-final. However, an example that was *not* possible to disambiguate despite the speaker-turn information is found in (15). Here, *I'm sure* could either be clause-final (*we went to Amsterdam for New Year, I'm sure*) or clause-initial (*I'm sure the only Christmas I did was the first year at uni*); tokens such as this one, where the extended context did not prove to be helpful for disambiguation, were classified as *unclear* and discarded.

- (14) [S0266:] no well no no no the really top top top people [S0271:] mm **I think** (BNC2014_S7KK)
 (15) [S0198:] because we no yeah two thousand and seven no it was two thousand and seven [S0192:] was it seven? [S0198:] cos we went to Amsterdam for New Year **I'm sure** the only Christmas I did was the first year at uni [S0198:] two thousand and six we were in Paris two thousand and seven we went I went home and then [S0192:] what I did I do? I don't think I really went home [S0198:] did you not go home? (BNC2014_SN98)

All tokens were normalised per million words to enable comparisons across the (sub)corpora. With regard to *I think*, the number of valid tokens and their distribution per 2000 tokens were used as an estimate of the number of valid tokens in the full set; these projected

⁴³Quirk et al., 490–501.

⁴⁴Cf. Quirk et al., 495[a].

Table 2. Overall frequency of *I'm sure* and *I think*.

<i>I'm sure</i>		<i>I think</i>	
BNC1994 pmw (raw)	BNC2014 pmw (raw)	BNC1994 Projected pmw (raw per 2000 tokens)	BNC2014 Projected pmw (raw per 2000 tokens)
113 (477)	145 (1650)	1598 (1,519)	2390 (1,547)

frequencies were subsequently normalised per million words. The UCREL log-likelihood and effect-size calculator was used to test the differences for statistical significance.⁴⁵

4. Results and Discussion

4.1. Overall Frequencies

In total, there were 5193 valid tokens of *I'm sure* and *I think*; the frequency per construction and the distribution across the different corpora are shown in Table 2.

While *I think* is about 25 times more frequent than *I'm sure*, Table 2 shows that both *I'm sure* and *I think* have increased in frequency over time; this increase is statistically significant.⁴⁶ *I'm sure* has gone from occurring 113 times per million words (pmw) to 145 pmw (28.2% increase), and *I think* has increased from 1598 times pmw to 2390 times pmw (49.5% increase when the projected frequencies pmw are compared). Assuming that there is a frequency effect in play here, in the sense that frequent constructions tend to become even more frequent over time,⁴⁷ it is perhaps not surprising that the more frequent construction, *I think*, has increased at a higher pace than the less frequent construction, *I'm sure*; nevertheless, it is interesting to note that the frequency of both constructions has increased considerably.

A closer look at the instances of *I'm sure* and *I think* respectively that are followed by a complement clause (CC)⁴⁸ reveals similarities between the constructions and further differences across time, as shown in Table 3.

As can be seen, the frequency with which an instance of *I'm sure* + CC and *I think* + CC occurs has gone up in relation to all other uses of SURE and THINK over time. While this is not in and of itself evidence of grammaticalisation, two prerequisites for *I'm sure* being a possible candidate for grammaticalisation seem to have been met here. First, the relative frequency of *I'm sure* + CC in the most recent data is almost identical to the relative frequency that *I think* exhibited in 1994. As *I think* is viewed as being grammaticalised, the fact that *I'm sure* now exhibits equally high frequencies suggests that the latter construction is found in an equally productive grammaticalisation environment as the former. Second, given that grammaticalised constructions are more entrenched in the language and that entrenched units tend to be used more frequently than non-entrenched units,⁴⁹ we can expect the relative frequency of grammaticalised constructions to increase over time, which is the case for both constructions (even if the increase is relatively small for *I'm sure*).

⁴⁵ucrel.lancs.ac.uk/llwizard.html.

⁴⁶*I'm sure*: LL = 23.90; $p < 0.0001$; BIC = 7.33 (cf. Wilson, for the interpretation of Bayes Factor, BIC); *I think*: LL = 951.30; $p < 0.0001$; BIC = 934.73.

⁴⁷Bybee, "Frequency of Use"; Hopper.

⁴⁸See Jaeger.

⁴⁹See, e.g., Ellis.

Table 3. Percentage of *I'm sure* + CC and *I think* + CC.

<i>I'm sure</i> + CC		<i>I think</i> + CC	
BNC1994 percentage	BNC2014 percentage	BNC1994 percentage	BNC2014 percentage
22.2	24.1	24.9	33.2

However, the fact that some prerequisites are in place does not, of course, have to mean that grammaticalisation has taken place. In order to investigate whether this indeed seems to be the case for *I'm sure*, and whether this construction thus could be argued to be on the same trajectory as *I think*, as hypothesised, we will now turn to the results of two tests used in previous studies, namely that of clausal position and that of *that*-omission vs. *that*-retention.⁵⁰

4.2. Clausal Position

In terms of the distribution across the clausal positions investigated—initial, medial and final—further similarities between the constructions emerge, as can be seen in Figure 1; the raw and normalised frequencies can be found in Appendix 1. As can be recalled from Section 2, one test of grammaticalisation that has been applied for *I think* in previous studies is whether the construction can occur outside its canonical, clause-initial position and thus occur in the positions that have previously been reserved for adverbs, namely clause-medial and clause-final position.⁵¹

As is shown, while the initial position is by far the most frequent clausal position for both constructions, the other positions, medial and final, are well represented in all corpora, which seems to support the hypothesis that *I'm sure* can now be considered to be grammaticalised. Examples of each position for both constructions can be found in (16)–(21).

- (16) [S0417:] oh **I'm sure** you'll have enough time (BNC2014_STA2)
- (17) [S0475:] you know it's erm erm she was so and yes my granny was **I'm sure** quite quite a demanding lady (BNC2014_SCSF)
- (18) [S0688:] it was more than that **I'm sure** [S0687:] I don't think it was more than that (BNC2014_SWZA)
- (19) [S0251:] **I think** I only know Gabriel as an individual angel (BNC2014_27QE)
- (20) Erm well it'll be useful **I think** for you (BNC1994_PS1JD_KDW)
- (21) [S0520:] yeah no it's quite good stuff **I think** [S0519:] mm (BNC2014_S2QU)

As is shown, the relative frequencies for the clause-initial position have increased over time. In addition, a closer look at the absolute frequencies (see Appendix 1) shows that there is a clear and statistically significant increase in the clause-initial position for both *I'm sure* and *I think*,⁵² which entails that the overall frequency increase noted in Table 2 can be attributed mainly to the increase in the clause-initial position.⁵³

⁵⁰See Van Bogaert, "A Constructional Taxonomy"; Van Bogaert, "*I think*"; Thompson and Mulac, "A Quantitative Perspective."

⁵¹See Van Bogaert, "A Constructional Taxonomy"; Van Bogaert, "*I think*."

⁵²*I'm sure*: LL = 26.04; $p < 0.0001$; BIC = 9.47; *I think*: LL = 950.13; $p < 0.0001$; BIC = 933.57.

⁵³However, there is, in fact, a statistically significant increase in terms of absolute frequencies for medial and final position for *I think* (medial: LL = 3.96; $p < 0.05$; BIC = -13.80; final: LL = 35.06; $p < 0.0001$; BIC = 18.50), even though there is a

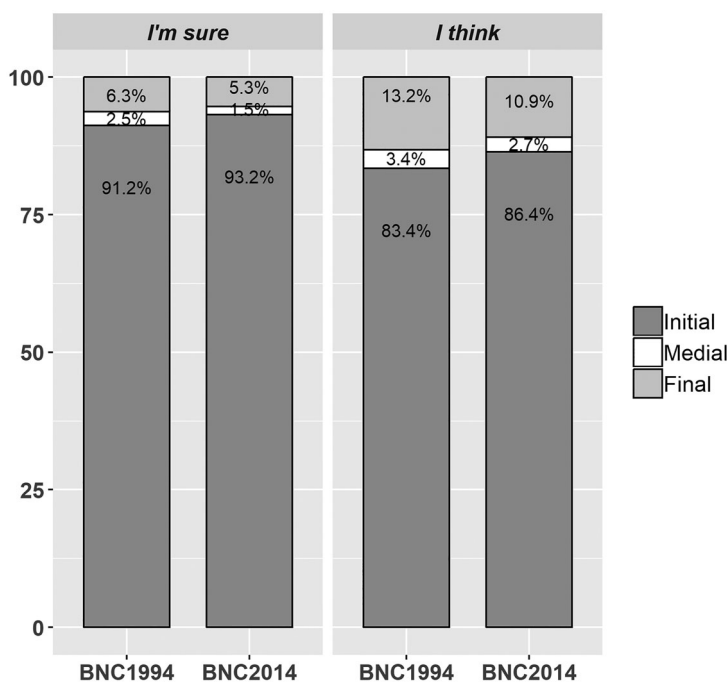


Figure 1. Percentages across clausal positions for *I'm sure* and *I think*.

However, if it is the case that an increase in terms of the relative frequencies of clause-medial and clause-final position for each position would indicate a higher degree of grammaticalisation, then it is perhaps somewhat unexpected that the relative frequencies of these non-canonical positions have gone down for both constructions (the relative frequencies for the medial and final positions have decreased from 2.5 to 1.5% and from 6.3 to 5.3% respectively for *I'm sure*, and from 3.4 to 2.7% and 13.2 to 10.9% respectively for *I think*). Put in another way, the fact that both *I'm sure* and *I think* show increasing frequency over time, coupled with the fact that clause-medial and clause-final positions are both decreasing in terms of relative frequency, suggests that the relative frequency with which clause-medial and clause-final position (the two “grammaticalised positions”) occur is not immediately correlated with the overall frequency of *I'm sure* and *I think*. It would thus seem that the frequency of these two positions has to be explained in other terms.

There are, as we see it, two plausible (and potentially interrelated) explanations for this decrease: either the two versions of the BNC are not fully comparable or the use of the English language has changed in the years that separate the two corpora. However, while no two corpora can ever be fully comparable, we have no reason to believe that the subcorpora used here differ in any significant way, as only conversation data are included in the present study (see Section 3.1).

decrease in terms of relative frequencies; this is due to the fact that the increase noted for the clause-initial position is proportionally larger than that of the other positions.

To explore whether the use of English could have changed, we focused our attention to two possible explanations: (i) as the use of clause-medial position has been found to be more frequent in more formal text categories than in less formal ones, spoken English might be becoming less formal overall;⁵⁴ (ii) the clause-medial and clause-final positions may be becoming increasingly less frequently used over time. The next sections will discuss these hypotheses in turn.

4.2.1. Formality

In order to test the first hypothesis, namely that the BNC2014 contains less formal language than the BNC1994 (for reasons of comparability or language change), the type-token ratio (TTR) was compared across the corpora. The rationale behind this approach is that this measurement has been shown to vary across level of formality, where production from registers that can be considered to be less formal, such as spoken production, tend to have a lower TTR and more formal registers, such as written production, tend to have a higher TTR.⁵⁵ Moreover, while we are here dealing with production from the same register (spoken production) and even the same subregister (face-to-face conversation), it could still be the case that there are differences at an even lower level, which is not unlikely considering that TTR has been found to be useful for distinguishing between low-level registers of written production.⁵⁶

Once punctuation and markers for unknown words or anonymised places or speakers had been removed, the TTR was calculated for both corpora. The results show that the TTR for the BNC1994 turned out to be slightly higher (0.81) than that of the BNC2014 (0.55),⁵⁷ which could suggest that the most recent data is somewhat more informal overall.

However, as the differences were relatively small, we went on to look at the most readily accessible features from Biber's first dimension (*involved* vs. *informational* production) of his multi-dimensional (MD) model of language analysis.⁵⁸ This dimension has been understood as being especially useful for distinguishing between production that is traditionally considered highly informal, such as face-to-face conversations, and more formal spoken registers, such as interviews, prepared speeches and broadcasts.⁵⁹ High frequencies for the negative features in this dimension suggest more informational production, which is here taken to mean more formal production; high frequencies for the positive features suggests more involved production (here considered as more informal). The following negative features were included: attributive adjectives, prepositions, nouns. The only positive feature included was the pronoun *it*; the results are shown in Table 4.

The results were somewhat inconclusive, as the finding that the BNC1994 seems to contain more formal production than the BNC2014 suggested by the TTR was only partly supported. Overall, the differences were not particularly large; nonetheless, the hypothesis that the BNC2014 contains less formal data cannot be dismissed outright.

⁵⁴See Kaatari, "On the Syntactic Status."

⁵⁵See, e.g., Baker, 52.

⁵⁶Chafe and Danielewicz.

⁵⁷For reasons of readability, these TTR scores were multiplied by 100. The original scores were 0.008106227 and 0.005535856 for the BNC1994 and the BNC2014 respectively.

⁵⁸See, e.g., Biber, *Variation across Speech* and Biber, "Using MD Analysis" for a description of this model and its dimensions.

⁵⁹Biber, *Variation across Speech*; see also Biber, "Using MD Analysis."

Table 4. Features of informational vs. involved production investigated.

Feature	BNC1994 (pmw)	BNC2014 (pmw)	Comment
Attributive adjectives	54,187 (12,798)	166,216 (14,551)	BNC2014 more informational
Prepositions	237,335 (56,055)	669,722 (58,631)	BNC2014 more informational
Nouns	435,895 (102,952)	1,160,422 (101,590)	BNC1994 more informational
Pronoun <i>it</i>	112,328 (26,530)	353,400 (30,939)	BNC2014 more involved (= BNC1994 more informational)

4.2.2. Adverb Placement

In order to investigate whether the frequency of adverbs in clause-initial position has gone up at the expense of the two other positions in general, two investigations were carried out. The first one involved a comparison of the frequency of adverbs that occur in “utterance-initial” position vs. “utterance-final” position (using CQPweb’s syntax). As we are mainly interested in adverbs that are potential competitors to the epistemic constructions *I think* and *I’m sure*, we investigated the class of adverbs that are in Jacobson’s terminology referred to as *modal adverbs*:⁶⁰ *maybe, perhaps, probably, surely, clearly, actually, apparently, definitely, certainly, evidently, obviously, possibly, really* and *simply*. The results can be found in Figure 2.

Interestingly, the results show that the relative frequency of the clause-final position has decreased for all adverbs except for the two adverbs that exhibited the lowest frequencies in clause-initial and clause-final position: *clearly* and *simply*.

Secondly, as this approach has its disadvantages (for example that the relative frequencies for the clause-medial position are not investigated and that clause-initial instances of adverbs that are preceded for example by an adverbial are not counted), we went on to manually code randomised samples of 100 tokens of the three adverbs from this list that can be considered to be most similar to *I think* and *I’m sure* both semantically and syntactically, namely *maybe, perhaps* and *surely* in both corpora. As can be seen in Figure 3, the results show a very similar trend, where the frequency of the clause-initial position has increased at the expense of the clause-final position in particular.

Assuming that the corpora are fully comparable, there seems to have been a general decrease of such adverbs in clause-final position in spoken British English over time. It is still conceivable that there is some interaction between the two factors of comparability and level of formality, but this must be left to future research.

Nonetheless, it appears that the clause-placement behaviour of modal adverbs is mirrored by that of *I think* and *I’m sure*. What follows from this is that both these constructions seem to be similar enough syntactically to adverbs to follow the same general trend in terms of clausal position,⁶¹ which could offer some support for the claim that *I’m sure*, like *I think*, has grammaticalised, as this suggests that these constructions now can function as adverbs.

4.2. That-Omission vs. That-Retention

With regard to the frequency with which *I’m sure* and *I think* occur with *that*-omission, it becomes apparent that they both favour omission of *that*, as shown in Figure 4; the raw and normalised frequencies can be found in Appendix 2. Assuming that a high proportion

⁶⁰Jacobson; see also Granath.

⁶¹See Thompson and Mulac, “A Quantitative Perspective.”

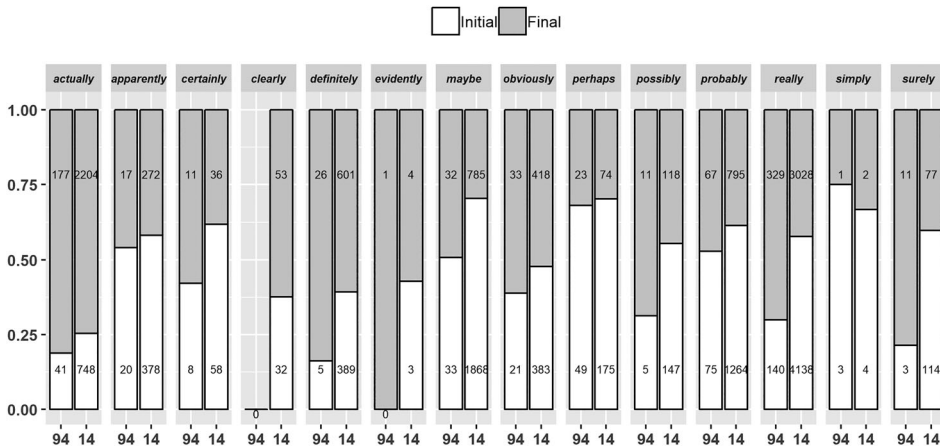


Figure 2. Percentages and raw frequency of modal adverbs across clausal positions and time (94 = BNC1994; 14 = BNC2014).

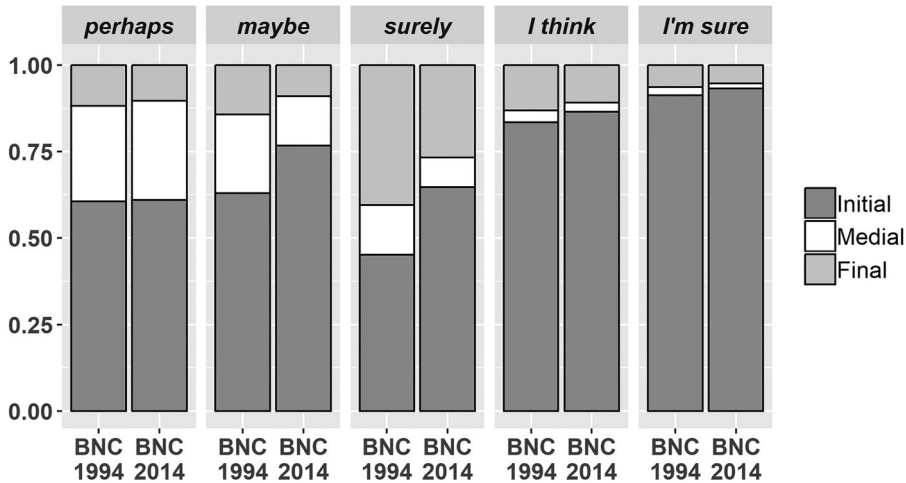


Figure 3. Percentages across clausal positions for *I think*, *I'm sure* and the adverbs investigated.

of *that*-omission is a sign of grammaticalisation,⁶² it seems that *I'm sure*, like *I think*, can be considered to be grammaticalised.

Interestingly, *I'm sure* (93.3% and 97.9%) exhibits a larger increase of *that*-omission over time than *I think* (98.5% and 98.8%). Nonetheless, in terms of overall frequencies (see Appendix 2), the increase in *that*-omission is statistically significant for both constructions.⁶³ There is no statistically significant difference for *that*-retention for either construction.⁶⁴ Examples for both constructions can be found in (22)–(25).

⁶²Cf. Van Bogaert, "A Constructional Taxonomy"; Van Bogaert, "*I think*"; Thompson and Mulac, "A Quantitative Perspective."

⁶³*I'm sure*: LL = 28.43; $p < 0.0001$; BIC = 11.86; *I think*: LL = 951.03; $p < 0.0001$; BIC = 934.46.

⁶⁴*I'm sure*: LL = 0.92; $p > 0.05$; BIC = -15.64; *I think*: LL = 3.21; $p > 0.05$; BIC = -19.64.

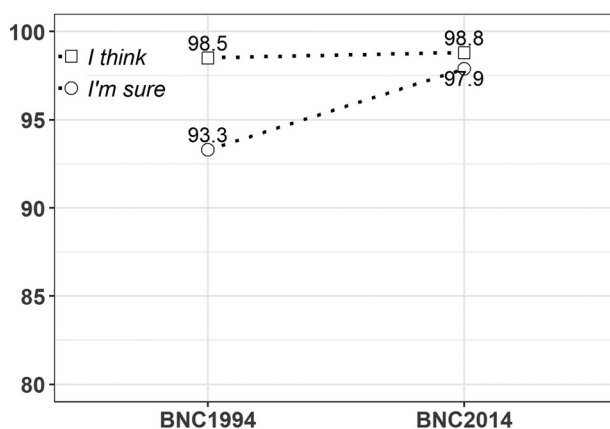


Figure 4. Percentages of *that*-omission for *I think* and *I'm sure* across time.

- (22) [S0571:] well well **I'm sure that** I'll end up paying your insurance for a bit (BNC2014_SVP4)
- (23) [S0008:] **I'm sure** you must've only done it on calm days (BNC2014_SZNG)
- (24) **I think that** he had a very rounded message that er had some interesting things in it that one could explore (BNC2014_S24A)
- (25) **I think** I'd better take these books back down again (BNC1994_PS0W2_KE2)

As *I'm sure* seemed not yet to have taken on an adverb function in the 1970s, when Hooper asserted that *I'm sure* could not be found outside its canonical clause-initial position, it could thus be that *I'm sure* is just only recently catching up to *I think*; based on this, we could expect the frequency of *that*-omission for *I'm sure* to keep increasing. However, based on the frequencies of *that*-omission for *I think*, it seems that the saturation point is somewhere around 99%, and it might not be realistic to expect the frequencies to ever reach 100%.

5. Concluding Discussion

The present study has investigated the degree to which *I'm sure* is on the same grammaticalisation trajectory as *I think* with regard to the frequency distribution across clausal positions (syntagmatic variability) and to what extent the complementiser *that* is omitted (paradigmatic variability). The study has also investigated (i) whether there are differences across time such that the development of *I'm sure* could be viewed to mirror that of *I think* and (ii) whether this mirror development can offer evidence to support the claim that *I'm sure*, like *I think*, is grammaticalised.

With regard to overall frequencies, both *I'm sure* and *I think* were found to have increased significantly in the years that separate the two corpora. Interestingly, the results showed that this was mainly due to a considerable increase in instances of the constructions found in clause-initial position. We were also able to further isolate this increase in frequency to contexts in which the complementiser *that* is omitted.

More generally, the results showed that the prerequisites of grammaticalisation previously noted for *I think* are in place for *I'm sure* as well. This claim is supported by

previous studies stating that *I'm sure* and *I think* exhibit very similar frequencies in relation to competing constructions, such as *I guess* and *I suppose* (for *I think*) and *I'm confident* and *I'm convinced* (for *I'm sure*), in that they are both by far the most frequent construction in relation to their competitors.⁶⁵ In fact, with regard to *I'm sure*, there are no clear competitors among the class of adjectives that come even close in terms of frequency. In a study focusing solely on adjectives complemented by *that*-clauses, Kaatari found that SURE accounts for 32% of all tokens (excluding extraposition).⁶⁶ In the second place, we find CONFIDENT with a share of 9%.⁶⁷

The results further suggest that *I'm sure* appears to follow along the same path as *I think* with regard to both syntagmatic and paradigmatic variability. With respect to the distribution across the clausal positions, it seemed that *I'm sure*, like *I think*, is indeed grammaticalised, as both constructions can occur in clause-medial and clause-final position, in addition to clause-initially. From this perspective, both constructions thus behave similarly syntactically to adverbs. This claim is further supported by the fact that the frequency with which in particular the clause-final position is used has gone down over time, both for the constructions and for the adverbs investigated. While this decrease might be a result of the level of formality having gone down over time and/or the fact that the two corpora used are not fully comparable, no conclusive evidence of either was found; this will therefore have to be left to future studies to investigate. With regard to the extent to which the complementiser *that* is omitted, further evidence was found in support of the claim that *I'm sure* is grammaticalised. In the relatively small number of years separating the corpora, there has been a clear increase in *that*-omission, such that the frequencies for *I'm sure* now closely resemble those of *I think*.

All in all, the two constructions exhibit remarkable similarity, especially in the most recent data, which suggests that *I'm sure* is on the same trajectory as *I think* and that both constructions can, indeed, be viewed as grammaticalised, despite the differences with regard to frequency and the word class of the predicate. The findings thus show that grammatical category (adjective vs. verb) is of less importance than syntactic function (both *sure* and *think* are predicates that can take clausal propositions as their complement). The fact that *I think* and *I'm sure* exhibit similar behaviour despite differences in frequency entrenchment could perhaps suggest that there is a superordinate construction, as predicted by Van Bogaert.⁶⁸ Both constructions could then be considered to be part of the same constructional grammaticalisation schema in which the frequency and entrenchment of *I think* could be seen to reinforce the grammaticalisation of both *I think* and *I'm sure*. The fact that predicates from two different word classes behave very similarly could act as further evidence of a need for a superordinate construction to explain the similarities between these constructions; this is, however, another aspect that will have to be left for future studies to look into.

⁶⁵Kaatari, "On the Syntactic Status"; Torres Cacoullós and Walker.

⁶⁶Kaatari, "On the Syntactic Status."

⁶⁷However, as previously discussed in this article, there are still differences between THINK and SURE: THINK is around eight times as frequent as SURE in the spoken component of the BNC 1994. Furthermore, Torres Cacoullós and Walker, 20, in a study including both verbs and adjectives followed by *that*-clauses, report that 43% of their tokens are represented by THINK and only 2% by BE + SURE. Thompson and Mulac, "A Quantitative Perspective" and Jaeger, both dealing only with verbs followed by *that*-clauses, report that THINK account for 53% and 52% of the total tokens.

⁶⁸Van Bogaert, "A Constructional Taxonomy."

Finally, it is noteworthy that even with this relatively short time span of present-day English, there were clear differences in use over time. These differences would not have been possible to detect had it not been for the compilation of this new corpus of spoken data. This study is thus evidence of the fact that we need large corpora of this kind to gain further insights into the use of the ever-changing English language.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Tove Larsson  <http://orcid.org/0000-0002-0489-2697>

References

- Aijmer, Karin. "I Think: An English Modal Particle." In *Modality in Germanic Languages: Historical and Comparative Perspectives*, edited by Toril Swan, and Olaf Jansen Westvik, 1–47. Berlin: Mouton de Gruyter, 1997.
- Baker, Paul. *Using Corpora in Discourse Analysis*. London: Continuum, 2006.
- Biber, Douglas. *Variation across Speech and Writing*. Cambridge: Cambridge University Press, 1988.
- . "Using MD Analysis to Explore Cross-Linguistic Universals of Register Variation." In *Genre- and Register-Related Discourse Features in Contrast*, edited by Marie-Aude Lefer, and Svetlana Vogeleer, 7–34. Amsterdam: John Benjamins, 2016.
- Boye, Kasper, and Peter Harder. "Complement-Taking Predicates: Usage and Linguistic Structure." *Studies in Language* 31, no. 3 (2007): 569–606.
- Burnard, Lou. *Reference Guide for the British National Corpus* (XML edition), 2007 [cited 11 January 2018]. Available from <http://www.natcorp.ox.ac.uk/docs/URG/>.
- Bybee, Joan. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.
- . *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press, 2007.
- Chafe, Wallace, and Jane Danielewicz. "Properties of Spoken and Written Language." In *Comprehending Oral and Written Language*, edited by Rosalind Horowitz, and S. Jay Samuels, 83–113. New York: Academic Press, 1987.
- Dor, Daniel. "Toward a Semantic Account of *that*-Deletion in English." *Linguistics* 43, no. 2 (2005): 345–82.
- Ellis, Nick. "Construction Grammar and Second Language Acquisition." In *The Oxford Handbook of Construction Grammar*, edited by Thomas Hoffmann, and Graeme Trousdale, 365–78. New York: Oxford University Press, 2013.
- Elsness, Johan. "That or Zero? A look at the choice of object clause connective in a corpus of American English." *English Studies* 65, no. 6 (1984): 519–33.
- Granath, Solveig. "The Position of the Adverb Certainly Will Make a Difference." *English Today* 18, no. 1 (2002): 25–30.
- Herriman, Jennifer. "Extraposition in English: A Study of the Interaction between the Matrix Predicate and the Type of Extraposed Clause." *English Studies* 81, no. 6 (2000): 582–99.
- Hooper, Joan B. "On Assertive Predicates." In *Syntax and Semantics*, edited by John Kimball, Vol. 4, 91–124. New York: Academic Press, 1975.
- Hopper, Paul. "On Some Principles of Grammaticalization." In *Approaches to Grammaticalization*, edited by Elizabeth C. Traugott, and Bernd Heine, Vol. 1, 17–35. Amsterdam: John Benjamins, 1991.

- Huddleston, Rodney, and Geoffrey K. Pullum. *The Cambridge Grammar of English*. Cambridge: Cambridge University Press, 2002.
- Jacobson, Sven. *Factors Influencing the Placement of English Adverbs in Relation to Auxiliaries: A Study in Variation*. Stockholm: Almqvist & Wiksell International, 1975.
- Jaeger, Florian. "Redundancy and Reduction: Speakers Manage Syntactic Information Density." *Cognitive Psychology* 61, no. 1 (2010): 23–62.
- Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. "Probabilistic Relations between Words: Evidence from Reduction in Lexical Production." In *Frequency and the Emergence of Linguistic Structure*, edited by Joan Bybee, and Paul Hopper, 229–54. Amsterdam: John Benjamins, 2001.
- Kaatari, Henrik. "Variation across Two Dimensions: Testing the Complexity Principle and the Uniform Information Density Principle on Adjectival Data." *English Language and Linguistics* 20, no. 3 (2016): 533–58.
- . "On the Syntactic Status of *I'm sure*." *Corpora* 13, no. 1 (2018): 1–25.
- Kaltenböck, Gunther. "Explaining Diverging Evidence: The Case of Clause-Initial *I think*." In *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*, edited by Doris Schönefeld, 81–112. Amsterdam: John Benjamins, 2011.
- . "Development of Comment Clauses." In *The English Verb Phrase: Investigating Recent Change with Corpora*, edited by Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis, 286–317. Cambridge: Cambridge University Press, 2013.
- Lee, David. "Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle." *Language Learning and Technology* 5, no. 3 (2001): 37–72.
- Lehmann, Christian. "Grammaticalization: Synchronic Variation and Diachronic Change." *Lingua e stile* 20, no. 3 (1985): 303–18.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and Tony McEnery. "The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations." *International Journal of Corpus Linguistics* 22, no. 3 (2017): 319–44.
- Mair, Christian. *Infinitival Complement Clauses: A Study of Syntax in Discourse*. Cambridge: Cambridge University Press, 1990.
- McEnery, Tony, Robbie Love, and Vaclav Brezina. "Compiling and Analysing the Spoken British National Corpus 2014." *International Journal of Corpus Linguistics* 22, no. 3 (2017): 311–18.
- Mindt, Ilka. *Adjective Complementation: An Empirical Analysis of Adjectives Followed by that-Clauses*. Amsterdam: John Benjamins, 2011.
- Perkins, Michael R. *Modal Expressions in English*. London: Frances Pinter, 1983.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. Harlow: Longman, 1985.
- Rohdenburg, Günter. "Cognitive Complexity and Increased Grammatical Explicitness in English." *Cognitive Linguistics* 7, no. 2 (1996): 149–82.
- Shank, Christopher, Koen Plevoets, and Julie Van Bogaert. "A Multifactorial Analysis of *that*/Zero Alternation: The Diachronic Development of the Zero Complementizer with *think*, *guess* and *understand*." In *Corpus-based approaches to Construction Grammar*, edited by Jiyoung Yoon, and Stefan Th. Gries, 201–40. Amsterdam: John Benjamins, 2016.
- Thompson, Sandra A., and Anthony Mulac. "A Quantitative Perspective on the Grammaticization of Epistemic Parentheticals in English." In *Approaches to Grammaticalization*, edited by Elizabeth C. Traugott, and Bernd Heine, Vol. 2, 313–39. Amsterdam: John Benjamins, 1991.
- . "The Discourse Conditions for the Use of the Complementizer *that* in Conversational English." *Journal of Pragmatics* 15, no. 3 (1991): 237–51.
- Torres Cacoullos, Rena, and James A. Walker. "On the Persistence of Grammar in Discourse Formulas: A Variationist Study of *that*." *Linguistics* 47, no. 1 (2009): 1–43.
- Traugott, Elizabeth C., and Bernd Heine. "Introduction." In *Approaches to Grammaticalization*, edited by Elizabeth C. Traugott, and Bernd Heine, Vol. 2, 1–14. Amsterdam: John Benjamins, 1991.

UCREL log-likelihood and effect-size calculator. Developed at Lancaster University. Available from ucrel.lancs.ac.uk/llwizard.html.

Van Bogaert, Julie. "A Constructional Taxonomy of *I think* and Related Expressions: Accounting for the Variability of Complement-Taking Mental Predicates." *English Language and Linguistics* 14, no. 3 (2010): 399–427.

———. "*I think* and other Complement-Taking Predicates: A case of and for Constructional Grammaticalization." *Linguistics* 49, no. 2 (2011): 295–332.

Wilson, Andrew. "Embracing Bayes Factors for Key Item Analysis in Corpus Linguistics." In *New Approaches to the Study of Linguistic Variability: Language Competence and Language Awareness in Europe*, edited by Markus Bieswanger, and Amei Koll-Stobbe, 3–11. Frankfurt: Peter Lang, 2013.

Appendices

Appendix 1

Table A1. Frequencies across clausal positions for *I'm sure* and *I think*.

Pos.	<i>I'm sure</i>						<i>I think</i>					
	BNC1994			BNC2014			BNC1994 (in 2,000 subset)			BNC2014 (in 2,000 subset)		
	Raw	%	pmw	Raw	%	pmw	Raw	%	Projected pmw	Raw	%	Projected pmw
Initial	435	91.2	103	1538	93.2	135	1267	83.4	1333	1336	86.4	2064
Medial	12	2.5	3	24	1.5	2	52	3.4	55	41	2.7	63
Final	30	6.3	7	88	5.3	8	200	13.2	210	170	10.9	263
Total	477	100	113	1650	100	145	1519	100	1598	1547	100	2390

Appendix 2

Table A2. Frequency of *that*-omission vs. *that*-retention for *I'm sure* and *I think*.

	<i>I'm sure</i>						<i>I think</i>					
	BNC 1994			BNC 2014			BNC 1994			BNC 2014		
	Raw	%	pmw	Raw	%	pmw	Raw	%	pmw	Raw	%	pmw
+ <i>that</i>	16	3.7	4	32	2.1	3	19	1.5	20	16	1.2	25
– <i>that</i>	419	93.3	99	1506	97.9	132	1248	98.5	1313	1320	98.8	2039
Total	435	100	103	1538	100	135	1267	100	1333	1336	100	2064