English Language and Linguistics 11.3: 475–505. © Cambridge University Press 2007
doi:10.1017/S1360674307002353 Printed in the United Kingdom

Epistemic verbs and zero complementizer

KATE KEARNS
University of Canterbury
(Received 9 January 2006; revised 10 January 2007)

This article addresses the question of why certain epistemic verbs, particularly (*I*) *think*, have very high rates of zero complementizer in their clausal complements. I consider and reject the influential proposal from Thompson & Mulac (1991b) that the occurrence of zero complementizer in these contexts is associated with syntactic reanalysis of the epistemic verb and its subject to an adverbial-like epistemic phrase. Instead, I propose that zero complementizer signals a shift in the informational structure of the utterance, in which the embedded clause conveys the main assertion, but the syntactic structure of the sentence need not be altered. Further routinization is evident with *I think* but not with other epistemic verbs. Results from a study of newspaper prose are presented in support of the analysis.

1 Introduction

In studies of the *that*/zero choice in clausal complements to verbs over time, *think* and *say* stand out as having high zero rates. Rissanen (1991) examines the increase in zero complementizer from Late Middle English to Early Modern English using the Helsinki Corpus. Rissanen's results (1991: 279) show *think* with consistently the highest zero rate, rising from 17 per cent zero in Late Middle English (texts 1350–1420) to 86 per cent zero in Early Modern English (texts 1640–1710). *Say* starts with the second highest zero rate, 16 per cent, in Late Middle English, then lags behind *know* for a period, and emerges as the second runner behind *think* in Early Modern English with a zero rate of 77 per cent. Elsness (1984) presents a study of zero complementizer in the Brown Corpus of American English texts published in 1961. In the news and light fiction texts, Elsness' results show *say* with a zero rate of 76 per cent and *think* with 86 per cent, in contrast to an average zero rate of 38.6 per cent for all the other verbs in the same texts that had at least five occurrences. Biber et al. present analyses of the Longman Spoken and Written English Corpus of twentieth-century British and American English. They observe (1999: 681) that three main grammatical factors favour the omission of *that*, one of which is 'the use of *think* or *say* as the main clause verb (v. other less common verbs)'. The same pattern is found in the corpus of twenty-first-century newspaper prose on which the present study is based: both *say* and *think* have zero rates of 90 per cent, in contrast to the overall 50 per cent zero rate of all other verbs. Generally, *say* and *think* are the most frequent verbs in the corpora studied, and this is considered likely to contribute to their high zero rates. This article chiefly addresses epistemic verbs such as *think*, although I shall also comment on the possible relevance of *say* in my concluding remarks.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

In an influential study of epistemic verbs, particularly *think*, Thompson & Mulac (1991a,b) propose that high zero rates are associated with the reanalysis of an epistemic verb and its subject as an epistemic phrase (EP). An epistemic phrase is an adverbial-like, dependent element, comparable in function and distribution to *maybe* or *perhaps*. The erstwhile embedded clause is reanalysed as the root clause. Accordingly, Thompson & Mulac claim that an EP has no syntactic complement-introducing properties (1991a: 318), and EPs are an instance of 'the type of grammaticization in which a governing or head element is reanalysed as a governed or dependent element' (1991a: 322). They argue that the reanalysis of an epistemic verb and its subject to a dependent element is most clearly shown where the reanalysed sequence can be reordered to appear in clause-medial or clause-final position.

Thompson & Mulac suggest that EP formation is the chief mechanism in the rise of zero with epistemic verbs. This point is most strongly expressed in Thompson & Mulac (1991a), where zero complementizer is treated as diagnostic for the presence of an EP. Their comments cited below relate to their examples shown in (1)–(3).

(1) I think *that* we're definitely moving towards being more technological.
(2) I think *0* exercise is really beneficial, to anybody.
(3) It's just your point of view you know what you like to do in your spare time *I think*.

> Examples such as (1)–(3) have generally been thought of in terms of a process of '*that*-deletion', i.e. an alternation between constructions like (1) and (2), with and without *that*. The purpose of this paper, however, is to suggest that what has been thought of as '*that*-deletion' is better understood as an alternation between constructions like (1), in which *I* and *think* are main subject and verb, with *that* introducing a complement clause, and constructions like (2) and (3) in which *I think* is an epistemic phrase, expressing the degree of speaker commitment . . . functioning roughly as an epistemic adverb such as *maybe* with respect to the clause it is associated with. (1991a: 313)

The assumption that zero indicates the presence of an EP is also clear in their data-coding strategy (1991a: 316–17) where all tokens of the form of (2), with zero complementizer, are coded as EPs, in contrast to tokens with *that* where the epistemic verb and its subject are classified as a complement-introducing main clause.

Thompson & Mulac identify three properties which they argue are required for EP formation. The first characteristic of EPs is that the subject is in the first or second person. Thompson & Mulac note that '[a] number of researchers have pointed out that speaker commitment can only be asserted for the speaker (*I*), or queried of the addressee by the speaker (*you*). It cannot be meaningfully asserted or queried about a third person' (Thompson & Mulac 19991a: 243). Second, the verb must be epistemic in order to serve in an EP. Third, the apparent matrix subject + verb sequence is simple, without auxiliary verbs, adverbs, or indirect objects. Here Thompson & Mulac observe that '[a]ll these verb phrase elements reduce the likelihood that the main subject and verb are being used as an epistemic phrase; their presence correlates with a literal use of the subject and verb as independent lexical items' (1991a: 248). The main exemplar

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

of the combination of these properties, and the main instance of EP formation, is the sequence *I think*, which is the most frequent subject + verb collocation in the data, followed by *I guess*. Thompson & Mulac point out that the dominance of *I* subjects and the verbs *think* and *guess* as the most frequent subjects and verbs in their data is consistent with the frequency conditions for grammaticalization.

To sum up so far – the central point of Thompson & Mulac's analysis is a strong hypothesis concerning the role of EP formation in the rise of zero with epistemic verbs, to the effect that zero complementizer is a consequence of the syntactic reanalysis involved in EP formation, illustrated in simplified form in (4)–(5) below.

(4) [I think [that it's going to rain]] ⇒
(5) [[I think] it's going to rain]/[It's going to rain [I think]]

Following reanalysis, the clause following the epistemic verb is actually the root clause, and so an overt complementizer *that* is impossible. The preconditions for reanalysis are that the verb be epistemic, that the epistemic verb have a first- or second-person subject, and that the epistemic verb clause be simple in form (no auxiliaries, adverbials, or other modification).

The different kinds of evidence that Thompson & Mulac discuss bear on different points in their proposal. Functionally, Thompson & Mulac argue that a first/second-person subject and a simple matrix are conditions for the expression of epistemic modification (in addition to the obvious requirement that the verb be epistemic). These conditions could in principle be fulfilled by any epistemic verb, regardless of frequency, and do not in themselves imply any reanalysis. The proposed syntactic reanalysis involved in EP formation is identified as a type of grammaticalization, and is thus subject to the usual conditions for grammaticalization, especially frequency and a fixed form. These conditions are met by the collocations *I think* and *I guess*, but would not be met by instances of comparatively infrequent verbs such as *feel* and *believe*, which might nevertheless be expressions of epistemic modification. From these considerations two stages of development are implied: first, certain uses of epistemic verbs express epistemic modification (in contrast to predicating mental states of their subjects), and second, a frequent, fixed-form subset of these verb uses undergoes further 'routinization' or grammaticalization, involving syntactic reanalysis.

Thompson & Mulac identify the second stage as the key to high zero rates – the erstwhile embedded clause, reanalysed as the root clause, cannot have an overt complementizer – but there are problems with this view. Assume, following Thompson & Mulac, that EP formation is syntactic reanalysis affecting constituency and subordination relations. Assume also that the input to this reanalysis has the proposed required characteristics: the subject of the epistemic verb is first- or second-person and the epistemic verb clause is simple. The reanalysis is illustrated below: a structure like (6) is rebracketed as in (7).

(6) [I think [that it's going to rain]]
(7) [[I think] it's going to rain]

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Note that the proposed reanalysis from (6) to (7) is not consistent with the general assumption that the *surface* form is not altered by the underlying structural change (see Harris & Campbell 1995: 50–1, 70–2, 93; Langacker 1977: 58; Timberlake 1977: 148). The input to reanalysis should have a surface form which is ambiguous between the two bracketings, but the subordinate status of *it's going to rain* is overtly marked in (6) by complementizer *that*, and so the sentence is not surface-ambiguous between the two bracketings. Given (6) as an input, EP formation would involve *that*-deletion as well as rebracketing. I conclude that if EP formation occurs, the basic reanalysis involved must be of the form shown in (8) and (9) – in particular, the input to reanalysis has zero complementizer.

(8) [I think [*0* it's going to rain]]
(9) [[I think] it's going to rain]

Given this conclusion, the proposed syntactic reanalysis involved in EP formation cannot itself be a major contributor to high zero rates with epistemic verbs, as zero would be a precondition for EP formation. The rise in zero rates with epistemic verbs must be initially triggered by other factors.[1] I shall argue in this article that the initial rise in zero is triggered by the effect on information structure of modifier-like uses of epistemic verbs.

The article is structured as follows. Section 2 clarifies the distinction, based on Urmson (1963), between the uses of epistemic verbs in what I shall call the *main verb use*, which predicates a mental state or activity of its subject, and what I shall call the *modifier use*, which is the use at issue in the proposed formation of epistemic phrases. In this section I shall show that the modifier use has a wider range of forms than Thompson & Mulac suggest – in particular, I show that the modifier use is not confined to first/second-person subjects or unmodified matrix clauses. Diagnostics for modifier and main verb uses are identified in section 2. Section 3 discusses the significance of epistemic phrases in clause-medial and clause-final position (*The blue is better, I think*), and concludes that these constructions do not provide strong evidence for syntactic reanalysis to a corresponding epistemic phrase in clause-initial position. The findings of sections 1–3 are summarized in section 4, and the issues to be explored in the corpus study are identified. Section 5 then reviews factors in the *that*/zero choice that have been previously identified in verb complements generally, not only with epistemic verbs. This section presents results of preliminary investigations of the main corpus on which the present study is based, and identifies the significant general factors which are taken into account in the epistemic verbs study. The epistemic verbs corpus is

[1] Aijmer (1997: 8–10) also argues that EP formation is an unlikely source for high zero rates with *think*, given that zero rates have fluctuated over time, and as Rissanen (1991: 283) observes, 'zero may have been the unmarked link in speech throughout the Old and Early Middle English period'. In addition, nonepistemic verbs, especially *say*, appear over time with high zero rates, but are not plausibly subject to a similar account in terms of reanalysis of the matrix subject and verb to a dependent phrase. Brinton (1996: chapter 8) also criticizes Thompson & Mulac's proposal, with particular reference to the significance of clause-medial and clause-final parentheticals. Brinton's observations are discussed in the next section.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

described in section 6, with results and discussion following in section 7. Concluding remarks are in section 8.

2 Modifier and main verb senses of epistemic verbs

As Thompson & Mulac (1991b: 238–9) outline, the proposed reanalysis of an epistemic verb and its subject to an epistemic modifier correlates with a modifier-like sense of the verb. The modifier sense of an epistemic verb correlated with its use in a medial or final parenthetical phrase has been commented on by other writers, generally after the influential discussion in Urmson (1963).[2] Urmson raises a number of points that have not figured in more recent work, but are of interest in the present study: Urmson's distinction between verb uses is outlined in this section.

Urmson draws a distinction between two modes of use of verbs such as *think*, *believe*, *suppose*, and so on, which apparently denote mental states or activities. Urmson's discussion intertwines two main themes. His chief purpose is to argue that certain uses of epistemic verbs do not actually predicate mental states or processes of their subjects – that is, *x thinks that p* is not always used to assert that x holds a certain belief. His main argument in support of this claim is that sequences such as *x thinks* may not be a part of the asserted statement at all, but rather function as framing devices 'to orient the hearer aright towards the statements with which they are associated' (1963: 233).[3] Urmson calls a verb used in a nonasserted, framing device way a 'parenthetical verb', but this term is confusing because it implies, contrary to his intention, that he assumes such uses instantiate grammatical parenthesis, or epistemic phrase formation. Unless citing Urmson directly, I shall continue to use the terms 'main verb use' and 'modifier use'.

As additional support for his argument that certain verb uses are not part of the main assertion, Urmson points out that the nonasserted verb uses typically correlate with grammatically parenthetical forms, as illustrated with his examples in (10).

(10) (a) I suppose that your house is very old.
(b) Your house is, I suppose, very old.
(c) Your house is very old, I suppose.

Note that the nonasserted, modifier sense of *I suppose* is present in all three versions, including the evident use of *I suppose* as matrix subject + verb with overt *that* in (10a). The epistemic modifier sense of the verb is most clearly shown by epistemic phrase formation but is not confined to it. Urmson stresses this point: the verb can be used 'followed by *that* and an indicative clause, or else can be inserted at the middle or end

[2] Thompson & Mulac (1991b: 238) also cite Emonds (1969, 1976), Hooper & Thompson (1973), and Hooper (1975) on this point.

[3] Urmson identifies the 'main verb use' vs. 'modifier use' contrast in three separate domains: epistemic (*I think*, *I believe*); interpersonal/politeness (*I'm sorry*, *I regret*); and logical/discourse structural (*I deduce*, *I conclude*). In my view the distinction operates rather differently in the three domains. I shall discuss only the epistemic verbs in this article.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

of the indicative sentence, . . . I do not wish to imply that these verbs are parenthetical in any sense except that they are sometimes used parenthetically in a purely grammatical sense; beyond that "parenthetical" is merely a convenient label' (1963: 221). Explicitly, the fact that (10b) and (10c) have the same communicative force as the target sense of (10a) shows the 'parenthetical verb' sense (in my terms the modifier sense) of *I suppose* in (10a), but does not show that *I suppose* in (10a) is itself a reanalysed epistemic phrase.

Urmson also argues that the nonasserted, modifier senses of certain verb uses are paraphrasable by adverbials. He suggests (1963: 227–8) epistemic adverbs such as *presumably*, *certainly*, *undoubtedly*, *probably*, and *possibly* as paraphrases for sequences such as *I believe*, *I think*, *I know*, but notes that the adverbs do not capture the subjectivity of the subject–verb sequences. I note that overtly subject-oriented adverbials such as *in my opinion*, *to the best of my knowledge* may be better candidates here, and still make Urmson's point, that the targeted verb senses are modifier-like rather than part of the main assertion.

Given that the modifier use of an epistemic verb is not part of the main assertion, the subject + verb sequence can be omitted without altering the basic asserted statement. Urmson implies this in his comment: 'we can either make our statement "neat", . . . or I can employ the warning device of a parenthetical verb "I believe it will rain"' (1963: 227).

Urmson notes that the modifier sense of epistemic verbs is most commonly expressed with a first-person singular subject and a simple present tense verb. However, he also argues for interesting exceptions to the generalization concerning the person of the subject. Urmson's discussion on this point is fairly terse: I quote his remarks in full here and shall comment further below.

> (A) Jones says, 'X is, I believe, at home.' Here Jones makes an implied guarded claim (that is the effect of adding 'I believe') to the truth and also an implied claim to the reasonableness, of the statement 'X is at home.' This is the case already examined.
>
> (B) Smith, reporting Jones, says 'X is, Jones believes, at home.' This is *oratio obliqua*, reporting Jones' parenthetical use of the verb. Smith, by uttering the sentence, implies the truth and reasonableness of the statement that Jones has made the statement that X is at home (Jones thereby implying its truth and reasonableness with the conventional warning signal about the evidential situation).[4] (1963: 230)

I note first that in the local context of Urmson's discussion, 'x makes a claim to the truth and reasonableness of p' is an expanded version of 'x makes the statement that p' or 'x claims that p.' Urmson observes that by the usual conventions of discourse, to make a statement is also to imply the claim that the statement is both true and reasonable – cf. Grice's Co-operative Principle and the general Quality implicature that an asserted

[4] In this example and throughout his paper Urmson uses *I believe* as a frequent example of a modifier verb, though *I think* in this usage is far more common. Throughout the example under discussion here I have retained *believe*, but some readers may find the examples easier to judge with *think* instead – the point is not affected by the choice of verb.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

statement is implicated to be true (Grice 1975). With that proviso, Urmson's paragraph (B) above interprets Smith's utterance as asserting that Jones has *made the statement* that X is at home – explicitly, *X is at home* is interpreted as reported speech. In my view, Urmson's use of displaced epistemic phrases in these examples confuses the issue, and it is easier to consider the first point of interest with the nondisplaced word order as in (11).

(11) (a) Jones: I believe X is at home.
(b) Smith: Jones believes X is at home.
(c) Smith: Jones says he believes X is at home.

On the interpretation described by Urmson, (11b) represents a reported modifier use in the third person. A key point here is that Urmson interprets Smith's statement to mean that Jones *stated* that X was at home – in other words, (11b) is interpreted to mean something like 'Jones says that X is at home (to the best of his knowledge).' A more explicit form of this sense would be (11c), in which *he believes* is also a reported modifier form.

Now reconsider Urmson's original examples with displaced epistemic verbs, repeated for convenience in (12).

(12) (a) Jones: X is, I believe, at home.
(b) Smith: X is, Jones believes, at home.

With the sequence *Jones believes* overtly subordinated, the main clause of Smith's statement in (12b) is *X is at home*. In this format, contra Urmson, I consider that Smith's statement most naturally conveys that Smith himself asserts 'X is at home' and cites Jones as a supporting authority, in which case *Jones believes* is an epistemic modifier-like element.[5] The difference between the two interpretations is more evident if the utterances are placed in context, as sketched in (13)–(14).

(13) So what did Jones have to say?
(a) Jones believes X is at home.
(b) ?? X is, Jones believes, at home.
(c) ?? X is at home, Jones believes.

(14) Where's X?
(a) Jones believes X is at home.
(b) X is, Jones believes, at home.
(c) X is at home, Jones believes.

Note particularly the oddness of (13b,c) as a report of what Jones said, in contrast with the naturalness of the identical (14b,c) as a report on Jones' whereabouts. The utterances in (13a) and (14a) without displacement of *Jones believes* are equally natural in both senses. In (13a) *Jones believes* reports Jones' epistemic modification of his assertion

[5] Thompson & Mulac (1991b: 239) cite an unpublished paper by Underhill (1988) that makes the same point, apparently in reference to third-person examples. Underhill claims that zero appears where the writer asserts or endorses the embedded clause.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

that X is at home, and in (14a) *Jones believes* modifies the speaker's assertion that X is at home, appealing to Jones as a reliable source of information.

Thus there are two kinds of modifier interpretations with third person subjects. First, there is the use reporting a modified utterance, as primarily discussed by Urmson, in which 'Jones believes X is at home' means that Jones *said* 'I believe X is at home', and *I believe* was a modifier in Jones' utterance. The reported modifier use also appears with the verb of saying included, as in 'Jones said he believed X was at home.' Second, there is a sense which is not primarily interpreted to assert that the subject of the epistemic verb made a hedged statement. Rather, the speaker makes a statement ('X is at home'), and appends to it an epistemic/evidential modifier which invokes the authority or sagacity of a third party (Jones always knows where X is).

Urmson also points out (1963: 222) that the modifier use of epistemic verbs is usually expressed with the simple present tense. I note, however, that the reported speech modifier use also appears in the past tense, as illustrated in (15). The past tense on *thought* and *was* in (15) can be construed as formally dependent on the past tense of *said*, but with present time reference (i.e. sequence of tense rule).

(15) What did Jones have to say?
Jones (said he) thought X was at home.

Finally, Urmson comments in passing that modifier uses are compatible with matrix adverbials, giving the examples *I rather suspect* and *I strongly believe*.

Main verb uses of epistemic verbs, called by Urmson 'mental verbs', express part of the main assertion of the utterance and predicate mental states or processes of their subject. Urmson's main project is to argue for the existence of modifier uses, taking main verb uses more or less for granted as already familiar. He has less to say about main verb uses in general, and concentrates on the main verb uses corresponding to what I have called interpersonal/politeness modifiers, such as *I'm sorry*, *I regret* (see footnote 3). The two chief clues to main verb uses implicit in his discussion contrast directly with identified characteristics of modifier uses. With a first-person main verb use, in contrast to a modifier use, the associated clause is *not* asserted by the speaker: this may be overtly signalled as in (16) (my example).

(16) I thought the key was lost forever but it was down the side of the sofa.

Urmson illustrates a third-person main verb use (1963: 230–1) with a scenario in which, during a railway stoppage, Smith sees Jones rushing towards the station and utters (17).

(17) Jones believes that the trains are working.

In this context Smith's utterance attributes a belief to Jones which Smith infers from observing Jones' behaviour: unlike the reported speech modifier use discussed above, the utterance is not interpreted to mean that Jones made a qualified statement that the trains were working. Urmson notes that the main verb use here is further signalled by the impossibility of a syntactically parenthetical use, as (18) cannot be substituted for (17) in the same context and with the same intended sense.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Table 1. *Signs of modifier and main verb uses of first-person epistemic verbs*

	Modifier use	Main verb use
1. Associated clause is asserted by the speaker	yes	no (or not directly[a])
2. Epistemic verb + subject can be omitted without altering the propositional content of the main assertion	yes	no
3. Epistemic verb + subject can be paraphrased with an epistemic adverb	yes	no
4. Epistemic verb + subject has corresponding clause-medial and/or clause-final forms	yes	no
5. Epistemic verb + subject determines the form of the corresponding tag question	no	yes

[a] An apparent main verb use of an epistemic verb may carry an implicature that the speaker would assert the associated clause, as in (i).

(i) Those closed curtains make me think we've come to the wrong house.

The speaker's utterance serves to put forward the proposition 'We've come to the wrong house' for consideration, but the form of (i) doesn't fit the usual criteria for a modifier use. The subject and verb cannot be omitted (without also omitting *those closed curtains make me*), cannot be paraphrased with an epistemic adverbial, and cannot be moved to medial or final position. The matrix *Those closed curtains make* expresses the cause of the expressed belief rather than the cause of coming to the wrong house. Accordingly, the verb is used in a main verb sense, but the embedded clause may be said to be (weakly) asserted by implicature.

(18) The trains, Jones believes, are working.

An additional clue to the main/modifier distinction is provided by Halliday (1985: 332–3), who observes that the modifier use does not determine the form of a tag question, as in (19); cf. the main verb use in (20).

(19) I think the trains are working, aren't they?/*don't I?
(20) Jones thinks the trains are working, doesn't he?/*aren't they?

The tag *aren't they* in (19) is based on the subject and verb of the associated clause which expresses the main assertion. In contrast, the tag question in (20) is formed on the asserted main verb use of *think*.

We have seen that the person of the subject, the presence of adverbial or auxiliary modification of the epistemic verb, and the tense of the epistemic verb are not in fact criterial for the modifier/main verb distinction. The remaining indicators of the distinction for any particular instance of an epistemic verb are based on intuitions concerning the communicative force of the utterance and possible related forms. These indicators for first-person modifiers are summarized in table 1.

Three different kinds of third-person modifier use were identified above. The two types of interest in the present study are the reported speech modifiers, with or without

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

an overt verb of saying. These types behave slightly differently according to the criteria in table 1 (or corresponding criteria), as I review below. I begin with the nonembedded reported speech modifier, as illustrated in (21).

(21) Simple reported speech modifier use (not embedded under verb of saying)
(a) He thinks the team have improved/He thought the team had improved.
(b) He said 'I think the team have improved.'

In simple reported speech the associated clause is understood to have been stated, with an epistemic hedge, by the subject of the epistemic verb. This interpretation is explicit in (21b). An additional sign of this interpretation is that the sentence can be paraphrased as a reported statement: *He says the team has improved*. The matrix (*he thinks/he thought*) stands in for the understood verb of saying and also expresses the reported speaker's epistemic modifier. This gives the matrix a portmanteau quality, as illustrated in (22), where the dual content represented by *he thinks/he thought* in (21) is italicized.

(22) *He said* the team had improved, *in his opinion*.

The content 'he said', attributing the assertion of the associated clause to the subject of the epistemic verb, is part of the speaker's main assertion, and so the sequence *he thinks/he thought* in (21a), which conveys this content, cannot be omitted without altering the main assertion. For the same reason the sequence *he thinks/he thought* in (21a) cannot be replaced with an epistemic adverbial without slightly altering the reported speech sense. This is illustrated with (23). Although we might infer from (23) that the opinion-holder had expressed his opinion at some point, (23) itself does not have the reported-speech interpretation as identified by Urmson.

(23) The team have improved, in his opinion.

As I discussed above with reference to examples (12)–(14), I consider (contra Urmson) that the reported-speech interpretation of *he thinks/he thought* does not occur with clause-medial and clause-final placement of the corresponding epistemic verb and its subject, as illustrated in (24).

(24) The team have improved, he thinks.

I consider that examples like (24), where the associated clause is unambiguously a root clause, are more saliently interpreted as assertions of the associated clause by the speaker. This interpretation appears mostly with clause-medial and clause-final epistemic verb phrases, which are not included in the present study.

Finally, I consider that the reported-speech modifier use does not determine the form of a tag question, as in (25).

(25) He thinks the team have improved, doesn't he?

In (25) the tag question forces the main verb sense of *think* – cf. 'It is his opinion that the team has improved, isn't it?'

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Table 2. *Characteristics of third-person epistemic verbs expressing epistemic modifiers in reported speech*

	Nonembedded reported epistemic modifier	Reported epistemic modifier embedded under verb of saying
1. Utterance is interpreted as reporting a statement made by the subject of the epistemic verb	yes	yes
2. Epistemic verb + subject can be omitted without altering the propositional content of the main assertion	no	yes
3. Epistemic verb + subject can be paraphrased with an epistemic adverb	no	yes
4. Epistemic verb + subject has corresponding clause-medial and/or clause-final forms	no	yes
5. Epistemic verb + subject determines the form of the corresponding tag question	no	no

The reported-speech modifier which is embedded under a verb of saying, as in (26), is less complex, because the sequence *he thinks/he thought* expresses only the reported epistemic modifier, and not the attributed speech.

(26) (a) He says he thinks the team have improved.
(b) He said he thought the team had improved.

Accordingly, this type of third-person modifier behaves more or less like the first-person type, as illustrated in (27). The epistemic modifier can be omitted without altering the propositional content, as in (27a), and paraphrased with an epistemic adverbial, as in (27b). The modifier corresponds to clause-final phrases, as in (27c). Although a tag question based on the same form as the modifier is well-formed, as in (27d), the form of the tag is actually determined by the matrix *he says* – as has been observed for first-person modifiers, this type of third-person modifier does not determine the form of a tag question.

(27) (a) He says the team have improved.
(b) He says the team have improved, in his opinion.
He says the team have surely improved.
(c) He says the team have improved, he thinks.
He said the team had improved, he thought.
(d) He says he thinks the team have improved, doesn't he?

The characteristics of epistemic modifiers in reported speech are summarized in table 2.

The properties of main verb and modifier uses of epistemic verbs reviewed here will also be discussed further in section 6, which describes the preparation of the epistemic verbs corpus for the present study.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

3 Clause-medial and clause-final epistemic phrases

Thompson & Mulac argue that the displacement of *I think*-like sequences to noninitial positions demonstrates that such sequences have been reanalysed to dependent elements, and thus have the variable position which is typical of similar epistemic adverbials such as *maybe*. The force of the argument rests on the assumption that clause-medial and clause-final *I think* (and similar adverbials) are derived from clause-initial sequences.

This derivation is not uncontested. Brinton (1996: chapter 8), examining data from Chaucer, shows that there is little support for a correlation between clause-medial and clause-final epistemic parentheticals and corresponding clause-initial sequences. First, the expected correlations in frequency between zero-marked complements to particular verbs and medial and final parenthetical uses of the same verbs are not generally shown. The most frequent parenthetical verb *guess* has only one zero-marked complement in the data (although *trow* and *woot*, the next most frequent verbs in parentheticals, have high rates of zero-marked complements). The verbs *leve*, *suppose*, and *undertake* appear in noninitial parentheticals, but do not appear with zero-marked complements at all. Second, Brinton argues that medial and final parentheticals are more likely to have developed from existing parenthetical clauses in Middle English, such as *as I guess* or *I trowe so*. Brinton proposes that the final (and medial) forms *I guess*, *I trowe* are derived by dropping the relative element referring to the content of the main clause (*as* or *so*).

I note also that Thompson & Mulac's grammaticization account of *I think*, *I guess*, is not easily extended to the range and complexity of noninitial epistemic parenthetical phrases, which do not show much formal restriction, as illustrated below.

(28) This government, *you could be forgiven for thinking*, has completely lost the plot and is hell-bent on building some kind of Plato's Republic.
(29) The only way out *I believe with all my heart and soul* is to plant the entire country in native forest.

Further evidence bearing on the source of clause-medial and clause-final epistemic phrases comes from so-called Neg-Raising constructions, as illustrated in (30). The central characteristic of Neg-Raising is that a negation expressed in the matrix clause, as in (30a), is construed with the embedded clause. The resulting interpretation of (30a) is much the same as the modifier use of an epistemic verb associated with a negated clause, as in (30b).

(30) (a) I don't think it will matter.
(b) I think it won't matter.

At first sight, *I don't think* does not fall under Urmson's criteria for modifier uses because the literal sense of the associated clause *it will matter* is explicitly denied by the speaker, not asserted. But on the Neg-Raising interpretation the associated clause includes the negation, and the negated proposition *it won't matter* is in fact asserted by the speaker.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Halliday (1985: 333–40) includes both *I think* and *I don't think* as instances of epistemic modification (what he calls 'metaphors of modality'), observing that both positive and negative variants share the characteristic of not determining the form of a tag question. Studies of Neg-Raising (see, for example, Prince 1976 and Horn 1989: 308–30, and references cited in both works) discuss the role of Neg-Raising as a device for hedging the force of a negative assertion. In other words, Neg-Raising has been categorized as an instance of what is here called the modifier use of epistemic verbs. In what follows I shall assume that modifier uses of epistemic verbs also include instances with negated verbs on a Neg-Raising interpretation.[6]

Recent studies concerning epistemic phrase formation have not addressed the Neg-Raising phenomenon directly, although the possible effect of negation is considered. Thompson & Mulac (1991b) hypothesised that matrix negation would behave like other elements, such as matrix auxiliaries and adverbials, in contributing to the overall complexity of the matrix clause, and thus making EP formation less likely. This effect was expected to be reflected by a greater occurrence of *that* with additional matrix components, including negation. However, Thompson & Mulac (1991b: 245) found that negation did not correlate with a higher rate of *that*, unlike matrix auxiliaries and adverbials, which did.[7] This finding is consistent with classing Neg-Raising as compatible with modifier uses.

Now consider the clause-final variants of modifier *I don't think*, as in (31).

(31) (a) I don't think it will matter.
(b) *It will matter I don't think.[8]
(c) It won't matter I don't think.[9]

[6] Prince (1976: 425) draws a strong connection between Neg-Raising and first-person, nonprogressive uses of epistemic verbs used metaphorically. This suggests that Neg-Raising is confined to the expression of epistemic modification. Horn (1989: 314) shows that the Neg-Raising interpretation also occurs in other contexts, with examples including those below. Note that (i) contains a nonmodifier use of the epistemic verb *suppose*.

(i) Nobody supposes that nuclear war is winnable.
'Everybody supposes that nuclear war is not winnable.'
(ii) Neither of them is likely to marry you.
'Either of them is unlikely to marry you.'

I conclude that Neg-Raising occurs with modifier uses of epistemic verbs but is not confined to those contexts. Consequently the presence of Neg-Raising is not diagnostic of a modifier use.

[7] Tagliamonte & Smith (2005) also hypothesized that matrix negation correlates with a higher incidence of *that*, citing the examples reproduced below (their (14), p. 295).

(i) And I couldnae believe *that* you have a granddauchter that age.
(ii) And I believe ∅ he was famous, like.

Tagliamonte & Smith found a weak but significant correlation between *that* and additional elements in the matrix verb phrase, including negation, but results for negation separately were not reported (2005: 301, 304).

[8] Example (31b) is not an instance of the ironical . . . *not!* idiom illustrated in (i), where the delayed *not* is understood to negate the preceding positive statement.

(i) Yeah, we all really love group work . . . not!

[9] Note also that displacement of *I don't think* from initial position cannot straightforwardly be the source for final *I don't think* in (31c), as *I don't think it won't matter* has a different meaning.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

I point out that *I don't think* in initial position in (31a) is required to express the negation construed with the associated clause. As (31b) shows, there is no corresponding version of (31a) with the modifier epistemic verb in clause-final position. The negated epistemic modifier can appear clause-finally, but the associated clause must be independently negated, as in (31c). This shows that clause-initial epistemic modifier *I don't think* cannot be the source (via rebracketing and repositioning) of clause-final *I don't think*.

I conclude that clause-medial and clause-final epistemic phrases are not necessarily derived from reanalysed matrix clauses, and the presence of medial and final epistemic phrases does not necessarily constitute evidence that the corresponding clause-initial subject–verb sequence is a syntactically dependent element, even where the complementizer is null. Rather, the relationship is simply paraphrase, as assumed in section 2 above. A clause-medial or clause-final epistemic phrase which corresponds in form with a clause-initial subject–verb sequence has the same significance as paraphrasing with *maybe* or *in my opinion* – both types of paraphrase serve to show that the (clause-initial) matrix sequence has an epistemic modifier sense.

4 Summary of preliminary matters

I began by reviewing Thompson & Mulac's strong hypothesis concerning high zero rates with epistemic verbs, notably *I think*. In that account, zero complementizer is largely a consequence of syntactic reanalysis from a matrix subject and verb with a clausal complement to an adverbial-like element within the associated clause: this change constitutes a type of grammaticalization.

I have reviewed certain problems with the account. First, assuming that *I think* and similar sequences have in fact undergone reanalysis, the reanalysis itself is unlikely to be a major contributor to high zero rates, as zero should appear in the input to reanalysis. The most persuasive evidence in support of reanalysis – the presence of corresponding forms in clause-medial and clause-final position – is debatable, as (i) there is evidence that clause-medial and clause-final epistemic phrases developed from existing parenthetical clauses rather than from matrix clauses, and (ii) negated clause-medial and clause-final epistemic phrases (*I don't think*) cannot be derived by simply moving a negated matrix. I also argued that the reanalysis account for *I think* and *I guess*, identified as a type of grammaticalization, is supported by the high frequency and fixed form of these two collocations, but does not easily extend to other modifier-like uses of less frequent epistemic verbs which also have corresponding clause-medial and clause-final forms (*I truly believe*, *I strongly feel*, etc.).

I have reviewed the properties proposed by Thompson & Mulac to be conditions for the expression of epistemic modification, mainly a first/second-person subject and a simple matrix, and drawing on Urmson's discussion, have shown that epistemic modification may also be expressed with third person subjects and matrix modification. Another property which has been observed to be typical of the modifier use of epistemic verbs is the simple present tense (Urmson 1963: 222), although as we saw in section 2,

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Table 3. *Canonical and required properties of epistemic modifier senses*

	Observed properties of *I think*, *I guess*	Required conditions for epistemic modifier sense of other verbs?
1st/2nd person subject	yes	no
simple matrix	yes	no
highly frequent verb	yes	no
simple present tense	yes	no

epistemic modification in reported speech contexts may also be in the past tense. The differences between canonical and required properties is summarized in table 3.

From these preliminary considerations, I conclude that fixed and frequent forms such as *I think* and *I guess* represent a further stage of development beyond the expression of epistemic modification in general. I assume provisionally a conservative view, that these highly frequent forms in construction with zero complementizer have become routines, but that this need not entail any syntactic reanalysis. Assuming (as above) that syntactic reanalysis would not explain the presence of zero, and probably is not the source of clause-medial and clause-final modifiers, syntactic reanalysis in addition to routinization is unmotivated.

I propose that the high zero rate with epistemic verbs is triggered (at least in part) by their use with modifier senses, in contrast to main verb senses. The study reported in this article investigates the hypotheses in (32).

(32) *Hypothesis 1:* The modifier sense of an epistemic verb and its subject in matrix position promotes zero in the complement clause.
Hypothesis 2: Modifier *I think* (and *I guess*) in construction with zero complementizer has become a fixed routine, resulting in higher zero rates than other modifier uses of epistemic verbs.

Apart from the factors specific to epistemic verbs at issue in this study, a number of general factors in *that*/zero choice have been identified in previous studies, and must be taken into consideration. I turn to these issues in the next section.

5 General factors in *that*/zero choice

Before describing the corpus of epistemic verbs, in this section I review more general factors for the *that*/zero choice in clausal complements to verbs. I also introduce the main corpus on which the present study is based, and report the results of preliminary investigations of the general factors discussed here. Only some of the proposed factors are found to be significant and carried forward into the epistemic verbs study.

The full corpus on which the present study is based comprises 8,000 tokens of embedded finite clauses, 1,000 tokens from each of the following broadsheet newspapers: the *Boston Globe* (US), the *Los Angeles Daily Times* (US), *The Age* (Australia),

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

the *Daily Telegraph* (Australia), the *Guardian* (UK), the *Daily Telegraph* (UK), the *Christchurch Press* (NZ) and the *Sunday Star Times* (NZ), in the period 2002–4. All the embedded finite clauses found were collected, except for relative clauses and clauses with *wh*-gaps. For each embedded clause, the full containing sentence was collected. The data collection and initial coding were carried out by research assistants, and all codes were checked by the author. Of the tokens collected, the final corpus contained 5,735 tokens of complements to verbs, including 2,178 complements to *say*. The very high overall zero rate of 90 per cent in complements to *say* tended to override the possible effects of specific factors in zero rates. Given the large number of *say* tokens, it was decided to remove them from the preliminary investigations reported here, to exclude interference from this particular verb. The remaining verb complement corpus comprised 3,557 tokens.

Influences on the rate of zero in complements to verbs include genre, style and channel, elements intervening between the matrix verb and the embedded subject, the form of the embedded subject, and regional variation. I shall discuss these factors in turn, and report on how they are taken into account in the present study, if found to be significant.

5.1 Style, genre, and channel

It is generally proposed that less formal texts have a higher rate of zero than more formal texts (Bolinger 1972: 22; Burchfield 1996: 773, para 2; Elsness 1984: 519; Finegan & Biber 1995: 256, fn. 6; Huddleston & Pullum 2002: 953; Poutsma 1929: 615; Quirk et al. 1985: 734–5; Storms 1966: 262; see also Kroch & Small 1978 for *that*/zero choice and prescriptivism). This observation is linked to a number of different ways of identifying formality, and specific corpus-based findings are often more precisely framed in terms of genre and channel. It is consistently found that conversational language has a higher zero rate than written language in general (Biber et al. 1999: 680; Bolinger 1972: 14; Finegan & Biber 1995: 242; among others), and in this case the difference in channel is considered to correlate with a difference in formality – conversation is overall less formal than written texts.

Differences in zero rate have also been reported in different text genres, with academic prose having the lowest zero rate, light fiction a fairly high zero rate, and newspaper reportage a rate intermediate between the two (Biber et al. 1999: 680–1; Elsness 1984; Finegan & Biber 1995: 242; Fowler 1926: 632; Jespersen 1932: 33). Again, these differences among genres correlate to some extent with different levels of formality.

The data for the present study were taken from a single major genre of broadsheet newspaper prose. Data were further identified for three textual subtypes: quoted direct speech, general reportage and columns, and letters to the editor. These textual subtypes were found to have different overall rates of zero:

quoted speech	65%
news and columns	58%
letters	27%

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

The relative zero frequencies here may also reflect differences in formality. Quoted speech transcribed from interviews reflects the generally more informal nature of spoken language, and is slightly less formal than general written reportage. Letters to the editor commonly reflect the writer's awareness of a kind of public performance, and are usually serious in tone.

The factor of textual subtype was included in the epistemic verbs study, as discussed in section 6.

5.2 Intervening elements

It is generally observed that elements in the matrix clause which intervene between the matrix verb and the embedded subject tend to inhibit zero (Bolinger 1972; Elsness 1984; Finegan & Biber 1995; Hawkins 2003; Jespersen 1932; McDavid 1964; Rissanen 1991; Rohdenburg 1999; Tagliamonte & Smith 2005; Thompson & Mulac 1991a). The main intervening elements the authors discuss are indirect objects and matrix adverbials, as illustrated in (33)–(34).

(33) He assured them (that) the damage would be made good.
(34) She discovered to her surprise (that) the basement was under water.

Elements intervening between the matrix verb and the embedded subject also include initial adverbials in the embedded clause, as in (35).

(35) He also admitted (that) had he opened negotiations with BMW sooner, the price would have been cheaper.

In the current corpus, verb complements separated from the matrix verb by an intervening adverbial were not included in the verb complement subcorpus, being separately coded as extraposed clauses. Consequently no data of this type were included in the investigations reported in this section, nor in the investigation of epistemic verbs.

It was found that verb complements preceded by an indirect object ($n = 294$) had an overall zero rate of 47 per cent, in comparison with the 57 per cent zero rate in complements to all other verbs other than *say* ($n = 3263$). The difference was significant (Pearson's GFX $p \leq 0.0011$). The issue did not arise in the epistemic verbs study, as none of the verbs studied take indirect objects.

Finally, embedded clauses with a clause-initial adverbial were compared with clauses in which the subject is the first element of the embedded clause. (This comparison was carried out using only nonextraposed complements to verbs other than *say*, and without indirect objects.) The clause-initial adverbial was found to inhibit zero significantly ($p < 0.0001$, Pearson's GFX):

clause-initial adverbial	8% zero
clause-initial subject	55% zero

Embedded clauses with a clause-initial adverbial constituted only 5 per cent of the data, and given the small number of tokens, it would be preferable to exclude them

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

from a multifactorial analysis. Accordingly, tokens with a clause-initial adverbial in the embedded clause were excluded from the analysis of epistemic verbs, and the zero-inhibiting effect of these adverbials is excluded.

5.3 The embedded subject

Characteristics of the embedded subject are invariably found to have an effect on *that*/zero distribution. A number of authors have found that the form of the embedded subject has a marked effect on *that*/zero choice (among others Elsness 1984; Finegan & Biber 1995; Hawkins 2003; Rissanen 1991; Rohdenburg 1999; Roland et al. 2005; Thompson & Mulac 1991a). The commonest finding is that the rate of zero is increased by a pronominal embedded subject, in contrast to a full noun phrase.[10] It has also been found that the length of the embedded subject is relevant, and a long NP inhibits zero more than a short NP (Elsness 1984; Hawkins 2003; Roland et al. 2005).

Some researchers propose that differences are found among different pronoun subjects. Hawkins (2003: 191) suggests that a subject with nominative case marking assists in parsing the onset of an embedded clause, and so reduces the likelihood of *that* more than a non-case-marked pronoun. Elsness (1984: 526) finds a higher rate of zero where the embedded subject is coreferential with the matrix subject, and that first- and second-person embedded subjects produce a higher zero rate than third-person embedded subjects.

Thompson & Mulac (1991b) suggest that a pronoun subject promotes zero because pronouns express definite reference and mark topics, thus increasing the overall topicality of the embedded clause. The proposed relevance of topicality is consistent with Elsness' (1984) observation that definite embedded subjects occur with a higher zero rate than indefinites, and also with his finding that an embedded subject which is coreferential with the matrix subject favours zero. Ferreira & Dell (2000) also found that zero was more likely to occur in speech production where the embedded subject was repeated from previous discourse or prompted by a cue to recall – both of these characteristics suggest a current or revived topic as referent.[11]

Proposed embedded-subject factors in *that*/zero choice are listed in (36).

(36) (a) A pronoun embedded subject promotes zero in comparison with a full NP.
(b) A shorter NP subject promotes zero in comparison with a longer NP subject.
(c) A subject marked with morphological nominative case may promote zero in comparison with a non-case-marked subject.
(d) A topical or discourse-familiar embedded subject promotes zero in comparison with a nontopical subject.
(e) A first- or second-person subject promotes zero in comparison with a third-person subject.

[10] A pronominal subject also promotes zero in relative clauses (Biber et al. 1999: 620; Jaeger & Wasow 2005; Race & McDonald 2003).

[11] Jaeger & Wasow (2005) also found that the likelihood of zero in relative clauses increases where the referent of the embedded subject is salient or given in the discourse context.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

A preliminary examination of the potential factors listed above was carried out in nonextraposed complements to verbs (excluding complements to *say*, complements preceded by indirect objects, and complements with clause-initial adverbials). I review the results briefly here, and show that only one of the proposed distinctions was evident in the present data.

5.3.1 Length of the embedded subject

A preliminary calculation of the relative zero rates with embedded subject pronouns, short NPs (one or two words, not including pronouns), and long NPs (three or more words) revealed significant differences as shown below, including a significant difference between short and long NPs ($p < 0.0001$, Pearson's GFX).

pronoun embedded subject	73% zero
short NP embedded subject	51% zero
long NP embedded subject	38% zero

Accordingly, the length of the embedded NP was retained as a factor in the epistemic verbs analysis.

5.3.2 Nominative case

Nominative embedded subjects (*he*, *she*, *we*, *they*) were compared with non-case-marked pronominal subjects (*you*, *it*, *this*).[12] The nominative subjects ($n = 626$) occurred with a 76 per cent zero rate, slightly *lower* than the non-nominative subjects ($n = 319$) zero rate of 80 per cent, although the difference was not statistically significant (Pearson's GFX $p \leq 0.2341$). Roland et al. (2005: 25) report a similar finding. The presence or absence of nominative case marking on pronominal embedded subjects was not retained as a factor in the epistemic verbs analysis.

5.3.3 Topicality

The general issue of topicality was tested in two ways. First, embedded subject pronouns *he* and *she* were coded for coreferentiality with the matrix subject. Coreferential *he*/*she* ($n = 166$) and non-coreferential *he*/*she* ($n = 123$) occurred with the same zero rate of 69 per cent. This is consistent with results from Finegan & Biber (1995: 254–5) and Torres Cacoullos & Walker (2003), who also found that coreferentiality does not affect zero rates.

Second, the potential relevance of definite reference or topicality was also tested by comparing zero rates with embedded subjects anaphoric (referential) *it* and dummy (nonreferential or cataphoric) *it*. Examples from *The Age* (Australia) are given in (37)–(40).

(37) Meanwhile, in the Federal Court yesterday, the Australian Securities and Investments Commission insisted that *it* did not oppose the administrators' decision to let Ansett

[12] The first person was excluded from this comparison in order to minimize possible interference from *I think*-like routines, on the assumption that first-person embedded subjects may be more likely to occur with these routines in comparison with other types of matrix. Subject *there* was also excluded from this comparison as it is a recognizable subject form despite not having morphological case.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

trade while they settled the Tesna deal even though Ansett was suffering hefty losses. 09.02.02. (*it* refers to the ASIC)

(38) Air New Zealand said yesterday it was prepared for a 'Qantas attack' on its domestic market after the Australian international airline confirmed *0 it* had plans to expand in New Zealand. 05.02.02. (*it* refers to Qantas)

(39) Sports hack Stuart (Iain Glen) and financial whiz Cara (Sharon Small) realise that *it*'s not just the newspapers' future that is at stake. 06.02.02. (nonreferential or cataphoric *it*)

(40) The parents defied the ban, insisting *0 it* was their religious obligation and personal right to maintain the modest dress code, generally applied after puberty, for their young daughters. 06.02.02. (nonreferential or cataphoric *it*)

Referential *it* (n = 249) occurred with 78 per cent zero, slightly higher than the 71 per cent zero rate with dummy *it* (n = 79), although the difference was not significant ($p \leq 0.2012$, Pearson's GFX). Topicality of the embedded subject was not retained as a factor in the epistemic verbs analysis.

5.3.4 Person features of the embedded subject

The possible effect of the person of the embedded subject was tested. Embedded first- and second-person subjects (*I*, *we*, *you*: n = 66) were compared with all third-person pronominal subjects (n = 358).[13] Both groups of pronouns had the same zero rate of 70 per cent. I note that Elsness' results on the effect of the person features of the embedded subject do not explicitly distinguish between third-person pronouns and third-person noun phrases. Consequently, the zero-promoting effect of *I*, *we*, *you* observed by Elsness may be due to the independently established effect of embedded subject pronouns in comparison with full noun phrases, assuming that noun phrases were included in his third-person subjects. The person of the embedded subject was not retained as a factor in the final analysis.[14]

Overall, none of the potential differences among pronominal embedded subjects was found to significantly affect the rate of zero complementizer – the main embedded subject effect, as discussed above, is the length of the embedded subject.

5.4 Regional variation

Some regional variation has been observed in the distribution of null complementizer in the presence of *wh*-gaps (e.g. Henry 1995 on Irish English), but regional variation in *that*/zero choice in other contexts has been little studied.[15] Staum (2005) found that *I think* occurred with a lower zero rate among speakers of the northeast United

[13] Despite the possible interference from *I think*-like routines mentioned in footnote 12, it was not possible to exclude *I* subjects for this test, as they constituted the majority of the first-person tokens.

[14] Although the person features of the embedded subject were not retained in the epistemic verbs study, the person features of the matrix subject were coded, as described in section 6.

[15] Biber et al. report on discourse and register factors affecting the frequency of complementizer and relativizer *that*, but do not report any differences between British and American English, although varietal differences are reported for other aspects of usage. It isn't clear whether no differences were found or the issue was not examined. See now Kearns (to appear).

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

States compared with other American speakers, but she notes that this may be due to interaction with other factors.

Unlike those in many existing studies, the present corpus is drawn from twenty-first-century publications and includes Australasian English as well as American and British English. The zero rates in complements to verbs (excluding *say*, extraposed complements, indirect objects, and complements with clause-initial adverbials) were New Zealand 73 per cent, Australia 63 per cent, United States 48 per cent, and Great Britain 37 per cent. All differences were significant (Pearson's GFX $p \leq 0.0005$). The regional source was included as a factor in the epistemic verbs analysis.

6 The epistemic verb corpus

Data for the epistemic verb corpus were taken from the verb complement subcorpus as introduced in section 5. The tokens had no elements intervening between the matrix verb and the embedded subject, and no extraposed clauses. The three most frequent epistemic verbs with both main verb and modifier uses were selected: these were *think* (n = 481), *believe* (n = 205), and *feel* (n = 65), making a total of 751 tokens.[16] Codes carried forward from the main corpus were regional source (New Zealand, Australia, United States, Great Britain); textual subtype (quoted speech, letters to the editor, general news & columns); and the form of the embedded subject (pronoun, short NP of 1–2 words, long NP of 3+ words). The data were also coded for the epistemic verb lexeme (*think*, *believe*, *feel*), the person of the epistemic verb subject, and modifier and main verb uses.

The epistemic verb subject was coded as first/second person or third person. First-person subjects included 15 tokens of *we*. Of the 20 *you* subjects, 17 had an impersonal referent rather than a true addressee (cf. impersonal *one*), and were coded as third person. This type of *you* subject is illustrated in (41a,b). In comparison, the *you* subject in (41c) had an individual identified referent and was coded as first/second person.

(41) (a) After that plot summary, you may feel *0* you could picture the whole video, especially as it's set in Mexico, where Tom Cruise went in Losin' It, as did so many others in the dozens of US clones before and since. (*Guardian*, 11.10.02)
(b) You'd think *0* this BCS freak of nature would offer a wild, high-scoring affair, but it's hard to get a real feel for what's going to happen. (*Los Angeles Daily News*, 02.01.03)
(c) Thanks Julietta, I'm glad you think *0* I'm as fascinating as I do. (*Daily Telegraph* (Aust.), 17.01.04)

Modifier and main verb uses were coded according to the criteria identified in section 2 and summarized in tables 1 and 2. The application of these criteria to selected data is illustrated here, beginning with modifier uses.

[16] In contrast to Thompson & Mulac's study of American conversational data, (*I*) *guess* was rare (n = 2). Tagliamonte & Smith's (2005) study of British dialectal spoken data also did not show frequent (*I*) *guess*. I suggest that this is because *I guess* is most frequent in conversation, and more frequent in American English than British English.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

A straightforward example of modifier *I feel* is illustrated in (42a). The sequence *I feel* can be omitted without altering the statement made (42b); *I feel* can be paraphrased by an epistemic adverbial with roughly the same sense as the original (42c); and a clause-final epistemic phrase *I feel* has the same sense (42d). Recall that the possibility of a corresponding clause-final epistemic phrase is not considered to be evidence that clause-initial *I feel* has undergone syntactic reanalysis – rather, the clause-final phrase is unambiguously a modifier use, and the modifier sense of *I feel* in (42a) is demonstrated by the fact that (42a) and (42d) have the same sense.

(42) (a) With Sydney facing the worst bushfire crisis in 30 years I *feel 0* the New Year's Eve fireworks should have been postponed. (*Daily Telegraph* (Aust.), 01.01.02)
(b) With Sydney facing the worst bushfire crisis in 30 years the New Year's Eve fireworks should have been postponed.
(c) With Sydney facing the worst bushfire crisis in 30 years *surely/in my opinion* the New Year's Eve fireworks should have been postponed.
(d) With Sydney facing the worst bushfire crisis in 30 years the New Year's Eve fireworks should have been postponed, I feel.

Modifier *I think* with Neg-Raising is illustrated in (43). With the negation placed in the associated clause, the signs of modifier use correspond to those in (42). The modifier epistemic verb can be omitted as in (43b), paraphrased with an epistemic adverbial as in (43c), or replaced with a corresponding clause-final epistemic phrase as in (43d). As discussed in section 3 above, a negated clause-final epistemic phrase must appear with an independently negated main clause.

(43) (a) 'But while the stock is very cheap at the moment, I don't think *0* the company's problems are going to go away quickly,' Kraft said. (*Los Angeles Daily News*, 27.11.02)
(b) 'But while the stock is very cheap at the moment, the company's problems are not going to go away quickly,' Kraft said.
(c) 'But while the stock is very cheap at the moment, (in my opinion) the company's problems are (surely) not going to go away quickly,' Kraft said.
(d) 'But while the stock is very cheap at the moment, the company's problems are not going to go away quickly, I don't think,' Kraft said.

The simplest type of reported modifier use appears embedded under a verb of saying, as in (44a). The main signs of modifier use are the same as for first-person modifiers: the modifier verb can be omitted (44b), paraphrased with an epistemic adverbial (44c), or replaced with a corresponding clause-final epistemic phrase (44d).

(44) (a) National Party backbencher Paul Neville said he *believed 0* the long-standing convention should be preserved. (*The Age*, 05.02.02)
(b) National Party backbencher Paul Neville said the long-standing convention should be preserved.
(c) National Party backbencher Paul Neville said surely/in his opinion the long-standing convention should be preserved.
(d) National Party backbencher Paul Neville said the long-standing convention should be preserved, he believed.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

A reported epistemic phrase with Neg-Raising (45a) shows the signs of modifier use in the same way as the first-person modifier illustrated in (43). With the negation placed in the associated clause, the modifier verb can be omitted (45b), paraphrased with an epistemic adverbial (45c), or replaced with a corresponding clause-final epistemic phrase.

(45) (a) The station's morning shift host, Derryn Hinch, said he did not *think 0* Mr Kennett was breakfast material. (*The Age*, 05.02.02)
(b) The station's morning shift host, Derryn Hinch, said Mr Kennett was not breakfast material.
(c) The station's morning shift host, Derryn Hinch, said Mr Kennett was not breakfast material in his opinion.
(d) The station's morning shift host, Derryn Hinch, said Mr Kennett was not breakfast material, he thought/he didn't think.

Finally, a reported speech modifier epistemic verb can be used without a verb of saying in a sentence which reports a statement made by the subject of the epistemic verb. This usage appears in news reports based on interviews, and would appear to be a strategy for varying the style of presenting a sequence of reported statements. Potential tokens of this kind were provisionally identified by detailed attribution, and checked by examining the wider context, as illustrated in (46) below. The second paragraph of the extract reports an interview with AMP fund manager Michael Evans. The use of direct quotation to report some of his views indicates that the report has been prepared from a tape recording. In this context we are to infer that the clause following *believes* reports content asserted by Evans.

(46) Sydney – Dramatic share price declines in ResMed, Cochlear and CSL for the first time in four years have severely dented the stellar performance of Australia's health-care sector. One of the market's best-performing industries, with a gain of 20 per cent last year, the sector has fallen almost 7 per cent since the start of this year. Over the same period, ResMed has fallen 25 per cent, CSL 10 per cent and Cochlear 9 per cent.

AMP fund manager Michael Evans ***believes*** *0* a change in investor sentiment in favour of more cyclical stocks, combined with the high price/earnings multiples of some health-care stocks, has triggered the dramatic fall from grace. 'The market is just going through premium stocks and finding reasons to sell them,' said Mr Evans. 'Then they're piling into the other end of the market, into stocks like Orica, which were previously unloved and on very low P/Es.' (*The Age*, 05.02.02)

The majority of third-person epistemic verbs that were not embedded under a verb of saying, such as (47) below, were classified as main verb uses. Tokens like (47) primarily report that the subject of the epistemic verb holds a particular view, even though we may also infer that at some stage the reported holder of the view has expressed it in a statement, orally or in print.

(47) Experts *believe 0* the record box office pace is also due to savvier marketing campaigns from the major studios as well as independents such as IFC Films, which stunned the industry with its well-executed word of mouth campaign for 'Greek Wedding', and

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

marketing by MGM, which drew a younger demographic than ever to the debut of its latest Bond film. (*Los Angeles Daily News*, 28.11.02)

Main verb uses are to some extent identified negatively, in that they do not show the signs of modifier uses. Another common positive indicator of a main verb use is that the reported view is presented as false, as in (48).

(48) (a) People get seduced by the idea of being a wine maker and *think that* it's all very romantic. (*The Age*, 06.02.02)
(b) They just *felt 0* video games weren't appealing to kids. (Uttered by representative of video games company.) (*Los Angeles Daily News*, 27.11.02)

Main verb uses are also indicated by modification which expresses the reason for the view to be held, as in (49), or the belief is presented as the reason for action taken, as in (50).

(49) (a) Having a lot of work experience, and an adaptability of skills, I *thought 0* I was a sitter for quick re-employment. (*Christchurch Press*, 25.09.02)
(b) And Vin Diesel *thinks 0* Anna Nicole Smith is hot because he's attracted to 'white trash'. (*Los Angeles Daily News*, 28.11.02)
(50) (a) Six women and children *thought 0* the hall was unsafe and made a 400-metre dash to shelter under the bridge at Fiery Creek. (*The Age*, 09.02.02)
(b) 'It took us two days to get back to camp; no one was there because they *thought 0* we were dead.' (*Daily Telegraph* (UK), 04.01.03)

To sum up this section, the tokens were coded for the following factors:

Complementizer:	zero, *that*
Subject of epistemic verb:	first/second person, third person
Length of embedded subject:	pronoun, short NP (1–2 words), long NP (3+ words)
Epistemic verb:	*think*, *believe*, *feel*
Verb use:	modifier use, main verb use
Regional source:	New Zealand, Australia, United States, Great Britain
Textual subtype:	news & columns, quoted speech, letters

7 Results and discussion

A logistic regression analysis was carried out using Goldvarb software (PC version of Varbrul, Cedergren & Sankoff 1974), with the results shown in table 4 (significance = 0.013). All the factors were retained as significant in the *that*/zero alternation except for the person of the matrix subject, which was eliminated.

In order to more closely compare the factors of particular relevance to epistemic verbs, it was desirable to exclude the effects of the more general factors as far as possible. The results in table 4 suggest that the significance of text type was mainly due to the higher rate of *that* in letters, and the difference between quoted speech and news/columns might not be significant. The zero rates with different regional sources were examined for each verb, and it was found that the greatest variation appeared

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Table 4. *Results of first logistic regression analysis*

Factors	Zero rate	Factor weight
Regional source		
NZ	90%	0.584
AU	83%	0.486
US	90%	0.575
GB	79%	0.346
Verb		
think	90%	0.566
believe	80%	0.424
feel	72%	0.268
Text type		
quoted speech	90%	0.504
news/columns	83%	0.531
letters	65%	0.243
Embedded subject		
pronoun	92%	0.644
short NP	80%	0.363
long NP	76%	0.289
Verb use		
modifier	90%	0.598
main	80%	0.362

with *believe*, where the British data showed a zero rate of 62 per cent, in contrast to rates of 82–93 per cent for the other varieties. It was hypothesized that the regional source differences in the remaining data might not be significant. A revised corpus was prepared by removing all tokens of text-type letters (n = 41) and all British tokens of *believe* (n = 52). In addition, the effect of the embedded subject was most marked in the zero-promoting effect of a pronominal subject, and this factor was recoded as a two-way contrast between pronominal and NP embedded subjects.

A logistic regression analysis was carried out on the revised corpus (n = 658) which confirmed the suspected interactions noted above. Regional source and text type were no longer retained as significant factors, leaving the lexical verb, the type of verb use, and the embedded subject as the remaining significant factors, with the values shown in table 5 (significance = 0.004).

So far the results confirm that modifier uses of epistemic verbs have significantly higher zero rates than main verb uses. However, the modifier uses include all occurrences of *I think*, which is hypothesized to constitute a fixed routine in collocation with zero complementizer, in addition to any effect of the modifier sense. To examine the effect of the modifier sense alone, the *I think* effect must be excluded. All tokens of modifier *think* in the revised corpus were then coded as *I think*, *I don't think*, and other. *I don't think* was initially coded separately as its status with respect to routinization was uncertain. Zero rates were then calculated for the different types of modifier

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Table 5. *Results of logistic regression analysis of revised corpus*

Factors	Zero rate	Factor weight
Verb		
think	91%	0.528
believe	88%	0.532
feel	73%	0.238
Embedded subject		
pronoun	93%	0.633
NP	83%	0.330
Verb use		
modifier	92%	0.580
main	85%	0.379

Table 6. *Routines, modifiers, and main verb uses*

	Zero	That	Zero rate
I think	60	4	94%
I don't think	16	0	100%
All other modifier *think*	27	5	84%
All modifier *feel*, *believe*	69	11	86%
All main verb uses	68	26	72%

think, and compared with other modifiers (i.e. *believe* and *feel*) and main verb uses. The strong zero-promoting effect of pronominal embedded subjects generally masked the differences among these groups: zero rates with embedded subjects were all in the 92–98 per cent range, except for modifier *feel/believe* with 86 per cent zero. Accordingly, the comparisons were made using only tokens with NP embedded subjects – thus the only factors at issue are the lexical verb, the verb use, and the forms *I think/I don't think*. The results are shown in table 6.

The results in table 6 group the data into three main groups. *I don't think*, with the highest zero rate, belongs with *I think*: I shall assume that these two collocations are both routines, giving an overall zero rate for routines of 95 per cent. With *I think* and *I don't think* removed, the 84 per cent zero rate with other modifier uses of *think* is about the same as the 86 per cent average rate for modifier *believe* and *feel*: modifier *believe*, *feel* and nonroutine *think* can be combined to represent nonroutine epistemic modifier senses, with an overall zero rate of 86 per cent. The final comparison is shown in figure 1. The differences in zero rates are significant: for main verb uses vs. nonroutine modifiers, $p \leq 0.0176$, for nonroutine modifiers vs. routines, $p \leq 0.0378$.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

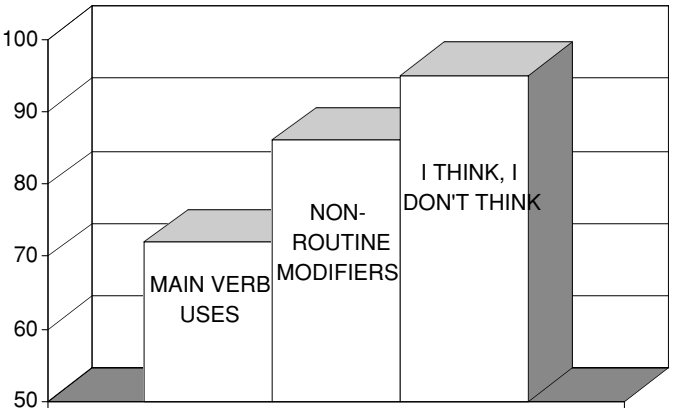


Figure 1. Zero rates with routines, modifier and main verb uses

8 Concluding remarks

The results support the hypotheses stated in section 4, repeated below, and also suggest that the routinization of *I think* with zero complementizer includes *I don't think*.

(32) *Hypothesis 1:* The modifier sense of an epistemic verb and its subject in matrix position promotes zero in the complement clause.
Hypothesis 2: Modifier *I think* (and *I guess*) in construction with zero complementizer have become fixed routines, resulting in higher zero rates than other modifier uses of epistemic verbs.

The main issue to be considered further here is why the modifier use of epistemic verbs should promote zero, assuming, as I argued above, that syntactic reanalysis is not involved. I propose a pragmatic account in terms of different degrees of prominence in the information conveyed by an utterance. I assume an account of the modifier use of an epistemic verb as an instance of Atlas & Levinson's (1981) information-strengthening implicature or Horn's (1984) R-implicature. An assertion of the general form *I believe (that) p* may be taken to mean literally that the speaker holds the belief that *p*. It is reasonable to infer from this that the speaker would assert *p* in appropriate circumstances – that is, where *p* is relevant to the current discourse. Consequently, if *I believe (that) p* is uttered and *p* is relevant, the utterance implicates *p* as the speaker's main assertion. The proposition *p* is the main assertion, and the accompanying information that the speaker believes that *p* is comparatively less prominent, though it is still communicated by the utterance (and accepted as true under Grice's Maxim of Quality). The interpretation of *I believe (that) p* as a hedged assertion of *p* follows from the fact that *p* is implicated rather than directly stated.

The implicature described here is equally available with or without the overt complementizer. However, the raised status of the embedded clause as the content of

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

the main assertion is more consistent with the absence of overtly marked subordination, and accordingly zero is more frequent in this type of context.

I suggest that a broadly similar effect, in terms of the relative backgrounding and foregrounding of information, may be seen with some uses of *say*, which, like *think*, has markedly high zero rates, as noted in the introduction. The information conveyed by *x said* has low informational prominence in many contexts, particularly in newspaper reports based on interviews. Following the introductory section the reader already knows who the interviewee is – all that is required is some indication that a statement is attributed to the interviewee, and is not a statement made by the writer. I observed in section 5 that *say* in the current corpus has the highest overall zero rate of all verbs, at 90 per cent. Although my remarks here are speculative, I suggest that the extremely high zero rate with *say* in newspaper reportage may reflect the low informational prominence of many instances of *x said*, in comparison with the greater prominence of the embedded clause as new content.

The relative unimportance of *x said* is also demonstrated by its omission where an alternative strategy is used to express attribution of the statement. This is illustrated in the extract below, where several attribution devices are used, presumably to vary the style of presentation.

(51) National MP Simon Power, who obtained the reports, said the incidents were [1] cause for concern. 'Asking untrained cops to deal with out-of-control prisoners and violent situations is [2] asking for trouble,' he said. . . . Police Association president Greg O'Connor said he was [3] reluctant to comment without knowing the full circumstances of the incidents. But training was [4] always a concern. That was [5] why it was [6] so important that people who used [7] pepper spray had [8] been properly trained and that they only used [9] it in the right circumstances, O'Connor said. It was [10] also essential that the public knew [11] that when they were [12] dealing with someone who looked [13] like a police officer, that they were [14] a police officer. (*Christchurch Press*, 02.11.06)

The verbs *were* [1] and *was* [3] are embedded under *say*, and show the sequence-of-tense shifting to past tense, in contrast to the present tense of directly quoted *is* [2]. The past tense can also be used alone to indicate sequence-of-tense, and thus to identify reported speech, in the remaining marked verbs. The identification of the speaker is already known, and the indication of reported speech is sufficient. In *was* [5], *was* [6], *used* [7], *had* [8], and *used* [9], the past tense signals reported speech in anticipation of the sentence-final *O'Connor said*, but in *was* [4], *was* [10], *knew* [11], *were* [12], *looked* [13], and *were* [14], the past tense is the only marker of reported speech in the sentence, and in the context is sufficient to attribute the statements to the previously identified interviewee, Greg O'Connor.[17] This phenomenon illustrates the low informational prominence of similar uses of *x said*, where the speaker is easily identified from the context, so long as *some* indication is given that a proposition is to be interpreted

[17] Explicitly, we infer that Greg O'Connor said: 'It is also essential that the public know that when they are dealing with someone who looks like a police officer, that they are a police officer.'

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

as an attributed statement. I suggest that the low informational prominence of *x said* illustrated here stems from the same kind of shift in informational prominence outlined above. The lower informational status of the *say* clause is shown by its omissibility in (51). The corresponding greater prominence of the embedded clause expressing the attributed statement may be generally reflected by the absence of overt subordinating *that*.

To sum up, the present study seeks to add to our understanding of the relationship between modifier uses of epistemic verbs, zero complementizer, and the frequent forms *I think/I don't think*. I have shown that although *I think* is the canonical instance of an epistemic modifier, it does not signal the required properties of modifier uses of epistemic verbs – in particular, modifier uses are compatible with third-person subjects, complex matrix clauses, and the past tense. I have argued that the broad pattern of increased zero rates with modifier uses of epistemic verbs is not indicative of syntactic reanalysis, and I have offered an alternative pragmatic account. The presence of zero in the embedded clause correlates with a shift in informational prominence, where the embedded clause becomes more prominent than the matrix. This shift is marked by the avoidance of overt marking of subordination by *that*, which is weakly inconsistent with the raised status of the embedded clause.

Author's address:
Linguistics Department
University of Canterbury
Private Bag 4800
Christchurch
New Zealand
kate.kearns@canterbury.ac.nz

References

Aijmer, K. 1997. *I think* – an English modal particle. In T. Swan & O. J. Westrik (eds.), *Modality in Germanic languages: Historical and comparative perspective*, 1–47. Berlin & New York: Mouton de Gruyter.

Atlas, J. D. & S. C. Levinson. 1981. *It*-clefts, informativeness, and logical form: Radical pragmatics (Revised Standard Version). In P. Cole (ed.), *Radical pragmatics*, 1–61. New York: Academic Press.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Ltd.

Bolinger, D. 1972. *That's that*. The Hague: Mouton.

Brinton, L. J. 1996. *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin & New York: Mouton de Gruyter.

Burchfield, R. (ed.) 1996. *The new Fowler's modern English usage*, 3rd edition. Oxford: Clarendon Press.

Cedergren, H. & D. Sankoff 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50, 333–55.

Elsness, J. 1984. *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65, 519–33.

Emonds, J. 1969. *Root and structure-preserving transformations*. Bloomington, IN: University Linguistics Club.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Emonds, J. 1976. *A transformational approach to English syntax*. New York: Academic Press.
Ferreira, V. S. & G. S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40, 296–340.
Finegan, E. & D. Biber. 1995. *That* and zero complementisers in Late Modern English: Exploring ARCHER from 1650–1990. In B. Aarts & C. F. Meyer (eds.), *The verb in contemporary English*, 241–57. Cambridge: Cambridge University Press.
Fowler, H. W. 1926. *A dictionary of Modern English usage*. Oxford: Clarendon Press.
Grice, H. P. 1975. Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Speech acts* (*Syntax and semantics*, vol. 3), 41–58. New York: Academic Press.
Halliday, M. A. K. 1985. *An introduction to Functional Grammar*. London: Edward Arnold.
Harris, A. & L. Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
Hawkins, J. A. 2003. Why are zero-marked phrases close to their heads? In G. Rohdenburg & B. Mondorf (eds.), *Determinants of grammatical variation in English*, 175–204. Berlin & New York: Mouton de Gruyter.
Henry, A. 1995. *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford: Oxford University Press.
Hooper, J. B. 1975. On assertive predicates. In John Kimball (ed.), *Syntax & semantics*, vol. 4, 91–124. New York: Academic Press.
Hooper, J. B. & S. A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4, 465–97.
Horn, L. R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (ed.), *Meaning, form and use in context: Linguistic applications* (Georgetown University Round Table on Languages and Linguistics 1984), 11–42. Washington: Georgetown University Press.
Horn, L. 1989. *A natural history of negation*. Chicago: University of Chicago Press.
Huddleston, R. & G. K. Pullum (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Jaeger, T. F. & T. Wasow. 2005. Processing as a source of accessibility effects on variation. Paper presented to 31st annual meeting of the Berkeley Linguistics Society. To appear in R. T. Cover & Y. Kim (eds.), *Proceedings of the 31st annual meeting of the Berkeley Linguistics Society: General session and parasession on prosodic variation and change*. Berkeley, CA: Berkeley Linguistic Society.
Jespersen, O. 1932. *A Modern English grammar on historical principles*. London: George Allen & Unwin.
Kearns, K. S. To appear. Regional variation in the syntactic distribution of null finite complementizer. *Language Variation and Change* 19.
Kroch, A. & C. Small. 1978. Grammatical ideology and its effect on speech. In D. Sankoff (ed.), *Linguistic variation: Models and methods*, 45–55. New York: Academic Press.
Langacker, R. 1977. Syntactic reanalysis. In C. N. Li (ed.), *Mechanisms of syntactic change*, 57–139. Austin, TX: University of Texas Press.
McDavid, V. 1964. The alternation of 'that' and zero in noun clauses. *American Speech* 39, 102–13.
Poutsma, H. 1929. *A grammar of Late Modern English*. Groningen: P. Noordhoff.
Prince, E. F. 1976. The syntax and semantics of NEG-Raising, with evidence from French. *Language* 52, 404–26.
Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
Race, D. S. & M. C. MacDonald. 2003. The use of 'that' in the production and comprehension of object relative clauses. Paper presented at the 26th Annual Meeting of the Cognitive Science Society.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press

Rissanen, M. 1991. On the history of *that*/zero as object clause links in English. In K. Aijmer & B. Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, 272–89. London & New York: Longman.
Rohdenburg, G. 1999. Clausal complementation and cognitive complexity in English. In F.-W. Neumann & S. Schülting (eds.), *Anglistentag 1998, Erfurt*, 101–12. Trier: Wissenschaftlicher Verlag.
Roland, D., J. L. Elman & V. S. Ferreira. 2005. Why is *that*? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98, 245–72.
Staum, L. 2005. When stylistic and social effects fail to converge: A variation study of complementizer choice. Ms., Stanford University.
Storms, G. 1966. *That*-clauses in Modern English. *English Studies* 47, 249–70.
Tagliamonte, S. & J. Smith. 2005. *No momentary fancy!* The *zero* 'complementizer' in English dialects. *English Language and Linguistics* 9, 289–309.
Thompson, S. A. & A. Mulac. 1991a. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15, 237–51.
Thompson, S. & A. Mulac. 1991b. A quantitative perspective on the grammaticization of epistemic parentheticals in English. In E. C. Traugott & B. Heine (eds.), *Approaches to grammaticalization*, vol. II, 313–29. Amsterdam and Philadelphia: John Benjamins.
Timberlake, A. 1977. Reanalysis and actualization in syntactic change. In C. N. Li (ed.), *Mechanisms of syntactic change*, 141–77. Austin: University of Texas Press.
Torres Cacoullos, R. & J. A. Walker. 2003. Taking a complement . . . variably. Paper presented at 32nd NWAV conference.
Underhill, R. 1988. The discourse conditions for that-deletion. Ms., San Diego State University.
Urmson, J. O. 1963. Parenthetical verbs. In C. E. Caton (ed.), *Philosophy and ordinary language*, 220–46. Urbana: University of Illinois Press.

https://doi.org/10.1017/S1360674307002353 Published online by Cambridge University Press