English Language and Linguistics 9.2: 289–309. © Cambridge University Press 2005
doi:10.1017/S1360674305001644 Printed in the United Kingdom

No momentary fancy! The *zero* 'complementizer' in English dialects[1]

SALI TAGLIAMONTE and JENNIFER SMITH
University of Toronto and University of York
(Received 9 June 2004; revised 3 December, 2004)

In this paper we analyse variable presence of the complementizer *that*, i.e. *I think that/Ø this is interesting*, in a large archive of British dialects. Situating this feature within its historical development and synchronic patterning, we seek to understand the mechanism underlying the choice between *that* and *zero*. Our findings reveal that, in contrast to the diachronic record, the *zero* option is predominant – 91 per cent overall. Statistical analyses of competing factors operating on this feature confirm that grammaticalization processes and grammatical complexity play a role. However, the linguistic characteristics of a previously grammaticalized collocation, *I think*, exerts a greater effect. Its imprint is visible in multiple internal factors which constrain the *zero* option in the other contexts. We argue that this recurrent pattern in discourse propels the *zero* option through the grammar. These findings contribute to research arguing for a strong relationship between frequency and reanalysis in linguistic change.

1 Introduction

Variable presence of the complementizer *that* in constructions, as in (1) to (4), is a widespread feature of English grammar.

(1) To prove Ø I could do it. Yes, I had to prove *that* I could do it. (CMK/b)[2]
(2) I wish *that* forty or fifty years ago I'd as much confidence. I wish Ø I'd had it then. (MPT/@)
(3) Somebody told me here lately Ø the trees is all cut down at the back . . . Somebody told me *that* the Macneils was up again for sale. (CLB/q)
(4) My grannie says *that* her da was very low, waiting on him, you know. She says Ø the Sunday nicht afore my granda was drowned, she was up at haim. (PVG/g)

According to the theoretical literature 'the complementiser *that* is optional' (Haegeman & Guéron, 1999: 173) and can be 'non-overt'(Haegeman & Guéron, 1999: 100), i.e. not realized in the surface form.

[1] The first author gratefully acknowledges the generous support of the Economic and Social Research Council of the United Kingdom (the ESRC) for research grant #R000239097, *Back to the Roots: The Legacy of British Dialects*. We are indebted to Helen Lawrence, the postdoctoral research assistant on this grant, for the gargantuan task of extracting and coding the data for this analysis and conducting a preliminary review of the literature. Hearty thanks to Minna Palander-Collin for lending her insights into the historical situation as well as engaging in informative discussion with us on this topic. We also benefited from two anonymous reviewers whose comments were most valuable. Any remaining errors are, of course, ours alone.

[2] Codes in parentheses refer to the community from which the example comes. CMK stands for Cumnock, MPT for Maryport, CLB for Cullybackey, and PVG for Portavogie. Single digit characters after the back slash refer to the individual speaker.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

In the descriptive literature, *that* is mostly considered a stylistic option which is deleted in informal contexts (e.g. Huddleston & Pullum, 2002: 953; Leech & Svartvik, 1975: 249; Quirk et al., 1972: 317; Zandvoort, 1969: 222). The overt form *that* is considered formal:

> less personal, less familiar, less warm, less friendly, less emotive. It is objective, factual, formal, official, sometimes tending to hostility. (Storms, 1966: 262)

On the other hand, Jespersen (1954: 38) suggests that the variable presence of *that* is simply the result of 'momentary fancy'.

In this article we consider the variation between *that* or *zero* by conducting a large-scale quantitative analysis of these options across a number of dialects in the British Isles and link our results to earlier findings in the literature. Specifically, we aim to address the following questions:

- What mechanism(s) underlie the variation?
- What can account for the patterns of use that are found?
- Are the patterns of use implicated in grammatical structure?

First, however, it is important to situate this feature of English grammar within its historical development and synchronic patterning. We first turn to a review of *that* or *zero* from a diachronic perspective.

2 The diachronic picture

A survey of the historical record reveals that the use of *that* or *zero* has changed radically in the history of English (Palander-Collin, 1997; Rissanen, 1991). In Old English, use of *that* is almost categorical (Mitchell, 1985: 1976ff), as in (5a). Examples with *zero*, as in (5b), are rare.

(5)(a) Swera me aþas þurh þone heofonlican God, *þæt* þu næfre ne nyme wif Isaace minum suna of þisum mennisce, þe ic mid wunige. (Genesis, 23:3, taken from Mitchell, 1992: 174)
'Swear by the lord, the God of heaven and the God of earth, *that* you will not get a wife for my son from the daughters of the Canaanites, among whom I am living.'

(b) Ic swerie þurh me selfne . . . *Ø* ic þe nu bletsie. (Genesis 22:16–17, taken from Mitchell, 1992:180)
'I swear by myself . . . *Ø* I will now bless you.'

Predominance of *that* continues in the Middle English period. For example, Warner (1982) finds 98 per cent use in his study of 571 complement clauses in the Wycliffe Sermons, as in (6a) and (6b). Here too *zero* is scant, as in (6c).

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

(6)(a) **ð**ei trowiden **ð***at* he was a trewe man, and lovede treufle, and dampnede synne (ii.53.18; Warner, 1982: 100)
(b) We taken as bileve **ð***at* epistlis of apostlis ben gospelis of Crist (ii.221.1; Warner, 1982: 102)
(c) and he seide *Ø* he ou**ʒ**te him an hundrid barrels of oyle. (Warner, 1982: 168)

However, by Late Middle English, more substantial variation is apparent, as in (7a–b), and thus examples of *zero* are much easier to find (Palander-Collin, 1997; Rissanen, 1991).

(7)(a) Pardee, ye knowen wel *that* ye maken no deffense (c. 1387, Chaucer, *Canterbury Tales: The Tale of Melibee*, line 1537)
(b) Thou knowest wel *Ø* thou art my sworen brother. (c. 1387, Chaucer, *Canterbury Tales: The Pardoner's Tale*, line 808)

By Early Modern English *zero* is common, as in (8b).

(8)(a) Fond man, remember *that* thou hast a wife. (1590, Shakespeare, *King Henry VI*, III.i)
(b) Now I remember, my Lord, *Ø* you gave good words the other day. (1604, Shakespeare, *Timon of Athens*, I.ii)

A critical backdrop for the diachronic perspective comes from the Helsinki corpus (Kytö, 1993).[3] There are two major studies which focus on this feature – Palander-Collin (1997) and Rissanen (1991). The frequency of *zero* from Rissanen's (1991: 279–82) seminal study of four high-frequency verbs (*know, think, say*, and *tell*) can be viewed in figure 1, which plots the trajectory of this option from Late Middle English (1350) to Early Modern English (1710), in context with Warner's study of the Wycliffe Sermons.

The figure reveals an astonishing trajectory of change. Given this unmistakable pathway of development, it is vital to consider *that* and *zero* in terms of linguistic change. Thus we also aim to address the following question: how does the use of *that* and *zero* in the historical literature compare with what is reported in contemporary varieties of English?

We now turn to consider the character of *that* and *zero* in contemporary varieties of English.

3 Contemporary research

The results of studies based on contemporary varieties of English are divided between those that study written sources and those which study spoken materials. Rates of *zero* are overwhelmingly high for spoken data. For example, Thompson & Mulac (1991a)

[3] The Helsinki corpus is a one and half million word collection of English texts spanning c.750–1710, a time period which encompasses Old to Early Modern English (Kytö, 1993: 2). The corpus is particularly useful for the study of diachronic change in the English language in that it provides a controlled sample for different periods, genres, and registers.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

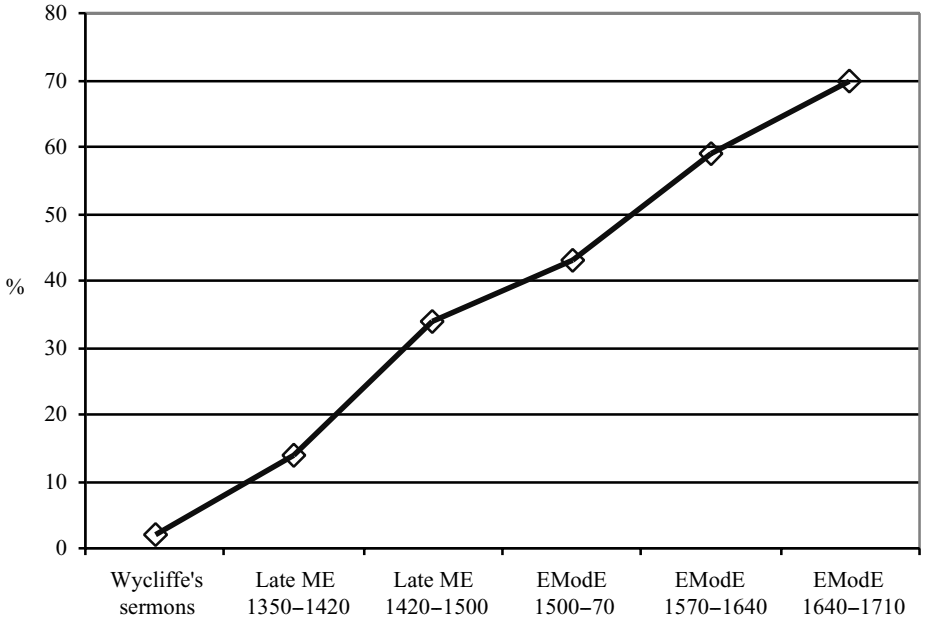


Figure 1. Frequency of *zero* from Late Middle English (1350) to Early Modern English (1710) (from Rissanen, 1991: 279–82 and Warner, 1982)

find up to 90 per cent *zero* in North American university students' speech,[4] while Walker & Cacoullos (2003) find over 90 per cent *zero* in Canadian English (Poplack & Walker, 2002, 2003).[5] Research on written data, on the other hand, reports considerably lower rates of *zero*. In a study of written American English (c. 1961), Elsness (1984) reports a little over 1 per cent *zero* in some cases.[6] Moreover, rates are highly dependent on the type of genre: learned science writing reflects the proportion of *zero* reported for Middle English: 1.3 per cent. In letters and biography, this rises to 14.6 per cent and there is another substantial rise with newspaper articles and fiction – 52.1 per cent and 58.1 per cent respectively.[7]

These differences were also apparent in the historical record: Rissanen (1991: 283–4) shows that 'speech related text types' such as trial records and comedies show higher rates of *zero* when compared to more formal writing genres such as scientific and educational treatises.

[4] This corpus comprises 116 eight-minute recorded conversations between university students at the University of California. The data total 240,000 words (Thompson & Mulac, 1991a: 240).

[5] The variety of Canadian English in their study is spoken in Québec City, which is predominantly French speaking.

[6] Elsness' data come from the Capital Syntax Data Corpus which consists of 64 of the 500 texts making up the Brown University Corpus of American English, 128,000 words (Elsness, 1984: 520).

[7] Finegan & Biber's (2001: 250) study of *that* omission in the London–Lund corpus does not provide comparable results as it is calculated per thousand words.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

In sum, the combined findings from both written and spoken data, contemporary and historical, offer two different perspectives on the use of *zero*: first, noteworthy increase in use of *zero* over time; second, in contemporary data, where both written and spoken English can be studied, the frequency of *zero* is much higher in spoken data.

Thus, we have the perspective from historical written sources (Rissanen, 1991; Warner, 1982), which can be compared with at least one contemporary written data set (Elsness, 1984). We also have the perspective from two spoken sources: American English studied by Thompson & Mulac (1991a, 1991b) and Canadian English (Walker & Cacoullos, 2003).

4 Explanation

The focus of our analysis will be on the two main language-*internal* explanations that have been put forward to account for the variation between *that* and *zero*. The first is grounded in grammaticization, and the second in syntactic complexity. In what follows, we review each of these proposals in turn.

4.1 Grammaticization

According to Thompson & Mulac (1991a, 1991b), grammaticization – the process by which grammatical forms and constructions arise from earlier lexical forms (e.g. Hopper & Traugott, 1993) – provides the most useful framework to explain the case of the *zero* complementizer in English.

Grammaticization theory predicts, in part, that lexical items in certain uses may be reanalysed into new functional material (Lehman, 1985). This is the pathway of change that Thompson & Mulac (1991b: 314) believe has taken place with the *zero* complementizer in English. They argue that there is actually no deletion in these constructions at all, but that the subject–matrix verb collocation has gradually grammaticized into an 'epistemic phrase'. The collocation is no longer a matrix clause; consequently, a complementizer is immaterial.

4.2 Grammaticization of parentheticals

Thompson & Mulac (1991a, 1991b) put forward a number of testable hypotheses which support the interpretation of the collocations without *that* as epistemic phrases.

The simplest definition of 'epistemic phrase' is 'modality concerned with propositions rather than events' (Palmer, 1979: 21). This is the difference between the collocation *I think*, in (9a), which involves the speaker's judgements about the possibility that something is or is not the case, in contrast to the collocation *I said*, in (9b), which involves an event.

(9)(a) *I think Ø* it was just the threat of it . . . (MPT/x)
(b) *I said that* they hardly would master an earthly weapon . . . (PVG/e)

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Because of this, certain lexical verbs favour the *zero* option. For example, matrix subject plus verb collocations such as *I think*, as in (10),[8] and *I guess*[9] rarely occur with *that*.

(10)(a) *I think Ø* she was lucky to get him. (CMK/f)
(b) *I think Ø* it was Wings for Victory week. (PVG/h)
(c) *I think Ø* John and me nears about the one age. (CLB/i)
(d) *I think Ø* he finished up in charge of the Scottish steel-works, did Tom. (MPT/n)

According to Thompson & Mulac (1991a, 1991b), these are constructions in which the relationship between the matrix and the complement has become eroded over time, and so the collocations themselves have become bleached of semantic content. In other words, they are acting as parentheticals rather than true clauses. Thus, *that* never occurs in these contexts because there is in fact *no* matrix clause. A similar proposal is found in Huddleston & Pullum (2002: 953) where *I think* is said to be more akin to modal qualification, taking on a background role which is 'informationally comparable to a parenthetical'.

Thompson & Mulac (1991a) test the hypothesis that these collocations are epistemic phrases in three ways, which we will now review.

4.2.1 Epistemic verb/Grammatical person

The *zero* option is favoured when the matrix verb is epistemic, as in (11a), as opposed to nonepistemic, as in (11b) (Thompson & Mulac, 1991a: 243).

(11)(a) *I believe Ø* I see a dent in Sean's door. (PVG/I)
(b) She *saw that* I had two friends coming in (CMK/i)

Further, epistemic parentheticals tend to occur with first- and second-person subjects (Thompson & Mulac 1991a: 242), as in (12a) and (12c), rather than third-person subjects, as in (12b) and (12d).[10]

(12)(a) *I mean Ø* things had got a wee bit better. (CMK/E)
(b) *It means that* I can't go anyplace you-know. (CLB/m)
(c) *You know Ø* John-Thompson'll preach the gospel to you a lot and all the rest but never listen. (CLB/f)
(d) *And Mammy reckoned that* she didn't make as good a butter as the plunger. (CMK/m)

Although Thompson & Mulac make no specific reference to tense, whether the matrix verb is present or past may be implicated too, since parentheticals are always present tense. Compare the examples (13a) and (13b).

[8] Thompson & Mulac (1991b: 315) point out that the collocation of epistemic verbs and their subjects, e.g. *I think*, behave like unitary epistemic morphemes in other languages as well.
[9] We note that these varieties do not use *I guess*.
[10] Quirk et al. (1985: 778–9) point out that first- and second-person pronouns are often used as 'stereotyped conversation fillers' when introducing comp clauses, e.g. *you know* and *I mean*.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

(13)(a) *I think Ø* they'll fettle you up. (MPT/r)
(b) *I thought that* I would end up singing. (MPT/%)

4.2.2 Additional elements

Another contextual constraint comes from whether or not the matrix verb phrase has additional elements. For example, negation or presence of modal auxiliaries favours the use of *that*. This is because the presence of such items reduces the likelihood that the subject and verb are functioning as an epistemic phrase (Thompson & Mulac, 1991a: 245). Compare the examples in (14a) and (14b).

(14)(a) And I *couldnae* believe *that* you have a granddauchter that age. (PVG/d)
(b) And I believe *Ø* he was famous, like. (MPT/Σ)

4.2.3 Complement subject

Finally, the subject of the complement clause also influences the choice of *that* or *zero*. According to Thompson & Mulac (1991a: 248), use of a pronominal subject, as in (15a), indicates that it is also the *topic* of the discourse. In their words, pronouns have 'high discourse topicality' (Thompson & Mulac, 1991a: 248). This, in turn, makes the *matrix* subject more likely to be epistemic, and hence the *zero* option more likely. Conversely, when this subject is a full NP as in (15b), the matrix functions as the topic, and hence more likely to be nonepistemic and to have an *overt* complementizer – *that*.

(15)(a) I think *Ø she's* gonnae come by bus. (CMK/a)
(b) I think *that some of his family* would be the same. (CLB/j)

In sum, Thompson & Mulac's (1991a, 1991b) hypothesis predicts more *zero* with epistemic verbs, first-person subjects, present tense, when there are no additional elements in the matrix verb phrase and when the subject of a complement clause is a pronoun.

We now turn to the second major proposal put forward in the literature for explaining the case of the *zero* complement clauses in English – syntactic complexity.

4.2.4 Syntactic complexity

Rohdenburg (1998: 101) states that the overriding influence in the appearance of *that* and *zero* is what he refers to as the *Complexity Principle* which holds that 'in the case of more or less explicit grammatical options the more explicit one(s) will be preferred in cognitively more complex environments'. In other words, more complex grammatical constructions, more use of *that*.

Rohdenburg (1998: 103) outlines 'two major complexity factors' which can test this principle. First, *zero* should be less likely to appear in contexts where adverbs or other adverbial elements intervene between matrix and complement clause, as in (16a), but more likely where there is no such intervening material, as in (16b):

(16)(a) I mind [*one time*] *that* you lift Mrs Macleary home. (CLB/g)
(b) I mind *Ø* she's a big place up there. (CLB/g)

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Table 1. *Explanations for that/Ø variability*

Grammaticization of epistemic phrases (Thompson & Mulac, 1991)	Complexity Principle (Rhodenburg, 1998)
Factors favouring zero:	Factors favouring zero:
Epistemic matrix verb	*No intervening adverbial or other elements between matrix and comp clause*
1st/2nd person pronouns	*'Simple' complement subject*
Present tense	
No additional elements in matrix verb phrase	
Pronominal subject in complement clause	

The influence of such intervening material on *that* or *zero* has also been mentioned by other researchers (e.g. Poutsma, 1928: 614; Quirk et al., 1972: 735; Rissanen, 1991: 286; Warner, 1982: 174–5).

Second, Rohdenburg (1998) argues that *zero* is more likely to appear when the complement's subject is a pronoun, rather than a full NP. We saw this earlier in (15). However, in contrast to Thompson & Mulac's (1991a, 1991b) formulation of this constraint, Rohdenburg appeals to the fact that the pronoun is less complex rather than its topicality.

This factor has also been mentioned by other researchers. For example, Rissanen (1991: 286) states that 'a pronominal subject beginning the object clause . . . increases the cohesion of the sentence' making *zero* more likely to appear (see also Warner, 1982: 171).

In sum, Rohdenburg's hypothesis predicts more *zero* with pronominal complement subjects and when there is no intervening material between matrix and complement clause.

The two major explanations for the *zero* option are outlined in table 1. It now remains to test these hypotheses in a relevant data set.

The current status of research on this phenomenon leaves two critical gaps. First, evidence from spoken varieties comes from relatively standard sources. Consequently, little is known about highly vernacular usage. Second, there is no evidence (to our knowledge) from spoken English outside of North America. Further, if linguistic change is occurring, it may well be important to determine whether it is proceeding in the same way or at the same rate across all varieties of English. Our archive of British dialect data, the Roots corpora, will provide appropriate evidence on each count.

5 The data

The data analysed for this study come from four communities on the periphery of the Irish sea: Maryport in Cumbria, Cumnock in southwest Scotland, and Cullybackey and Portavogie in Northern Ireland (Tagliamonte, 2001–3). The location of the communities is shown in figure 2.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press



Figure 2. Communities in Britain

These extensive materials totalling more than one million words provide a valuable resource for exploring questions of linguistic variation and change in situations of dense and multiplex social networks, in contexts which are geographically remote. In other words, these materials are highly informal and so enable us to tap into vernacular norms of English, as well as into the (recent) history of the language (see also Tagliamonte & Smith, 2002; Tagliamonte et al., 2005, to appear). In the analyses that follow we will abstract to the supralocal level, as we found that the individual communities all patterned very similarly with respect to the feature under consideration.

All speakers in the sample are from the oldest generation (60+) and were born and raised in the community. The speaker sample is shown in table 2.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Table 2. *The 'Roots' corpora (Tagliamonte 2001–3)*

Corpus	Male	Female	Total	Words
Cumnock, Scotland	18	22	40	346,346
Maryport, England	19	22	41	401,271
Culleybacky & Portavogie, Northern Ireland	22	7	29	316,493
Total	59	51	110	1,064,110

As we discussed earlier, there is a dramatic pathway of change in the use of *that* and *zero*, a trajectory which begins with near categorical use of an overt complementizer *that* to near categorical *zero* use. Given this perspective, we can ask: what linguistic changes have taken place in this area of grammar such that English complement clauses no longer require an overt complementizer? Or, asked another way, has there been a change in the matrix verb complex that has permitted the construction with no overt complementizer to increase so substantially over time?

6 Method

We extracted all (apparent) matrix + complement constructions where *that* or *zero* was possible, i.e. where the target matrix verb occupied this canonical slot. Such tokens are indicated as 'included' in (17a-c). As we outlined earlier, Thompson & Mulac (1991a, 1991b) argue that some of the same matrix verbs may appear in other contexts where they function entirely as 'parentheticals'. These are constructions that have been bleached 'to the point where the phrase acts as an adverbial' (Thompson & Mulac, 1991a: 239–40), as indicated by the fact that they are 'transportable to positions other than that which they could occupy if they were only functioning to introduce a complement'. Such constructions do appear in the data. However, following Thompson & Mulac (1991a: 241) none of these parenthetical forms have been included in the analyses that follow, as indicated as 'not included' in (17a–c).

(17)(a) [021] *I think*$_{[\text{included}]}$ he met me, actually, yeah. [Interviewer] Right. How did you come across one another? [021] Well, I was sitting in the – I would be about twenty-five, *I think*,$_{[\text{not included}]}$ and I was sitting in the Empire-Theatre in Maryport, on my own, watching the picture. (CMK/021)

(b) [019] *I think*$_{[\text{included}]}$ she won the waterloo-cup of some big, prestigious thing. I hae nae much an idea about greyhound-racing, but- We called the old boat Silver-Anna. She was an old [inc], *I think*$_{[\text{not included}]}$. She stood about two birds clear of the water. (MPT/019)

(c) [002] Sure, even the people in the south o this island *I think*$_{[\text{not included}]}$ really, if they were getting without being intimidated and all, they would hae went in under the Commonwealth. You-know. *I think*$_{[\text{included}]}$ they would. (PVG/002)

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Table 3. *Overall distribution of* zero

%	N	Total N
91	3731	4106

Table 4. *Distribution of* zero *in* I think, you know, I mean

Matrix collocation	%	N	Total N
I think	98.5	974	989
You know	99	535	541
I mean	100	428	428

In all, this procedure provided over 4,000 tokens. As discussed in section 4.2, each token was then coded for all the factors implicated in the use of *that* and *zero*. In the analyses that follow, we establish which of these factors explain the options in vernacular British dialects.

7 Results

Table 3 shows the overall distribution of *that* and *zero* in the data. The frequency of *zero* is very high – 91 per cent.

Recall Thompson & Mulac's (1991b: 320) findings regarding the collocations *I think* and *I guess*. In their data, these collocations had high rates of *zero*: *I think* 91 per cent *zero* [N = 683] and *I guess* 99 per cent *zero* [N = 150]. Similar high rates for these lexical verbs are reported for Canadian English (Walker & Cacoullos, 2003) (*I think* 95 per cent N = 754, *I guess* 97 per cent N = 150).

In addition to *I think*, as in (18a), closer examination of our data revealed a number of other collocations of this calibre as well: *you know*, as in (18b) and *I mean* as in (18c).

(18)(a) *I think Ø* the secretary's job is the hardest, I think. (CMK/b)
(b) *You know Ø* he never got a pension or anything. (CMK/c)
(c) *I mean Ø* I was married when I was twenty. (CMK/M)

Table 4 shows the overall distribution of forms in these contexts only. From a total of 989 *I think* forms in the data, 98.5 per cent occurred with *zero*. This result mirrors both Thompson & Mulac's (1991a, 1991b) analysis of American English as well as patterns in Canadian English (Walker & Cacoullos, 2003). Such parallel findings across three major varieties of English (American, Canadian, and British) suggest that the mechanism behind the use of *that* vs *zero* in the collocation *I think* is the same across the board.

In our data, two other frequent collocations have overwhelming use of *zero*: the collocation *you know* is 99 per cent *zero* and *I mean* is categorical, 100 per cent.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Table 5. *Overall distribution of* zero *(excluding* I think, you know, I mean*)*

%	N	Total N
84	1798	2148

Recall Thompson & Mulac's argument that this particular distribution of *I think* provides evidence that the underlying explanation for *zero* in these contexts is that there is in fact no matrix clause in these contexts. Instead, this form is said to function as a formulaic 'chunk' – an epistemic parenthetical, not a matrix verb with a subordinate complement clause at all – so no complementizer is required. The fact that *zero* is virtually categorical in these constructions in the dialect data provides striking confirmation of Thompson & Mulac's hypothesis in highly vernacular English usage. Thus, we conclude that British English corroborates their hypothesis that the use of *that* and *zero* with *I think*, *you know*, and *I mean* is the result of grammaticization of these collocations.

However, Thompson & Mulac's analysis stops here. There are still innumerable matrix verb + complement clauses in our data to be accounted for. Hence, there is yet another avenue to pursue in this endeavour. Indeed, if collocations such as *I think*, *you know* and *I mean* have been grammaticized 'it would certainly be worth looking for other verbs or pronoun and verb combinations' (Palander-Collin, 1997:393). Taking up Palander-Collin's suggestion, we now turn to this type of investigation.

Due to the categorical status of the collocations *I think*, *you know* and *I mean* as we have just seen, we now remove these from consideration.[11] Table 5 now shows the overall distribution of *that* and *zero* in the remaining contexts in the data.

Despite the fact that the categorical contexts have been removed, *zero* occurs 84 per cent of the time. How can this be accounted for? Is there ongoing grammaticization of *other* matrix subject plus verb collocations, as might be predicted from Thompson & Mulac's research? Can the variation be accounted for by the complexity constraints outlined by Rohdenburg? Or is there some other explanation?

In the remainder of the article we explore these questions. Due to the competing explanations outlined in the literature and their numerous concomitants, we carry forward previous investigations of this phenomenon by conducting a multivariate analysis of the variation between *zero* and *that* by means of GOLDVARB 2.0 (Rand & Sankoff, 1990). This method permits us to model the combined contribution of all the contextual factors simultaneously, thus enabling us to address the problem of whether or not each of the factors 'independently influences the occurrence of *that*' (see Thompson & Mulac, 1991a: 249). Furthermore, this type of analysis will enable us to determine which of the putative factors contribute statistically significant effects to the

[11] Removal of categorical and near-categorical contexts is a standard methodological procedure in conducting a statistical analysis of variable linguistic phenomena.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Table 6. *Multivariate analyses of the factors conditioning* zero, *excluding tokens of* I think, I mean *and* you know

Input	.87		
Total N	**2148**		
	FW	%	N
LEXICAL VERB IN MATRIX CLAUSE			
Think	**.70**	93	461
Say	**.54**	85	647
Know	**.46**	85	196
Other	**.38**	80	680
Tell	**.30**	64	164
Range	*40*		
MATRIX SUBJECT			
1st person singular	**.60**	88	985
Other pronoun	**.45**	83	894
Other	**.30**	67	267
Range	*30*		
ADDITIONAL ELEMENTS IN MATRIX VERB PHRASE			
Nothing	**.56**	87	1512
Something	**.37**	76	636
Range	*19*		
COMPLEMENT CLAUSE SUBJECT			
Personal pronoun	**.54**	86	1703
Other	**.37**	74	443
Range	*17*		
INTERVENING MATERIAL			
None	**.52**	85	1963
Some	**.35**	71	185
Range	*17*		
VERB TENSE			
Present	**.56**	87	917
Past	**.44**	84	861
Range	*12*		

appearance of *that* and *zero* forms as well as to determine the direction and relative magnitude of the effects.

We test each of the contextual factors outlined earlier in section 5.2, which are thought to condition the use of *that* or *zero*. Recall that the factors expected to function if the underlying mechanism is ongoing grammaticization are: matrix verb, matrix subject, additional elements in the matrix clause, complement clause subject, and verb tense. On the other hand, the factors expected to function if the underlying mechanism is complexity are: complement clause subject and intervening material.

Table 6 presents all the necessary information for interpreting the evidence provided by the multiple regression analysis. The INPUT indicates the overall tendency of the

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

dependent variable (in this case the *zero* relative marker) to surface in the data. The TOTAL N records the denominator of the total number of contexts treated in the analysis. Each of the factor groups that have been considered in the analysis is listed with the results for each factor. Point-form numbers are FACTOR WEIGHTS. These indicate the probability of the dependent variable occurring in that context. The closer these numbers are to 1, the more highly favouring the effect is; the closer they are to *zero* the more disfavouring the effect is. The RANGE provides a nonstatistical measure of the relative strength of the factor. The higher this number is, the greater the contribution of that factor to the probability of the form.

Table 6 shows a multivariate analysis of the factors contributing to use of *zero* in these data (recall that the categorical contexts *I think*, *you know* and *I mean* are excluded). We now interpret these results taking each factor in turn.

7.1 Lexical verb in the matrix clause

First, consider the lexical verb in the matrix clause which tests for grammaticization. Thompson & Mulac (1991a: 244) argue that verbs occurring without *that* are 'heavily involved in the expression of epistemicity'. However, which verbs can be considered epistemic and which not is not straightforward.[12] In addition, it is important to keep in mind that the most frequent matrix verbs implicated in this variation (*think, say, know*, and *tell*) comprise the vast majority of the data in English, and thus in any analysis. Moreover, innumerable researchers have previously noted that particular verbs tend to favour *zero* regardless of their epistemicity (e.g. Jespersen, 1954: 33; Poutsma, 1928: 615; Rissanen, 1991: 280). Indeed, consider once again Rissanen's (1991) quantitative results from 1350–1710 in figure 3 which now plots the rate of *zero* for each of these frequently occurring verbs separately. The ranking of these frequent verbs is quite constant over time. Once the *zero* option has reached a threshold of robust occurrence by the Early Modern English period (1500–70), *think* has taken a lead in the use of *zero* and thereafter remains the frontrunner. The other verbs follow along, with *know* the second-ranked, then *say*, then *tell*.[13] Thus, these individual verbs have always had highly differentiated patterns which are consistent across a long trajectory of change. Consequently, in the variable rule analysis we configure this factor group so as to test these highly frequent lexical verbs individually – *think, say, know,* and *tell*, alongside all others.[14]

[12] Givón (1979: 134), for example, situates verbs within a continuum of 'subjective certainty': *I know* is the most certain and *I guess* and *they say* the least, with a number of other verbs in between. Thus, verbal epistemicity is highly gradient.

[13] The only exception to this ranking is a drop in the frequency of *know* from 1500 to 1640.

[14] Intervening material (adverbs, etc.) and additional elements in the matrix verb phrase (i.e. negation, modals, etc.) are independent, noninteracting factors. One tests the effect of additional material within the verb phrase, (i); the other tests intervening material between the matrix and complement clause, (ii). See also (14a) and (16a).

(i) That was the only way that they [could] *prove that* they had this pneumoconiosis (CMK/b)
(ii) I think the catholic community *realised* [eventually] *that* I had no sides (CMK/q)

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

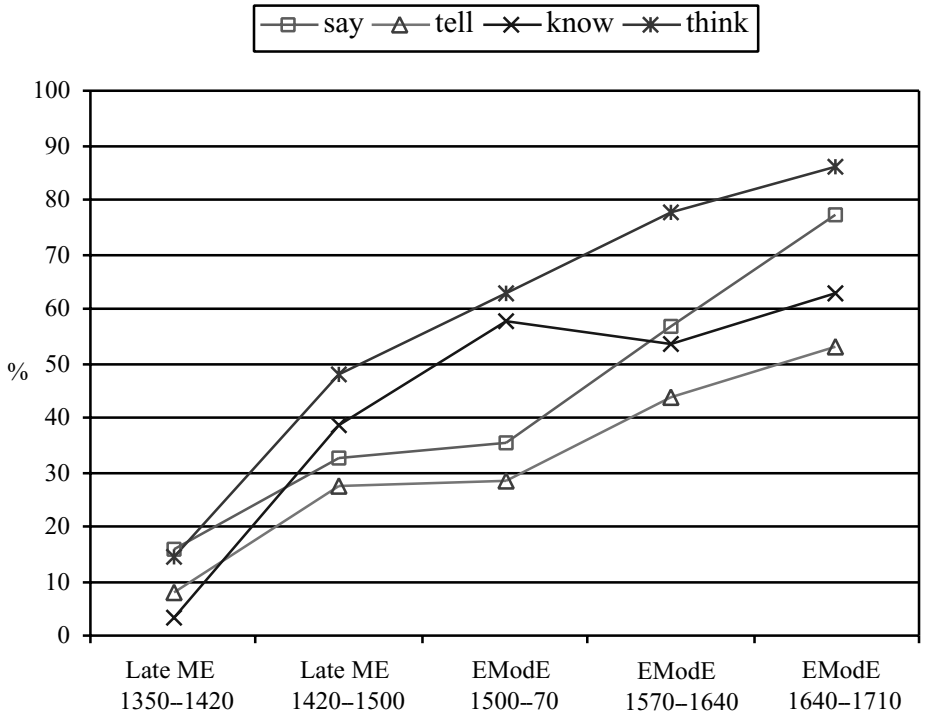


Figure 3. Frequency of *zero* from Late Middle English (1350) to Early Modern English (1710) with individual verbs separated (Rissanen, 1991: 279–82)

In fact, this factor exerts the strongest constraint on the variation between *that* and *zero* in the British dialect data, with a range of 40. Most extraordinary is that the ranking of verbs precisely reflects the historical trajectory in Rissanen's (1991) study – *think* > *know* > *say* > *tell*. Note that the verb *think* highly favours *zero*, with a factor weight of .70.[15]

7.2 Matrix subject

Consider next the results for matrix subject, another test for grammaticization. According to Thompson & Mulac's analysis, if grammaticization is the relevant explanation, first-person-singular pronouns should favour *zero*. In our data, this grammatical person stands out well ahead of the rest, favouring *zero* at .60. Note too that this is the second strongest effect, with a range of 30.[16]

[15] *Tell* is set apart from the other frequently occurring verbs in that it strongly disfavours *zero*. However, this verb is syntactically distinct from *think*, *know*, and *say* as it takes a direct object.

[16] Second-person pronouns have been combined with other pronouns because they patterned in the same way, contra Thompson & Mulac 1991: 242 (see section 4.2.1)

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

The remaining four factors are also statistically significant, but are much weaker. They all exert approximately the same strength.

7.3 Additional elements in the matrix verb phrase

If grammaticization is explanatory here we should observe that *zero* is favoured when there are no additional elements in the matrix verb phrase. This is exactly what the results show: the simpler constructions favour *zero* at .56.

7.4 Subject of the complement clause

The test for subject of the complement clause is a factor which is compatible with either a grammaticization or complexity explanation. The result for the complement clause subject shows that when it is a personal pronoun, *zero* is favoured, at .54. When it is not, *zero* is disfavoured at .37. This result is consistent with both a grammaticization and a complexity explanation.

7.5 Intervening material between matrix and complement clause

Intervening material between the matrix and complement clause tests Rohdenburg's main hypothesis for complexity: intervening material is said to introduce complexity to the construction and so promote *that*. This is demonstrated here where intervening material disfavours *zero* at .35, and no intervening material shows little effect at .52.[17]

7.6 Verb tense

Lastly, the grammaticization hypothesis predicts that present-tense verbs should favour zero. In these data, present tense favours *zero* at .56.

8 Discussion

How can these results be explained in the light of the two main hypotheses put forward for variable *that* or *zero*: ongoing grammaticization and complexity? First, all the factors we tested were selected as significant, and with the constraint ranking predicted by grammaticization and complexity. Thus, both explanations can account for the observed variability. But if both processes are operating, what explanation provides a synthesis?

As we outlined earlier, according to Thompson & Mulac, epistemic parentheticals in English such as *I think* have no overt complementizer because they have grammaticized away from their original function of subject plus verb introducing a complement

[17] Note that a common statement found in variationist methodology is that anything over .5 favours and anything under .5 disfavours. While this generalization is often true, the more important observation is the *relative* difference between the factor weights. In this case, when there is some intervening material, there is a clear favouring effect on the use of *that.*

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

clause. Moreover, they argued that it was the *frequency* of *I think* without *that* in these constructions that led to their reinterpretation as epistemic phrases.[18] How and when did this happen?

If we consider the historical trajectory of the *zero* complementizer we find that *I think* has long been cited as a favouring location for *zero* (Rissanen, 1991: 280). Indeed, we can gain considerable perspective on the history of this use by considering more closely the research of Palander-Collin (1997), who traces the development of the particular collocations *me thinks* and *I think* in the Helsinki corpus from Middle to Early Modern English.

The original meaning of the collocation *I think* was to express the mental function of thinking or having something in mind. The impersonal collocation *me thinks* on the other hand was the form used to express an opinion.[19] However, '*me thinks* was fast disappearing in the 15th century' (Palander-Collin, 1997: 383). The interesting fact for our purposes here is that *I think* took over *me thinks* and its speaker-oriented function of opinion. In Palander-Collin's (1997: 383) words: 'The situation is properly reversed at the beginning of the Early Modern English period' (1500–70). In other words, *I think* replaces *me thinks*. Thus, *I think* in its contemporary use had a precursor in the impersonal phrase *me thinks*.

Interestingly, *me thinks* had already long been propitious to the *zero* complementizer, and this even when the *zero* complementizer was extremely rare elsewhere. Note that the example in (19) comes from Old English and it has *that*. Palander-Collin (1997) finds only one example of *zero* from this time period (1150–1250), but nineteen examples with *that*.

(19) Meþinc *þæt* hit hæbbe geboht sume swie leaslice mære. (eOE ÆLFRED tr. Boethius *De Consol. Philos*. (Otho) xxiv. 54)

However, in later periods the situation changes. Recall that Warner's (1982) study of the Wycliffe Sermons (c. 1390) found only 2 per cent *zero*, yet examples of *me thinks* from approximately the same period in (20) show that it begins to appear with *zero* during this time period.

(20)(a) He of mi loue vnder-nome were. *Me þenke Ø* y no myt it him nout werne. (*OED*: c. 1330 *Guy of Warwick* (Auch.) 613)
(b) *Me thinc Ø* it es noght sua. (*OED*: a1400 (a1325) Cursor Mundi (Vesp.) 16332)
(c) And yet *me thinkyth Ø* it were pete That my werk were hatyd for me. (*OED*: a1393 GOWER Confessio Amantis (Fairf.) IV. 2785 1447 O. BOKENHAM Lives of Saints 2)

In late Middle English (1420–1500) the frequency of *zero* rises with *me thinks* to about 50 per cent of the available tokens. Indeed, from 1500 onwards there are virtually no examples with *that*, as in (21).

[18] Recall that the *I think* collocation represented nearly 1000 tokens in the original data set as we saw in table 4.

[19] These forms are derived from two Old English verbs *þencan* and *þyncan*. '*þencan* was used in the sense "to think" in SVO constructions, where the subject denotes the thinker, while *pyncan* was used impersonally in the sense "to seem"' (Palander-Collin, 1997: 376).

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

(21)(a) *Me thinketh Ø* it somwhat more ciuilytye . . a litle merely to mok him, then . . seriouslye to preace vpon him. (*OED*: 1533 T. MORE *Debellacyon Salem & Bizance* in *Wks*. (1557) 951/1)
(b) *Me thinketh Ø* this motiue should not be without effect. (*OED*: 1607 S. HIERON *Wks*. I. 439)

Referring to Thompson & Mulac's (1991a, 1991b) conclusions, Palander-Collin (1997: 396) proposes that *me thinks* is the result of grammaticization too, albeit a medieval case. Like the contemporary *I think* in Thompson & Mulac's (1991a) data, she suggests *me thinks* came to be used as an 'adverb indicating point of view which could also be used parenthetically'. This means that Thompson & Mulac's (1991a, 1991b) suggestion that phrases such as *I think* are grammaticized epistemic parentheticals is actually a phenomenon that is embedded in a trajectory that can be traced way back in the history of the English language (see also Palander-Collin, 1997).

Indeed, the roots of this grammaticization process may well have began as early as Middle English with the use of *zero* with *me thinks*, which later culminated in a surge in the frequency of *zero* with *I think* from Early Modern English onwards. Note the spike in frequency of *zero* from 1350 to 1420 in figure 3. Given this early model, the longitudinal trajectory of change it provides, and the overwhelming proportion of *zero* in our data, we speculate that the collocational pattern of *me thinks/I think* may have had an effect elsewhere in the grammar. How can this hypothesis be tested? We now return to the results of the multivariate analysis in table 6 and re-examine the evidence it provides in the light of the characteristics of the collocation *I think*.

First of all, the collocation *I think* obviously implicates the verb *think*. Notice that this is the verb that favours *zero* the most. Even though the collocation *I think* is excluded from the data being analysed, the verb *think* more generally is a highly favoured *zero* context. Second, the collocation *I think* is first person. Once again, despite the absence of the *I think* collocation in the data, first-person-singular subjects are the most highly favouring matrix subject for *zero*. Further, *I think* has got no additional elements in its verb phrase. This factor favours *zero* too. *I think* is present tense, and present tense favours *zero* as well.

In sum, even when the most frequent epistemic parenthetical in the data is *removed* from the analysis we still have an overwhelming preference for *zero* in precisely the same types of contexts. In fact, four out of the six conditioning factors exactly mirror the characteristics of *I think*. We suggest, therefore, that the model of this collocation (and its precursor *me thinks*) has actually had a tremendous influence on the rise of *zero* in the *rest* of the grammar.

9 Conclusion

To conclude, the presence or absence of *that* in these data is not entirely explained by processes of grammaticization. While the grammaticization of parenthetical expressions is proceeding just as Thompson & Mulac (1991a, 1991b) outlined, it cannot explain *all* the contexts of *zero* in the data. Not all of them are epistemic; not all

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

of them are parenthetical. Yet *zero* prevails. Similarly, the complexity principle does not entirely explain the use of *zero* either. While constructions that are more complex have a greater probability of an overt complementizer, just as Rohdenburg predicted, this too does not explain the complex array of factors impinging on the variation. Indeed, the complexity factors are amongst the weakest in the analysis. Moreover, they have operated since the very beginning, as even in 1390 when *zero* represented only 2 per cent of the data (in the Wycliffe Sermons), complexity effects were visible (Warner, 1982). Further, complexity cannot explain the long course of historical change in which *that* and *zero* are embedded.

Instead, it would appear that there are three inter-related facets to the explanation for *zero*. Complexity operates as an over-arching processing constraint – a universal we might expect to operate in all areas of grammar. Grammaticization, on the other hand, is like an arrowhead in the initial motivation of this change. However, the frequent lexical collocation that grammaticizes first actually propels the new option, in this case *zero*, through the grammar. Indeed, its imprint leaves visible tracks in the multiple constraints on the variation. Thus, our results, like those of Thompson & Mulac (1991b: 314), contribute to the larger body of research which has argued for a strong relationship between frequency and reanalysis in linguistic change. Our findings provide empirical confirmation, from vernacular data, that 'recurrent patterns in discourse tokens exert pressure on linguistic types' (Dubois, 1985: 359–60). The rise of the *zero* option can be explained as the dramatic imposition of a frequent lexical collocation rippling through the grammar.

Authors' addresses:
Department of Linguistics
University of Toronto
Toronto, Ontario M5S 3H1
Canada
sali.tagliamonte@utoronto.ca
st17@york.ac.uk

Department of Language and Linguistic Science
University of York
York YO10 5DD
United Kingdom
js40@york.ac.uk

References

Dubois, J. (1985). Competing motivations. In Haiman, J. (ed.), *Iconicity in grammar*. Amsterdam: John Benjamins. 343–65.

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.

Finegan, E. & D. Biber (2001). Register variation and social dialect variation: the register axiom. In Eckert, P. & J. Rickford (eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 235–67.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Haegeman, L. & J. Guéron (1999). *English grammar: a generative perspective*. Oxford: Blackwell.
Hopper, P. J. & E. C. Traugott (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
Huddleston, R. & G. K. Pullum (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Jespersen, O. H. (1954). *A modern English grammar on historical principles: Part III: Syntax* (second volume). London: George Allen & Unwin.
Kytö, M. (1993). *Manual to the diachronic part of the Helsinki Corpus of English Texts: coding conventions and lists of source texts*. Helsinki: University of Helsinki, Department of English.
Leech, G. N. & J. Svartvik (1975). *A communicative grammar of English*. London: Longman.
Lehman, W. (1985). Grammaticalization: synchronic variation and diachronic change. *Lingue e Stile* **20**: 303–18.
Mitchell, B. (1985). *Old English syntax,* vols. I, II. 2. Oxford: Clarendon Press.
Mitchell, B. C. & F. C. Robinson (1992). *A guide to Old English*. Oxford: Blackwell.
Palander-Collin, M. (1997). A medieval case of grammaticalization, *methinks*. In Rissanen, M., M. Kytö & K. Heikkonen (eds.), *Grammaticalization at work* (Studies of Long-term Developments in English 24). Berlin and New York: Mouton de Gruyter. 371–403.
Palmer, F. R. (1979). *Modality and the English modals*. New York: Longman Group Ltd.
Poplack, S. & J. Walker (2002). A majority language in minority guise: the future of Quebec English. Paper presented at Canadian Linguistics Association (CLA). Dalhousie University, Halifax, Nova Scotia, Canada. June 2003.
Poplack, S. & J. Walker (2003). An English '*like no other*'? Language contact and change in Quebec. Paper presented at New Ways of Analyzing English (NWAV) 31. Stanford University, Palo Alto, California. October 2002.
Poutsma, H. (1928). *A grammar of Late Modern English*. Groningen: P. Noordhoff.
Quirk, R., S. Greenbaum, G. Leech et al. (1972). A grammar of contemporary English. New York: Harcourt Brace Jovanovich, Inc.
Quirk, R., S. Greenbaum, G. Leech et al. (1985). *A comprehensive grammar of the English language*. New York: Longman.
Rand, D. & D. Sankoff (1990). *GoldVarb: a variable rule application for the Macintosh*. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal. Version 2.
Rissanen, M. (1991). On the history of *that*/*zero* as object clause links in English. In Aijmer, K. & B. Altenberg (eds.), *English corpus linguistics: studies in honour of Jan Svartvik*. London and New York: Longman. 272–89.
Rohdenburg, G. (1998). Clausal complementation and cognitive complexity in English. In Neumann, F.-W. & S. Schülting (eds.), *Anglistentag Erfurt*. Trier: Wissenschaftlicher Verlag. 101–12.
Storms, G. (1966). That-clauses in Modern English. *English Studies* **47**: 249–70.
Tagliamonte, S. A. (2001–3). Back to the roots: the legacy of British dialects. Research Grant, Economic and Social Research Council of the United Kingdom (ESRC). #R000239097.
Tagliamonte, S. A. & J. Smith (2002). Either *it isn't* or *it's not*: NEG/AUX contraction in British dialects. *English World Wide* **23**: 251–81.
Tagliamonte, S. A., J. Smith & H. Lawrence (2005). No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* **17**: 75–112.
Tagliamonte, S. A., J. Smith & H. Lawrence (to appear). Disentangling the roots: the legacy of British dialects in cross-variety perspective. In *The proceedings of METHODS XI*. Joensu, Finland.
Thompson, S. & A. Mulac (1991a). The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics* **15**: 237–51.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press

Thompson, S. & A. Mulac (1991b). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In Traugott, E. C. & B. Heine (eds.), *Approaches to grammaticalization*. Amsterdam and Philadelphia: John Benjamins. 313–29.
Walker, J. & R. Cacoullos (2003). Taking a complement . . . variably. Paper presented at New Ways of Analyzing Variation (NWAV) 32, October 11, 2003. Philadelphia.
Warner, A. (1982). *Complementation in Middle English and the methodology of historical syntax*. London and Canberra: Croom Helm.
Zandvoort, R. W. (1969). *A handbook of English grammar*. London: Longmans, Green & Co. Ltd.

https://doi.org/10.1017/S1360674305001644 Published online by Cambridge University Press