The discourse conditions for the use of the complementizer *that* in conversational English

Sandra A. Thompson and Anthony Mulac*

Received March 1989; revised version February 1990

The complementizer *that* in English, as in *I heard (that) you were sick*, has widely been regarded as optional. Our research demonstrates that the use of *that* in such utterances in conversation is highly related to various other features in the discourse. First and second person subjects, the verbs *think* and *guess*, and auxiliaries, indirect objects, and adverbs in the main clause, and pronominal complement subjects are all significant in predicting the use of *that*. As seemingly disparate as these factors are, their influence finds a unified explanation in the acknowledgement that certain combinations of main clause subjects and verbs in English (such as *I think*) are being reanalyzed as unitary epistemic phrases. As this happens, the distinction between ‘main’ and ‘complement’ clauses is being eroded, with the omission of *that* a strong concomitant. Our findings show that the factors most likely to contribute to this reanalysis are precisely those which relate either to the epistemicity of the main subject and verb or to the topicality of the complement at the expense of the main clause.

1. Introduction and background

It is well known that the complementizer *that* is omissible in English. Two examples from our data illustrate this; (1) below has *that*, while (2) does not:

(1) I mean like, my little sister, you can tell *that* she’s growing up so much faster than I did. (26.89)

(2) I heard *that* your food’s a lot better than our food. (25.163)*

* Authors’ address: S.A. Thompson, Dept. of Linguistics, and A. Mulac, Dept. of Communication, University of California, Santa Barbara, CA 93106, USA.

We would like to thank the following people for helpful comments on earlier versions of this paper: Dwight Bolinger, Joan Bybee, Wallace Chafe, Simon Dik, John Du Bois, Mark Durie, Julie Gerhardt, David Hargreaves, Paul Hopper, C. Douglass Johnson, Ronald Macaulay, Tsuyoshi Ono, William Paśluka, Gisela Redeker, Nathan Salmon, Janine Scancarelli, Arthur Schwartz, Dan Slobin, R. McMillan Thompson, and Elizabeth Traugott. None of these people necessarily agrees with the interpretation we’ve given to their input. We also appreciate Bonnie Seligmann’s and Yuka Koike’s excellent senior essays on this topic.

* The numbers in parentheses indicate dyad number and line number in our data base (see section 2, Methodology, below).

0378-2166/91/$03.50 © 1991 — Elsevier Science Publishers B.V. (North-Holland)
Most traditional and modern formal approaches to English have taken the view that this omission represents an option for speakers of English, but relatively few have concerned themselves with determining what the conditions underlying its appearance might actually be.

Our purpose in this paper is to refute the absolute optionality of this alternation, and to show that, in probabilistic terms, there are very clear conditions underlying the decision to use or to omit the complementizer that in ordinary talk. We will compare our findings to earlier studies on the use of that in written English.

We are not the first to propose explanations for the fact that the complementizer that is sometimes missing and sometimes present. The work of Dwight Bolinger challenged the idea of ‘optional’ ‘rules’ almost as soon as this concept appeared in the writings of grammarians oriented towards transformational descriptions. Thanks in large part to his efforts, it is now widely accepted that what had been thought of as ‘stylistic options’ are better seen as genuine choices of the speaker based on such factors as attitude, emotional stance, information flow, and discourse structure.

Bolinger (1972) suggested that the choice of whether to use or not use a complementizer that is likely to be determined in large part by considerations involving subconscious factors such as the attitudes of the subject of the main verb towards the content of the complement clause and the extent to which its content is known/important.

Three studies have investigated the question of the alternation between complementizer that and its omission for written English. McDaid (1964), using a corpus of modern non-fiction, concludes that “the conditions under which that may be omitted seem partly stylistic and partly grammatical” (p. 113). Elsness (1984) considers the alternation across several genres, and concludes that “the conditioning process governing the choice of connective [i.e., that or ‘zero’ – ST & AM] in English object clauses is more complex than has usually been acknowledged” (p. 533). According to him, a number of factors are involved, including style, potential ambiguity, structural complexity, weight, and closeness of clause juncture.

Underhill (1988) is the first to relate the use of that to the relationship between the ‘main’ verb and the ‘complement’. Emonds (1969, 1976); Hooper and Thompson (1973), and Hooper (1975) had observed that the operation of ‘complement preposing’, as illustrated in (3a) and (3b), is determined by the semantics of the ‘main’ verb.

(3a) I believe it’s going to snow.
(3b) It’s going to snow, I believe. (complement preposed)

These studies point out that this alternation is possible only with certain verbs, those whose meaning has essentially been ‘bleached’ to the point where the phrase acts as an adverbial. We will call the use of a subject and verb phrase illustrated in (3b) its parenthetical use, after Emonds (1969) and Urson (1963).

Halliday (1985) suggests that the epistemic sense of subject + verb phrases as in (2) is in a metaphorical relationship with the main verb use, the metaphor “functioning as an expression of modality” (p. 332). He also notes certain arguments for the reanalysis cited in Emonds and in Hooper and Thompson, including the argument that the tag agrees with the apparently embedded clause, rather than with the apparently main subject and verb:

(4) I think it’s going to rain, isn’t it? (*don’t I?)

Underhill (1988) is also the first to demonstrate that the use of that is related to discourse structure; his research provided the inspiration for the study we are reporting here. He makes a strong case in favor of the following partially overlapping hypotheses for journalistic English:

(a) That is deleted when the subject of the lower sentence is the topic of the utterance; that is retained when the subject of the higher sentence is the topic.

(b) That is deleted when the writer makes or endorses the assertion of the lower sentence; that is retained when the writer does not (necessarily) endorse the assertion, but attributes it to someone else. In such a case the speaker may deny the assertion, doubt it, evade responsibility, or simply be non-committal.

Underhill is thus claiming that these discourse facts can be accounted for in terms of his hypothesis: where the embedded clause loses much of its embeddedness, its subject, rather than the main clause subject, tends to be the topic of the discourse, and its content, rather than that of the main clause, tends to be what the writer is endorsing. The main clause subject and verb, then, become secondary to the content of the lower sentence, which the writer endorses, and to the subject of the lower sentence, which is the topic.

We would suggest that this line of reasoning can be taken one step further: these bleached-out main verbs and their subjects behave very much like single epistemic morphemes in other languages to the point of being ‘transportable’

2 Emonds’ term ‘complement preposing’, was introduced at a time when grammatical phenomena tended to be described in terms of metaphors involving movement from one part of a structural diagram to another. In (3b), of course, as pointed out by Emonds, the erstwhile complement takes on the role of an independent assertion and the erstwhile main clause becomes a tag.
to positions other than that which they could occupy if they were only functioning to introduce a complement, as is shown in (3b). 3

Our study of the use of that moves away from written English to consider a large database of spoken conversational English. In addition, we have employed a quantitative methodology, which reveals patterns of use not found by earlier researchers.

2. Methodology

The transcripts analyzed for this study come from 116 8-minute recorded conversations between university students at the University of California, Santa Barbara. For the purposes of a separate series of studies (see, e.g., Mulae et al. 1987), each student was randomly paired with two individuals he or she did not know well, creating a mixed-sex and same-sex dyad. In this way, 58 male/female, 29 male/male, and 29 female/female dyads were formed. They were asked to discuss at least one of several topics provided them, such as 'What should be the University's policies on drug/alcohol abuse?' and, 'What are the advantages and disadvantages of a liberal arts education?'

The recordings were transcribed orthographically and verified by trained observers. A computerized word count showed that the nearly 15½ hours of conversation resulted in more than 240,000 words spoken.

The first author analyzed the transcripts for instances of the target construction, with or without that. 4 The second author provided a reliability check by independently coding 20% of the transcripts. The point-by-point percent of agreement between the two of us was 97%. In all, 1287 instances of the target form were identified and coded for selected parameters.

These parameters included the following:

(a) whether or not a that occurred;
(b) whether or not the main clause contained an auxiliary, an adverb, an indirect object, or a passive, or was an interrogative or a negative;
(c) whether or not the complement subject was a pronoun or was co-referential with the main clause subject.

Later, we considered to be the target construction, then, were all occurrences of 'main' subject and verb, as in (1) and (2), which could occur with that, whether or not the that was present. We also counted and coded separately all epistemic parentheticals, as in (3).

3 For discussion of the grammaticization of epistemic parentheticals, see Thompson and Mulae (in press).

4 We are using the term 'target construction' simply to refer to the 'construction in question'.

Because of their function as pragmatic expressions, 5 we did not count occurrences of you know or I mean, as in:

(5) An' I'll get everything accomplished and this and that and I may ever do good. You know but it's just all the work that you realize that you have to do before Monday morning at 8 o'clock. (14,375)

(6) B Food's terrible, huh?
A: Well, it's not actually that bad. It all depends on your, on your taste. I mean, how well, what you're used to I mean you're used to really fine cuisine at home you're not going to be able to tolerate dorm food. (15,31)

3. Terminology

For the remainder of this paper, we will refrain from using the term 'that'-deletion because it connotes an inappropriate processual metaphor. For ease of reference, we will continue to refer to 'main' and 'complement' clauses without quotation marks. These should be taken as simply mnemonic labels, however, as our central point is that the blurring of the distinction between 'main' and 'complement' clause is precisely what is involved in creating the conditions which are giving rise to epistemic parentheticals. When we use the terms 'subject' and 'verb', we will be referring to the 'main' clause subject and verb, not the 'complement' clause subject and verb. As mentioned above, we will use the term 'target' to refer to subjects and their verbs in pre-complement positions, whether or without that (illustrated by (1) and (2)), where they can be analyzed as introducing the complement.

The term 'verb type' will occasionally be used in implicit opposition to 'verb token'.

4. Results and discussion

Our findings from conversational English strongly support Underhill's hypothesis from newspaper English that the use of the complementizer that is correlated with the degree of 'embeddedness' of the complement clause. That is, when there is no that, the main clause subject and verb function as an epistemic phrase, not as a main clause introducing a complement.

We will support this claim by showing that:

1 For discussions of such discourse particles in conversation, see Goldberg 1980; Osman 1981; Schurup 1985; Erma 1986; and Schiffrin 1987.
(1) the most frequent main verbs and subjects are just those which typically occur without that, and which are characteristically associated with epistemicity;

(2) a pronominal subject in the complement clause, which is an indication of the 'topicality' of that subject in the discourse, also tends to occur without that, as suggested by Underhill’s results;

(3) other elements in the main verb phrase, such as adverbs, which reduce the likelihood that the subject and verb are functioning as an epistemic phrase, significantly favor the use of that.

Our most striking finding is that the use of that is strongly related to the use of various other elements in the utterance, as described in the discussion of methodology above. An overall chi-square analysis supports our contention of differences in the use of that depending of the presence of five eight elements (chi-square \(7, N = 1287\) = 79.77, \(p < 0.0001\)). With this result as the warrant for further analyses, we will consider each of these in turn.

4.1. Subjects

Table 1 shows that there is a much greater likelihood of omitting that when the main subject is I or second person you than when it is any other subject. 6

Table 1: Occurrence of that with I vs. you vs. other subjects.

<table>
<thead>
<tr>
<th></th>
<th>- that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>955</td>
<td>110</td>
<td>1065</td>
</tr>
<tr>
<td>you</td>
<td>55</td>
<td>6</td>
<td>61</td>
</tr>
<tr>
<td>other</td>
<td>02</td>
<td>59</td>
<td>161</td>
</tr>
</tbody>
</table>

Chi-square \(2, N = 1287\) = 79.19, \(p < 0.0001\)

Table 2 shows that I and you are also the subjects occurring with the highest frequency in the data.

These tables make is clear that the most frequent subjects in our data are also used most frequently without that.

Interestingly, as has often been noted, these are also the subjects which can express epistemicity or degree of speaker commitment (Palmer 1986: 51; Traugott 1987). A number of researchers have pointed out that speaker commitment can only be asserted for the speaker by the speaker (I), or queried of the addressee by the speaker (you). It cannot be meaningfully asserted or queried about a third person.

For example, Woodbury (1986: 192), in a discussion of evidentiality in Sherpa, states that:

“The first person vs. non-first person distinction is widespread in Sherpa, but the term ‘first person’ is something of a misnomer. In the interrogative all so-called first person phenomena are associated with second person.” (emphasis added [ST & AM])

In his discussion of subjectivity in language, Benveniste (1971: 228–229) notes that certain verbs do not have the property of "permanence of meaning" when the person of the subject is changed; in particular verbs of "mental operators", such as believe, when used with I, "convert into a subjective utterance the fact asserted impersonally".

Similarly, in his discussion of 'parenthetical verbs' in English, Urmson (1963: 222) notes:

"Part of what I design to show is how differently these verbs [i.e., what he calls 'parenthetical' verbs, e.g., think and believe – ST & AM] are used in the first person present and in other persons and tenses."

Neither Benveniste nor Urmson considered interrogative you with the verbs in question, but our data suggest that it forms a class with I in the present tense, and that the two together must be distinguished from all other persons and tenses.

In our data, 55 out of 67 (82%) of the target clause you's are in questions. In the examples below, using intonation as the criterion, (7) illustrates a target clause question, and (8), just for comparison, illustrates a target clause which is not a question:

7. We follow the standard transcription practice of indicating rising intonation with a question mark and falling intonation with a period.
(7) So what do you think you’re going to major in now that you’re down here? (108.133)

(8) (talking about dormitory regulations on alcohol)
   A: ... they probably shouldn’t have any rules at all but
   B: [laughs] you think it should be wide open.
   A: [If they’re legally responsible] (95.274)

So the fact that I and you, the subjects which can express epistemicity, are the most frequent main clause subjects is a strong factor favoring the reanalysis of these main subjects and their verbs as epistemic phrases.

4.2. Verbs

A very similar situation exists for main verbs in our data. The most frequently occurring verbs in the data are also those occurring most frequently without that, and they are also verbs heavily involved in the expression of epistemicity.

Table 3 indicates that think and guess (in the present tense) also occur significantly more frequently without that than any other verb in the data base.9

<table>
<thead>
<tr>
<th></th>
<th>− that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>622</td>
<td>61</td>
<td>683</td>
</tr>
<tr>
<td>guess</td>
<td>148</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>other</td>
<td>342</td>
<td>112</td>
<td>454</td>
</tr>
</tbody>
</table>

Chi-square (2, N = 1287) = 79.24, p < 0.0001

Table 4 shows the relative frequency of verbs; think and guess far outnumber any other verb in the data. Altogether there are 44 different target verb types in our data; it is striking that think and guess account for 65% of all the target verb tokens, while the other 42 verbs are spread out remaining 35%

Once again, it is no accident that these most frequent verbs, think and guess, are highly epistemic verbs, expressing a mid degree of speaker commitment. They are not the only verbs in the data which express speaker commitment, but it is significant that the verbs occurring most frequently without that are epistemic.10

Table 4
The most frequent ‘main’ verbs out of 44 main verb types.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| think| 683    | 53%
| guess| 150    | 12%
| other| 454    | 35%

Total 1287 (100%)

4.3. Elements of the main clause

As indicated above in our methodology discussion, we considered a range of main clause variables which we hypothesized might predict the occurrence of that. The variables can be divided into two groups, motivated by two different hypotheses.

4.3.1. Group I: Interrogation and negation

Considering negation and negation as variables was motivated by the hypothesis that the use of that is correlated with the greater syntactic complexity of an interrogative or negative verb phrase. This hypothesis was not supported by chi-square analyses; neither of these variables was significant in predicting the use of that.11

4.3.2. Group II: Other main clause elements

As mentioned in the discussion of our methodology, in addition to interrogation and negation, we also considered the following main clause verb phrase elements:

- auxiliaries
- passive morphology
- adverbs
- indirect objects

---

8 Our transcription system uses brackets to indicate overlapping or simultaneous portions of utterances.
9 Not only is the three-way comparison significant at the level of p < 0.001, but chi-square pairwise comparisons also show that think and guess are each significantly more likely (at the level of p < 0.001) to occur without that than is the complement set of all other verbs.
10 Dwight Bolinger and Gillian Sankoff have pointed out to us that the nature of our data base, conversations in which speakers were expressing opinions, may have influenced the proportions of verbs used. One might have expected more occurrences of believe, bet, figure, or suppose in other conversations, for example. We acknowledge this possibility, but note that our hypothesis is strongly confirmed by the fact that the two verbs which are far and away the most frequent in the data base are both highly epistemic and are found significantly more often without that than the other verbs in the data base.
11 In section 4.1 we showed that you in interrogatives was a predictor of the absence of that. The failure of interrogatives to be significant in predicting the occurrence of that is not inconsistent with this finding; the chi-square analysis simply shows that interrogatives and declaratives do not differ significantly in their ability to predict the occurrence of that. We hypothesize that this is because I in declaratives is as likely to occur without that as you in questions is.
The rationale for considering these variables was our hypothesis that any of these elements reduces the ability of the main clause subject and verb to function as an epistemic phrase, since they each add semantic content to the verb phrase qua main verb phrase, and thereby reduce the likelihood that that verb phrase could be taken as an epistemic phrase.

This hypothesis was strongly supported; all these variables predict the use of that to a significant degree, except for passive, of which there were only two instances. We will discuss each of the others in turn.

4.3.2.1. Auxiliaries. In this category we included any example containing either a modal or the auxiliaries have or progressive be.\(^\text{12}\) (9) is an example with the modal should:

(9) (talking about the perception that Californians know a lot of movie stars)
... people who should know you know that you don't just walk down the street and shake hands with them. \( (63.249) \)

An example of a main verb with the progressive be from our data is given in (10):

(10) ... he was saying that by the year 2000 there's, there will be, we will, the Caucasians, or the whites will be a minority as far as in California ... \( (60.312) \)

Table 5 shows that utterances whose main verb phrases contain auxiliaries are significantly more likely to occur with that than those which do not. Verb phrases with auxiliaries occur with that 27.5% of the time, but verb phrases without auxiliaries occur with that only 12.5% of the time.

Table 5
Occurrence of that with and without auxiliary.

<table>
<thead>
<tr>
<th></th>
<th>- that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>With aux</td>
<td>86</td>
<td>35</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>(71.1%)</td>
<td>(28.9%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Without aux</td>
<td>1026</td>
<td>140</td>
<td>1166</td>
</tr>
<tr>
<td></td>
<td>(85.0%)</td>
<td>(12.0%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Chi-square \((1, N = 1287) = 26.81, p < 0.0001\)

4.3.2.2. Indirect objects. In this category we included any main clause containing a noun phrase playing a recipient or beneficiary semantic role coming directly after the verb, as exemplified in (11):

(11) I wanted to show her that I ... she got the points. \( (8.257) \)

As can be seen in table 6, verb phrases with indirect objects occur with that 53.3% of the time, whereas verb phrases without indirect objects occur with that only 13.1% of the time.\(^\text{13}\)

Table 6
Occurrence of that with and without indirect object.

<table>
<thead>
<tr>
<th></th>
<th>- that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>With ind. obj.</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(46.7%)</td>
<td>(53.3%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Without ind. obj</td>
<td>1105</td>
<td>167</td>
<td>1272</td>
</tr>
<tr>
<td></td>
<td>(86.9%)</td>
<td>(13.1%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Chi-square \((1, N = 1287) = 20.4, p < 0.0001\)

4.3.2.3. Adverbs. The situation with adverbs is very similar. An example of an adverb in the main verb phrase is given in (12):

(12) Uh, I just figured that I'd walk to class at that time, just stay in the library ... \( (68.473) \)

Table 7 shows that main verb phrases with an adverb (such as just or only) occur with that 35% of the time, while main verb phrases without adverbs occur with that only 12.2% of the time.\(^\text{14}\)

Table 7
Occurrence of that with and without adverbs.

<table>
<thead>
<tr>
<th></th>
<th>- that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>With adverbs</td>
<td>52</td>
<td>28</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>(65%)</td>
<td>(35%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>Without adverbs</td>
<td>1060</td>
<td>147</td>
<td>1207</td>
</tr>
<tr>
<td></td>
<td>(87.8%)</td>
<td>(12.2%)</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

Chi-square \((1, N = 1287) = 33.26, p < 0.0001\)

4.3.3. Summary of main clause variables

We have seen that certain main verb elements, namely auxiliaries, indirect objects, and adverbs, significantly increase the use of that to introduce the complement. This fact can be explained if we hypothesize that the more likely the

\(^{13}\) Here the dependence among our variables does appear to play a role. The results in tables 5 and 6 are undoubtedly influenced by the fact that think and guess, in the relevant usages, do not take indirect objects and typically do not occur in the progressive.

\(^{14}\) Simon Dik has suggested to us that some adverbs, such as personally or strongly would in fact seem to reinforce, rather than weaken, the epistemic force of the utterance, and hence should be expected to occur with greater absence, rather than presence, of that. It is possible that this is so, but the verb phrase adverbs in our data base consisted almost entirely of qualifying, intensifying, and hedging adverbs such as just, really, and kind of.
main subject and verb are to be taken as a unitary epistemic phrase, the less likely is that to occur. All these verb phrase elements recuce 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.

4.4. Complement subject

So far, we have considered main clause elements which are relevant in the use of that. We have seen that the main clause subject, the main clause verb, and other elements of the main verb phrase play a role. Intriguingly, the choice of complement subject also figures in the use of that. When the complement clause subject is a pronoun, that is less likely to be used than when it is a full noun phrase, as shown in Table 8.

Table 8
Occurrence of that with complement subject full NP vs. pronoun.

<table>
<thead>
<tr>
<th></th>
<th>– that</th>
<th>+ that</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full NP</td>
<td>205</td>
<td>54</td>
<td>259</td>
</tr>
<tr>
<td>Pronoun</td>
<td>861</td>
<td>108</td>
<td>969</td>
</tr>
</tbody>
</table>

*Chi-square (1, N = 1287) = 16.81, p < 0.0001*

As Table 8 indicates, when the complement subject is a full noun phrase, that occurs 20.7% of the time, but when the complement subject is a pronoun, that is used only 11.2% of the time.

Why should this be so? Recall that Underhill’s hypothesis (a) above predicts that that will tend to be omitted in journalistic written English when the subject of the complement clause is the topic of the discourse. Our findings bear out this prediction for conversational English, since pronouns indicate high discourse topicality.

Our findings and Underhill’s have the same explanation. In Underhill’s data from newspaper English, complement subjects which are topics of the discourse tend to occur without that. In our data, pronominal complement subjects (which can be considered topical by virtue of being pronouns) tend to occur without that. These two facts can both be explained, once again, if we assume that the reanalysis of main subject and verb as an epistemic phrase correlates with the topicality of the complement subject in the discourse. The more the main clause subject is taken as part of a unitary epistemic phrase, the more likely is the complement subject to be the discourse topic.

4.5. Summary of results

In this section, we have seen that the use of complementizer that is highly related to the use of various other features of the discourse. First and second person subjects, the verbs think and guess, pronominal complement subjects, and auxiliaries, indirect objects, and adverbs are significant in predicting the use of that. As seemingly disparate as these factors are, their behavior finds a unified explanation in the acknowledgement that certain combinations of main clause subjects and verbs in English are being reanalyzed as unitary epistemic phrases. As this happens, the distinction between ‘main’ and ‘complement’ clause is being eroded, as suggested by Emonds and Underhill, with the omission of that a strong concomitant. That is, the more the ‘main’ subject and verb are taken as an epistemic phrase, the less the ‘complement’ is taken as a ‘complement’, and the less likely is the complementizer that to be used.

Our analyses have shown that the factors most likely to contribute to this reanalysis are precisely those which relate either to the epistemicity of the main subject and verb or to the topicality of the complement at the expense of the main clause. As we have seen, the frequency of the subjects I and you and the verbs think and guess provides a strong push towards the reanalysis because these are just the subjects and verbs most likely to express epistemic meanings: speakers assert speaker commitment with I or question it with you, and think and guess express a mild degree of speaker commitment.

Before turning to our conclusions, we would like to address one objection which might be raised at this point. Readers might wonder whether each of the variables discussed in this section independently influences the occurrence of that. That is, in assessing the role of a given variable, should we not also control for the possible influence of another co-occurring variable?

In an attempt to determine the answer to this question, we created a modified database consisting of those utterances which contained only one of the variables we have discussed. That is, if an utterance contained two or more of the variables, for example, a modal and a negative morpheme, it was not included in the modified database. The result of this ‘cleaning’ process, however, was to create a modified database containing only about 20% of the data. In other words, the great majority of the utterances in our original database contain more than one of the variables we have considered.

The modified database showed tendencies in the predicted direction, but only three of the variables influenced the use of that to a significant extent, as determined by chi-square tests.13

This exercise raises an intriguing methodological question: do we learn more about the phenomenon we are investigating if we report findings from a greatly diminished database in which variables are considered strictly independently, or do we learn more if we report findings from a full database?

13 These three were ‘auxiliaries’, ‘adverbs’, and ‘indirect objects’; these were all significant at p < 0.05, with chi-square values greater than 5.00 (Ns < 260).
with variables confounded such that it is indeterminate which one is responsible for the occurrence of *that?*

The goals of this study strongly motivate us to favor the latter option. That is, as stated at the outset, we seek to refute the claim that the use of *that* in conversational English is random or 'optional', and to ascertain the conditions under which *that* is used. We are thus interested in predictors of *that*, rather than in causes. The evidence clearly supports our claim that the use of *that* is not random, and that the conditions influencing its use include the person of the main clause subject, whether the verb is think or guess, the presence of one of the main clause variables we have considered, and whether the complement subject is a pronoun. Given that we are not interested in determining a unique factor which determines the use of *that*, we are free to utilize a methodology which does not insist on controlling for interaction among variables.

5. Conclusions

As we attempt to understand why grammatical patterns work the way they do, we must look to the way language is used by speakers in ordinary conversation.

Our study of the apparently random use of the 'complimentizer' *that* in English illustrates this in two ways.

From a methodological point of view, it is crucial to note that speakers are entirely unaware of the factors influencing their own use of *that*. This means that attempts to understand the grammar of the use of *that* cannot be based on speakers' judgments about when they use *that* or what sentences with or without *that* mean.

But even more important, it is only when we look at naturally-occurring discourse, specifically conversational discourse, that we can see distributional patterns which directly bear on the question of how the grammatical patterns we are interested in come about. As Du Bois (1985) has put it, "*grammars code best what speakers do most*". That is, grammatical patterns tend to be isomorphic with discourse patterns (Du Bois, 1987), and this is entirely natural, since the particular set of subsystems which we think of as 'grammar' can only arise from patterns in the way language is used by speakers. Our findings confirm this view of grammar as discourse-dependent in the strongest possible way. The factors determining the use of *that* have exclusively to do with such discourse and interactional parameters as expression of epistemicity and topic. Neither the grammatical 'optionality' of *that* nor the grammatical categorial blurring of the distinction between 'main' and 'embedded' clause make any sense unless we take into account the patterns in the way speakers express epistemicity and topic management in discourse.

References


Hegel, Bernd and Elizabeth Taugott, in press. Grammaticalization. Amsterdam: Benjamins.


---

16 For discussion of the 'emergence' of grammar from discourse, see Hopper (1987, 1988).