English Language and Linguistics 10.2: 371–396. © Cambridge University Press 2006
doi:10.1017/S1360674306001961 Printed in the United Kingdom

'. . . *That* is the question': complementizer omission in extraposed *that*-clauses[1]

GUNTHER KALTENBÖCK
University of Vienna
(Received 17 October 2005; revised 18 April 2006)

This article investigates the omission of the *that*-complementizer in extraposed subject clauses (e.g. *It is obvious (that) she did it*), which has so far received very little attention in the literature. Using corpus data from the British component of the International Corpus of English (ICE-GB) various potential conditioning factors are analysed. Of these the following are found to have a major impact on the choice of zero *that*: informality of the text category, type of matrix predicate, and information value of the complement clause. On the basis of these corpus results the article then proposes a unified explanation for the use of *that* or zero by positing an underlying abstract feature of 'distance' for the *that*-complementizer, a semantic residue of its original demonstrative use, which gives rise to different pragmatic interpretations, depending on the actual use of *that* in context.

1 Introduction

The question of choice between two constructional variants is a crucial one in grammatical analysis: why does a speaker choose one grammatical option rather than another? This question becomes even more interesting when there is no apparent difference in meaning between them, as is typically claimed for *that*-clauses and zero *that*-clauses (e.g. Biber, Johansson, Leech, Conrad & Finegan, 1999: 680). In fact, the choice between retention and omission of the *that*-complementizer has traditionally been treated as a prime example of free variation (Dor, 2005: 345). This view has been challenged by Bolinger (1972) and others, and will be subjected to further scrutiny in the present article.

That-clauses allowing omission of the *that*-complementizer (subordinator) may of course occur in various different constructions, with the *that*-clause assuming a variety of different syntactic functions, as illustrated in (1).

(1) (a)	She thought *(that) John was in London.*	(direct object)
(b)	The problem is *(that) the weather is not very good.*	(subject complement)
(c)	I'm glad *(that) I found you again.*	(adjectival complement)
(d)	Your criticism, *(that) other factors have not been taken into consideration*, is fully justified.	(apposition)
(e)	It's on Monday *(that) they leave*.	('relative clause' of *it*-cleft)
(f)	It's obvious *(that) she did it.*	(extraposed subject)

[1] I would like to thank Bas Aarts and the two anonymous *ELL* reviewers for their valuable comments and suggestions. I am also grateful to Nikolaus Ritt, Herbert Schendl, and Henry Widdowson for their comments on an earlier version of this paper. Special thanks are also due to the *VIEWS* team, Bryan Jenner and the MW-group, Christiane Dalton-Puffer, Julia Hüttner, Barbara Mehlmauer, Angelika Rieder-Bünemann, Ute Smit. I also wish to thank Gerald Nelson and Sean Wallis for their help with retrieving the corpus data.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

The present study investigates the choice between omission and retention of *that* in only one syntactic type, viz. extraposed subject clauses, as in example (1f), with the aim of identifying possible conditioning factors for the use of a zero *that*-clause (also referred to as 'contact clause', 'bare clause', 'asyndeton'). The reason for focusing on this constructional type is that, unlike object clauses, extraposed subject clauses have so far received hardly any attention with regard to their use of *that* and zero (cf. section 2). Formally, this construction consists of the dummy pronoun anticipatory *it* typically followed by a copula and the matrix predicate (which can be verbal, adjectival, nominal, or prepositional) and the complement clause (i.e. the extraposed subject) with or without *that*.[2]

The database for the investigation is provided by the British component of the *International Corpus of English* (ICE-GB), a one-million-word corpus of contemporary British English, including both spoken and written texts (cf. e.g. Nelson, Wallis & Aarts, 2002). This corpus yields a total of 731 instances of subject *that*-clause extraposition, of which the overwhelming majority, viz. 655 instances (89.6 per cent), take the complementizer; only 76 instances (10.4 per cent) have *that*-omission. These figures already clearly indicate that zero *that*-clauses represent the statistically marked variant (cf. Givón's 1995 criteria of markedness), and as such require particular attention in the analysis. In fact, the overwhelming predominance of *that*-connectives raises the question why *that* omission is used at all.

The article consists of three parts. After a brief overview of previous research in section 2, section 3 analyses the corpus data with regard to possible conditioning factors for the choice of the marked form zero *that*. Based on factors discussed in previous research, this part identifies informality of text category, type of matrix predicate, and information status of the complement as decisive for *that*-omission in extraposed subject clauses. These factors are shown to differ from the ones previously identified for object clauses. The third part, section 4, is mainly explanatory. It discusses the interaction of the different factors identified in section 3 and suggests a unified account by positing the underlying feature of 'distance' for *that*-complementizers.

2 Previous studies

Previous studies of the *that*/zero alternation have almost exclusively concentrated on object *that*-clauses; only Ellinger (1933), McDavid (1964), and Bolinger (1972) include brief remarks on extraposition.[3] Ellinger (1933: 106), for instance, simply observes that omission of *that* seems to be more frequent in object clauses than (extraposed) subject

[2] For a different view of anticipatory *it* cf. Kaltenböck (2003).

[3] There are also studies which have focused on the impossibility of *that*-deletion in subject position and other positions. These operate with the Empty Category Principle (ECP) (e.g. Kayne, 1981; Stowell, 1981; cf. also Doherty, 2000 for a different view) or in an optimality framework (e.g. Grimshaw, 1997; Pesetsky, 1998) and have been disregarded here. This overview of previous research also does not include accounts of logical form of indirect speech reports and related cases (e.g. Davidson, 2001; Hand, 1991, 1993; Lepore & Loewer, 1989; Segal & Speas, 1986). They are, however, referred to in section 4.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

clauses. Moreover, the database for most studies comprises written texts, mainly of American English. A notable exception in this respect is Bryant (1962: 209), who reports several studies, including an analysis of a taped interview. She observes that the complementizer occurs most often in formal written texts, zero *that* most often in informal spoken texts, with relatively formal speech falling in between.

McDavid (1964: 113), using a corpus of modern American nonfiction, notes that 'the conditions under which *that* may be omitted seem partly stylistic and partly grammatical', with omission being more common in informal writing. Of a total of 41 investigated subject extrapositions only 6 omitted *that*.

The influence of stylistic factors on the choice between the two variants has also been emphasized by Storms (1966: 262–5), who points out that *that*-omission introduces 'an element of subjectivity or emotiveness' (*ibid.*: 256), while an object clause with *that* tends 'to be less personal, less familiar, less warm, less friendly, less emotive. It is objective, factual, formal, official, sometimes tending to hostility' (*ibid.*: 262). Storms, in other words, implies an informal, spoken use for zero, as for him the amount of subjectivity is 'responsible for the distinction between colloquial, spoken English and so-called formal, written or elevated speech' (*ibid.*: 262).

Whereas most studies attempt to explain the omission/retention of *that* by pointing at register and discourse factors (see also below), Bolinger (1972) detects a genuine semantic contrast between the two types. He argues that the complementizer *that*, which has developed out of a demonstrative, has retained some of its original deictic or anaphoric meaning (*ibid.*: 10), and as such tends to be selected in contexts where it points back to some previously mentioned or known state-of-affairs. Zero, on the other hand, is more likely to be preferred in other contexts.

A detailed list of conditioning factors has been established by Elsness (1984), who explored the use of 671 object *that*-clauses in four text types of the Brown Corpus, consisting of American writing published in 1961. He identifies the following factors as favouring the choice of zero *that*: (i) informality of style, (ii) lack of structural complexity of the subject in the object clause (i.e. realization as pronoun), (iii) no deviation from the common weight distributional pattern of light–heavy in either the matrix or the object clause. Most importantly, Elsness points out that (iv) zero marks a closer link between the matrix and the object clause, which he deduces from his observation that zero is more frequent with object clauses whose pronominal subject is coreferential with the subject of the matrix clause, with definite (anaphoric) object clause subjects, and when the subject of the matrix clause or the subject of the object clause is a first or second person pronoun, e.g. *I know I'm not responsible for this*.

Thompson & Mulac (1991) studied the conditioning factors of the use of *that* in spoken English. Using a 240,000, word corpus of conversation among American college students (yielding a total of 1,287 instances of object *that*-clause), they found that zero *that* occurred most frequently with the matrix predicates *think* and *guess* in the first or second person singular and with a pronominal subject in the object clause, e.g. *I think you're right*. On the other hand, the use of *that* is linked to the occurrence of an auxiliary verb (e.g. *He was saying that . . .*), an indirect object (*I wanted to show*

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

her that . . .), or an adverb in the matrix clause (e.g. *I just figured that* . . .), and a full NP (rather than pronominal; e.g. *I know that many people loved the film*) subject in the object clause. As a unified explanation for these seemingly disparate factors they suggest that certain combinations of main clause subjects and verbs (such as *I think*) 'are being reanalyzed as unitary epistemic phrases. As this happens, the distinction between "main" and "complement" clause is being eroded . . . with the omission of *that* a strong concomitant' (*ibid*.: 249).

As part of a larger study of the variability of English complement clauses, Biber (1999) shows that omission of the *that*-complementizer is closely linked with informal register. In his corpus data the complementizer is deleted in about 85 per cent of all complement clauses in conversation; in fiction 60 per cent are deleted, in news 30 per cent, and in academic prose about 3 per cent (*ibid*.: 145).

Most recently, Dor (2005) argues that *that*-deletion in English is determined by the semantic property 'truth claim'. In this view, predicates which allow *that*-deletion are those which entail that 'a cognitive agent [i.e. typically the subject] has made an epistemic claim concerning the truth of the proposition denoted by the embedded clause' (*ibid*.: 347). The truth claim of these predicates takes the form of either a public expression (e.g. *say*, *claim*) or a mental representation (e.g. *believe*, *think*). Conversely, predicates disallowing *that*-deletion do not have this entailment. Consequently, Dor distinguishes two main classes of predicates plus an intermediate third class which subsumes predicates that do not normally entail the notion of a truth claim but whose meanings may be pragmatically or conventionally extended to approximate that of a truth claim predicate (*ibid*.: 348). In this extended meaning these predicates (e.g. manner of speaking verbs such as *whisper*, *scream*) may therefore allow *that*-deletion. Dor's distinction of essentially truth-claim predicates, which allow *that*-deletion, and non-truth-claim predicates, which disallow *that*-deletion, is confirmed by random corpus data from the *British National Corpus*. As an explanation for this correlation pattern Dor (*ibid*.: 348) suggests that *that*-clauses and bare clauses have different meanings: *that*-clauses denote propositions, whereas bare clauses denote so-called 'asserted propositions'. Only truth-claim predicates (and those which can be pragmatically used as such) are compatible with 'asserted propositions'.

Finally, two diachronic studies trace the historical development of *that* and zero complementizers. Using data from the *Helsinki Corpus*, Rissanen (1991) examines the use of *that* and zero as object clause links in Late Middle and Early Modern English and reports a steady increase of *that*-omission from 14 per cent (in the period 1350–1420) to 70 per cent (in the period 1640–1710) after the matrix verbs *say*, *tell*, *know*, *think*. As factors favouring the choice of zero he identifies a pronominal subject in the object clause, the lack of intervening elements between matrix and object clause, and occurrence in spoken texts (*ibid*.: 286). Finegan & Biber (1995) investigate the development of zero and *that* in three registers of British English between 1650 and 1990, using data from the *ARCHER Corpus*. They show that the trend – reported by Rissanen – towards an increased preference for zero in Late Middle and Early Modern English is reversed in the text types sermons, medical articles, and, eventually, also

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

letters. Their findings also confirm the conditioning factors for the use of *that* and zero identified by previous studies.

3 Analysis of possible conditioning factors

Apart from occasional references to extraposed subject clauses, all previous studies have focused on the analysis of object *that*-clauses. The factors identified for the choice of zero *that* therefore have to be reviewed for applicability in the present investigation. In the following I briefly summarize the main factors established by these studies with a view to identifying those that are of immediate relevance for our purpose. The factors quoted as favouring zero complementizer in object clauses are thus:

(a) Spoken language, informal style.
(b) Absence of intervening elements between matrix and complement clause, making explicit boundary marking (disambiguation) with *that* unnecessary.
(c) Pronominal subject of the complement clause (e.g. *I think you should leave now*).
(d) Matrix predicates are high-frequency lexemes (e.g. *think*, *guess*).
(e) The complement clause expresses information that is nonanaphoric or asserted.
(f) Matrix clause subjects are *I* or *you.*
(g) Absence of additional elements in the matrix clause (e.g. auxiliaries, indirect objects, adverbials) which reduce the ability of the matrix clause to function as an epistemic phrase by adding additional semantic content (cf. Thompson & Mulac, 1991: 246).
(h) Appropriate light–heavy weight distribution pattern in matrix and complement clause.
(i) Pronominal subject of the complement clause is coreferential with the matrix clause subject (e.g. *I know I'm late*).

Of these potential factors only (a)–(e) are immediately relevant. Factor (f) is clearly not applicable as extraposed subject clauses, by definition, only occur with a main clause subject in the third person, viz. anticipatory *it*. Factor (g) only applies in the case of a purely epistemic interpretation of the matrix clause (e.g. *I think*), and is therefore not immediately relevant here as the matrix clause of extraposed subjects has been shown to convey a wide range of speaker comment going beyond the expression of mere epistemic modality (cf. e.g. Gómez-González, 1997: 102; Herriman, 2000: 584–6). As for factor (h), a light–heavy distribution of the matrix clause is ensured by the pronominal subject anticipatory *it*; while in the complement clause it is taken care of by factor (c) (pronominal subject) and factor (b) (absence of adverbials at clause boundary). Factor (i), finally, is also not applicable with 'coreferentiality' being excluded by anticipatory *it* in the main clause. Factors (a)–(e), however, must be taken as potential influences on the choice of zero or *that* and will be investigated in turn in what follows below. Section 3.1 examines the distribution of extraposed *that*-clauses in different spoken and written text types. Section 3.2 addresses the question of intervening adverbials between matrix and complement clause. Section 3.3 looks at the syntactic category of the complement clause subject. Section 3.4 homes in on the semantic and syntactic categories of the matrix predicate, and section 3.5 investigates the information status of the complement clause. Section 3.6, finally, offers a brief comparison of the

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Table 1. *Frequency of extraposed* that- *and zero* that-*clauses in the ICE-GB text categories*

		+*that*		−*that*		Total	
Text category	No. of texts[a]	n	10,000 words	n	10,000 words	n	10,000 words
Private dialogue S1A	(100)	28	1.4	16	0.8	44	2.2
Public dialogue S1B	(80)	135	8.4	10	0.6	145	9.1
Unscripted monologue S2A	(70)	104	7.4	7	0.5	111	7.9
Spoken	(250)	267	5.3	33	0.7	300	6.0
Scripted monologue S2B	(50)	69	6.9	17	1.7	86	8.6
Non-professional writing W1A	(20)	44	11.0	2	0.5	46	11.5
Correspondence W1B	(30)	39	6.5	3	0.5	42	7.0
Academic writing W2A	(40)	95	11.9	1	0.1	96	12.0
Non-academic writing W2B	(40)	60	7.5	2	0.3	62	7.8
Reportage W2C	(20)	19	4.8	10	2.5	29	7.3
Instructional writing W2D	(20)	18	4.5	0	0	18	4.5
Persuasive writing W2E	(10)	30	15.0	3	1.5	33	16.5
Creative writing W2F	(20)	14	3.5	5	1.3	19	4.8
Written	(250)	388	7.8	43	0.9	431	8.6
Total	(500)	655	6.6	76	0.8	731	7.3

[a]A text in ICE-GB comprises roughly 2,000 words.

conditioning factors for extraposition with those identified in previous studies for object clauses.

3.1 Text types

The distribution of extraposed *that* and zero *that*-clauses across the different text categories in ICE-GB is given in table 1 above. To facilitate comparison of the different text types (which are unequal in size), the counts have been normalized per 10,000 words.[4] As already indicated in section 1, the corpus yields 655 instances of *that*-complementizer (89.6 per cent; 6.6 instances per 10,000 words) as opposed to only 76 instances of zero (10.4 per cent; 0.8 per 10,000 words). Interestingly, this overall distribution pattern is largely the same for both modes: spoken texts show a ratio of 89 per cent *that* vs 11 per cent zero (5.3 vs 0.7 instances per 10,000 words) and written texts 90 per cent *that* vs 10 per cent zero (7.8 vs 0.9 instances per 10,000 words). This can be taken as indication that omission or retention of *that* in extraposed clauses is

[4] The spoken and the written text categories are equal in size, each containing 250 texts. This differs slightly from the text classification in ICE-GB itself, which follows the delivery mode and groups 50 texts of scripted speech (S2B, 'written to be spoken') with the spoken texts. In terms of production, however, this category differs from other orally delivered texts in that it is not produced 'online', as it were, but with a degree of premeditation and planning typical of written texts. Following the arrangement of other corpora (e.g. the Survey of English Usage Corpus) 'scripted speech' is therefore grouped here with the written text category.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

fairly independent of mode, an assumption that is also confirmed by tests of statistical significance: $\chi^2 = 0.351 <$ crit. (1, 0.05).

What does seem to have an influence, however, is the text type. In the spoken texts the frequency of *that*-complementizers is unusually low in private dialogue, with as few as 1.4 occurrences per 10,000 words (as opposed to 8.4 occurrences in public dialogue). This small figure has to be seen in the light of an overall low number of extraposed *that*-clauses in this text category, but seems to suggest that the *that*-complementizer is a feature which is less compatible with informal spoken texts. By comparison, the use of zero is, in absolute terms, relatively unaffected by the formality/informality of a spoken text.[5] In the written mode four text types stand out: on the one hand, reportage and scripted monologue with unusually high figures of zero linkage, on the other hand instructional and academic writing with unusually low figures. To examine whether these fluctuations in the use of zero are linked to the frequency of the *that*-complementizer, i.e. whether an increase of zero is at the expense of *that*, figure 1 below gives the relative frequencies of each variant.

This direct comparison of zero and *that*-complementizer confirms the influence of formality/informality and the independence from mode: public dialogue and unscripted monologue conform to the general pattern found in the written mode (roughly 93 per cent with *that*, 7 per cent without), while private dialogue stands out with a ratio of 63.6 per cent *that* vs 36.4 per cent zero, i.e. a much more balanced proportion.[6] The considerable increase of zero connectives in private dialogue may be attributed to the lack of formality in this text type in comparison to the more formal style of public speech. As discussed in section 2, formality of style has variously been identified as a major factor promoting the use of a *that*-complementizer. Bryant's (1962: 209) assertion, however, that *that* appears most frequently in formal writing, zero most frequently in informal speech, with relatively formal speech falling in between, is not entirely borne out by extraposed subject clauses: while there is a clear difference between private dialogue and written texts, formal speech does not take an intermediate position but behaves exactly like formal writing.

In the written texts, too (cf. columns 4 to 9 in figure 1), there is considerable variation in the distribution of *that* and zero. The lowest proportions of zero occur in instructional writing (0 per cent), academic writing (1 per cent), nonacademic writing (i.e. popular

[5] The concept of formality/informality is, of course, difficult to pin down. In the spoken domain it can be linked to the setting in which the interaction takes place. The basic difference between private dialogues (i.e. direct conversations and telephone calls) and public dialogues (e.g. legal cross-examinations, parliamentary debates, broadcast interviews/discussions) in ICE lies in the presence/absence of an audience: in private dialogue the speakers address each other and do not speak for the benefit of anyone else. Public dialogue, on the other hand, is intended to be heard by others who are not participating in the exchange (cf. Nelson et al., 2002: 6). This entails a certain 'performance quality', which public dialogue shares with the category monologue (e.g. lectures, speeches, commentaries). Being aware of a wider audience is likely to result in a higher degree of control and self-monitoring on the part of the speaker and therefore in more 'careful' speech. It is in this highly restricted sense that the concept of formality has to be understood here.

[6] The stronger preference for 'zero' in private dialogue (compared to public dialogue) was also statistically highly significant: $\chi^2 = 12.47 >$ crit. (1, 0.001).

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

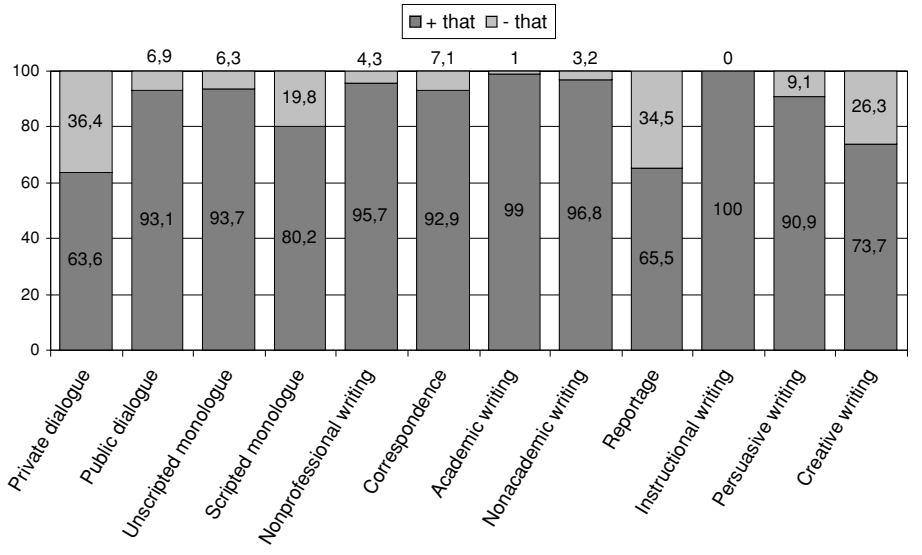


Figure 1. Relative frequency of extraposed *that*-clauses and zero *that*-clauses (in percentages) according to text categories

nonfiction) (3.2 per cent), and nonprofessional writing (i.e. student essays and exam scripts) (4.3 per cent), all of which are well below the written average of 10 per cent. The figures for instructional writing have to be taken with some caution though, owing to the small overall number of extraposed *that*-clauses in this category.[7] On the whole, however, the low occurrence of zero in these four text types can be attributed to the high degree of formality of these categories and, more specifically, the widespread claim that zero connectives are particularly rare in scientific writing (e.g. Ellinger, 1933: 107; Elsness, 1984: 520–1; Greenbaum, Nelson & Weitzman, 1996: 84; Storms, 1966: 264), even more so as the overall frequency of complement clauses in absolute terms (see table 1) is unusually high in academic writing and nonprofessional writing (i.e. student essays). With the text category correspondence, on the other hand, the formality criterion does not seem to apply at first glance: despite the generally acknowledged informality of letters (cf. Finegan & Biber, 1995: 247; Greenbaum et al., 1996: 84) the count of *that* connectives is relatively high. It is necessary, however, to differentiate between the two subcategories social and business letters, with the latter being more formal in style. Business letters yield exclusively *that* connectives, while social letters, on the other hand, contain 75 per cent *that*-clauses and 25 per cent zero *that*-clauses.

At the other end of the written scale, the text types with the highest ratios of zero connectives are reportage (34.5 per cent), creative writing (26.3 per cent), and scripted

[7] The small number of *that*-clauses (with and without complementizer) in this category is not really surprising. As a text type which focuses mainly on the performance of physical actions and processes, it is more likely to attract infinitival complement clauses rather than finite complements, which tend more towards the expression of facts (cf. Collins, 1994: 18; Mair, 1990: 25).

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

monologue (19.8 per cent). While in creative writing (i.e. novels and stories) the unusually high percentage of asyndeton can be attributed to the frequent use of direct speech in this category,[8] and hence to the factor of (in)formality, the situation is less clear with the other two categories. The style of scripted monologue, i.e. broadcast news and (broadcast) talks, as well as reportage, i.e. news reports, can hardly be classified as informal. Scripted monologue, of course, is close in style to the spoken mode, but clearly the more formal end of it, as it shows a proportion of zero connectives which is as low as the average for written texts. These categories, in other words, do not fit in a neat distinction between formal and informal. The reason for the high percentage of zero complementizers in the two text categories reportage and scripted monologue must therefore be sought elsewhere.

3.2 Intervening adverbials

Adverbials occurring at the clause boundary between matrix and complement clause may be an important factor for retaining a *that*-complementizer because of their potential to obscure the boundary between the two clauses if interpretable as (semantically) belonging to either. A *that*-complementizer clearly marks the boundary between the two constituents and as such can have an important disambiguation function.

An analysis of the 76 zero *that*-clauses in the corpus shows that there is only a single instance with an adverbial PP at clause boundary, given in (2) below, which is, however, unmistakably identifiable as part of the complement clause in semantic terms.

(2) It seems to me *in financial terms* you could have had a hard up father (s1a-075-83)

This seems to suggest that absence of intervening material between matrix and complement clause is indeed an important prerequisite for *that*-omission. Also, there is no indication that the function of explicit boundary marking is taken over by other means, i.e. in spoken language by the use of fillers, hesitation noises or pauses. There is only a single instance of an intervening filler, given in (3).

(3) It is strange *you know* so many people just don't bother to shop anywhere else for those (s1a-017-329)

However, since the overwhelming majority of *that*-clauses is also free from intervening adverbials (viz. 81.3 per cent, 594 instances), this cannot be taken as a decisive factor for the omission of *that*.

3.3 Syntactic category of the complement clause subject

It has already been pointed out above that coreferentiality of the subject of the complement clause with that of the main clause (as found by Elsness, 1984 with

[8] Cf. also Tannen's (1980: 214) assertion that written fiction, as opposed to other categories of written language, uses strategies that are typically associated with the spoken mode: 'it builds upon the immediacy function of spoken language – "imageability" and "involvement"'.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

zero *that*) is in our case excluded by the use of anticipatory *it* as (dummy) subject of the matrix clause. Consequently, the use of a pronominal subject in the complement clause to establish a close semantic link with the matrix clause is not an issue here. However, the number of pronominal subjects (as opposed to fully fledged NPs) may be indicative of yet another factor put forward by Elsness (1984) as favouring *that*-omission, which is that of a balanced, i.e. not 'nose-heavy', weight distribution in the complement clause (possible adverbials in pre-subject position have already been excluded in section 3.2). Thus one would expect a high proportion of pronominal subjects in zero *that* complements, as illustrated in (4).

(4) It's a shame *he*'s not going to be back here (s1a-042-32)

The corpus data, however, show that this is not the case: with 43 instances of pronominal subjects (as opposed to 32 instances of full NPs and 1 clausal subject) they account for only a slight majority (57 per cent) and cannot be taken as a decisive factor for the selection of zero *that*.[9] By comparison, extraposed clauses with *that* have a somewhat smaller proportion of pronominal subjects (42.4 per cent, 310 instances; 56.4 per cent NPs, 0.9 per cent clauses, 0.3 per cent PPs), but this difference may be attributed to the more frequent occurrence of this syntactic type in the written text categories (cf. section 3.1). A lengthy complement clause subject, in any case, does not automatically preclude *that* omission, as is illustrated by the following example with a clausal subject.

(5) It is admitted *what one observer has called an economic experiment which lasted for four generations after the Revolution* is slowly and painfully given up (s2bB-039-13)

3.4 Matrix predicate

As a further possible conditioning factor for the choice between *that* and zero connective let us now examine the matrix predicate governing the complement clause. With regard to its syntactic category the corpus yields the results given in table 2 below. An example of each category is given in (6).

(6) (a) It's *thought* most of the volunteers would serve in Britain helping to release other personnel (s2b-012-4)
(b) It is *probable* she will need a further operation during the period September to December of this year (s2a-062-137)
(c) Oh it's *a shame* the other thing didn't come out (s1a-091-110)
(d) So it's *to our advantage* that the Greenhouse Effect exists (s2a-043-115)

If we compare *that*- and zero *that*-clauses with regard to their choice of matrix predicate, we notice relatively high proportions of VPs and NPs with zero *that*-clauses (14.6 per cent and 16.3 per cent, respectively), and low proportions of AdjPs and PPs (5.5 per cent and 0 per cent respectively). While the high percentage of VPs is relatively stable

[9] Of the 32 NPs all but three are definite, implying givenness of the subject (cf. Elsness, 1984: 530–1). Neither this nor the predominance of pronominal subjects is, however, in any way surprising, as givenness and light weight are prototypically associated with subject (topic) position.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Table 2. *Matrix predicates governing extraposed* that-*clauses and zero* that-*clauses*

		VP	AdjP	NP	PP	Total
Spoken	+*that*	86.2%	93.0%	85.7%	100%	89.0%
(250 texts)		(112)	(107)	(42)	(6)	(267)
	−*that*	13.8%	7.0%	14.3%	0	11.0%
		(18)	(8)	(7)		(33)
Written	+*that*	85.0%	95.5%	90.6%	100%	90.0%
(250 texts)		(181)	(170)	(29)	(8)	(388)
	−*that*	15.0%	4.5%	9.4%	0	10.0%
		(32)	(8)	(3)		(43)
Total	+*that*	85.4%	94.5%	87.7%	100%	89.6%
(500 texts)		(293)	(277)	(71)	(14)	(655)
	−*that*	14.6%	5.5%	16.3%	0	10.4%
		(50)	(16)	(10)		(76)

Table 3. *Verbal predicates of extraposed* that-*clauses and zero* that-*clauses according to transitivity*

		intransitive		transitive		Total
Spoken	+*that*	60	80.0%	52	94.5%	112
(250 texts)	−*that*	15	20.0%	3	5.5%	18
Written	+*that*	60	83.3%	121	85.8%	181
(250 texts)	−*that*	12	16.7%	20	14.2%	32
Total	+*that*	120	81.6%	173	84.4%	293
(500 texts)	−*that*	27	18.4%	23	15.6%	50

in both the spoken and the written mode, the percentage of NPs is considerably lower in the written texts (9.4 per cent). In absolute terms the number of NP predicates is of course very low and therefore has to be taken with caution.

With a total of 50 instances (out of 76) it is, however, the class of VPs that accounts for the clear majority of the predicates governing zero *that*-clauses (viz. 65.8 per cent) and as such requires a more detailed analysis. In table 3 the category of VP is divided into transitive and intransitive subclasses.[10] We can see that with zero connectives the ratio of intransitive predicates is slightly higher than with transitive ones. However, a comparison of intransitive predicates governing *that* and zero *that*-clauses (cf. table 4) reveals that there are no striking differences between the two lexical sets: all lexical items used with zero are also used with *that*-connectives, even in the same order of frequency. In other words, all intransitive predicates with a zero connective could just as well have occurred with a *that*. The choice of lexical item therefore does not constitute a conditioning factor for zero linkage. Note, however, that the reverse is not necessarily

[10] The category intransitive also includes copular verbs (e.g. *seem*, *appear*).

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Table 4. *Intransitive verbal matrix predicates governing extraposed* that- *and zero* that-*clauses*

+that		*−that*	
frequency	predicate	frequency	predicate
49	*seem*	16	*seem*
20	*be*[a]	4	*be*
18	*appear*	4	*appear*
9	*follow*		
7	*turn out*	2	*turn out*
5	*occur*		
3	*happen*		
3	*emerge*	1	*emerge*
2	*transpire*		
1	*exist*		
1	*sound*		
1	*sink in*		
1	*dawn*		
Total 120		Total 27	

[a]E.g. *It may be that only slighly better quality bait will be necessary* (w2d-017-24). Examples such as this, where the matrix clause consists of *it* + *be* with *be* being modified by a modal, have been classified as marginal cases of *it*-extraposition (for a discussion cf. Kaltenböck, 2004: 56–7).

true: predicates taking a *that*-connective do not necessarily have this freedom of choice. Predicates of occurrence such as *occur*, *happen* disallow *that*-deletion. According to Dor (2005: 363), this can be explained by the fact that these predicates refer 'directly and objectively to an event in the real world' and as such do not imply a cognitive agent which has made a truth claim about this event.[11] But what about transitive predicates? Table 5 further divides them into active and passive subtypes. Here we can see that zero *that*-clauses, unlike the other variant, exclusively take passive predicates without an agent, with the relative proportion being higher in written texts than in spoken. While the lack of zero linkage with passive verbs + agent can be explained by the necessity to clearly mark the clause boundary with *that* after a prepositional *by*-phrase (cf. *It has been agreed by the parties* ***that*** *they want to part* – s2b-019-82), the use of zero with agentless passives is somewhat surprising in the light of the findings reported by Biber et al. (1999: 682–3) for object clauses, who note an exclusive use of *that* (i.e. 100 per cent) with passive main verbs in written texts (cf. also Biber, 1999: 146–7).

Taking a closer look at the occurrence of these zero *that*-clauses governed by agentless passive predicates, we can note a clear preference for this combination in

[11] Although not a typical example of a predicate of occurrence, the same explanation seems to apply for *follow*. *Turn out*, on the other hand, may look like a predicate of occurrence, but has a close affinity with reporting verbs (cf. *I have been told/informed*) and therefore seems to allow *that*-deletion (cf. the discussion of reporting verbs below).

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Table 5. *Frequency of transitive matrix predicates governing extraposed* that- *and zero* that-*clauses*

		active	passive + agent	passive – agent	Total
Spoken	+*that*	7	2	43 (93.5%)	52
(250 texts)	−*that*	0	0	3 (6.5%)	3
Written	+*that*	7	3	116 (85.3%)	126
(250 texts)	−*that*	0	0	20 (14.7%)	20
Total	+*that*	14	5	159 (87.4%)	178
(500 texts)	−*that*	0	0	23 (12.6%)	23

the text types scripted speech and reportage, which have been identified in section 3.1 as containing unusually high proportions of zero linkage: these two text categories account for all except three occurrences of this constructional type. The unusually high frequency of zero *that*-clauses in scripted monologue and reportage can therefore be attributed to the use of a special type of matrix predicate. A closer inspection of the semantic category of these agentless passive verbs shows that they are all assertive predicates (Hooper, 1975), typically reporting verbs, as illustrated in (7).[12]

(7) (a) *It is claimed* he may have been tricked into carrying the bomb on board (w2c-001-53)
(b) *It is anticipated* a final decision will be made in the New Year (w2c-011-96)
(c) *It was said* there was blood on British coal (w2c-007-96)
(d) *It's thought* most Iraqi airfields have now been knocked out (s2b-015-41)
(e) *It's understood* Margaret Thatcher made up her mind (s2b-020-8)

As a special type of evidentials (e.g. Chafe, 1986) the function of these matrix predicates is different from the ordinary matrix clause in extraposed structures:[13] rather than expressing a personal speaker comment on the state-of-affairs in the extraposed clause (e.g. assessing likelihood, difficulty, desirability; cf. Gómez-González, 1997: 102; Herriman, 2000: 587–9; Kaltenböck, 2005: 136–8), the function of these message-conveying verbs (*verba dicendi*) is simply to assert the content of the complement clause by attributing it to some undefined external source. While it is not surprising to find a considerable number of message-conveying verbs in press news reports (reportage) and broadcast news/talk (scripted speech), whose main function is precisely to report rather than evaluate or convey speaker stance, the question remains why these matrix predicates are especially prone to co-occur with zero *that*-clauses. For an explanation we need to turn to yet another possible conditioning factor for the choice between *that* and zero, viz. the information value of the complement clause.

[12] The lexical items are *say*, *think*, *claim*, *admit*, *anticipate*, *argue*, *discover*, *know*, *hope*, *understand*.

[13] The term evidential is used here in its narrowest sense as 'markers of speaker's information source' (Traugott, 1989: 32), and hence distinguished from epistemics.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

3.5 *Information value of the complement clause*

In contrast to a contact clause, the complementizer of a *that*-clause not only serves as a conspicuous clause boundary marker, it also makes the subordination more explicit. Subordination, in turn, has been shown (e.g. Mackenzie, 1984; Sadock, 1984) to correlate with backgrounding of the information conveyed by the subordinate clause. Such downgrading of the information of the complement clause is, however, not particularly compatible with text categories such as broadcast news/talk (scripted monologue) and press news reports (reportage), whose very purpose is that of conveying the facts of the complement clause. No doubt, reportability is closely linked to, even contingent on, the newsworthiness of the information.

At the same time, syntactic downgrading of the newsworthy information (i.e. the complement clause) may also have the effect of indirectly foregrounding the reporting verb of the matrix clause, which likewise is not desirable as it places undue emphasis on a reporting phrase which is highly unspecific (cf. agentless passive, impersonal *it*). Compare the adapted version of (7d) in (8).

(8) *It's thought* (that) most Iraqi airfields have now been knocked out

Clearly, what is being asserted here is the complement clause rather than the reporting clause.[14] The function of the reporting clause is simply to introduce or 'stage' the information expressed by the complement clause. It may also act as a hedging device since it attributes the information expressed by the extraposed clause to some unspecified external source (agentless passives), which reduces speaker responsibility with regard to the actual truth of the statement made by the complement clause.[15] As such, it comes close to a type of hedge referred to by Prince, Frader & Bosk (1982) as 'Attribution Shield', which weakens the relationship between propositional content and speaker as it 'attributes the belief in question to someone other than the speaker, the speaker's own degree of commitment being only indirectly inferrable' (Prince et al., 1982: 89).

Omission of *that*, on the other hand, presents the information of the complement clause as more on a par with that of the matrix clause.[16] The distinction between matrix and complement clause is being eroded, as suggested by Thompson & Mulac (1991), and both units are reanalysed as main clauses in a supplementation relation (cf. Huddleston & Pullum et al., 2002: 952), with the 'reporting clause' functioning as a parenthetical 'stager' (Hannay, 1990: 13). Like typical parentheticals the 'reporting clause' is not normally restricted to just one position but allows moving around (cf. Ziv, 1985: 181–3), as illustrated by the examples in (9) and (10).[17]

[14] Cf. Noonan (1985: 86), who makes a similar point with object complements of predicates such as *believe*, *think*, *suppose*, and *regret*.

[15] Paradoxically, rather than weakening the statement, this attribution to an undefined external 'authority' may have the effect of strengthening the statement, presenting it as an accepted fact (cf. e.g. example (7e) above).

[16] Cf. e.g. Aijmer (1972: 52), who points out that '*that*-deletion destroys the subordinate relation between the matrix and its complement'.

[17] Instances such as these have not been counted as *it*-extrapositions.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

(9) The new structures *it was said* wouldn't themselves conceal policy differences (s2b-013-22)
(10) The Government's plans to extend its rent-to-mortgage scheme to council tenants will have an adverse effect on rural areas where housing opportunities are scarce, *it is claimed today* (w2c-015-25)

The stylistic effect of *that*-omission in connection with message-conveying verbs can be seen as increasing the 'immediacy' of a text by turning indirect reported speech into something not unlike direct reported speech, with the complement clause being reanalysed as a nonembedded clause, cf. *It was said: 'They have no money'*, *It was thought: 'They might be sleeping rough'*. A direct quote, it has been claimed, 'communicates a more authentic piece of information than an indirect quote in the sense that a direct quote implies a greater fidelity to the source of information' (Li, 1986: 41). Moreover, the 'colon' style of direct reported speech also reflects the more fragmented style typical of spoken language (Chafe, 1979, 1982). Compare (11) and the adapted version in (12).

(11) It is now known ∅ they took a rucksack of clothes with them But police say ∅ they have no money and it's thought ∅ they might be sleeping rough. (s2b-009-43/44)
(12) It is now known *that* they took a rucksack of clothes with them But police say *that* they have no money and it's thought *that* they might be sleeping rough.

Omission of *that* with message-conveying verbs as an attempt to avoid explicit syntactic downgrading (backgrounding) of the complement clause can also be linked to Bolinger's (1972: 56–71) view of an implicit anaphoric function of *that* (cf. also Dor, 2005: 376). Thus, it would seem that an implied backward reference is particularly inappropriate with message-conveying verbs used for reporting news items, which, by their very nature, are supposed to be 'new' to the listener/reader. Rather than referring back to information that is already 'known', their purpose is precisely to convey information that is new, to inform of 'news'. In fact, a systematic analysis of the information status of zero *that*-complements of message-conveying verbs reveals that all invariably convey a state-of-affairs that is 'new' to the discourse in the sense that it is irretrievable from the preceding co(n)text.[18] A comparison of the information status of both constructional types, *that*-clause and zero *that*-clause, shows a clear preference

[18] According to the concept of retrievability (e.g. Firbas, 1992; Geluykens, 1991) information is 'given' if it is retrievable (recoverable) from the preceding co(n)text, and 'new' if it is not thus derivable. To account for the various degrees of givenness/newness, the present study makes use of Prince's (1981, 1992) taxonomic model of 'assumed familiarity' and adapts it to the concept of retrievability, i.e. a discourse-based model of 'discourse familiarity' (cf. also Miller, 2001). It distinguishes between two types of new (irretrievable) information, viz. 'brand-new' and 'new-anchored', and three types of givenness, viz. 'inferrable', 'textually evoked', and 'situationally evoked'. The following two examples illustrate the categories of 'textually evoked' (i) and inferrable (ii) (for a detailed discussion of the taxonomy of information status used and further examples see Kaltenböck, 2004: 158–65, 2005: 126–35).

(i) [soccer commentary] Good ball by uh Merson *Smith* heads it in but *it's offside* The flag was up very early it was a lovely ball into the box by Mason but it was uh fairly clear *that Smith was in an offside position* (s2a-015-172)

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Table 6. *Information status of extraposed complements with* that *and zero linkage*

	Given (retrievable) information		New (irretrievable) information		Total	
+*that*	98	15.0%	557	85.0%	655	100%
−*that*	6	7.9%	70	92.1%	76	100%
Total	104		627			

of zero *that*-complements for the expression of new (irretrievable) states-of-affairs (cf. Table 6). In the rare cases where 'given' complements do occur without *that*, corpus evidence shows that there are stylistic reasons for *that* omission, e.g. *It was imperative Ø you had to get back that morning that day* (s1a-028-78), where *that* suppression avoids an accumulation of three *that*s in a row. Conversely, new complements are also often introduced by *that*-complementizers, as can be seen from the high number of *that*-clauses conveying irretrievable states-of-affairs. These are particularly interesting cases as here we get a mismatch of information status and mode of presentation: the new information is not only expressed in a subordinate clause with explicit subordinator, but also as the topic (rather than comment) of the construction (cf. Lutz, 1981), and often as complement of a factive matrix predicate, which lexically presupposes its complement. This has the rhetorical effect of presenting a new state-of-affairs as if it were known, a strategy particularly useful for persuasive writing. Compare, for instance:

(13) The last thing I want to do is to put you to any more trouble personally. I am therefore prepared to write off this amount to experience although I think *it is a pity that LB is the only major corporation I have worked for where this has been a problem* ⟨/p⟩ (w1b-020-26)

Omission of *that*, on the other hand, may cancel the presupposition effect of a factive predicate (prosody permitting) and present the content of the complement clause as an assertion rather than a presupposition, as in the following example (where the complement clause introduces new information).

(14) It's a shame ∅ they haven't got anything on Bonhöffer (s1a-053-178)

3.6 Extraposed clauses and object clauses contrasted

As pointed out at the beginning of section 3, some conditioning factors identified for object clauses are not applicable to extraposed clauses because of structural differences. With regard to the remaining potential conditioning factors, extraposed subject clauses differ from object clauses in a number of ways.

(ii) *Neither we nor, almost certainly, President Bush and Mr Major know* how much of Iraq's armoury remains intact, the extent of Sadam Hussein's capacity still to impede the allied ground forces, and how many of his chemical and biological weapons have escaped destruction. It is the nature of the fog of war *that such doubts exist* (w2e-005-81)

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

One difference lies in the relative unimportance of intervening adverbials at the clause boundary and the categorial status of the subordinate clause subject. The reason for the former may lie in the clear bipartite structure of extraposition constructions signalled by their unusual comment (matrix clause) before topic (subject clause) arrangement (cf. Lutz, 1981; e.g. *It's surprising # John has moved to London*). Unlike object clause constructions, which are in line with the prototypical topic-before-comment structure (e.g. *The jury # thinks that the defendant is guilty*), the marked thematic structure of extrapositions (with its clear functional division of comment and topic) emphasizes the bipartite structure, which, in turn, may reduce the need for explicit structural boundary marking with *that*. As for the categorial status of the subordinate clause subject, we need to recall that, according to Elsness (1984), the reason for the high percentage of pronominal subjects in zero *that* object clauses lies in their coreferentiality with the matrix clause subject and the resulting closer link between matrix and object clause, which in turn favours *that*-omission. In extraposition such coreferentiality is excluded owing to the use of impersonal *it* as matrix clause subject.

A further difference between extraposed clauses and object clauses can be detected in their occurrence in different text types. While formality of the text category seems to play an important role for both constructions (at least in spoken texts), *that*-omission in extraposition is unaffected by mode (i.e. spoken–written), a factor that has been identified for object clauses. Instead, *that*-omission in extraposition is highly dependent on text type (cf. the frequent omission in scripted monologue and reportage).

Most crucially, however, *that*-omission in extraposition is favoured by a specific communicative function, viz. that of conveying new (i.e. irretrievable) information in the extraposed clause. Comparison with object clauses in this respect is not possible, as there is no systematic analysis of their information status to date. Bolinger (1972) and Dor (2005) merely speculate on the potential influence of 'anaphoric' or 'asserted' clauses when *that* is omitted or retained, without, however, carrying out a corpus-based investigation of the matter. The results provided by the present study for the information status of the extraposed subject, which are based on the concept of retrievability from the previous co(n)text, are nonetheless largely compatible with Bolinger's and Dor's views (cf. sections 3.5 and 4.1).

The corpus analysis thus shows that the conditioning factors for *that*-omission in extraposed subject clauses differ noticeably from those previously identified for object clauses. The present investigation also differs from previous studies of object clauses in that it examines the close interaction of the individual conditioning factors and provides a unified account of these seemingly unconnected features (cf. section 4).

4 Discussion of corpus results

This section tries to provide a coherent and unified explanation for the seemingly unrelated corpus findings for extraposed clauses in section 3. Section 4.1 discusses the close interaction of the conditioning factors identified in section 3 as being responsible

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

for the high percentage of *that*-omission in reportage and scripted monologue. Section 4.2 proposes a way of accounting for the different conditioning factors in a unified explanation. Section 4.3, finally, offers a brief observation on *that*-omission in a diachronic context.

4.1 Conditioning factors for that*-omission*

The above discussion of various potential conditioning factors has shown that the choice of asyndeton over *that*-linkage in extraposed subjects is largely affected by two parameters: formality of text type and a particular reporting style. As already indicated in previous studies of object clauses, formality, or rather lack of it, was confirmed as an important factor for the omission of *that*. Two text types, however, did not conform to this pattern, showing unusually high ratios of zero: reportage and scripted monologue. Mode (i.e. spoken vs written), although closely connected with formality, could not be identified as a decisive factor. Other factors investigated were the categorial status (pronominal vs fully fledged NP) of the complement subject and adverbials occurring at the boundary of matrix and subordinate clause, both of which were found not to contribute to the decision to use zero, although absence of intervening adverbials at clause boundaries seems to be a prerequisite for *that*-omission.

To investigate the high proportion of zero linkage in reportage and scripted monologue the matrix clause of zero *that*-clauses was subjected to some scrutiny and it was found that in the majority of cases (65.8 per cent) it is verbal, typically realized by a passive reporting verb without agent (of the type *It is said*). Not surprisingly, these reporting verbs frequently occur in the text types reportage (press news reports) and scripted monologue (broadcast news/talks), where they account for unusually high percentages of zero *that*-complements. The reason for *that*-omission in these cases is largely functional in that assertive predicates such as reporting verbs do not seem to square well with the syntactic backgrounding (signalled by the *that*-subordinator) of information they are supposed to report and which is typically new. Omission of *that* solves this inherent incompatibility by doing away with the subordinator and hence eroding the hierarchical syntactic difference between matrix and complement clause. This results in reanalysis of the complement clause as a nonembedded constituent in supplementation relation (Huddleston & Pullum, 2002) to the reporting frame, which in turn is reinterpreted as a separate parenthetical (as discussed by Thompson & Mulac, 1991).

Omission of *that* can therefore be seen as a direct result of the need to assert the proposition of the complement clause. This view is supported by Dor (2005) (cf. section 2), who argues that in clauses without *that*, unlike those with explicit complementizer, 'the proposition is structurally marked as an asserted proposition' (*ibid*.: 375). This, in turn, allows the speaker to mark the proposition of the complement clause as new information and as something that is taken to be true by the cognitive agent (i.e. the speaker) (*ibid*.: 375–6).

It is, however, not only the assertion of the new or newsworthy information in the complement clause that favours omission of the complementizer. *That*-deletion is also

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

facilitated by the specific type of matrix predicate found in reportage and scripted monologue. Whereas the complement clause conveys the 'information proper', the matrix clause with its reporting verb simply has an evidential function attributing the information of the complement clause to an unspecified external source (cf. the use of agentless passive and impersonal pronoun *it*). This has the pragmatic effect of reducing speaker responsibility with regard to the truth of the complement clause. The parenthetical 'matrix clause', in other words, functions as a hedge, which allows the speaker to distance him/herself from the proposition of the complement clause. This distancing effect has also been attributed to the complementizer *that*. Hand (1993), building on Davidson (2001) and Lepore & Loewer (1989), argues that the *that*-complementizer has a demonstrative-like quality.[19] This has the pragmatic effect of creating some illocutionary distance between the speaker and the complement clause (Hand, 1993: 497), supported by iconic syntactic distance. This distancing or indeed hedging effect of the *that*-complementizer, in other words, is similar to that of the reporting predicates of the matrix clause, which also reduce the speaker's commitment in relation to the assertion (cf. section 3.5). It may therefore be assumed that the hedging function of parenthetical reporting clauses makes the use of *that* redundant, as the illocutionary distance of the reporter is already signalled by the reporting clause.

4.2 A unified account

The discussion above shows that the frequent omission of *that* in the text types reportage and scripted monologue involves a variety of interacting factors: semantic (assertive predicate, avoidance of lexical presupposition of the subordinate clause), functional (information value of the complement clause), and structural (lack of explicit subordinator). It is possible, however, to unite the different factors under one common heading and arrive at a more unified explanation by positing an underlying semantic feature for the *that*-complementizer from which more specific pragmatic meanings may be inferred depending on its actual use in context.[20] As a suitable candidate for this rather abstract 'smallest common denominator' the notion of 'distance' offers itself. This characteristic has already been noted by Langacker (1991: 447) and Hand (1993: 497), but can be given a more central role as an underlying semantic feature

[19] Davidson's (2001) paratactic analysis of indirect speech interprets the *that*-complementizer as a demonstrative, viz. *Galileo said that the earth moves = Galileo said that. The earth moves*. This analysis has been extended to all propositional attitude verbs like *believe*, *think* by Lepore & Loewer (1989) and criticized as syntactically unworkable by Segal & Speas (1986) and Hand (1991). Hand (1993) resolves the incompatibility of syntactic and semantic analyses by proposing an account which is analogous to Davidson's (2001) but formulated as a pragmatic analysis.

Anecdotal evidence in favour of a demonstrative quality of the *that*-complementizer is provided by the following notice in a London pub:

(i) Please note that:
Thieves operate in this area.

[20] Although the present discussion relates to the *that*-complementizer in extraposed clauses, the proposed explanation, because of its general cognitive nature, also applies to object clauses. However, see also footnote 25.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

of the complementizer which feeds into its various instantiations in actual use. This implicit distancing effect of *that* is a feature that derives quite naturally out of a view which takes *that* to have retained some of its original meaning as demonstrative pronoun, from which it has historically developed (cf. Bolinger, 1972: 56; Hand, 1993: 497).[21] This semantic residue can be specified as follows: unlike demonstrative *this*, which signals proximity, *that* – in its original demonstrative use – indicates distance (i.e. nonproximity), both spatial and temporal. By way of metaphorical extension this distancing effect of *that* can also be construed to relate to the distance between speaker (conceptualizer) and the proposition expressed in the complement clause.[22] From this abstract notion of distance it is possible to derive the following more concrete pragmatic uses of the *that*-complementizer in actual discourse.

(a) 'Illocutionary distance' between the speaker (conceptualizer) and the proposition of the subordinate clause 'allowing the speaker to disavow any illocutionary responsibility to the embedded sentence' (Hand, 1993: 497). Accordingly, omission of *that* correlates with assertion of the complement clause (as illustrated, for instance, in section 3.5), while the use of *that* 'disposes the matrix to carry the illocutionary force' (*ibid.*: 500). This view relates to Dor's (2005), who describes the difference between *that*-clauses and bare clauses in terms of truth claim: *that*-clauses are part of a 'propositional act', where 'some proposition was acquired, or is being stored, by a cognitive agent', while bare clauses express an 'assertive act', where 'some proposition is taken by the cognitive agent to be true' (Dor, 2005: 375).

(b) 'Temporal (anaphoric) distance' in the sense that the state-of-affairs expressed in the complement clause has already been talked about and as such represents old (given, known) information, i.e. information which has already been processed prior to the present moment and is therefore more 'remote' or 'distant' than new information.[23] As argued by Bolinger (1972: 10) and Dor (2005: 376–7), *that*-clauses are particularly prone to express old information, whereas bare clauses tend to 'assert' new information.

(c) 'Emotional distance' between the speaker (conceptualizer) as an indication of a lack of personal involvement on the part of the speaker (cf. Storms' 1966: 256 mention of a lack of subjectivity or emotiveness). This use of the *that*-complementizer has sometimes been referred to as 'objectifying' the conception of the proposition expressed

[21] The distancing effect of *that* can also be linked to its function as complementizer. As pointed out by Langacker (1991: 447), '*that* itself refers to the subordinate process as a unitary entity (its e-site)', with the resulting distancing effect being 'reflected iconically in the phonological distance it introduces between the two clauses'.

[22] In this function, as well as its metaphorical extension of meaning, *that* resembles past-tense forms, which have a similar distancing effect on a temporal, social (e.g. *I wondered if* . . .), and hypothetical level (e.g. *If I was you* . . .). Note also that even in its use as demonstrative pronoun, the distancing effect of *that* extends beyond the spatio-temporal domain, including also distance in discourse (i.e. anaphoric uses) and emotional distance (e.g. *That President Bush has caused a lot of trouble*) (cf. Lakoff, 1974: 345–6).

[23] Compare, for instance, the following example of a 'given' complement:

(i) And so *there's a Natural Greenhouse Effect* which is really useful to us because if it weren't there then the surface of the earth would be at about two hundred and fifty or two hundred and sixty degrees absolute that's uh say fifteen degrees below the freezing point of water and the whole of the surface of the earth would be frozen over So *it's to our advantage that the Greenhouse Effect exists* (s2a-043-115)

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

(Langacker, 1991: 447; cf. also Bolinger, 1972: 56; Wierzbicka, 1988: 164–5).[24] In terms of cognitive grammar this refers to a conception which makes no reference to the conceptualizer (i.e. speaker). The conceptualizer, in other words, is seen as outside the scope of predication (cf. Langacker, 1991: 215–16).[25] *That*-omission, on the other hand, can be seen as a form of subjectivization which signals greater personal involvement of the speaker (cf. also the stylistic device of greater 'immediacy' noted in section 3.5).

It is clear that all these pragmatic uses of the *that*-complementizer are closely related and partly overlapping, a fact that is reflected by the common underlying semantic feature of 'distance': different contextual uses give rise to different pragmatic interpretations, all of which are, however, rooted in the same underlying abstract feature. Of course, as pragmatic realizations, the different meanings of *that* interact with other conversational forces and may be outweighed by other factors (e.g. little semantic weight of the matrix predicate may be conducive to assertion of the complement, despite the use of *that*).

The account outlined above provides an explanation for the omission of *that* with reporting clauses, but what about the factor of formality, which has also been found to represent an important conditioning factor for *that* omission? Informal spoken language has variously been characterized as exhibiting a greater degree of 'immediacy' and 'involvement' (cf. e.g. Chafe, 1982; Tannen, 1980, 1982), features that are perfectly compatible with the lack of illocutionary and emotional distance attested for *that*-omission above. As an iconic reflection (cf. Haiman, 1983, 1985) of their immediacy, informational (spoken) texts are, not surprisingly, much less hierarchical (hypotactical) in their syntactic structure, leaning more towards linearity (parataxis) (cf. e.g. Brazil's 1995 linear grammar), with speaker comment occurring more frequently in parenthetical comment clauses, rather than in the form of superordinate matrix clauses (cf. also the more fragmented style of spoken language; Chafe, 1979, 1982).

4.3 A diachronic view

As a final observation, let me briefly consider the omission of *that* from a diachronic perspective which takes parenthetical clauses such as *I think* as having developed out of a corresponding matrix clause by grammaticalization or pragmaticalization

[24] Hand (1993: 497) also speaks of objectifying the complement, but in the sense of making it 'an object of the discourse – a linguistic part of the discourse *context* – rather than part of the discourse itself', i.e. more in terms of assertion (see pragmatic use (a)).

[25] In this process of objectivization *it*-extrapositions seem to differ from object clauses (e.g. *He knows (that) she likes him*), as the matrix clause of an extraposition does not have a personal subject (cf. impersonal *it* and frequent use of agentless passives) to which the conceptualization of the subordinate clause could be attributed – a typical feature of objectivization (cf. Langacker, 1991: 447). Achieving the desired objectivization of an extraposed subject may therefore, it could be argued, depend to an even greater extent on the presence of the *that*-complementizer than in the case of object clauses. Conversely, absence of *that* may result in an even clearer case of subjectivization (despite the use of impersonal *it*), i.e. attribution of the conceptualization of the complement clause to the speaker (conceptualizer), owing to the unavailability of a main clause participant to which the conceptualization could be attributed, who could be used for construing an objectivized viewing relationship along the main clause–subordinate clause axis (cf. Langacker, 1991: 447).

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

(e.g. Aijmer, 1997; Erman & Kotsinas, 1993). *That*-omission can be seen as an intermediate stage in this grammaticalization process, which, in its final stage, produces fully 'pragmaticalized' discourse markers (such as *I mean*, *you know*). The corpus data show that it is indeed possible to construe a structural link between matrix clauses of *it*-extraposition and parenthetical clauses. As outlined in section 3.4, the majority of all matrix predicates with asyndetic link are verbal (65.8 per cent). They comprise two subsets: intransitive verbs typically denoting 'existence' or 'appearance', such as *it seems/appears/turns out,* and agentless passive reporting verbs, such as *it is claimed/was said/is thought*. Interestingly, it is precisely these two classes of verbs that are the only matrix predicates disallowing a nonextraposed counterpart, presumably because they are semantically too 'light', i.e. do not carry a high degree of Communicative Dynamism (Firbas, 1992). Compare, for instance, (15) and the adapted versions of (7) in (16):

(15) *That John went to London *seems.*
(16) (a) *That he may have been tricked into carrying the bomb on board *is claimed.*
(b) *That a final decision will be made in the new year *is anticipated.*
(c) *That there was blood on British coal *was said*.

Thus it can be speculated that severing the syntactic link between matrix and subordinate clause by *that*-omission, and the ensuing reanalysis of the former as a parenthetical, is an attempt to break up the rigid pattern of an extraposition construction that does not offer the alternative of a nonextraposed variant to accommodate, for instance, a different information structure or present a different constituent in focus or theme position. Reanalysis as parenthetical, indeed, dramatically increases flexibility of the construction in terms of mobility of the parenthetical, as can be seen in (17) and (18).

(17) (a) John, *it seems*, went to London.
(b) John went to London, *it seems.*
(18) (a) He may have been tricked, *it is claimed*, into carrying the bomb on board.
(b) He may have been tricked into carrying the bomb on board, *it is claimed.*

5 Conclusion

The present study has investigated *that*-omission in extraposed subject clauses using corpus data from the ICE-GB corpus, which revealed a noticeable difference from the previously investigated object clauses (cf. section 3.6). The analysis of various potential conditioning factors has shown, first of all, a strong influence of the text type (section 3.1) to the effect that informal texts favour the use of asyndetic links. The distinction between spoken and written mode did not prove to be a decisive factor; what counted was merely the degree of informality. Two text categories (reportage, scripted monologue), however, did not conform to this overall pattern. Other potential factors were investigated, viz. the syntactic category of the complement clause subject (section 3.2) and adverbials at the matrix/complement clause boundary (section 3.3),

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

but these were not found to affect the choice between *that* and zero, although absence of intervening adverbials at the clause boundary seems to be a necessary structural precondition for *that*-omission. What could be identified as a decisive factor for the deletion of *that*, however, was the type of matrix predicate (section 3.4), which in the two text categories in question frequently takes the form of a reporting verb in the passive (without agent; e.g. *It is said*). The investigation of yet another factor, the information status of the complement clause (section 3.5), provides a possible answer for the frequent combination of reporting clauses with asyndetic complements. As assertive predicates, reporting verbs have the function to convey ('assert') some new or newsworthy information in their complement clause. As such, they are more compatible with zero *that*-clauses, which, owing to their lack of a subordinator, do not overtly signal syntactic backgrounding and generally show a stronger propensity towards expressing new (i.e. irretrievable) information. Leaving out the *that*-complementizer can thus be seen as a direct result of the assertive function of reporting predicates together with the high information value of their complements. The ensuing erosion of the hierarchical syntactic relation between matrix and complement clause leads to an 'upgrading' of the complement clause as a nonembedded constituent and to a 'downgrading' of the matrix clause, which is reinterpreted as independent parenthetical (cf. section 4.1).

As a unifying principle for the different forces at work in these 'reporting extrapositions', it seemed appropriate to assume an underlying abstract feature of 'distance' for the *that*-complementizer, which can be seen as a semantic residue of its original status as demonstrative pronoun out of which the complementizer has evolved. Depending on the actual use of the complementizer in context, this feature will give rise to different (but related and overlapping) pragmatic interpretations identified as 'illocutionary distance', 'temporal (anaphoric) distance', and 'emotional distance' (cf. section 4.2). In reporting texts, where some degree of authorial dissociation from the state-of-affairs reported may indeed be a desirable stylistic feature implying objectivity, the distancing effect is already taken care of, as it were, by the evidential (hedging) function of the reporting clause. Retention of *that* is therefore no longer required (indeed the assertion of the complement clause – as pointed out above – makes it even undesirable).

A unified explanation along these lines is also compatible with the frequent use of *that*-omission in informal texts, a typical feature of which is their greater degree of 'immediacy' and involvement. Structurally, the greater immediacy of informal texts is iconically reflected in a tendency towards linearity (parataxis, supplementation) rather than hierarchical syntactic relations (hypotaxis).

Author's address:
University of Vienna
Department of English
Spitalgasse 2–4
A-1090 Vienna
Austria
gunther.kaltenboeck@univie.ac.at

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

References

Aijmer, K. (1972). *Some aspects of psychological predicates in English*. Stockholm: Almqvist & Wiksell.

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

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

Biber, D. (1999). A register perspective on grammar and discourse: variability in the form and use of English complement clauses. *Discourse Studies* **1**: 131–50.

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

Brazil, D. (1995). *A grammar of speech*. Oxford: Oxford University Press.

Bryant, M. M. (ed.) (1962). *Current American usage*. New York: Funk & Wagnalls.

Chafe, W. L. (1979). The flow of thought and the flow of language. In Givón, T. (ed.), *Syntax and semantics*, vol. 12: *Discourse and syntax.* New York, San Francisco and London: Academic Press. 159–81.

Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In Tannen, D. (ed.), *Spoken and written language: exploring orality and literacy.* Norwood, NJ: Ablex. 35–53.

Chafe, W. L. (1986). Evidentiality in English conversation and academic writing. In Chafe, W. L. & J. Nichols (eds.), *Evidentiality: the linguistic coding of epistemology*. Norwood, NJ: Ablex. 261–72.

Collins, P. (1994). Extraposition in English. *Functions of Language* **1**: 7–24.

Davidson, D. (2001). *Inquiries into truth and interpretation*, 2nd edition. Oxford: Clarendon.

Doherty, C. (2000). *Clause without 'that': the case for bare sentential complementation in English.* New York and London: Garland.

Dor, D. (2005). Toward a semantic account of *that*-deletion in English. *Linguistics* **43**: 345–82.

Ellinger, J. (1933). Substantivsätze mit oder ohne *that* in der neueren englischen Literatur. *Anglia* **57**: 78–109.

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.

Erman, B. & U.-B. Kotsinas (1993). Pragmaticalization: the case of *ba'* and *you know*. *Studier i modern språkvetenskap*. Stockholm: Almqvist & Wiksell. 76–93.

Finegan, E. & D. Biber (1995). *That* and zero complementisers in Late Modern English: exploring ARCHER from 1650–1990. In Aarts, B. & Ch. M. Meyer (eds.), *The verb in contemporary English: theory and description*. Cambridge: Cambridge University Press. 241–57.

Firbas, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.

Geluykens, R. (1991). Information flow in English conversation: a new approach to the given–new distinction. In Ventola, E. (ed.), *Functional and systemic linguistics: approaches and uses*. Berlin: Mouton de Gruyter. 141–67.

Givón, T. (1995). *Functionalism and grammar*. Amsterdam and Philadelphia: John Benjamins,

Gómez-González, M. Á. (1997). On subject *it*-extrapositions: evidence from Present-Day English. *Revista Alicante de Estudios Ingleses* **10**: 95–107.

Greenbaum, S., G. Nelson & M. Weitzman (1996). Complement clauses in English. In Short, M. & J. A. Thomas (eds.), *Using corpora for language research*. London: Longman. 76–91.

Grimshaw, J. (1997). Projection, heads and optimality. *Linguistic Inquiry* **28**: 373–422.

Haiman, J. (1983). Iconic and economic motivation. *Language* **59**: 781–819.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Haiman, J. (1985). *Natural syntax: iconicity and erosion*. Cambridge: Cambridge University Press.
Hand, M. (1991). On saying that again. *Linguistics and Philosophy* **14**: 349–65.
Hand, M. (1993). Parataxis and parentheticals. *Linguistics and Philosophy* **16**: 495–507.
Hannay, M. (1990). Pragmatic function assignment and word order variation in a functional grammar of English. *Working Papers in Functional Grammar* 38 (Pragmatic functions: the view from the V.U.): 1–23.
Herriman, J. (2000). Extraposition in English: a study of the interaction between the matrix predicate and the type of extraposed clause. *English Studies* **6**: 582–99.
Hooper, J. B. (1975). On assertive predicates. In Kimball, J. P. (ed.), *Syntax and semantics*, vol. 4. New York: Academic Press. 91–124.
Huddleston, R. & G. K. Pullum (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Kaltenböck, G. (2003). On the syntactic and semantic status of anticipatory *it*. *English Language and Linguistics* **7**: 235–55.
Kaltenböck, G. (2004). *It-extraposition and non-extraposition in English: a study of syntax in spoken and written texts.* Vienna: Braumüller.
Kaltenböck, G. (2005). *It*-extraposition in English: a functional view. *International Journal of Corpus Linguistics* **10**: 119–59.
Kayne, R. (1981). ECP extensions. *Linguistic Inquiry* **12**: 93–133.
Lakoff, R. (1974). Remarks on This and That. *Chicago Linguistic Society* **10**: 345–56.
Langacker, R. W. (1991). *Foundations of Cognitive Grammar*, vol. II. Stanford, CA: Stanford University Press.
Lepore, E. & B. Loewer (1989). You can say *that* again. In French, P. et al. (eds.), *Midwest studies in philosophy*, vol. XIV. Notre Dame, IN: University of Notre Dame Press. 338–56.
Li, C. N. (1986). Direct speech and indirect speech: a functional study. In Coulmas, F. (ed.), *Direct and indirect speech*. Berlin: Mouton de Gruyter. 29–45.
Lutz, L. (1981). *Zum Thema 'Thema'. Einführung in die Thema-Rhema Theorie*. (HALT 1). Hamburg: Hamburger Buchagentur.
Mackenzie, J. L. (1984). Communicative functions of subordination. In Mackenzie, J. L. & H. Wekker (eds.), *English language research: the Dutch contribution I*. Amsterdam: Free University Press. 67–84.
Mair, C. (1990). *Infinitival complement clauses. a study of syntax in discourse*. Cambridge: Cambridge University Press.
McDavid, V. (1964). The alternation of *that* and zero in noun clauses. *American Speech* **39**: 102–13.
Miller, P. H. (2001). Discourse constraints on (non)extraposition from subject in English. *Linguistics* **39**: 683–701.
Nelson, G., S. Wallis & B. Aarts (2002). *Exploring natural language. working with the British Component of the International Corpus of English.* Amsterdam and Philadelphia: John Benjamins.
Noonan, M. (1985). *Complementation.* In Shopen, T. (ed.), *Language typology and syntactic description*, vol. II. Cambridge: Cambridge University Press. 42–140.
Pesetsky, D. (1998). Some optimality principles of sentence pronunciation. In Barbosa, P. et al. (eds.), *Is the best good enough?* Cambridge, MA: MIT Press. 337–83.
Prince, E. F. (1981). Toward a taxonomy of given–new information. In Cole, P. (ed.), *Radical pragmatics*. New York: Academic Press. 223–55.
Prince, E. F. (1992). The ZPG letter: subjects, definiteness, and information status. In Mann, W. C. & S. A. Thompson (eds.), *Discourse description: diverse analyses of a fund-raising text*. Amsterdam and Philadelphia: John Benjamins. 295–325.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press

Prince, E. F., J. Frader & C. Bosk (1982). On hedging in physician–physician discourse. In Pietro, R. J. (ed.), *Linguistics and the professions. Proceedings of the Second Annual Delaware Symposium on Language Studies.* Norwood, NJ: Ablex. 83–97.
Rissanen, M. (1991). On the history of *that*/zero as object clause links in English. In Aijmer, K. & B. Altenberg (eds.), *English corpus linguistics*. London: Longman. 272–89.
Sadock, J. M. (1984). The pragmatics of subordination. In Geest, W. de & Y. Putseys (eds.), *Sentential complementation*. Dordrecht: Foris. 205–13.
Segal, G. & M. Speas (1986). On saying ðət. *Mind and Language* **1**: 124–32.
Storms, G. (1966). *That*-clauses in Modern English. *English Studies* **47**: 249–70.
Stowell, T. (1981). Origins of phrase structure. Unpublished doctoral dissertation, MIT.
Tannen, D. (1980). Spoken/written language and the oral/literate continuum. *Proceedings of the 6th Annual Meeting of the Berkeley Linguistics Society*. 207–18.
Tannen, D. (1982). The oral/literate continuum in discourse. In Tannen, D. (ed.), *Spoken and written language: exploring orality and literacy*. Norwood, NJ: Ablex. 1–16.
Thompson, S. A. & A. Mulac (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* **15**: 237–51.
Traugott, E. C. (1989). On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* **65**: 31–55.
Wierzbicka, A. (1988). *The semantics of grammar*. Amsterdam: John Benjamins.
Ziv, Y. (1985). Parentheticals and functional grammar. In Bolkestein, A. M. et al. (eds.), *Syntax and pragmatics in functional grammar*. Dordrecht: Foris. 181–99.

https://doi.org/10.1017/S1360674306001961 Published online by Cambridge University Press