Language Variation and Change, **19** (2007), 295–336. Printed in the U.S.A.
© 2007 Cambridge University Press 0954-3945/07 $16.00
DOI: 10.1017/S0954394507000117

Regional variation in the syntactic distribution of null finite complementizer

KATE KEARNS
University of Canterbury

ABSTRACT

This article explores syntactic and regional variation in the choice between declarative (nonrelativizer) *that* and zero complementizer. Using a corpus of contemporary prose from New Zealand, Australian, American, and British newspapers, the study examines complementizer choice in complements to verbs and adjectives, extraposed complements to verbs, *it*-subject constructions (*It is obvious* (*that*)), and copula constructions (*The trouble is* (*that*), *It could be* (*that*), *What matters is* (*that*), *It was only later* (*that*)). The form of the embedded subject (pronoun, short NP (noun phrase), long NP) is also taken into account. It is shown that significant regional differences in zero rates are to some extent syntactic. The New Zealand and Australian data show less inhibition of zero in clauses that are not adjacent to the clause-selecting lexical head than the American and British data.

This article reports a study of regional variation in the syntactic factors bearing on the choice between *that* and zero complementizer in nonrelative embedded clauses, as illustrated in (1)–(9).[1]

(1) *Complement to verb*
He heard (that) the pies were good at that café.
(2) *Complement to adjective*
They were sure (that) the door had not been locked.
(3) *Verb complement following an indirect object*
The firm had warned customers (that) the product was not waterproof.
(4) *Extraposed complement to verb*
He announced the same night (that) the chairman would resign.
(5) It-*subject construction*
It was a pity (that) he couldn't be here to see it.
Copula construction
(6) What we were told was (that) the council would pay.
(7) It could be (that) the soil is too dry.

I wish to thank the University of Canterbury for a Summer Scholarship in 2002/2003 and a Linguistics Department research grant in 2004, both of which greatly assisted in the collection and preparation of the data. I also thank audiences at the University of Canterbury Linguistics Research Seminar, the New Zealand Linguistic Society Conference 2005, and the Sydney University Linguistics Colloquium for comments on presentations related to this article. Any errors are entirely my own.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

(8) The trouble is (that) the soil is too dry.
(9) It was only later (that) we heard the news of his death.

The *that*/zero choice in English has been extensively studied in the complements to verbs, where there is general agreement that the zero complementizer is grammatical in principle. Because the *that*/zero choice does not carry an obvious semantic or pragmatic distinction, the phenomenon is useful for the study of extra-linguistic factors in language variation.[2] A number of factors influencing the choice of *that* or zero in the complements to verbs have been identified or proposed in corpus-based studies, and the main factors will be reviewed in more detail later.

Syntactic theorists, on the other hand, are more interested in categorical patterns of *that* and zero distribution rather than in variation, and focus on the contrast between complements to verbs and adjectives, in which a zero complementizer is grammatical in principle, and contexts in which a zero complementizer is commonly agreed to be rare or impossible, such as in the complements to nouns. Although corpus-based studies occasionally confirm that zero is excluded in certain syntactic contexts, the study of syntactic factors in zero distribution is mainly based on data provided by intuition. A goal for syntactic theory is to account for the syntactic constraints on the distribution of zero complementizer, which generally has the more limited distribution.[3] The main syntactic characteristics of contexts considered to license a zero complementizer will also be reviewed in more detail later.

The present study combines aspects of the two traditional approaches to complementizer distribution. The main focus is on a comparison of the distribution of *that*/zero choice in different syntactic contexts in addition to verb complement clauses, but the study is based on corpus data and also takes nonsyntactic factors into account, chiefly regional variation.[4] I shall present different syntactic patterns of zero distribution in data from New Zealand, Australian, American, and British English, which indicate that adjacency to a lexical head plays a greater part in licensing zero in American and British English, and a lesser part in Australian and New Zealand English.

FACTORS IN *THAT*/ZERO CHOICE IN THE COMPLEMENTS TO VERBS

The main nonsyntactic influences on the rate of zero in complements to verbs are genre, style, and channel, the form of the embedded subject, clause-initial adverbials in the embedded clause, characteristics of the matrix verb, and regional variation. I shall discuss these factors in turn and comment on how they relate to the present study.

Style, genre, and channel

It is generally proposed that less formal texts have a higher rate of zero than more formal texts (Bolinger, 1972:22; Burchfield, 1996:773, para 2; Elsness, 1984:519;

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Finegan & Biber, 1995:256, fn. 6; Huddleston & Pullum, 2002:953; Poutsma, 1929:615; Quirk, Greenbaum, Leech & Svartvik 1985:734–735; Storms 1966:262; see also Kroch & Small, 1978 for *that*/zero choice and prescriptivism). This observation is linked to a number of different ways of identifying formality, and specific corpus-based findings are often more precisely framed in terms of genre and channel. It is consistently found that conversational language has a higher zero rate than written language in general (Biber, Johansson, Leech, Conrad & Finegan, 1999:680; Bolinger, 1972:14; Finegan & Biber, 1995:242, among others), and in this case the difference in channel is considered to correlate with a difference in formality—conversation is overall less formal than written texts.

Differences in zero rate have also been reported in different text genres, with academic prose having the lowest zero rate, light fiction a fairly high zero rate, and newspaper reportage a rate intermediate between the two (Biber et al., 1999:680–681; Elsness, 1984; Finegan & Biber, 1995:242; Fowler, 1926:632; Jespersen, 1932:33). Again, these differences among genres correlate to some extent with different levels of formality.

The embedded subject

Characteristics of the embedded subject are invariably found to have an effect on *that*/zero distribution. A number of authors have found that the form of the embedded subject has a marked effect on *that*/zero choice (among others Elsness, 1984; Finegan & Biber, 1995; Hawkins, 2003; Rissanen, 1991; Rohdenburg, 1999; Roland, Elman & Ferreira, 2005; Thompson & Mulac, 1991a). The commonest finding is that the rate of zero is increased by a pronominal embedded subject, in contrast to a full noun phrase.[5] It has also been found that the length of the embedded subject is relevant, and a long NP inhibits zero more than a short NP (Elsness, 1984; Hawkins, 2003; Roland et al., 2005).

Some researchers have proposed that differences are found among different pronoun subjects. Hawkins (2003:191) suggested that a subject with nominative case marking assists in parsing the onset of an embedded clause, and so reduces the likelihood of *that* more than a noncase-marked pronoun. Elsness (1984:526) found a higher rate of zero where the embedded subject is coreferential with the matrix subject, and that 1st and 2nd person embedded subjects produce a higher zero rate than 3rd person embedded subjects.

Thompson & Mulac (1991a) suggested that a pronoun subject promotes zero because pronouns express definite reference and mark topics, thus increasing the overall topicality of the embedded clause. The proposed relevance of topicality is consistent with Elsness's (1984) observation that definite embedded subjects occur with a higher zero rate than indefinites, and also with his finding that an embedded subject which is coreferential with the matrix subject favors zero. Ferreira & Dell (2000) also found that zero was more likely to occur in speech production in cases in which the embedded subject was repeated from previous discourse or prompted by a cue to recall—both of these characteristics suggest a current or revived topic as referent.[6]

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Intervening elements

It is generally observed that elements in the matrix clause which intervene between the matrix verb and the embedded subject tend to inhibit zero (Bolinger, 1972; Elsness, 1984; Finegan & Biber, 1995; Hawkins, 2003; Jespersen, 1932; McDavid, 1964; Rissanen, 1991; Rohdenburg, 1999; Tagliamonte & Smith, 2005; Thompson & Mulac, 1991a). The main intervening elements the authors discuss are indirect objects and matrix adverbials, as illustrated in (10)–(11).

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

Elements intervening between the matrix verb and the embedded subject also include initial adverbials in the embedded clause, as in (12).

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

Properties of the matrix verb

It has long been established that individual verbs have varying rates of zero complementizer, many being biased towards zero or *that*. Epistemic verbs (*think, believe, know, suppose*, etc.) have high zero rates, and this is frequently explained in terms of the altered informational status of many uses of epistemic verbs as epistemic modifiers on the embedded clause, rather than as expressions of the main proposition (Thompson & Mulac, 1991a, 1991b; Tagliamonte & Smith, 2005; Urmson, 1963). Lists of verbs that allow the zero complementizer, grouped according to broad semantic classes, are given in Storms (1966:266–267) and McDavid (1964:109).

The zero rate with individual verbs is often claimed to correlate positively with frequency, the most frequent verbs having the highest zero rate. Bolinger (1972) claimed that verb frequency was important in predicting the appearance of zero, but cited no quantitative results to support this. Highly frequent verbs, particularly *say, think, know*, and *tell*, have been the main focus of historical studies in *that*/zero distribution (Finegan & Biber, 1995; Rissanen, 1991), possibly in part because only frequent verbs can return enough data from the available historical texts for quantitative analysis. Studies of conversational data have also noted that the verbs with the highest zero rates, particularly *think*, are also the most frequent in the data (Tagliamonte & Smith, 2005; Thompson & Mulac, 1991a). However, none of these studies had enough data for individual infrequent verbs to assess their actual frequency and zero rate.

The only study I am aware of that quantifies the frequencies and zero rates for a wide range of less frequent verbs is Roland et al. (2005). Roland et al. examined a very large corpus (142,956 tokens) of 100 verbs that may take either a nominal object or a sentential complement, as identified by Garnsey, Pearlmutter, Myers &

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Lotocky (1997). They found that verb frequency was the weakest factor of those studied in predicting *that*/zero choice, but was still significant.

Regional variation

Some regional variation has been observed in the distribution of null complementizer in the presence of *wh*-gaps (e.g., Henry, 1995, on Irish English), but regional variation in *that*/zero choice in other contexts has been little studied.[7] Staum (2005) found that *I think* occurred with a lower zero rate among speakers of the northeast United States compared with other American speakers, but she noted that this may be a result of interaction with other factors. The present study is the first that I am aware of to examine regional variation in different syntactic contexts for *that*/zero choice where *wh*-gaps are not involved.

To sum up this section, it has been observed previously that zero complementizer is more frequent in more informal registers—zero is more frequent in conversation that in written texts generally, and the lowest rate of zero is found in highly formal texts such as academic prose. The subject of the embedded clause is a factor in zero rates, and the following effects have been observed or hypothesized. Zero complementizer is more frequent with (i) a pronominal embedded subject (cf. a full NP); (ii) a shorter NP subject (cf. a longer NP); (iii) a morphologically marked nominative subject (cf. no case morphology); (iv) an embedded subject with a discourse-familiar or topical referent (cf. a new referent or no referent); (v) a 1st or 2nd person embedded subject (cf. 3rd person). Finally, material intervening between the matrix verb and the embedded subject has been found to inhibit zero complementizer. Intervening elements include indirect objects, matrix adverbials, and clause-initial adverbials in the embedded clause. Later I report how these factors were taken into account in the present study.

SYNTACTIC FACTORS IN *THAT*/ZERO DISTRIBUTION

A full discussion of the syntactic theoretical literature on zero complementizer is well beyond the bounds of this paper, and my approach here will be very focussed on present concerns. My purpose is to identify the main components of syntactic contexts that have been identified as contributing to the ungrammaticality of a zero complementizer in certain contexts, contrasted with the complements to verbs and adjectives, in which zero is generally agreed to be grammatical in principle.

One of the most consistent points of agreement is the observation that zero complementizer does not occur in the complements to nouns, as in (13).

(13) *John's belief he would win the race was misguided (Stowell, 1981:398).

Syntactic theorists continue to assume that zero in the complement to a noun is ungrammatical (e.g., Bošković & Lasnik, 2003; Pesetsky & Torrego, 2004).

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Corpus-based studies occasionally confirm this (Biber et al., 1999:645), but it is sometimes noted that zero can occur in a noun complement clause (Bolinger, 1972:19–20). Syntactic theoretical accounts simply stipulate that zero is ungrammatical in noun complement clauses, and the potential zero-licensing syntactic factors of interest in this article are not appealed to, so noun complement clauses are not included in the present study.[8]

The characteristics to be considered here are the complement status of the embedded clause, the adjacency of the embedded clause to a lexical head, and the major or minor status of an adjacent lexical head.

Complement status

Stowell (1981:398–400) pointed out that manner-of-speaking verbs (*sigh, mutter, whisper*, etc.) cannot license a null complementizer, in contrast with other verbs of saying. The contrast is illustrated in (14)–(15).

(14) Ben sighed *(that) he was sick of not getting fed.
(15) Ben suggested (that) he was sick of not getting fed.

Stowell attributed this contrast to a difference in status between the two types of embedded clauses, arguing that the manner-of-speaking verb takes a clause in apposition, whereas the clause following the other verbs of saying is a true complement. This contrast in the status of the embedded clause is also supported, Stowell argued, by different passivization possibilities, as in (16)–(17). The embedded clause following the manner-of-speaking verb *sigh* in (14) is not a complement to the verb and the corresponding passivized sentence in (16) is ungrammatical. The embedded clause with *suggest* in (15) is a true complement to the verb and the sentence can be passivized, as in (17).

(16) *That he was sick of not getting fed was sighed by Ben.
(17) That he was sick of not getting fed was suggested by Ben.

Stowell argued that these (and related data) illustrate the general principle that a null complementizer must be in a clause that is selected as an argument by a lexical head—in the terms of Government and Binding (GB) theory, a null complementizer must be properly governed.

Adjacency

Within the GB framework, Stowell's (1981) analysis also predicted that the embedded clause must generally be adjacent to the selecting head for a null complementizer to be licensed. Briefly, the proper government relation between the selecting head and the clause with the null complementizer requires the verb to assign a theta-role to the clause; theta-role assignment, in turn, requires that the verb assign case to the clause under adjacency. The importance of adjacency is illustrated in (18)–(19). The embedded clause in (18) is separated from its selecting head

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

believed by the adverbial, and the null complementizer is judged ungrammatical, in contrast with (19), where the verb and clausal complement are adjacent.

(18) They believed at the time *(that) the gate was locked.
(19) At the time they believed (that) the gate was locked.

The main exception to the adjacency requirement is the grammaticality of null complementizer in a clause that is separated from its selecting verb by an indirect object, as in (20).

(20) John told his mother (that) he'd be home late.

Stowell proposed that the class of verbs with this complementation pattern are exceptional, in that they can assign a theta-role to a nonadjacent argument, and therefore, in this case, proper government of a null complementizer does not require adjacency.

Stowell's account of the licensing of null complementizer by proper government (adopted and modified by Pesetsky 1991) was framed within GB theory. With the development of the Minimalist Program (MP) the notion of government was no longer available, and the insights of earlier work were reframed. Bošković and Lasnik (2003) proposed that a null complementizer is an affix that must adjoin to a suitable lexical head at PF (phonetic form), and that this operation requires adjacency. Like Stowell, Bošković and Lasnik proposed a lexical work-around for the problem of indirect objects, suggesting that the clausal complement with these verbs may be an IP rather than a CP, and thus no complementizer, null or otherwise, would be present to require licensing.

Note that whether null complementizer is licensed under proper government, which generally entails adjacency, as in the GB account, or null complementizer is licensed by affixation to a lexical head, which also entails adjacency, as in the MP account, the descriptive prediction is the same: Null complementizer must be adjacent to the head that selects the embedded clause, except when the intervening element is an indirect object.

Bošković and Lasnik also differ from earlier accounts (e.g., Stowell, 1981:397) in judging the null complementizer to be grammatical in *it*-subject constructions, as in (21).

(21) It is a pity__John doesn't have any friends.

Here Bošković and Lasnik (2003:538) suggested that the null complementizer at issue is able to adjoin to the adjacent lexical head (in this case *pity*). In this structure the lack of a selection relationship between *pity* and the clause is immaterial, and adjacency is the only consideration.

Finally, in both government-based and affixation-based accounts, the licensing head—the proper governor of a null complementizer or the host to a complementizer affix—must be of a major lexical category. In particular, copula

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

be is not a major lexical verb and thus it is not a potential licenser of null complementizer (Bošković & Lasnik, 2003:529, fn. 5), and this accounts for the ill-formedness of (22).

(22) *What the students believe is__they will pass the exam.

Descriptively, the main potential syntactic factors in the licensing of null complementizer (setting aside noun complement clauses) are as follows:

1. Complement status:
   Null complementizer is licensed in a complement to a lexical head.
2. Adjacency:
   Null complementizer is licensed where it is adjacent to a lexical head:
   (a) a lexical head that selects the clause; or
   (b) the predicate in an *it*-subject construction.
3. Lexical category:
   Null complementizer is licensed in the complement to or adjacent to a major lexical category, but not a minor lexical category.

The data representing these factors in the present study will be introduced in the next section.

THE CORPUS

The full corpus on which the present study is based is comprised of 8,000 tokens of embedded finite clauses, 1,000 tokens from each of the following broadsheet newspapers: the Boston Globe (US), the Los Angeles Daily Times (US), The Age (Australia), the Daily Telegraph (Australia), the Guardian (UK), the Daily Telegraph (UK), the Christchurch Press, (NZ) and the Sunday Star Times (NZ), in the period 2002–2004. All the embedded finite clauses found were collected, except for relative clauses and clauses with *wh*-gaps. For each embedded clause, the full containing sentence was collected. The data collection and initial coding were carried out by research assistants, and all codes were checked by the author. As noted earlier, noun complement clauses were not included in the present study, because the syntactic factors of interest here are not predicted to be involved in noun complement clauses.[9] Also excluded from the present study were embedded clauses in noninitial conjoined positions or parataxis, as illustrated in (23)–(24).

(23) "It is absolutely my commitment that no company gets treated like a dinosaur, [that every job is fought for] and [no job becomes extinct]," Romney added. *Boston Globe*, October 8, 2002.

(24) The plan is that roaster companies pay farmers a decent price; [that stocks of coffee are cut partly by destroying at least 5 million bags in an effort funded by rich countries and roasters]. *Christchurch Press*, September 23, 2002.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Constructions like (23)–(24) were omitted because the syntactic status of the embedded clauses is not entirely clear, and their significance for the issues under discussion is difficult to judge.

Coding for nonsyntactic factors

Following the initial preparation of the corpus, and before proceeding to an analysis of the main syntactic construction types, preliminary analyses were carried out to assess the effects of the proposed nonsyntactic factors in *that*/zero choice as reviewed earlier. The issues identified were style, genre, and channel, properties of the embedded subject, the presence of indirect objects, matrix adverbials between the main verb and a verb complement clause, clause-initial adverbials in the embedded clause, and properties of the matrix verb. I will address these issues in turn.

Style, genre, and channel. Data were taken from a single major genre of broadsheet newspaper prose. Data were further identified for three textual subtypes: quoted direct speech, general reportage and columns, and letters to the editor. These textual subtypes were found to have different overall rates of zero: quoted speech, 65%; news and columns, 58%; and letters, 27%.

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

Of the three textual subtypes, quoted speech comprised 18% of the data, letters 5%, and news and columns 77%. Given the relatively small numbers of quoted speech and letters tokens, it would not be possible to carry out multifactorial analysis including all three types.[10] It was decided to exclude quoted speech and letters from further analysis. Consequently, the zero-promoting effect of quoted speech and the zero-inhibiting effect of letters are excluded.

The embedded subject. Recall that several properties of the embedded subject have been observed or hypothesized to influence the *that*/zero choice, as cited below.

1. A pronoun embedded subject promotes zero in comparison with a full NP.
2. A shorter NP subject promotes zero in comparison with a longer NP subject.
3. A subject marked with morphological nominative case may promote zero in comparison with a noncase-marked subject.
4. A topical or discourse-familiar embedded subject promotes zero in comparison with a nontopical subject.
5. A 1st or 2nd person subject promotes zero in comparison with a 3rd person subject.

A preliminary examination of the potential factors mentioned in points 1–5 was carried out on a subset of the data. To exclude the effects of syntactic construction, only verb complements adjacent to the verb were included. In addition, for reasons

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

discussed below under properties of the matrix verb, complements to *say* were excluded, and tokens in which an adverbial in the embedded clause preceded the pronominal subject were also excluded. I review the results briefly here, and will show that only one of the proposed distinctions was evident in the present data.

Length of the embedded subject. A preliminary calculation of the relative zero rates with embedded subject pronouns, short NPs (1–2 words, not including pronouns), and long NPs (3 or more words) revealed significant differences, including a significant difference between short and long NPs ($p \leq .0000$, Pearson's chi square Goodness of Fit (GFX)).

pronoun embedded subject 73% zero
short NP embedded subject 51.4% zero
long NP embedded subject 38% zero

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

Nominative case. Nominative embedded subjects (*he, she, we, they*) were compared with noncase-marked pronominal subjects (*you, it, this*).[11] The nominative subjects ($n = 626$) occurred with a 76% zero rate, slightly *lower* than the nonnominative subjects ($n = 319$) zero rate of 80%, although the difference was not statistically significant (Pearson's GFX, $p \leq .2341$). Roland et al. (2005:25) reported a similar finding. The presence or absence of nominative case marking on pronominal embedded subjects was not retained as a factor in the final analysis.

Topicality. The general issue of topicality was tested in two ways. First, embedded subject pronouns *he* and *she* were coded for coreferentiality with the matrix subject. Coreferential *he/she* ($n = 166$) and noncoreferential *he/she* ($n = 123$) occurred with the same zero rate of 69%. This is consistent with results from Finegan & Biber (1995:254–255) and Cacoullos & Walker (2003), who also found that coreferentiality does not affect zero rates.

Second, the potential relevance of definite reference or topicality was also tested by comparing zero rates with embedded subjects anaphoric (referential) *it* and dummy (nonreferential or cataphoric) *it*. Examples from *The Age* (Australia) are given in (25)–(28).

(25) Meanwhile, in the Federal Court yesterday, the Australian Securities and Investments Commission insisted that *it* did not oppose the administrators' decision to let Ansett trade while they settled the Tesna deal even though Ansett was suffering hefty losses. February 9, 2002. (*it* refers to the ASIC)

(26) Air New Zealand said yesterday it was prepared for a "Qantas attack" on its domestic market after the Australian international airline confirmed_*it* had plans to expand in New Zealand. February 5, 2002. (*it* refers to Qantas)

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

(27) Sports hack Stuart (Iain Glen) and financial whiz Cara (Sharon Small) realise that *it*'s not just the newspapers' future that is at stake. February 6, 2002. (nonreferential or cataphoric *it*)

(28) The parents defied the ban, insisting_*it* was their religious obligation and personal right to maintain the modest dress code, generally applied after puberty, for their young daughters. February 6, 2002. (nonreferential or cataphoric *it*)

Referential *it* ($n = 249$) occurred with 78% zero, slightly higher than the 71% zero rate with dummy *it* ($n = 79$), although the difference was not significant ($p \leq .2012$, Pearson's GFX). Topicality of the embedded subject was not retained as a factor in the final analysis.

Person features of the embedded subject. The possible effect of the person of the embedded subject was tested. Embedded 1st and 2nd person subjects (*I, we, you*: $n = 66$) were compared with all 3rd person pronominal subjects ($n = 358$).[12] Both groups of pronouns had the same zero rate of 70%. I note that Elsness's results on the effect of the person features of the embedded subject do not explicitly distinguish between 3rd person pronouns and 3rd person noun phrases. Consequently, the zero-promoting effect of *I, we, you* observed by Elsness may be due to the independently established effect of embedded subject pronouns in comparison with full noun phrases, assuming that noun phrases were included in his 3rd person subjects. The person of the embedded subject was not retained as a factor in the final analysis.

To sum up, none of the potential differences among pronominal embedded subjects was found to significantly affect the rate of zero complementizer in the present data. In what follows, distinctions will not be maintained among pronominal embedded subjects.

Intervening elements. Elements intervening between the selecting head (typically the matrix verb) and the subject of the embedded clause include indirect objects, matrix adverbials, and clause-initial adverbials in the embedded clause. Verbs that select an indirect object and a clausal complement are treated as an exceptional lexical class in the syntactic theoretical approaches previously reviewed, and are coded as a separate construction type in the present study.

Verb complement clauses separated from the head by an intervening matrix adverbial are predicted by syntactic theory not to be possible environments for zero, as the complementizer position is not adjacent to the selecting head. Clauses of this type, traditionally called extraposed clauses, are also coded as a separate construction type in this study.

Embedded clauses with a clause-initial adverbial were compared with clauses in which the subject is the first element of the embedded clause. The clause-initial adverbial was found to inhibit zero significantly ($p \leq .0000$, Pearson's GFX): clause-initial adverbial occurred with 8% zero, and clause-initial subject with 55% zero. Embedded clauses with a clause-initial adverbial constituted only 5% of the data, and given the small number of tokens, it would be impossible to include them in a multifactorial analysis. Accordingly, tokens with a

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

clause-initial adverbial in the embedded clause were excluded from the final analysis, and the zero-inhibiting effect of these adverbials is excluded.

Properties of the matrix verb. Data for the present study were identified and coded manually,[13] and it was not possible to collect a large enough corpus for the effects of lexical type frequency to be consistently calculated. Preliminary analysis of verbs adjacent to their clausal complements showed that *say* was by far the most frequent of all verbs in the corpus (38% of all verb complement tokens), and also had an extremely high zero rate of 90%, in comparison with the overall zero rate of all other verbs of 57%. However, *say* was not equally frequent in the data from the regional sources. Of all (adjacent) verb complements, *say* tokens comprised 39% of Australian data, 44.5% of American data, 46% of New Zealand data, but only 28% of British data. If the *say* tokens are included in the data representing the complements to verbs, the strong zero-promoting effect of *say* is under-represented in the British data in comparison to the other varieties. For this reason the *say* tokens were removed from the general verb complement data for the final analysis.

Overall, the main purpose of the study is to compare zero rates across constructions and regional sources, so the main issue is whether or not the most frequent lexical types are fairly evenly distributed across the compared groups. I will comment on this issue as it arises in the presentation and discussion of results.

Summary. To sum up, the influence of textual subtypes within the genre of newspaper prose was reduced by removing two less frequent subtypes: quoted speech, which had a higher zero rate, and letters to the editor, which had a low zero rate. The strong zero-inhibiting effect of clause-initial adverbials in the embedded clause was countered by removing these tokens. Proposed contrasts in zero rates among different types of pronominal subject were found not to be significant in the present data, and pronominal embedded subjects were coded alike as "pronoun." The proposed effect of embedded subject length (pronoun, short NP of 1–2 words, long NP of 3 or more words) was found to be significant, and embedded subject length was retained as a factor. The very strong zero-promoting effect of *say* with complements to verbs was unequally represented across regional sources, and *say* tokens were removed to counter the potential distortion of zero rates for complements to verbs. The remaining corpus comprised 3,145 tokens. Tokens were also coded for the regional source, Australia, New Zealand, America, and Britain.

Coding for syntactic factors

All tokens were coded into six major construction types, identified in accordance with the syntactic licensing of zero complementizer, discussed in the earlier section titled "Syntactic Factors in *That*/Zero Distribution". Next, these constructions are discussed further.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Complement to verb. Complements to verbs included all complements to verbs (other than *say*) that immediately followed the selecting verb. In the majority of tokens the embedded clause was a complement to an active verb, as in (29).

(29) But Oxfam's research suggests_we're drinking deep from the cup of inequity. *Christchurch Press*, September 23, 2002.

Complements to passive verbs or raising verbs with dummy *it* subjects as in (30)–(31) were initially coded separately.

(30) Although it was assumed_McCarty had played his last game for Boston when the Celtics lost in the Eastern Conference finals, the versatile forward signed a one-year, veteran-minimum contract worth $762,435. *Boston Globe*, October 11, 2002.
(31) It seems *that* every nation but the English has a one-pot stew. *The Guardian*, January 4, 2003.

These tokens were then compared with active verb complements like (29) and with other *it*-subject constructions as discussed in the subsection below on *it*-subject construction. It was found that complements to raising and passive verbs did not differ significantly in zero rates from complements to active verbs. Raising/passive verbs with *it* subjects had a zero rate of 57%, similar to the 59% zero rate with active verbs. On the other hand, embedded clauses in the remaining *it*-subject constructions had a significantly lower zero rate of 18%. Syntactic theory also classes examples like (30) and (31) with examples like (29), in that the clause is analyzed as the selected complement to the verb in all these structures. Accordingly, complements to passive and raising verbs with *it* subjects were coded as verb complements.

Complement to adjective. Complements to adjectives included basic adjectives, as in (32), and also participial adjectives, as in (33).

(32) Are you watching BBC? sang the fans, still angry *that* the match hadn't been televised. *Daily Telegraph* (UK), January 6, 2003.
(33) He was particularly annoyed *that* his elderly mother in Sweden became embroiled. *The Guardian*, October, 11, 2002.

As with raising verbs, complements to raising adjectives with *it* subjects, as in (34), were initially coded separately, and were then compared with both nonraising adjectives and other *it*-subject constructions.

(34) It's unlikely_the Andersen brand would survive. *The Age*, February 9, 2002.

The raising adjectives had a zero rate of 44%, closer to the overall 55% rate for complements to other adjectives than to the 18% zero rate for the remaining *it*-subject constructions. As was noted for passive and raising verbs with *it*

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

subjects, syntactic theory analyzes examples like (32)–(33) and (34) alike, in that the embedded clause is a selected complement to the adjective. Accordingly, examples like (34) were coded as adjective complements.

Complement to verb preceded by an indirect object. The syntactic theories reviewed briefly earlier identify verbs that select an indirect object and a clausal complement, as in (35), as a special case. Although the embedded clause is not adjacent to the selecting head, zero complementizer is considered to be grammatical. In contrast, zero complementizer is generally considered to be ungrammatical in a complement clause that is separated from its selecting head by a matrix adverbial (see the following subsection).

(35) Lawyers acting for Lundy advised the court_he did not contest Mrs. Weggery's application. *Christchurch Press*, September 25, 2002.

Complements to passive verbs, as in (36), were also included in this group. In a transformational analysis the surface subject is moved from the position immediately following the verb, leaving a trace that intervenes between the verb and clause complement.

(36) All were told_they were going to Australia. *The Age*, February 5, 2002.

Extraposed complement to verb. In traditional syntactic discussion a number of different sentence types may be described as instances of extraposition, including a range of sentence types with a dummy *it* subject. However, sentences that may be described as extraposition may differ in the general syntactic factors of interest here, particularly in whether or not the embedded clause is a complement selected by a major lexical head. To focus on the issue of adjacency to the selecting head, I have included in this class only verb complements that are separated from the verb by a matrix adverbial, as illustrated in (37).

(37) Rosenthal said yesterday_he is convinced authorities could have caught the sniper by now if they had been able to match spent projectiles from the first shooting to a ballistic fingerprint database of newly made guns. *Boston Globe*, October 24, 2002.

It-subject construction. As discussed, the complements to passive and raising verbs with *it* subjects, and the complements to raising adjectives with *it* subjects, were coded respectively as verb complements and adjective complements. The remaining *it*-subject constructions included only constructions in which the clause is interpreted as the subject of the main predicate, as illustrated in (38)–(39).

(38) The dust is an irritant so it's possible_it could aggravate existing respiratory problems. *Los Angeles Daily News*, November 28, 2002.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

(39) It is evident *that* other passengers feel the same way. *Daily Telegraph* (Australia), January 17, 2004.

Copula constructions. Copula constructions include several types of constructions in which the main verb of the matrix clause is copula *be*. In the first type, illustrated in (40)–(41), the embedded clause gives the content of a fact- or proposition-denoting nominal, which is the matrix subject.

(40) The only problem is_certain individuals don't quite live up to your hopes for them. *Daily Telegraph* (Australia), January 1, 2002.
(41) Bucksbaum said a sure sign that 2002's batch of films are being received better overall is *that* more of the films have what the industry calls "legs" meaning their popularity is sustained week after week. *Los Angeles Daily News*, November 28, 2002.

In the second type of copula construction, illustrated in (42), an *it*-subject frame serves to express a modal modification on the embedded clause.[14]

(42) It may be *that* you need to follow this military strategy if you expect to do any entertaining this fall. *Boston Globe*, October, 8, 2002.

Copula constructions also include *wh*-clefts, as in (43)–(44). The predicational structure of these examples is similar to that of (40)–(41), in that the *wh*-clause matrix subject refers to a proposition that is spelled out in the embedded clause.

(43) What we didn't say was_this was a policy we wanted to impose. *The Guardian* (UK), November 22, 2002.
(44) What people haven't been told is *that* this delay was the result of a decision to expand the catalogue to become a major and definitive publication of McCahon and his work. *Sunday Star Times* (NZ), October 27, 2002.

Finally, copula constructions include *it*-clefts, as in (45)–(46). Here the *it*-subject format focuses an adverbial which is interpreted within the embedded clause.

(45) It is here_he has practised three days a week since April and it is at the centre of the dispute. *Sunday Star Times* (NZ), October 27, 2002.
cf. He has practised here three days a week since April.
(46) But it was under Mr. Helms *that* the first cracks in the agency's reputation appeared. *Boston Globe*, October 24, 2002.
cf. The first cracks in the agency's reputation appeared under Mr. Helms.

For these constructions, the question arises (see also endnote 14) as to whether they should be classified as *it*-subject constructions, which represent the potential zero-licensing factor proposed by Bošković & Lasnik—that is, the null complementizer can be licensed by affixation to the adjacent matrix predicate even though the embedded clause is not a complement to the predicate. Bošković & Lasnik (2003) do not specifically discuss *it*-cleft constructions. However, their analysis for *it*-subject constructions appeals to the idea that the matrix predicates

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 1. *Syntactic construction types in the study*

	Clause is selected complement to a head	Clause is adjacent to a head	Selecting and/or adjacent head is major lexical category
Complement to verb (other than *say* or with indirect object)	Yes	Yes	Yes
Complement to adjective	Yes	Yes	Yes
Complement preceded by indirect object	Yes	No	Yes
Extraposed complement to verb	Yes	No	Yes
It-subject construction	No	Yes	Yes
Copula construction	No	Yes (except some *wh*-cleft)	No

in *it*-subject constructions form a special class, and that this underlies their ability to host an affixal null complementizer. The generalization does not extend to *it*-clefts, in which the expression following the copula (*here* in (45), *under Mr Helms* in (46)) is a focused element construed within the embedded clause.[15]

Overall, the examples I have grouped as copula constructions share the characteristic that the embedded clause is not a complement or subject of any major lexical predicate, but is nevertheless a required component of the sentence type (e.g., it is not an optional adjunct).

The six construction types and the syntactic factors they represent are summarized in Table 1.

ANALYSIS AND MAIN RESULTS

In this section I present three groups of results. The first subsection reviews the results for each of the factor groups of regional source, construction type, and embedded subject type. The results show that the overall zero rates differ significantly across regional sources with the ranking: New Zealand > Australia > United States > Britain. In addition to the canonical syntactic licensing contexts, zero complementizer also occurs fairly frequently in contexts in which it is predicted not to be licensed. Nevertheless, the overall ranking of construction types by zero rates is broadly consistent with the syntactic licensing factors that the constructions represent. The highest zero rates occur in cases in which all licensing factors are present—the embedded clause is adjacent to and complement to a major lexical head (verb or adjective). The lowest zero rates occur in cases in which none of the licensing factors are present—in the copula construction the embedded clause is not adjacent to nor complement to a major lexical head. The predicted effect of the embedded subject type is robust, and appears across construction types.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

The second subsection examines the regional differences in zero rates for the construction types individually, taking into account the possible effects of embedded subject type and lexical type. The overall ranking by zero rates (New Zealand > Australia > United States > Britain) is *broadly* preserved across the separate constructions. New Zealand is always in the group of regional sources with the highest zero rate, and Britain is always in the group of regional sources with the lowest zero rate. If Australia and the United States differ, the difference is that Australia has a higher zero rate than the United States. However, the individual construction types do not show the same patterns of differences in zero rates among regional sources. The constructions differ in whether or not the regional sources do have significantly different zero rates at all, and if so, how the regional sources are grouped. This suggests that the regional varieties do not group the constructions by zero rate in the same way—the syntactic factors influencing zero rates may operate differently across regional sources.

This issue is explored in the third subsection, in which the within-region patterns of zero rates across construction types are identified and compared. Here I show that in cases in which the embedded clause is not adjacent to a potential licensing head, the zero-inhibiting effect of nonadjacency is weaker in the New Zealand and Australian data than in the United States and British data.

The factors of regional source, construction type, and embedded subject

A logistic regression analysis was performed using GOLDVARB (2001) software (Cedergren & Sankoff, 1974) for the factors of construction type, embedded subject, and regional source, with the results shown in Table 2. All factors were significant.

Regional source. The regional sources for the data showed differences in zero rate: New Zealand 62%, Australia 51%, United States 40%, and Britain 26%. All differences were significant ($p \leq .0000$, Pearson's GFX).

The question arises as to whether the differences here may reflect differences in style among the sampled newspapers; for example, *The Guardian* and *The Telegraph* may be more formal than the other newspapers sampled, and this might partly account for the lower zero rate in the British data. Differences in formality within a main register type such as broadsheet newspaper reportage are difficult to examine in the absence of finely tuned independent measures of formality. One possible measure of formality is the frequency of contraction of auxiliary verbs and negation (*he's, isn't*).[16] To check for evidence of differences in formality, the data were coded for the contraction of contractible auxiliaries and *not*, and the rate of contraction for each regional subcorpus was calculated. The results – 0.06% of New Zealand tokens with contraction, Australia 9.3%; United States 11.7%, and Britain 6.29% – do not produce the same ranking as the overall zero rates. Although the New Zealand data have the highest zero rates, they have a very low rate of contraction, and the United States data have a higher contraction rate than the Australian data, but a lower zero rate. I conclude

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 2. *Results of logistic regression analysis*

		No. of tokens	Zero rate	Factor weight
Construction type	Verb complement	2179	55%	0.596
	Adjective complement	176	55%	0.604
	Complement preceded by indirect object	207	48%	0.457
	Extraposed verb complement	271	31%	0.289
	It-subject construction	106	20%	0.214
	Copula construction	206	9%	0.080
Embedded subject	Pronoun	1237	62%	0.698
	Short NP	1275	39%	0.448
	Long NP	1111	29%	0.320
Regional source	NZ	864	62%	0.681
	AU	914	51%	0.624
	US	753	40%	0.447
	GB	1092	26%	0.289

that the regional differences in zero rates in the present data do not primarily reflect different levels of formality.

Construction types. The zero rates across construction types show that a clear-cut distinction into grammatical and ungrammatical contexts for zero is not easy to maintain. Syntactic theory predicts that zero complementizer is not licensed in extraposed complements to verbs, but the overall zero rate for this construction is 31%. The 31% zero rate in extraposed complements to verbs is actually higher than the 20% zero rate in *it*-subject constructions, although Bošković & Lasnik (2003) predicted that zero is ungrammatical in extraposed clauses and grammatical in *it*-subject constructions. Syntactic theory also predicts that zero complementizer is not licensed in copula constructions. The observed zero rate of 9% for copula constructions is the lowest for all the construction types, but still high enough to indicate that zero is not categorical in this type of context.

Nevertheless, the relative zero rates give a ranking that does reflect the syntactic factors proposed in the licensing of zero. The three construction types with the highest overall zero rates (complement to verb 55%, complement to adjective 55%, and complement preceded by an indirect object 48%) are all instances of an embedded clause that is a selected argument of a major lexical head. The lowest zero rates are found in copula constructions (9%), which is consistent with expectations, given that copula constructions contain no suitable lexical head that might serve as a licenser of zero. In all the copula constructions except *wh*-clefts, the null complementizer is adjacent to the copula, but the copula is not a major lexical category and cannot license zero.

The intermediate frequency of zero in extraposed complements to verbs (31%) and in *it*-subject constructions (20%) suggests that lack of adjacency or lack of complement status both have an effect. This issue is explored further in the subsection on varietal patterns of potential syntactic zero-licensing factors.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Embedded subject. The effect of the embedded subject is consistent with previous findings. A pronoun embedded subject strongly promotes zero, and a long NP inhibits zero more than a short NP does. The difference between short and long NPs is significant ($p \leq .0000$, Pearson's GFX). This supports the view that the embedded subject effect depends largely on the length of the embedded subject, and I shall assume Hawkins' (2003) account of this phenomenon in terms of processing efficiency.

Briefly, Hawkins (2003) argued that the parsing load in recognizing embedded clauses is proportionate to the amount of material from the left boundary of the clause that must be processed before the clause is unambiguously recognized as an embedded clause. If the finite complementizer *that* is present, this may promote immediate recognition of the embedded clause onset. Alternatively, in cases in which *that* is absent, an embedded finite clause is recognized when the subject and finite verb are processed. The delay in reaching the finite verb is proportional to the length of the embedded subject[17]. In cases in which the embedded subject is longer, the processing load in the absence of *that* is increased. Accordingly, a longer embedded subject promotes the presence of *that* to increase parsing efficiency, whereas a pronominal subject causes little delay in reaching the finite verb, and is easily processed with a zero complementizer.

The embedded subject effect is fairly consistent across construction types, as shown in Figure 1. In all construction types a pronominal embedded subject occurs with a higher zero rate. The difference between short and long NPs is smaller, and is not evident in extraposed complements to verbs or copula constructions. The general trend in Figure 1 suggests that the embedded subject

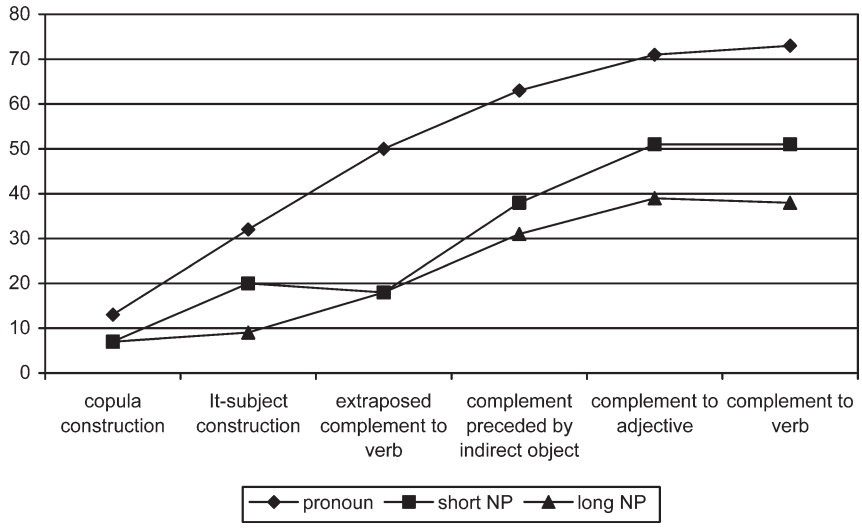


FIGURE 1. Zero rates across construction types and embedded subject types.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

effect is largely confined to pronoun subjects in contexts in which zero is more infrequent, and the difference between short and long NPs appears mainly in cases in which the overall zero rate is around 50% or more. This is consistent with a parsing efficiency account of the embedded subject effect. In contexts such as the complements to verbs and adjectives where other constraints, such as syntactic constraints, play a lesser part in determining *that*/zero choice, effects of parsing efficiency are more prominent, and the length of an NP subject has a greater effect.

Varietal differences in particular constructions

Although the zero rates for the different construction types do not support a categorical division into licensing and nonlicensing contexts for null complementizer, the relative zero rates do broadly reflect the presence of the potential zero-licensing factors identified in syntactic theory. Adjacent complements to verbs and adjectives have the overall highest zero rates, and copula constructions have the lowest zero rates. The constructions with some, but not all, licensing factors present—extraposed verb complements, verb complements with indirect objects, and *it*-subject constructions—have zero rates intermediate between the other two groups.

A conservative assumption would be that any effects of syntactic licensing factors would operate uniformly across the regional data, and thus that the overall regional ranking by zero rates of NZ > AU > US > GB would be preserved across the individual constructions. If so, it would appear that the regional differences reflect different regional sensitivities to nonsyntactic factors in the *that*/zero choice. To check this hypothesis, I now compare the zero rates of the four regional subcorpora in individual constructions. The possible effects of lexical type frequency and types of embedded subject are also considered. I shall show that significant differences in zero rates across regional sources are not constant across construction types, although the overall ranking is never reversed. The results for each construction will be presented in turn, and then summarized at the end of the section.

Complements to verbs. We saw in the previous section that the total zero rates of the regional sources were significantly different, and ranked as follows: New Zealand > Australia > United States > Britain. The most frequent construction type in the data is complement to verb, representing 60% of the corpus, and the overall ranking of the regional sources reflects the relative zero rates in verb complements, as shown in Table 3. The probability values shown (Pearson's GFX) indicate the significance levels of the difference between the row region and column region, for example, the second cell in the top row shows that the difference between New Zealand at 73% and Australia at 63% is significant, at $p \leq .0005$.

To control for potential interference from lexical type frequency, the zero rates were also calculated excluding the four most frequent lexical types *think, believe, suggest*, and *know*, which also had higher than average zero rates. The relative

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 3. *Significant regional differences in zero rates in complements to verbs*

	New Zealand	Australia	United States	Great Britain
	zero *that*	zero *that*	zero *that*	zero *that*
	404 150 73%	330 190 63%	230 245 48%	236 394 37%
New Zealand		$p \leq .0005$	$p \leq .0000$	$p \leq .0000$
Australia	$p \leq .0005$		$p \leq .0000$	$p \leq .0000$
United States	$p \leq .0000$	$p \leq .0000$		$p \leq .0014$
Great Britain	$p \leq .0000$	$p \leq .0000$	$p \leq .0014$	

TABLE 4. *Significant regional differences in zero rate in complements to adjectives*

	New Zealand	Australia	United States	Great Britain
	zero *that*	zero *that*	zero *that*	zero *that*
	27 16 63%	29 9 76%	24 11 69%	16 44 27%
New Zealand		$p \leq .2826$	$p \leq .8280$	$\mathbf{p \leq .0000}$
Australia	$p \leq .2826$		$p \leq .4394$	$\mathbf{p \leq .0000}$
United States	$p \leq .8280$	$p \leq .2826$		$\mathbf{p \leq .0000}$
Great Britain	$\mathbf{p \leq .0000}$	$\mathbf{p \leq .0000}$	$\mathbf{p \leq .0000}$	

rankings were unchanged and the differences were still significant: New Zealand 67%, Australia 57%, United States 40%, and Britain 29%, showing that the differences in zero rate across regional sources were not produced by unequally represented frequent lexical types.

The proportions of pronominal embedded subjects in verb complements for each regional source were also calculated, and were New Zealand 36%, Australia 35%, United States 44%, and Britain 34%. The rates of pronominal embedded subjects were uniform except for the slightly higher rate in the United States data. However, the United States zero rate ranked second lowest, and the slightly higher pronoun subject rate does not appear to have altered the relative ranking of regional sources by zero rates in verb complements.

Complements to adjectives. The zero rates in adjective complements were not significantly different among New Zealand, Australia, and the United States, but the British zero rate was significantly lower than each of the other regional sources, as shown in Table 4 (significance levels calculated with Pearson's GFX). The significant differences are shown in bold.

The zero rates were also calculated excluding lexical types with ten or more tokens (*aware, concerned, confident, convinced, sure*). The results showed the same grouping by region: New Zealand 55%, Australia 76%, United States 54%, and Britain 7%. The British data had a significantly lower zero rate than the

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 5. *Significant differences in zero rates in verb complements preceded by indirect objects*

	New Zealand			Australia			United States			Great Britain		
	zero *that*			zero *that*			zero *that*			zero *that*		
	40	24	62%	35	10	78%	11	27	29%	14	46	23%
New Zealand				$p \leq .2144$			**$p \leq .0005$**			**$p \leq .0000$**		
Australia	$p \leq .2144$						**$p \leq .0000$**			**$p \leq .0000$**		
United States	**$p \leq .0005$**			**$p \leq .0000$**						$p \leq .1929$		
Great Britain	**$p \leq .0000$**			**$p \leq .0000$**			$p \leq .1929$					

other regional sources, which were not significantly different from each other. The large drop in zero rates for the British data (all adjectives 27%, excluding high-frequency lexical types 7%) indicates that zero is almost confined to the most frequent lexical types in the British data.

The rates of pronominal embedded subjects in complements to adjectives were New Zealand 33%, Australia 39%, United States 54%, and Britain 34%, showing a higher rate for the United States data. Excluding all tokens with pronominal embedded subjects, the zero rates were New Zealand 62%, Australia 74%, United States 60%, and Britain 15%, showing the same grouping of regional sources. The British data had a significantly lower zero rate than the other regional sources, which were not significantly different from each other.

Verb complements preceded by indirect objects. In verb complements preceded by indirect objects we see a different grouping of regional sources, as shown in Table 5 (significance levels calculated with Pearson's GFX).

The most frequent lexical type in this construction is *tell*, representing 71% of tokens. Analyzing tokens with *tell* only, the zero rates are New Zealand 73%, Australia 82%, United States 31%, and Britain 31%, showing the same grouping of regional sources, with the same significant differences.

It was hypothesized that the length of the indirect object might affect zero rates in this construction, as well as the length of the embedded subject. The forms of indirect object were initially coded as pronoun, "trace" (with a passive verb), short NP, and long NP, with the following zero rates: pronoun 60%, trace 57%, short NP 41%, and long NP 15%. A logistic regression analysis was performed including the factors of regional source, form of embedded subject, and form of indirect object, and all factors were retained as independently significant.

A cross-product analysis of subject and indirect object type was carried out. Given the small number of tokens ($n = 207$) the embedded subjects were coded as pronoun or NP, and the indirect objects as pronoun (including trace) or NP, giving four combinations. Results are shown in Table 6.

As expected, the combination of two pronouns has a higher zero rate than other conditions, significant compared to the combination of two NPs, and approaching

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 6. *Zero rates in verb complements preceded by an indirect object, comparing combinations of embedded subject type and indirect object type*

	Zero	*That*	Zero rate
Subject pronoun, object pronoun	35	14	71%
Subject pronoun, object NP	29	25	54%
Subject NP, object pronoun	23	20	53%
Subject NP, object NP	12	48	20%
Comparison of zero rates			
Highest zero rate with pronoun/pronoun condition:			
Subject pronoun/object pronoun > subject pronoun/object NP:	$p \leq .0640$		
Subject pronoun/object pronoun > subject NP/object pronoun:	$p \leq .0752$		
Subject pronoun/object pronoun > subject NP/object NP:	$p \leq .0000$		
Same zero rates with pronoun/NP condition:			
Subject pronoun/object NP > subject NP/object pronoun:	$p \leq .9831$		
Lowest zero rate with NP/NP condition:			
Subject pronoun/object NP > subject NP/object NP:	$p \leq .0001$		
Subject NP/object pronoun > subject NP/object NP:	$p \leq .0004$		

TABLE 7. *Significant regional differences in zero rate in verb complements preceded by indirect objects, controlling for embedded subject type and indirect object type*

	New Zealand			Australia			United States			Great Britain		
	zero	*that*		zero	*that*		zero	*that*		zero	*that*	
	23	8	74%	16	4	80%	5	13	28%	8	20	29%
New Zealand				$p \leq .450162$			**$p \leq .001952$**			**$p \leq .000504$**		
Australia	$p \leq .450162$						**$p \leq .001564$**			**$p \leq .000511$**		
United States	**$p \leq .001952$**			**$p \leq .001564$**						$p \leq .612623$		
Great Britain	**$p \leq .000504$**			**$p \leq .000511$**			$p \leq .612623$					

significance compared with the two pronoun/NP conditions. The combination of two NPs has a zero rate significantly lower than the other three conditions. Given the demonstrated effects of both the embedded subject and the indirect object on zero rates, both factors must be taken into account in comparing zero rates across regional sources.

The two pronoun/NP conditions have the same zero rate and comprise almost half the available data. Accordingly, the cross-regional differences can be checked by comparing the zero rates for these two conditions only. The results, as shown in Table 7, give the same grouping of regional sources as in Table 3. The significance levels here are calculated with Fisher's Exact text.

Extraposed complements to verbs. Zero rates in extraposed complements to verbs also group the New Zealand and Australian data separately from the

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 8. *Significant regional differences in zero rates in extraposed verb complements*

	New Zealand	Australia	United States	Great Britain
	zero *that*	zero *that*	zero *that*	zero *that*
	24 26 48%	32 42 43%	22 56 28%	5 64 7%
New Zealand		$p \leq .6015$	$\mathbf{p \leq .0227}$	$\mathbf{p \leq .0000}$
Australia	$p \leq .6015$		$\mathbf{p \leq .0528}$	$\mathbf{p \leq .0000}$
United States	$\mathbf{p \leq .0227}$	$\mathbf{p \leq .0528}$		$\mathbf{p \leq .0010}$
Great Britain	$\mathbf{p \leq .0000}$	$\mathbf{p \leq .0000}$	$\mathbf{p \leq .0010}$	

United States and British data, as shown in Table 8. The significance levels were calculated using Pearson's GFX.

Excluding the most frequent lexical types of the selecting verbs (*admit, announce, confirm, know, say, tell*, $n \geq 10$) yields the same pattern of significant differences among regional sources, except for the difference between New Zealand (zero rate = 48%) and the United States (zero rate = 28%), which approaches significance ($p \leq .0828$).

The rates of pronominal embedded subjects were New Zealand 43%, Australia 37%, United States 38%, and Britain 29%, showing a slightly higher rate for the New Zealand data and a lower rate for the British data. The zero rates were recalculated excluding all tokens with a pronominal embedded subject, producing the same grouping of regional sources. The zero rates were New Zealand 34%, Australia 29%, United States 11%, and Britain 2%. The New Zealand and Australian zero rates remained significantly higher than the United States rate (NZ > US, $p \leq .0167$; AU > US, $p \leq .0391$).

It-subject constructions and copula constructions. There were no significant regional differences in zero rate in *it*-subject constructions, the rates being New Zealand 35%, Australia 21%, United States 13%, and Britain 15%. The zero rates in copula constructions were New Zealand 8%, Australia 14%, United States 13%, and Britain 3%, the only significant difference being that between the Australian and British data ($p \leq .019778$, Fisher Exact).

Summary. The significant regional differences in zero rates in the individual constructions are summarized next: ">" represents a significant difference in zero rates, and "=" represents no significant difference.

Complement to verb	NZ > AU > US > GB
Complement to adjective	NZ = AU = US > GB
Complement clause preceded by an indirect object	NZ = AU > US = GB
Extraposed complement to verb	NZ = AU > US > GB
It-subject construction	NZ = AU = US = GB
Copula construction	NZ = AU = US = GB

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Complements to verbs and adjectives represent the canonical syntactic environment in which *that*/zero alternation is free—any constraints on *that*/zero alternation in this environment are expected to be nonsyntactic. I take it that the ranking of different regional zero rates in the complements to verbs reflects regional differences in the sensitivity to nonsyntactic zero-promoting and zero-inhibiting factors. No regional differences appear for *it*-subject constructions and copula constructions, nor for adjective complements, with the exception of the lower zero rate for the British data.

The differences that may bear on regional syntactic differences are those for complement clauses with indirect objects and extraposed verb complements. Here the New Zealand and Australian data do not differ significantly from each other, though both have significantly higher zero rates than the United States and British data. Bearing in mind that these two construction types represent complement clauses that are not adjacent to the licensing head, it would appear that the New Zealand and Australian data reflect a lower sensitivity to the zero-inhibiting effect of nonadjacency. More generally, the regional sources may differ in the relative strength of the syntactic factors in the *that*/zero choice. This issue is explored further in the next section.

Varietal patterns of potential syntactic zero-licensing factors

To assess the significance of licensing factors for each regional variety, the relative zero rates across construction types were compared *within* each regional subcorpus. The significance of interconstructional differences in zero rates was calculated using Pearson's GFX, or Fisher's Exact test in cases in which there were fewer than five tokens in one of the cells.[18] The results for each regional source, showing the significant differences in zero rates across constructions, are given in the next subsection. These results are then summarized, showing how the regional sources compare in assigning the construction types to three broad categories. Then I argue that the differences reflect a greater zero-inhibiting effect from nonadjacency in the New Zealand and Australian data, compared to the United States and British data.

Differences in zero rates across constructions in individual regional sources. Table 9 shows the patterns of difference in zero rates for the New Zealand data. Significant differences are shown in bold. Cells with nonbold text show which constructions are not significantly different from each other in zero rates.

The New Zealand data fall into three main blocks. The main contexts for zero are complements to verbs, complements to adjectives, and complements preceded by an indirect object, although the difference between verb complements and complements preceded by an indirect object approaches significance (starred in the table). The lowest zero rate occurs as expected in copula constructions. Extraposed verb complements and *it*-subject constructions have similar zero rates in the range of 35–40%, intermediate between the two other blocks.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 9. *Significant differences in zero rates across constructions in the New Zealand data*

	Verb complement 73%	Adjective complement 63%	Complement preceded by indirect object 62%	Extraposed verb complement 48%	*It*-subject construction 35%	Copula construction 8%
Verb complement		$p \leq 0.1530$	$p \leq 0.0791$*	**$p \leq 0.0000$**		
Adjective complement	$p \leq 0.1530$		$p \leq 0.9756$	$p \leq 0.1529$	**$p \leq 0.0537$**	
Complement preceded by indirect object	$p \leq 0.0791$*	$p \leq 0.9756$		$p \leq 0.1215$	**$p \leq 0.0441$**	**$p \leq 0.000000$**
Extraposed verb complement	**$p \leq 0.0000$**	$p \leq 0.1529$	$p \leq 0.1215$		$p \leq 0.7254$	**$p \leq 0.000264$**
It-subject construction		**$p \leq 0.0537$**	**$p \leq 0.0441$**	$p \leq 0.7254$		**$p \leq 0.016959$**
Copula construction			**$p \leq 0.000000$**	**$p \leq 0.000264$**	**$p \leq 0.016959$**	

In the New Zealand data the different constructions had pronominal embedded subjects in the range of 33–36%, except for a slightly higher rate with extraposed verb complements (43%) and a slightly lower rate with copula constructions (23%). To check for possible interference from the unequally distributed pronominal embedded subjects, zero rates excluding tokens with pronominal subjects were compared for verb complements, extraposed verb complements, and copula constructions. The relationships were unchanged: extraposed verb complements still had a significantly lower zero rate than verb complements, but a significantly higher zero rate than copula constructions.

Results for the Australian data are shown in Table 10. The Australian data also have the highest zero rates in verb complements, adjective complements, and complements preceded by indirect objects. Surprisingly, the highest overall zero rate is in complements preceded by indirect objects, with a zero rate significantly higher than in verb complements. Unlike the New Zealand data, *it*-subject constructions do not have significantly higher zero rates than copula constructions.

In the Australian data the pronominal embedded subject rates were in the range of 31–37%, except adjective complements with 39%, and *it*-subject constructions with 18%. Zero rates were compared, excluding tokens with pronominal subjects, for verb complements, adjective complements, *it*-subject constructions, and copula constructions. Adjective complements had a higher zero rate than verb complements, the difference approaching significance ($p \leq .0753$). There was no significant difference between *it*-subject constructions and copula constructions.

Table 11 shows results for the United States data. The pattern in the United States data is quite different: adjective complements and verb complements have the highest zero rates, but all the other construction types, including complements preceded by indirect objects, have zero rates that are not significantly higher than in copula constructions.

Rates of pronominal embedded subjects in the United States data were in the 38–44% range, except for a higher rate of 54% in adjective complements and a lower rate of 26% in copula constructions. Excluding tokens with pronominal subjects, zero rates were checked for adjective complements, verb complements, complements preceded by an indirect object, extraposed verb complements, and *it*-subject constructions. Adjective complements still had a significantly higher zero rate than verb complements ($p \leq .0401$), and the zero rates with complements preceded by indirect objects, extraposed verb complements, and *it*-subject constructions are still not significantly higher than in copula constructions.

Finally, results for the British data are shown in Table 12. As in the United States data, verb and adjective complements have the highest zero rates. Unlike the United States data, complements preceded by an indirect object and *it*-subject constructions have intermediate zero rates, lower than those for verb and adjective complements but higher than those for copula constructions. Extraposed verb complements have zero rates not significantly higher than copula constructions.

In the British data the rates of pronominal embedded subjects across constructions were 33–34%, except for a slightly lower rate of 29% in extraposed verb complements and a slightly higher rate of 45% in *it*-subject

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 10. *Significant differences in zero rates across constructions in the Australian data*

	Complement preceded by indirect object 78%	Adjective complement 76%	Verb complement 63%	Extraposed verb complement 34%	*It*-subject construction 21%	Copula construction 14%
Complement preceded by indirect object		$p \leq 0.8744$	$\mathbf{p \leq 0.0540}$	$\mathbf{p \leq 0.0002}$		
Adjective complement	$p \leq 0.8744$		$p \leq 0.1102$	$\mathbf{p \leq 0.0008}$		
Verb complement	$\mathbf{p \leq 0.0540}$	$p \leq 0.1102$		$\mathbf{p \leq 0.0008}$		
Extraposed verb complement	$\mathbf{p \leq 0.0002}$	$\mathbf{p \leq 0.0008}$	$\mathbf{p \leq 0.0008}$		$\mathbf{p \leq 0.0228}$	$\mathbf{p \leq 0.0000}$
It-subject construction				$\mathbf{p \leq 0.0228}$		$p \leq 0.3484$
Copula construction				$\mathbf{p \leq 0.0000}$	$p \leq 0.3484$	

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 11. *Significant differences in zero rates across constructions in the United States data*

	Adjective complement 69%	Verb complement 48%	Complement preceded by indirect object 29%	Extraposed verb complement 24%	*It*-subject construction 13%	Copula construction 13%
Adjective complement		**$p \leq 0.0213$**	**$p \leq 0.0007$**	**$p \leq 0.0000$**		
Verb complement	**$p \leq 0.0213$**		**$p \leq 0.0206$**	**$p \leq 0.0008$**	**$p \leq 0.005945$**	**$p \leq 0.000555$**
Complement preceded by indirect object	**$p \leq 0.0007$**	**$p \leq 0.0206$**		$p \leq 0.9337$	$p \leq 0.204869$	$p \leq 0.131049$
Extraposed verb complement	**$p \leq 0.0000$**	**$p \leq 0.0008$**	$p \leq 0.9337$		$p \leq 0.191318$	$p \leq 0.111120$
It-subject construction		**$p \leq 0.005945$**	$p \leq 0.204869$	$p \leq 0.191318$		$p \leq 0.702130$
Copula construction		**$p \leq 0.000555$**	$p \leq 0.131049$	$p \leq 0.111120$	$p \leq 0.702130$	

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 12. *Significant differences in zero rates across constructions in the British data*

	Verb complement 37%	Adjective complement 27%	Complement preceded by indirect object 23%	*It*-subject construction 15%	Extraposed verb complement 5%	Copula construction 3%
Verb complement		$p \leq 0.0970$	$\mathbf{p \leq 0.0296}$			
Adjective complement	$p \leq 0.0970$		$p \leq 0.6732$	$p \leq 0.1676$	$\mathbf{p \leq 0.0028}$	
Complement preceded by indirect object	$\mathbf{p \leq 0.0296}$	$p \leq 0.6732$		$p \leq 0.3074$	$p \leq 0.0101$	
It-subject construction		$p \leq 0.1676$	$p \leq 0.3074$		$p \leq 0.1952$	$\mathbf{p \leq 0.025998}$
Extraposed verb complement		$\mathbf{p \leq 0.0028}$	$\mathbf{p \leq 0.0101}$	$p \leq 0.1952$		$p \leq 0.214586$
Copula construction				$\mathbf{p \leq 0.025998}$	$p \leq 0.214586$	

TABLE 13. *Groups of constructions by zero rates across regional sources*

	Zero rate not significantly lower than in verb complements	Intermediate	Zero rate not significantly higher than in copula constructions
United States	VERB ADJ		IOBJ EXTR ITSUB COP
Britain	VERB ADJ	IOBJ	EXTR ITSUB COP
Australia	VERB ADJ IOBJ	EXTR	ITSUB COP
New Zealand	VERB ADJ IOBJ	EXTR ITSUB	COP

Note. VERB = verb complement (not extraposed, no indirect object), ADJ = adjective complement, IOBJ = verb complement preceded by indirect object, EXTR = extraposed verb complement, ITSUB = *it*-subject construction, COP = copula construction.

constructions. Excluding tokens with pronominal embedded subjects, only one relationship shown in Table 12 was changed: the zero rate with *it*-subject constructions was not significantly higher than in copula constructions ($p \leq .165473$). I shall assume that the high subject pronoun rate in *it*-subject constructions contributes to the difference from copula constructions shown in Table 12, and that *it*-subject constructions should be grouped with copula constructions for the British data.

With the adjustment to the British data as noted, the groupings of constructions by zero rate across the regional sources can be summarized as shown in Table 13. The table displays the regional patterns against three main divisions according to the potential syntactic licensing factors discussed previously. The left column represents contexts in which zero is predicted to be fully licensed. The typical construction for this column is taken to be verb complements, as verb complements are the canonical context in which zero is considered to be syntactically licensed. Constructions are listed in this column if they do not have a significantly lower zero rate than verb complements. The right column represents contexts in which zero is predicted not to be licensed. The typical construction for this column is taken to be copula constructions, as none of the potential syntactic licensing factors appears in a copula construction. Constructions are listed in this column if they do not have a significantly higher zero rate than copula constructions. The middle column represents constructions of intermediate status that have lower zero rates than verb complements and higher zero rates than copula constructions.

In all four varieties, adjective complements have a zero rate no lower than verb complements, as syntactic theory would predict; that is, the embedded clause in both verb and adjective complements is adjacent to and selected by a major lexical head. In all four varieties copula constructions have the lowest zero rate, once again consistent with syntactic predictions, that is, the embedded clause in a copula construction is neither adjacent to nor selected by a major lexical head. However, the regional rankings differ with respect to the intermediate constructions, which have at least one syntactic licensing factor, but not all,

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

including extraposed verb complements, complements with indirect objects, and *it*-subject constructions.

The effects of complement status and adjacency on zero rates: adjacent and nonadjacent complements. Extraposed verb complements have a lower zero rate than verb complements for all regional sources, showing that the intervening adverbial inhibits zero to some extent in all varieties. However, extraposed verb complements have a significantly higher zero rate than copula constructions for the New Zealand and Australian data, but not for the United States and British data. This suggests that the lack of adjacency between the verb and the extraposed complement clause inhibits zero to a much weaker extent in the New Zealand and Australian data.

A similar regional pattern appears in complements with indirect objects. In the New Zealand and Australian data the intervening indirect object does not significantly inhibit zero, and the construction has a zero rate not significantly lower than in adjacent verb complements. In contrast, complements with indirect objects in the United States and British data have significantly lower zero rates than in adjacent verb complements, and in the United States data, zero in complements with indirect objects is not significantly more frequent than in copula constructions.[19]

The effect of adjacency may also be investigated by comparing zero rates in the complements to heads such as the conjunctions *except, now*, and purpose *so* with zero rates in the complements to degree head *so, such*.[20] Complements to these heads are illustrated in the examples in (47)–(52). The complements to conjunctions are adjacent to their selecting heads, as shown in (47)–(50), whereas complements to degree *so* are separated from the selecting head, as shown in (51)–(52).

Conjunction *except*

(47) It's his version of pre-game jitters, except_they usually occur days before any USC game. *Los Angeles Daily News*, November 28, 2002.

(48) It can be partially alleviated by rejetting or fitting a Weber carb—except *that* the latter could cost more than the car is worth. *Daily Telegraph* (UK), January 4, 2003.

Purpose *so*

(49) The lifting of US sanctions against Iran so_aid can flow more freely has been welcomed by the Iranian regime, with officials in Tehran stating they hope the thaw in relations can be permanent. *Daily Telegraph* (Australia), January 3, 2004.

(50) Boardwalks were erected through the sanctuary so *that* visitors could have a good view of the trees and birds while allowing the undergrowth to regenerate. *Christchurch Press*, September 24, 2002.

Degree *so*

(51) The "good" Victorians of the novel are so busy spouting warnings about maintaining moral purity_you'd think they were refugees from the 700 Club. *Boston Globe*, October 8, 2002.

(52) According to Stealth, central London's airwaves are so overcrowded *that* the suburbs are the best option for a new station. *The Guardian*, November 22, 2002.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

TABLE 14. *Zero rates with conjunctions and degree heads*

	Conjunctions (clause adjacent to head)			Degree heads (clause not adjacent to head)		
	zero	*that*	zero rate	zero	*that*	zero rate
New Zealand	14	14	50%	15	19	44%
Australia	9	14	39%	11	14	44%
United States	20	14	59%	11	21	34%
Great Britain	21	30	41%	3	34	8%

NZ: head-adjacent zero rate > head-nonadjacent zero rate, $p = .6441$;
AU: head-adjacent zero rate > head-nonadjacent zero rate, $p = .7324$;
US: head-adjacent zero rate > head-nonadjacent zero rate, $p = .0467$;
GB: head-adjacent zero rate > head-nonadjacent zero rate, $p = .000424$.

Constructions of these types were selected from the main corpus from which the present study is drawn[21], and all were found to show *that*/zero variation for all regional sources, that is, neither *that* nor zero was categorical. However, zero rates in adjacent clauses were significantly higher than in nonadjacent clauses for the United States and British data, but not for the New Zealand and Australian data. The results are shown in Table 14.

To sum up so far, with complement clauses the New Zealand and Australian data reflect much less sensitivity to adjacency as a requirement for licensing zero complementizer than the United States or British data. The only zero-inhibiting effect of nonadjacency in the New Zealand and Australian data is the significant difference in zero rates between extraposed verb complements and adjacent complements to verbs and adjectives. There is no significant difference in zero rates in verb complements with and without indirect objects, nor in the (adjacent) complements to selected conjunctions and the (nonadjacent) complements to degree heads.

The British and United States data, on the other hand, show a strong adjacency effect in zero rates. The presence of an indirect object significantly lowers the zero rate, in comparison with complements adjacent to the selecting verb or adjective. Zero rates are also significantly higher in the (adjacent) complements to selected conjunctions than in the (nonadjacent) complements to degree heads.

The results for *it*-subject constructions are more difficult to interpret. Recall that *it*-subject constructions are included here to represent the syntactic predictions in Bošković & Lasnik (2003). In their analysis, zero complementizer in *it*-subject constructions (their "extraposed clauses") is licensed by affixation to the adjacent matrix predicate, even though the clause is not a complement to the matrix predicate. The potential effect of adjacency to the matrix predicate in an *it*-subject construction would be expected to appear in a comparison with copula constructions, where the adjacent matrix predicate is copula *be*, not a major lexical head, and not a potential host for an affixal null complementizer. As noted previously, this prediction is consistent with the overall zero rates found in

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

the present study: that is, the total zero rate of 20% in *it*-subject constructions is significantly higher than the total zero rate of 9% in copula constructions ($p \leq .0050$).

However, considering the regional sources individually, *it*-subject constructions have a significantly higher zero rate than copula constructions only for the New Zealand data, a pattern that does not appear for any other construction. I draw no conclusions concerning regional differences from the present results of *it*-subject constructions, and leave this particular issue for further research.

CONCLUDING REMARKS

The distribution of *that* and zero complementizer in the complements to verbs has been the target of numerous corpus-based studies, but there has been little research into *that*/zero choice in a wider range of syntactic contexts. In addition, although considerable research has revealed regional differences in the distribution of *that* and zero in clauses with a *wh*-gap, there has been little or no study of regional differences in other contexts.

This paper aims to increase our understanding of both of these issues. I have shown that in the genre of contemporary newspaper prose the overall rates of zero complementizer are significantly different across regional varieties, giving the ranking: New Zealand English > Australian English > American English > British English. This pattern is preserved in the complements to verbs, where according to the predictions of syntactic theory zero complementizer is fully licensed. I have suggested that the overall regional ranking by zero rates indicates different degrees of sensitivity to nonsyntactic factors in the *that*/zero choice.

I have also shown that the regional rankings by zero rates are not uniform across different syntactic types of context, and that cross-regional differences are stronger in some contexts than in others. The fact that the differences pattern according to syntactically identified contexts suggests that the regional differences are at least partly a result of differences in the syntax of null complementizer. In particular, I have argued that the chief regional syntactic difference concerns the importance of adjacency between a lexical head and the clause it selects, which contains the complementizer concerned. New Zealand and Australian English show less inhibition of zero where the clause is not adjacent than American and British English.

Overall, complement status appears to be a more powerful licenser than adjacency to a lexical matrix predicate. Given complement status, whether or not adjacency is also required is the main syntactic regional difference to emerge here. Among noncomplements, the contrast between *it*-subject constructions, in which the embedded clause is adjacent to a lexical matrix predicate, and copula constructions, where it is not, is much weaker. Adjacency correlates with a significantly higher zero rate only for the New Zealand data.

The results reviewed here also raise issues for syntactic theory. One obvious issue, identified earlier, is that the categorical predictions of formal syntactic theory are not upheld, in that the grammatical/ungrammatical dichotomy is not

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

found. However, the factors identified by syntactic theory—the complement or noncomplement status of the embedded clause, and adjacency or nonadjacency to a lexical matrix predicate—do correlate with the probability of zero complementizer. Zero complementizer is most frequent where the embedded clause is a complement and is adjacent to the selecting lexical head; it is least frequent where the embedded clause is not a complement and not adjacent to a lexical matrix predicate; and intermediate rates of zero complementizer occur where the embedded clause is a complement but not adjacent to a lexical matrix predicate, or adjacent to a lexical matrix predicate but not a complement. Although a full discussion of the issues would go beyond the bounds of this paper, the present results indicate that a probabilistic approach to the syntax of zero complementizer would provide a superior account.

Finally, I note that the results suggest an interesting puzzle for the Minimalist goal of eliminating government from syntactic explanation, particularly in accounting for complements preceded by indirect objects and extraposed verb complements. Recall that the syntactic accounts reviewed in the first section propose that verbs which select an indirect object and complement clause form a special class, with the consequence that nonadjacency does not degrade null complementizer. Stowell argued that this class of verbs can theta-mark, and thus properly govern, a nonadjacent complement, and Boškovic & Lasnik proposed that this class of verbs can select a clause which lacks the complementizer position altogether. Both analyses predict that a null complementizer preceded by an indirect object should have the same acceptability as a null complementizer adjacent to the clause-selecting verb. Although this pattern is seen in the New Zealand and Australian data, in the British and United States data the null complementizer is significantly less frequent in verb complement clauses where an indirect object intervenes. I shall assume that the verbs selecting an indirect object and clausal complement do not form a special class of the proposed type in these varieties.

If no special verb class is involved with indirect objects, we would expect the intervening indirect object to have the same effect as an intervening adverbial in disrupting the licensing of a null complementizer. This is plausible for the United States data, in which constructions with an indirect object or adverbial intervening between the verb and its clausal complement have zero rates not significantly higher than in copula constructions, where none of the proposed licensing factors are available. But the British data present a different pattern. The intervening indirect object produces a zero rate that is significantly lower than in an adjacent verb complement, but significantly higher than in an extraposed clause with an intervening adverbial. The constructions are illustrated with the simplified examples in (53).

(53) zero rates: (a) > (b) > (c)

a. He decided [$_{CP}$ (that) . . .]
b. He told his friend [$_{CP}$ (that) . . .]
c. He said yesterday [$_{CP}$ (that) . . .]

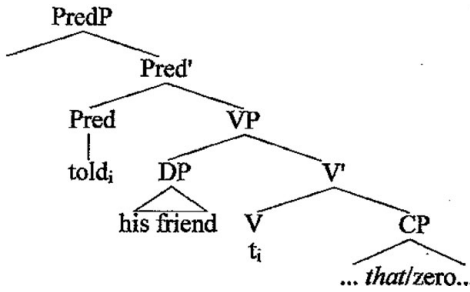
https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

The difference between (a) and (b) indicates that (b) does not involve a special class of verbs, and therefore a special verb class cannot be appealed to for the difference between (b) and (c). In addition, the embedded clause in (b) and (c) is a complement clause and nonadjacent to the selecting head in both contexts, so these two factors cannot characterize the difference between the two constructions as a context for null complementizer.

It is possible, however, to differentiate the relationship between the verb and complement clause using the notion of government in combination with current versions of phrase structure, as I shall briefly outline here. In current versions of phrase structure it is possible to argue that the verb governs its complement "through" an indirect object but not "through" an adverbial, without appealing to a special lexical class with indirect objects, as in Stowell's account.

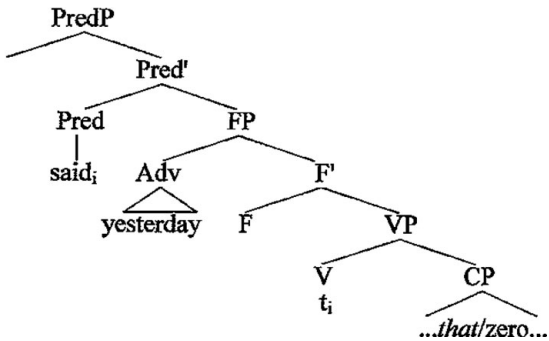
First, for an indirect object followed by a clausal complement I shall assume the structure shown in (54).

(54)



For extraposed clauses I shall assume for convenience the analysis introduced by Cinque (1999) in which adverbs are specifiers of functional heads. On this approach, and assuming the same surface position for the verb, it follows that the functional projection hosting the adverb is between PredP and VP, as shown in (55). The default assumption is that the complement clause is in its base-generated position.

(55)



https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

The relationship between the CP and the base position of the verb is the same in both structures, so the difference between the structures presumably concerns the surface position of the verb in Pred. Applying concepts of government based on the Barriers framework (Chomsky 1986), it can be shown that the null complementizer is properly governed by the verb in Pred in (54) but not in (55). The relevant structural relations I assume are defined in (56)–(59).

(56) *C-command:*
X c-commands Y if X does not dominate Y and every node that dominates X also dominates Y.

(57) *Government:*
X governs Y if X c-commands Y and neither (a) nor (b) holds.

(a) There is an element Z intervening between X and Y such that Z is a potential governor for Y of the same type as X.
(b) There is an element Z intervening between X and Y such that Z is a maximal projection (other than S/IP/TP) and Z is not selected by a lexical head.

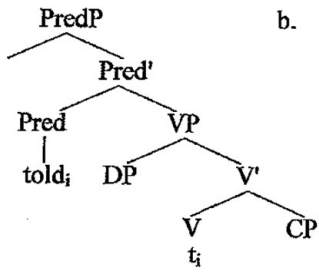
(58) *Proper government:*
X properly governs Y if X governs Y and either (a) or (b) holds.

(a) X is the antecedent of Y.
(b) X head-governs Y.

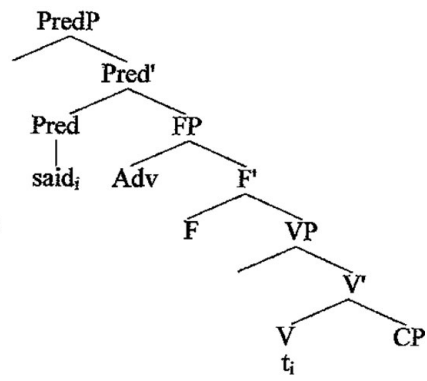
(59) *Head government:*
X head-governs Y if X governs Y and X is a lexical head which selects or case-marks Y.

For convenience, simplified representations for the structures at issue are given next. (57a) represents the construction with an indirect object and (57b) represents an extraposed clause.

(57) a.



b.



https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

First, the raised verb c-commands the CP in both structures (see (56)), and in both structures the verb is a potential head-governor and thus a potential proper governor, in that the verb is a lexical head that selects CP (see (59). The remaining points to be considered are the presence of any intervening elements to block government, as defined in clauses (a) and (b) of (57). In (57a) the maximal projection VP intervenes between the verb in Pred and the CP, but VP is selected by Pred. Although Pred is functional rather than lexical, it is lexicalized by the raised verb *told*, and I will assume that the lexicalized Pred can then count as a lexical head in the terms of clause (b) of (57), and so VP does not block government of the CP by *told*. Given that the verb is a proper governor, the CP and its head zero/*that* are properly governed by the verb.

In (57b) there are two elements that might block government of CP by the verb in Pred. The intervening head F appears to be a possible element Z as defined in clause (a) of (57)—F is a head intervening between Pred and CP. However, F is not lexical, so it cannot be the same kind of governor as the verb in Pred, and I assume F does not block government of CP by the verb in Pred. In (57b) there are also two maximal projections, FP and VP between the verb in Pred and the CP that might fit clause (b) in (57). Applying the same reasoning as for (57a), FP is selected by Pred, Pred is lexicalized by the raised verb, and so FP does not satisfy clause (b) of (57). VP, on the other hand, is selected by F, which is functional, not lexical, and is not lexicalized by any raised element. VP then fits the description in clause (b) of (57) and blocks government of CP by the verb in Pred. However the details are formalized, the analysis sketched here illustrates the general idea that the structural relationship between the verb in Pred and the CP is interrupted by the presence of the functional projection FP, which, unlike Pred (or AGR), is not one of the functional projections considered to be a kind of extended projection of VP. FP, in some sense, introduces a more distinct domain between PredP and VP.

In the context of moves to eliminate government from syntactic explanation, the licensing of a null complementizer is of particular interest in that it does not appear to involve locality constraints on movement or binding-like relations, both of which fell under antecedent government in the period of largely government-based theory. The more recent Minimalist developments seeking to eliminate government have generally focused on antecedent government. The main head-government-related issue to be addressed in detail has been the case properties of DPs, now reanalyzed in terms of feature-checking (with or without movement) between the DP and a case-related functional projection. However, S′/CP has always been considered to contrast strongly with NP/DP in not requiring, or indeed allowing, the assignment of case, as argued by Stowell (1981). It is not clear that the feature-checking model for the licensing of DP arguments, based on case, can be extended to CP arguments in such a way as to capture the contrast between intervening indirect objects and intervening adverbials.

In conclusion, the findings presented here represent a fairly small study in an area that has had little previous investigation. The cross-regional differences found in this study suggest directions for further research before firm

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

conclusions can be drawn. In particular, further study of indirect object constructions and a wider range of extraposition constructions would give a clearer picture of the role of adjacency, and it is hoped that work in progress on the distribution of null complementizer in noun complements will clarify the importance of the lexical category of the selecting head. The discussion of syntactic issues, particularly the possible need for some relation corresponding to head government, is speculative and incomplete—in particular, the combination of the notion of government with more recent versions of phrase structure raises a number of questions which could not be addressed here. It is to be hoped that further research will shed more light on these issues.

NOTES

1. Relativizer *that* also alternates with zero in relative clauses where the *wh*-gap is not in subject position. This alternation is not included in the present study, as my focus here is on the conditions for zero proposed in transformational syntactic theory.

Transformational theory treats complementizer *that* and relativizer *that* as structurally the same element C, but the proposed conditions for null complementizer and null relativizer are different. Briefly, as discussed in more detail later, the transformational account of null complementizer centers on the status of the embedded clause as a complement selected by a lexical head. Relative clauses, on the other hand, are analyzed as adjuncts, and the conditions for null relativizer are generally defined in terms of the licensing conditions for the *wh*-gap within the relative clause. Some of the nonsyntactic factors in *that*/zero choice are common to both complementizers and relativizers, and I shall comment further on these issues as they arise.

2. The *that*/zero choice is not generally considered to carry an obvious, established semantic or pragmatic distinction, although such distinctions have been proposed. Notably, Bolinger (1972) argued that the complementizer *that* carries a meaning derived from its original source as a demonstrative, and is used to mark propositions that are offered as true or in some way presupposed, or describe situations that serve as causes for psychological responses. His arguments are not based on actual data, and his discussion suggested a wide range of different instances of the proposed contrast between demonstrative *that* and nondemonstrative zero. Although many of his intuitional data are persuasive, it is not clear whether the proposed distinctions are true of actual usage, and the relevance of his individual arguments to the overall claim that the complementizer *that* has demonstrative meaning are difficult to evaluate. Yaguchi (2001) reported a corpus-based study developing Bolinger's thesis, arguing that *that* expresses "a state of knowledge, fixed belief, inference based on evidence, conscious mental activity, remote events, public information, old information, indirectness, and formality," whereas zero indicates "yet unproven information, groundless hearsay, inference without evidence, new(ly learned) information, nearby events, directness, immediacy, solidarity, emotion, and informality." It is not clear to me that a single, obvious, semantic/pragmatic distinction is at work in these phenomena. I also note that a number of the observations concerning the epistemic status of the embedded clause have been given an alternative explanation in which zero signals a different functional (and possibly syntactic) clausal structure. Briefly, it has been argued that the absence of *that* in utterances like *I think it'll be fine* indicates that *I think* serves as an epistemic modifier to the main assertion *It'll be fine* (see especially Thompson & Mulac, 1991a, 1991b). On this account, the *that*/zero alternation signals differences in functional and possibly syntactic clause structure, but this is quite distinct from Bolinger's claim that *that* has demonstrative meaning.

Dor (2005) offered a semantic account for the obligatoriness of *that* in contexts like *John mumbled *(that) Bill was an undercover agent*. Dor argued that zero is possible only in cases in which the embedded clause expresses a proposition concerning which a cognitive agent makes a truth claim. Of the predicates identified by Dor as requiring *that*, only one, decide in the sense of "decree," occurred in the present data, but five of the fifteen tokens had zero complementizer, indicating that the association of *decide* 'decree' with *that* is not categorical.

Finally, style guides commonly adjure writers to avoid zero in cases in which the resulting local structure is more difficult to parse, as in *She realized her dream was a sign of impending disaster*, where the reader may initially parse the NP *her dream* as the direct object of *realize*. An

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

experimental study by Ferreira & Dell (2000) found that speakers do not include *that* to avoid temporary ambiguity. In an extensive corpus study of structurally ambiguous sentences, Roland et al. (2005) found that the sentence context contains numerous clues to resolve the ambiguity independently of *that*, and that the *that*/zero choice is not apparently related to the avoidance of ambiguity.

3. Generally wherever zero can appear in an embedded clause, *that* is also possible. The main exception to this is in clauses with a subject gap, as in **Who did they think* (**that*) *let the dog out?* Clauses with *wh*-gaps are not included in this study.

4. Factors included as "nonsyntactic" include sentence properties that are partly identified in syntactic terms, for example, one zero-inhibiting factor is the presence of a clause-initial adverbial in the embedded clause, as in *We knew that if all else failed we could reach them by helicopter*. However, the requirement for *that* in such contexts is generally attributed to processing considerations, and the predictions of syntactic theory at issue in this article do not predict a formal syntactic requirement for categorical *that* in these contexts.

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

6. Jaeger & Wasow (2005) also found that the likelihood of zero in relative clauses increases in cases in which the referent of the embedded subject is salient or given in the discourse context.

7. Biber et al. (1999) reported on discourse and register factors affecting the frequency of complementizer and relativizer *that*, but do not report any differences between British and American English, although varietal differences are reported for other aspects of usage. It isn't clear whether no differences were found or the issue was not examined.

8. Noun complement clauses are included in the complete corpus for the present study, and were found to have an overall zero rate of 15%. This indicates that zero is not categorically ill-formed in noun complement clauses, and suggests that the low zero rate in this environment may be largely a result of factors that are not strictly syntactic. The *that*/zero alternation in noun complement clauses will be addressed in further work, now in preparation.

9. In a noun complement clause (e.g., *his belief that the fields of Italy would be full of waving macaroni*) the embedded clause is selected by and adjacent to the noun, which is a major lexical head, but nevertheless, zero complementizer is traditionally considered to be ungrammatical in this context.

10. A preliminary check confirmed that the inclusion of the letters and quoted speech text subtypes produces a number of empty cells in the multifactor analysis.

11. The first person was excluded from this comparison to minimize possible interference from *I think*-like routines (see Tagliamonte & Smith, 2005; Thompson & Mulac, 1991a), on the assumption that first-person embedded subjects are more likely to occur with these routines in comparison with other types of matrix. Subject *there* was also excluded from this comparison, as it is a recognizable subject form despite not having morphological case.

12. Despite the possible interference from *I think*-like routines mentioned in endnote 2, it was not possible to exclude *I* subjects for this test, as they constituted the majority of the first-person tokens.

13. To my knowledge, there is no parsed corpus available with the desired properties of sampling completely contemporary texts of comparable genre, with a degree of syntactic complexity allowing for a range of types of embedded clause, and including Antipodean English.

14. Note that examples like (42) are not the same as the *it*-subject constructions (*It is obvious (that)*. . .) in which the embedded clause is interpreted as the subject of the main clause lexical predicate. I also assume that the zero-licensing mechanism proposed by Bošković & Lasnik (2003) for *it*-subject constructions—that is, the null complementizer can be affixed to the adjacent matrix predicate even though the embedded clause is not selected by that predicate—does not apply here, as the matrix predicate is copula *be*, and not a major lexical head.

15. The classification of *it*-clefts as copula constructions was also checked during analysis by comparing the rates of zero complementizer. *It*-cleft constructions in which the embedded clause was adjacent to a major lexical head present at the end of the clefted phrase had an overall zero rate of 8%, similar to the 8.7% zero rate of the other copula constructions, and lower than the overall 20% zero rate of *it*-subject constructions.

16. Biber et al. (1999:1128–1132) found that overall contraction rates of negation and auxiliary verbs across registers were ranked as follows: conversation > light fiction > news > academic prose. This ranking of the register types broadly reflects different levels of formality. The relative informality of contracted forms is also occasionally noted in usage guides (Burchfield, 1996:180).

17. Hawkins (2003) and Elsness (1984) used the number of words as a measure of the length of the embedded subject, though the length of the embedded subject for parsing effects may be more precisely measured in terms of some other units, such as syllables, morphemes, or characters. Roland et al. (2005) used the number of characters as a measure of NP length.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

18. For these comparisons, there were not sufficient data to confine the analysis to one embedded subject type, and tokens grouped by regional source and construction included all three subject types.
19. The intent of Bošković & Lasnik's (2003) analysis of verbs that select an indirect object and a clausal complement as a special lexical class is to assign such verbs to the same group as verbs with adjacent complements. This is the pattern found in the New Zealand and Australian data, and we could say that the special rule for this class of verbs operates in these varieties but not in the United States and British varieties. However, this would miss the generalization that the New Zealand and Australian data show less sensitivity to adjacency, not only with indirect objects but also with extraposed clauses.
20. I note that a comparison of zero rates in the adjacent and nonadjacent complements to conjunctions and degree heads presupposes that these heads are potential zero licensers in principle, and thus are assumed to be major lexical categories, unlike the copula. Although it isn't clear how these heads would be specified in terms of the major lexical category features [±V, ±N], I consider that these heads select clauses both syntactically and semantically, and thus are reasonably classed as lexical rather than functional heads.
21. Tokens of degree and conjunction complements were taken from all three text subtypes of quoted speech, letters to the editor, general news, and columns, although the types were fairly evenly distributed across regional sources and the majority of tokens were from general news and columns.

REFERENCES

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finegan, Edward. (1999). *Longman grammar of spoken and written English*. Essex, UK: Pearson Education.
Bolinger, Dwight. (1972). *That's that*. The Hague: Mouton.
Bošković, Željko, & Lasnik, Howard. (2003). On the distribution of null complementizers. *Linguistic Inquiry* 34:527–546.
Burchfield, Robert. (ed.). (1996). *The new Fowler's modern English usage*. (3rd ed.). Oxford: Clarendon.
Cacoullos, Rena Torres, & Walker, James A. (2003). *Taking a complement . . . variably*. Paper presented at the 32nd NWAV conference, Philadelphia.
Cedergren, Henrietta, & Sankoff, David. (1974). Variable rules: Performance as a statistical reflection of competence. *Language* 50:333–355.
Chomsky, Noam. (1986). *Barriers*. Cambridge, MA: MIT Press.
Cinque, Guglielmo. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
Dor, Daniel. (2005). Toward a semantic account of *that*-deletion in English. *Linguistics* 43:345–382.
Elsness, Johan. (1984). *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65:519–533.
Ferreira, Victor S., & Dell, Gary S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40:296–340.
Finegan, Edward, & Biber, Douglas. (1995). *That* and zero complementisers in Late Modern English: Exploring ARCHER from 1650–1990. In Bas Aarts & Charles F. Meyer (eds.), *The verb in contemporary English*. Cambridge: Cambridge University Press. 241–257.
Fowler, H.W. (1926). *A dictionary of Modern English usage*. Oxford: Clarendon.
Garnsey, S.M., Pearlmutter, N.J., Myers, E., & Lotocky, M.A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37:58–93.
Hawkins, John A. (2003). Why are zero-marked phrases close to their heads? In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*. New York: Mouton de Gruyter. 175–204.
Henry, Alison. (1995). *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford: Oxford University Press.
Huddleston, Rodney, & Pullum, Geoffrey K. (eds.). (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Jaeger, T. Florian, & Wasow, Thomas. (2005). Processing as a source of accessibility effects on variation. Paper presented at the 31st annual meeting of the Berkeley Linguistics Society. To appear in *Proceedings of the 31st annual meeting of the Berkeley Linguistics Society*.
Jesperson, Otto. (1932). *A Modern English grammar on historical principles*. London: George Allen & Unwin.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press

Kroch, Anthony, & Small, Cathy. (1978). Grammatical ideology and its effect on speech. In David Sankoff (ed.), *Linguistic variation: Models and methods*. New York: Academic Press. 45–55.
McDavid, Virginia. (1964). The alternation of 'that' and zero in noun clauses. *American Speech* 39:102–113.
Pesetsky, David. (1991). *Zero syntax*, vol. 2. Manuscript, MIT.
Pesetsky, David, & Torrego, Esther. (2004). Tense, case, and the nature of syntactic categories. In Guéron, Jacqueline & Lecarme, Jacqueline (eds.), *The syntax of time*. Cambridge, MA: MIT Press. 495–537.
Poutsma, H. (1929). *A grammar of Late Modern English*. Groningen: Noordhoff.
Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, & Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. London: Longman.
Race, David S., & MacDonald, Maryellen C. (2003). The use of 'that' in the production and comprehension of object relative clauses. Proceedings of the 25th Annual Meeting of the Cognitive Science Society. 946–951.
Rissanen, Matti. (1991). On the history of *that*/zero as object clause links in English. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman. 272–289.
Rohdenburg, Günter. (1999). Clausal complementation and cognitive complexity in English. In Fritz-Wilhelm Neumann & Sabine Schülting (eds.), *Anglistentag 1998, Erfurt*. Trier: Wissenschaftlicher Verlag. 101–112.
Roland, Douglas, Elman, Jeffrey L., & Ferreira, Victor S. (2006). Why is *that*? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98:245–272.
Staum, Laura. (2005). When stylistic and social effects fail to converge: A variation study of complementizer choice. Manuscript, Stanford University.
Storms, G. (1966). *That*-clauses in Modern English. *English Studies* 47:249–270.
Stowell, Timothy. (1981). Origins of phrase structure. Doctoral dissertation, MIT.
Tagliamonte, Sali, & Smith, Jennifer. (2005). No momentary fancy! The *zero* complementizer in English dialects. *English Language and Linguistics* 9:289–309.
Thompson, Sandra A., & Mulac, Anthony. (1991a). The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics* 15:237–251.
_____ (1991b). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization* (Vol. 2). Amsterdam: Benjamins. 313–329.
Urmson, J. O. (1963). Parenthetical verbs. In Charles E. Caton (ed.), *Philosophy and ordinary language*. Urbana: University of Illinois Press. 220–246.
Yaguchi, Michiko. (2001). The function of the non-deictic *that* in English. *Journal of Pragmatics* 33:1125–1155.

https://doi.org/10.1017/S0954394507000117 Published online by Cambridge University Press