A Corpus-based Analysis of THAT-deletion in Complement Clauses after the Verbs of Saying in English and Lithuanian

Judita Giparaitė
Lithuanian University of Educational Sciences, Faculty of Education in Humanities, Studentų St. 39, LT-2034 Vilnius, judita.giparaite@leu.lt

Julija Ritčik
Modern School Centre, Žygimantų St. 4, LT-01102 Vilnius, j.ritcik@gmail.com

Abstract. The aim of the present research is to make a contrastive analysis of that-deletion in complement that-clauses in English and Lithuanian on the basis of corpus data. All the data for analysis are collected from the British National Corpus and the Corpus of the Contemporary Lithuanian Language. That-deletion in complement that-clauses is studied in terms of distribution across registers, the properties of the matrix verbs and the structural properties of the clauses.

Key words: that-deletion, complement that-clauses, verbs of saying, a corpus-based analysis, a contrastive study.

Introduction

Complement that-clauses with the complementizer omitted are equally grammatical in English as those joined by that. The omission of the complementizer is in most cases considered a free choice and does not change the meaning of the whole sentence.
(Biber et al., 1999, 680; Biber, 1999, 144), which means that the structures “can be used interchangeably” (Elsness, 1982, 1):

(1) a. I thought it was a good film.
    b. I thought that it was a good film. (Biber et al., 1999, 658)

With this in mind, the factors favouring the choice of one structure over the other have been of considerable interest for modern linguists.

The criteria for the omission or retention of the complementizer that can be defined as “partly stylistic and partly grammatical” (McDavid, 1964, 113). The grammatical factors include the syntactic function of the complement that-clause and structural ambiguity (McDavid, 1964; Elness, 1982; CGEL, 1995; CAGEL, 2002; Hudson-Ettle, 2002). The mode of communication (Finegan & Biber, 1995; Greenbaum et al., 1996; Kirk, 1997; Hudson-Ettle, 2002; Kaltenböck, 2004; Kearns, 2007), common (highly frequent) matrix verbs (McDavid, 1964; Thompson & Mulac, 1991; Finegan & Biber, 1995; Kearns, 2007), intervening elements between the matrix verb and the subject of the complement clause (Elseness, 1982, 1984; Thompson & Mulac, 1991; Rissanen, 1991; Finegan & Biber, 1995), and pronominal subjects in the main clause (Elseness, 1982, 1984; Thompson & Mulac, 1991; Rissanen, 1991; Finegan & Biber, 1995) are among the most influential stylistic and contextual factors determining the omission or retention of that in complement clauses.

The present research is concerned with the issue of that-deletion in verb complement that-clauses in the English and Lithuanian languages. The problem has at length been investigated in modern English, especially applying corpus-based methods as such methods allow to detect and evaluate different factors influencing the choice simultaneously. However, though the omission of the complementizer that (kad/jog) is possible in complement clauses in the Lithuanian language, the research into the factors which influence the choice of the structure with the zero complementizer has been insufficient and hardly any Lithuanian studies can account for this phenomenon.

The study aims at presenting a contrastive analysis of that-deletion in complement that-clauses in English and Lithuanian on the basis of corpus data. For this purpose, the contextual factors of that-deletion, such as the register, the properties of the matrix verbs and the structural properties of the clauses are analyzed in the two languages.

Due to the limited size of the article, only object clauses have been chosen for the investigation. The choice is influenced by the fact that, on the one hand, this type of that-clauses is reported as the most frequent type of complement clauses in both English and Lithuanian, and, on the other hand, these structures, as well as the conditions under which the zero complementizer can be used in them in the English language have been mostly studied and described in corpus linguistics. The latter argument is of particular importance, taking into consideration the lack of theoretical background in

---

1 Other terms used as synonyms for that-deletion are zero that, elision or omission.
Lithuanian research on the deletion of the complementizer in complement that-clauses. Another limitation concerns the verbs under consideration: only the verbs of saying are explored. This choice is determined by the fact that the verbs of saying belong to one of the semantic domains that are reported as those taking zero-that complements most often not only in English, but also in Lithuanian.

**That-deletion in English and Lithuanian**

As previously mentioned, there have been numerous investigations on the factors determining the omission or retention of that in complement clauses in English, whereas the issue has been only briefly discussed in Lithuanian.

An early corpus-based study on alternation of that and zero in complement clauses was conducted by McDavid (1964). In the research carried out on a 100,000-word corpus of American fiction, the author focuses, firstly, on the types of constructions using that or zero to introduce a complement clause, and, secondly, on the circumstances of that deletion. Among the constructions under consideration detected in the corpus the complement object clause of a transitive verb is reported as the most frequent in general, as well as with the zero that. Among the stylistic factors favouring the deletion of that, McDavid names common matrix verbs (especially know and say) and informality of language (McDavid, 1964, 113).

The influence of style is considered one of the basic factors influencing the omission or retention of that in complement clauses by Storms (1966). According to the author, that deletion adds an ‘element of subjectivity’, while that retention makes the clause “less personal, less familiar, less warm, less friendly, less emotive”, but, instead “objective, factual, formal, official” (Storms, 1966, 262). In the scholar’s view, it is the level of subjectivity that distinguishes colloquial spoken language from “formal, written or elevated speech” (op. cit.: 262).

A thorough examination of factors conditioning that deletion is made by Elseness (1982; 1984) in his works on the choice of that or zero–that complement clauses and, in particular, object that-clauses studied in four types of texts in 128,000-word Syntax Data Corpus, a part of Brown University Corpus of American English published in 1961. Elness emphasises that a number of factors can simultaneously have an impact on the choice of the form of the complementizer (1984, 533). For complement clauses in general, the factors favouring that–deletion include informality of style, syntactic function of the object of the matrix verb, the absence of potential ambiguity, lack of structural complexity near clause boundary, and closeness of the clause juncture (1982, 39–41).

Thompson & Mulac (1991) study the discourse conditions of the use of the complementizer that in spoken American English on the basis of the 240,000 word
corpus of American college students’ conversations. The results of this quantitative study reveal that object \textit{that}-clauses with the zero \textit{that} are most frequent with the matrix verbs \textit{think} and \textit{guess} in the 1\textsuperscript{st} or 2\textsuperscript{nd} person singular and a pronominal subject in the complement clause, while the retention of \textit{that} is more likely when the matrix clause includes an adverbial, an auxiliary or an indirect object. Thompson & Mulac point out that certain combinations of the subject and verb in the main clause like, for instance, \textit{I think} should be reanalyzed as “unitary epistemic phrases”, with the main idea of the sentence being expressed by the complement clause\(^2\) (1991, 237).

Besides synchronic investigations, there have also been diachronic studies tracing the historical development of the zero and full forms of the complementizer \textit{that}. Rissanen (1991) investigates the zero-\textit{that} and \textit{that} as object clause links on the basis of data from Helsinki Corpus of English Texts covering the period from Late Middle English (1350–1420) to Early Modern English (1640–1710). It is claimed that both zero-\textit{that} and \textit{that} as object clause links have existed in English for centuries and represent two variant links (ibid, 288). The author expresses the intuition that in spoken mode zero “may <...> have been the unmarked object clause link throughout the history of English” (op. cit.: 287), and thus, the very term ‘omission’ of the complementizer is incorrect. However, the growing frequency of zero in written English, where \textit{that} used to be “the unmarked link”, allows to interpret this novel tendency “as the ‘omission’ of the expressed conjunction” (op. cit.: 288). In his study, Rissanen also reports a consistent general increase of the zero-\textit{that} over the period of time under consideration after the matrix verbs \textit{say}, \textit{tell}, \textit{know}, and \textit{think} and lists the factors, such as a pronominal subject in the object clause, the lack of intervening elements between the matrix and the object clauses, and spoken mode of communication as favouring the choice of the zero \textit{that} (op. cit.: 286).

Another diachronic study on the two forms of the complementizer \textit{that} in complement object clauses is carried out by Finegan & Biber (1995). They analyze the use of the complementizer in three registers of British English in the period from 1650 to 1990 on the basis of data from approximately 1.7 million-word ARCHER Corpus and demonstrate that the tendency of the increasing use of the zero-\textit{that} in late Middle and early Modern English reported by Rissanen is opposite in some registers, viz. sermon, medical articles, and letters (1995, 247). The findings of this research generally confirm the conditioning factors of the use of \textit{that} and zero-\textit{that} mentioned in previous studies by McDavid (1964), Elsness (1984), and Thompson & Mulac (1991), except for the

\(^2\) In both English and Lithuanian, the matrix clauses that select zero-\textit{that} resemble parentheticals. However, the latter differ from the matrix clauses in that verbs in parentheticals undergo semantic bleaching and become either epistemic modifiers or pragmatic and discourse markers (Thompson & Mulac, 1991, 249; Brinton, 2008, 2; Kaltenböck, 2008, 124; Kaltenböck, 2009, 67). For a detailed account on the problem in Lithuanian refer to Sirtautas & Grenda (1988), Labutis (2002), Akelaitis (2002), Ruskan (2010).
fact that the results of this particular analysis do not show the coreferentiality between the subjects of the main and the complement clause as influential in the choice of the zero-\textit{that}, as it was earlier suggested by Elsness (1984) (op. cit.: 255)\textsuperscript{3}.

A more recent corpus-based study by Kearns (2007) analyses the use of the zero-\textit{that} in complement clauses after epistemic verbs. The investigation is conducted on the basis of the corpus of 2,178 tokens of complements to epistemic verbs, collected from eight broadsheet newspapers from the USA, the UK, Australia and New Zealand. The results based on the statistical data obtained in the process of this research correlate with most findings of previous studies on \textit{that}-complement clauses, as well as reveal some novel facts about regional differences in the use of zero-\textit{that} complements: the rate of zero-\textit{that} appeared to be the highest in the texts from New Zealand, significantly lower in Australian and American texts, and the lowest in the texts from the UK (op. cit.: 495).

There have been a considerable number of current corpus-based studies on subordinate and complement clauses confirming the majority of the earlier results discussed above. Greenbaum et al. (1996) report formality in both spoken and written modes as a factor influencing \textit{that} retention in complement clauses. Kirk (1997), Hudson-Ettle (2002), and Kearns (2007) observe that the zero-\textit{that} is generally more frequent in spoken mode of communication (Kirk, 1997, 360; Hudson-Ettle, 2002, 261; Kearns, 2007, 490). Biber (1999) distinguishes two types of factors influencing the choice of the form of the complementizer: \textit{stylistic} (zero-\textit{that} prevails in conversation while \textit{that} in academic writing), and \textit{textual} (zero-\textit{that} is typical of \textit{that}-clauses with the matrix verbs \textit{think} or \textit{say}\textsuperscript{4} and coreferential subjects in the matrix and complement clauses, while \textit{that} is more frequently chosen in coordinated \textit{that}-clauses, as well as when the matrix verb is in the passive form or there is an intervening noun phrase between the matrix verb and the complement clause) (1999, 144–147). The same factors are reported by Biber et al. (1999, 680–681). The view is supported by Kearns (2007) who claims that with a particular class of epistemic verbs intervening elements between the main and the complement clause, such as indirect objects and matrix adverbials also favour \textit{that} retention, while pronominal subjects in the complement clause favour zero-\textit{that} (2007, 491–492).

In the few studies on \textit{that}-deletion in complement clauses in Lithuanian, zero-\textit{that} clauses are referred to as asyndetic clauses. LKG III (1976) briefly discusses the factors influencing the choice of asyndetic versus syndetic complement structures. According to LKG III, asyndetic clauses can have different meanings and it is often difficult to differentiate them, thus, asyndetic clauses are hardly ever used in academic and formal language and are more typical of spoken mode of communication, as well as of fiction

\textsuperscript{3} The results of a corpus-based study by Kearns (2007) also show that coreferentiality does not affect zero rates in complement clauses.

\textsuperscript{4} Also proved in Rissanen (1991) and Kearns (2007).
and folklore (1976, 917). LKG III also lists the verbs of *saying*, *thinking*, and *perception* as the most frequent matrix verbs with asyndetic object clauses, especially when the verb is positioned at the end of the main clause and emphasises the link between the clauses (op. cit.: 937).

One more research on Lithuanian asyndetic complex clauses has been performed by Drotvinas (1961). In his study, the scholar states that asyndetic complex clauses are infrequent when compared to their syndetic equivalents and often lack explicit criteria for classification (1961, 185)⁵. Asyndetic complement clauses functioning as objects are claimed to be most often found with verbs of *saying* and *perception* in present or past tense first and third person singular form (op. cit.: 191). Such clauses can also be used to express indirect speech (op. cit.: 191).

Lithuanian asyndetic clauses could be said to differ from syndetic ones in their punctuation. In Lithuanian all syndetic complement clauses are separated by commas, whereas asyndetic clauses are usually separated by a colon or a dash, and less often by a comma (Dambriūnas, 1963, 69; LKG III, 1976, 937).

Similarly to English zero-*that* complement clauses, Lithuanian asyndetic complement clauses are considered more typical of spoken and less formal language and are reported to be frequently used with the verbs related to speech acts and mental states or activities. The present study is the first attempt to research Lithuanian asyndetic complement clauses on the basis of corpus data, which might reveal more about the features of these clauses and the factors favouring their choice over syndetic clauses.

**Data and Methodology**

The data for analysis were collected from the *British National Corpus* (BNC) (http://corpus.byu.edu/bnc/) and the *Corpus of the Contemporary Lithuanian Language* (LLC) (http://tekstynas.vdu.lt). Both corpora are representative electronic databases of spoken and written English and Lithuanian. They differ in size and the date of compilation. The BNC corpus contains almost 100 million words and was compiled in the period from the 1980s to the early 1990s, whereas the size of the LLC is more than 140 million words and its compilation period is from 1992 to 2011. Due to the different design of the investigated corpora, only particular text categories are incorporated into the study. The following subsections of the BNC and the LLC were searched in the process of collecting the data (as these subsections are of relative stylistic correspondence in both corpora): *spoken*, *fiction*, *newspaper* and *magazine* of the BNC and *sakytinė kalba* “spoken”, *grožinė literatūra* “fiction”, and *publicistika* “journalism” of the LLC. The registers under investigation are unequally represented in both corpora.

⁵ Sirtautas & Grenda (1988, 143) also consider asyndetic complex clauses as a minor class.
Due to the tremendous size of the corpora and absence of specific software simplifying the search for evidence, quantitative analysis is excluded from the present research.

The data collection was carried out in the following stages. First, the lists of English and Lithuanian verbs of saying taking complement that-clauses were produced. The initial list of English verbs was taken from CGP (1998, 98–99), as this list of verbs is the result of a detailed corpus-based study of verb complements in English. To produce the list of the Lithuanian verbs of saying, the English verbs were translated into the Lithuanian language and the Lithuanian equivalents were checked in the verb valency dictionary (Sližienė, 1994, 1998, 2004) in order to select the verbs which can be followed by complement that-clauses in Lithuanian as well as to detect synonyms which were not included in the list produced by the translation. In both languages only the verbs that can take complement that-clauses in their main (the first and second) meanings were selected.

Second, the English and Lithuanian verbs were checked in the corpora, namely, the BNC and the LLC, to ensure that the verbs occur in these corpora with that-complements. The final lists of matrix verbs under investigation comprised 87 English and 75 Lithuanian verbs of saying. After that, the examples of complex sentences with these matrix verbs followed by zero-that object complements were collected.

Finally, the collected data were analysed, and the cases of zero-that complement clauses in English and Lithuanian were described and contrasted. In analyzing and interpreting the data special attention was paid to structural and semantic peculiarities of English and Lithuanian zero-that complement clauses as well as registers typical of this type clauses.

### Zero that in Complement Clauses after the Verbs of Saying in English and Lithuanian: Findings

The verbs of saying under consideration “are concerned with speaking, writing and other forms of communication” (CGP, 1998, 98). The corpus evidence collected reveals

---

The registers of magazine and fiction of the BNC are considered together as one, as there is one section of journalism in the LLC, which can be comparable with these two.
the following facts about the omission of the complementizer *that* in verb complement clauses in the two languages under consideration.

The most obvious observation is that the omission of the complementizer *that* is by far more common in the English language than in Lithuanian. Out of 87 English verbs of *saying* studied, examples of 29 verbs that take zero-*that* complements were found in all registers under investigation: *admit, allege, announce, claim, complain, confess, declare, decree, deny, emphasize, explain, hint, imply, indicate, insist, maintain, mention, pray, predict, promise, propose, recommend, report, reveal, say, stress, suggest, swear, warn.*

Structurally, the selected examples show that most of these verbs exhibit a considerable variety of tense forms and types of subjects in the main as well as the complement clauses, which can be illustrated with the following examples from the BNC:

(2) a. Sergeant Allen Bovington Cox says **specialist teams are involved.** *(spoken)*
   b. And they said **it was our car.** *(spoken)*
   c. And I said **I’m not talking about the council.** *(spoken)*
   d. The hotelier said **it was Italy.** *(fiction)*
   e. You said you were thrown out by your landlord. *(fiction)*
   f. But he could <...> have said **one is enthusiasm and the other is cynicism.** *(fiction)*
   g. At least one member of the jury said **he would not have given a sentence of death.** *(magazine)*
   h. No one’s saying the Escort 1.4LX is a bad car <...>. *(magazine)*
   i. <...> Antonia said **she was upset and worried.** *(newspaper)*

As the examples demonstrate, the subjects in both the main and the complement clauses can be pronominal as in (2b), (2c), (2e), and (2f), expressed by a noun as in (2i), or a noun phrase as in (2a) and (2g). As the present study excludes frequency analysis, it is difficult to assess how significant the type of the subject in the main or the complement clause is for the choice of the zero-*that* complement. However, it should be noted that the corpus evidence suggests that pronominal subjects in both the main and the complement clauses, as well as coreferential subjects, as in (2c), (2e), (2g), and (2i), are frequent with the matrix verbs of *saying* followed by zero-*that* complements, which is in complete accord with the observations made in corpus-based studies of English zero-*that* complement clauses (Elsness, 1984, 533; Thompson & Mulac, 1991, 237; Rissanen, 1991, 288; Finegan & Biber, 1995, 255; Biber, 1999, 144).

In respect to contextual factors, some verbs have shown preference to be used with zero-*that* complements in certain registers. For instance, the verbs *advise, boast, brag,*

---

7 If not specified, all English examples in the research are taken from the BNC and all Lithuanian examples from the LLC corpora.
concede, remark, signify, stipulate, and wager are not used with zero-that complements in the spoken register, but occur in the fiction and magazine and newspaper registers:

(3) a. She was waved on by a sharp-eyed young officer, who boasted he could smell a smuggler from fifty yards away. (fiction)
   b. <...> we visit his home... still filled with hundreds of beautiful artefacts which proudly boast they were Hand Made in the Cotswolds. (newspaper)
   c. I’ll wager you’ve forgotten. (fiction)
   d. I’ll wager you would need to be pointing almost vertically downhill to achieve such a mind-numbing speed. (newspaper)

Other verbs, such as assert, beg, proclaim, prophesy, and request, are only found with zero-that complements in the fiction register:

(4) a. I beg you’ll send me out of my way to my poor parents. (fiction)
   b. Suppose he needed you there to assert he’d walked into a trap? (fiction)
   c. They had one likeness of Oreste, sent at Christmas when Wilson had requested it should be taken. (fiction)

Different distribution of these verbs across the registers can be explained by the fact that they belong to more formal style and are generally infrequent in spoken language and, on the other hand, that some of them are more typical of fiction than journalism.

There are two more substantial groups of matrix verbs that only occasionally occur with zero-that complements (a) or do not take zero-that complements at all (b): (a) advocate, affirm, argue, assert, beg, brag, caution, command, comment, concede, confide, contend, demand, dictate, disclose, forecast, grumble, insinuate, instruct, joke, moan, ordain, order, plead, pledge, preach, proclaim, prophesy, remark, request, signal, signify, specify, stipulate, submit, threaten, underline, wager; (b) attest, aver, divulge, enthuse, foretell, grouse, lament, opine, posit, postulate, profess, quip, remonstrate, sneer, testify, underscore.

The two groups include the most infrequent verbs of saying in the BNC corpus (the total number of occurrences of each of these verbs across the corpus in most cases is lower than 1,000). This observation generally confirms the findings of previous corpus-based researches which claim that rare verbs, typically stylistically marked and more formal, usually do not take zero-that complements while common matrix verbs favour the deletion of that (McDavid, 1964, 113; Biber, 1999, 147; Biber et al., 1999, 680).

Last but not least, the corpus data reveal one more distinct feature of zero-that complements with matrix verbs of saying in English. Following Elsness (1982), it can be labelled as closeness of the clause juncture8 (1982, 41). As seen in examples (2)–(4),

---

8 Elsness (1982) describes closeness of the clause juncture in terms of three features: the subject in the object complement clause is pronominal and coreferential with the subject in the matrix clause, the subject in the object clause has definite, often anaphoric, reference, and subjects in both matrix and object clauses are expressed by 1st and 2nd person pronouns (1982, 39–41).
zero-\textit{that} clauses typically follow the matrix verb immediately, and pronominal subjects in the complement clause have definite anaphoric reference.

Regarding the Lithuanian verbs of \textit{saying} followed by zero-\textit{that} complements, it can be stated that rare verbs do not show any tendency for the omission of the complementizer \textit{kad/jog}, either. The verbs \textit{pagrasinti} “threaten”, \textit{įspėti} “warn”, \textit{komentuoti} “comment”, \textit{postuluoti} “postulate”, \textit{prisiekėti} “swear”, \textit{susitarti} “agree” occur with zero-\textit{that} in rare cases, and the most infrequent verbs in the LLC (the total number of occurrences across the corpus does not exceed 2,000 for each verb) listed below have no instances of zero-\textit{that} complements in the corpus: \textit{anonsuoti} “announce”, \textit{bambėti} “grouse”, \textit{bėdoti} “moan”, \textit{burbtelėti} “grumble”, \textit{dievagotis} “swear”, \textit{dievažytis} “swear”, \textit{ginčytis} “argue”, \textit{įsakyti} “command”, \textit{pamokslauti} “preach”, \textit{propaguoti} “advocate”, \textit{protestuoti} “remonstrate”, \textit{sielotis} “lament”, and \textit{signalizuoti} “signal”.

As mentioned above, Lithuanian verbs of \textit{saying} are generally more infrequent with zero-\textit{that} complements than their English counterparts. In comparison, 29 out of 87 English verbs of \textit{saying} found in the BNC select zero-\textit{that} complements in all registers, whereas in the LLC only 6 out of 75 Lithuanian verbs of \textit{saying}, namely \textit{pripažinti} “admit”, \textit{prisipažinti} “confess”, \textit{pažadėti} “promise”, \textit{kalbėti} “speak/say”, \textit{šnekėti} “speak/say”, and \textit{sakyti} “say” take complement clauses with the complementizer \textit{that} omitted in all three registers under consideration. Only the verb \textit{sakyti} “say” followed by zero-\textit{that} complements is similar to its English counterpart in respect to the variety of tense forms and types of subjects, as illustrated in example (5):

(5) a. - Ai, greis 	extit{sakai}, 	extit{ta mergaitė} kažkaip tai mirs kaip ir reklamava <...>.

(speaking)

[- Oh, you correctly \textit{say the girl is going to somehow die, as it was announced} <...>.]

b. Graboriai \textit{sako}, 	extit{žmonės gimsta tam, kad mirtų} <...>.

(fiction)

[\textit{Coffin makers say people are born to die} <...>.]

c. Aš tu 	extit{sakau}, 	extit{stalas šoko}! (spoken)

[I’m telling you \textit{the table was dancing}!]

d. Tu 	extit{sakai}, 	extit{kėlei ranką prieš ją}. Kaip? Kodėl?

(speaking)

[You \textit{say (you) raised your hand against her}. How? Why?]

e. Taigi, \textit{ie sako}, \textit{nereikia jo spausti prie sienos} <...>.

(journalism)

[So, they \textit{say (we) shouldn’t push him to the wall} <...>.]

f. Ir \textit{jinai sakė}, \textit{mes pabūsim pas ją, išgersim šampano} <...>.

(speaking)

[And she \textit{said we’d stay at hers, have some champagne} <...>.]

The English verb \textit{say} is reported as one of the most frequent matrix verbs taking zero-\textit{that} complements. The Lithuanian verb \textit{sakyti} “say” is to a certain extent similar in this respect.

Another surprising difference between the English and Lithuanian zero-\textit{that} complement clauses after the verbs of \textit{saying} is that only a few cases of zero-\textit{that}
complements were found in the *spoken* register of the LLC, which is completely opposite to the tendencies reported in numerous corpus-based studies of English complement clauses, where the spoken mode of communication, less formal than the written mode, is claimed as one of the most influential factors on the omission of *that* (Storms, 1966, 262; Elsness, 1982, 39; Rissanen, 1991, 286; Kirk, 1997, 360; Biber, 1999, 144; Hudson-Ettle, 2002, 261; Kearns, 2007, 490). This also contradicts the general assumption that Lithuanian asyndetic sentences are more common in the spoken mode of communication (LKG III, 1976, 917; DLKG, 2005, 721). However, the small number of zero-*that* complements in the *spoken* register in the LLC might be explained by the fact that the *spoken* section of the LLC is relatively new and still small in size. Many verbs, especially less frequent ones, cannot be found in the *spoken* section of the LLC at all, and this is the reason why zero-*that* complements in Lithuanian can hardly be properly studied in the *spoken* register of the LLC and compared to zero-*that* complements in the BNC, where the *spoken* section is much bigger (refer to Table 1). With this in mind, it would be precipitated to claim that the omission of *that* is not characteristic of the spoken mode of communication in Lithuanian.

On the whole, the examples of zero-*that* complements in Lithuanian that resemble the English counterparts confirm the tendency observed by Drotvinas (1961): the matrix verbs are typically used in the present or past tense first or third person singular form (1961, 191). They also conform to the pattern where the matrix verb is positioned at the end of the main clause, which is described as a feature of asyndetic object clauses in LKG III (1976, 937). The pattern that resembles the *closeness of the clause juncture* defined by Elsness (1982, 41) can be illustrated by the following examples:

(6) a. Jis *sako*, *mudu tikrai galėtume tai padaryti* <...>. (*fiction*)
   [He *says* *we two* could really do that <...>.]
b. Aš *prisipažįstu*, *aš kaltas*. (*spoken*)
   [I *admit* I’m guilty.]
c. Pats *pripažįstu*, *tuomet aš buvau lyg apakęs*. (*journalism*)
   [I *admit* I was quasi blind then.]
d. Aš tau *byloju*, *kai kurios iš jų reformų yra tiesiog pasišlykštėtinos* <...>. (*fiction*)
   [I’m telling you *some of their reforms are just revolting* <...>.]
e. Esu garbingas žmogus, todėl *pažadu*, *jais nepasinaudodiu*. (*journalism*)
   [I’m a respectable man, therefore (I) *promise* I won’t use them.]
f. Tik įspėju, *ta būtybė nepakenčiama!* (*fiction*)
   [But I *warn* the *creature* is intolerable!]

The sentences in examples (6b), (6c), and (6e) also illustrate that the subject of the main and the zero-*that* complement clauses in Lithuanian can be coreferential. However, in general, the corpus evidence from the LLC does not suggest that this is a typical feature of Lithuanian zero-*that* complement clauses after the verbs of *saying*. The subject
in the complement clause is often not coreferential with the subject in the main clause, as seen in (6a) and (6d) and often is not even pronominal, as in examples (6d) and (6f) or (5a)–(5c). In English, on the contrary, pronominal subjects in the complement clauses are reported as a factor favouring that deletion. On the other hand, in Lithuanian the subject of the complement clause, even if not pronominal, usually implies anaphoric reference, as in examples (6e) and (6f), which corresponds to the tendency reported by Elsness (1982, 41). However, only a quantitative corpus-based study could reveal how significant the type of the subjects in the main and complement clauses is for the choice of the zero-that complementizer in Lithuanian.

Furthermore, the corpus evidence suggests that in Lithuanian the deletion of that in complement clauses after the verbs of saying is influenced by structural peculiarities of the clause. There is a clear tendency for the choice of zero-that complementizer in cases where the main clause is preceded by the particles esą or neva (7a) and (7b), which both can be translated as “ostensibly/ allegedly/supposedly” or a parenthetical girdi “(you) hear” (7c), which is very close in meaning to the mentioned particles.

(7) a. Dar kiti teigė, neva vaikus parduoda vergais turtingiems totoriam. (fiction)
   [Still others claimed (?) the children were sold as slaves to rich Tatars.]
b. Jaunuoliai <...> ligoninėje melavo, esą juos sumušė chuliganai. (journalism)
   [In the hospital the young men <...> lied (?) they had been beaten by hooligans.]
c. Kai kurie galbūt pakomentuos, girdi, sekso reikalauose svarbu vien meistriškumas. (fiction)
   [Some might comment (?) only excellence matters in sex affairs.]

Esą is defined as a particle used to introduce the indirect speech (LKŽ; Wiemer, 2007, 177). The latest edition of DLKŽ (2012) classifies it as a modal lexical element (2012:153). Neva is defined as a particle expressing doubt (LKŽ; DLKG, 2005, 435; DLKŽ, 2012, 423), whereas girdi is described as a parenthetical (DLKŽ, 2012, 177) or as a particle (LKŽ) with the meaning synonymous to esą and is used to report somebody’s words (Wiemer, 2007, 177). On the whole, in more recent studies all these three lexical items have been ascribed to lexical means of expressing evidentiality⁹ (Ruskan, 2010, 7). Wiemer (2007) describes the particles esą and neva, as well as the parenthetical girdi as hearsay or reportive evidential markers (2007, 177–180).

It has been generally assumed that the particles esą and neva can function as complementizers (Būda, 1986, 55; Wiemer, 2007, 178; Ruskan, 2010, 8). Būda (1986)

---

⁹ According to Ruskan (2010), in contemporary linguistics evidentiality is considered a semantic (cognitive) grammatical category used to indicate the cognitive or/and communicative basis of the speaker’s utterance.
also considers the parenthetical girdi as a type of complementizer (Būda, 1986, 56). This would explain the omission of that in complement clauses preceded by esą, neva, and girdi. However, corpus evidence shows that complement clauses preceded by these lexical items can also be found together with the complementizer kad:

(8)  
  a. O paskui <...> pareiškė, kad neva aš ir jos vyras esam kažkuo panašūs!  
      (fiction)  
      [And then <...> claimed that (?) I and her husband were similar in a way!]  
  b. <...> pacifistai susirūpinusiais veidais aiškina, kad esą negalima erzinti Rusijos, reikia su ja draugauti.  
      (journalism)  
      [<...> anxious-faced pacifists comment that (?) we mustn’t tease Russia, we have to get on well with it.]  
  c. <...> Viktorijos tėvai į tokius žodžius atrėžė, kad, girdi, šis jau niekados nebeparvyks  
      (fiction)  
      [<...> Victoria’s parents retorted to these words that (?) he would never return again.]  

Esą, neva, and girdi in (8) can be interpreted in two ways. First, they can be described as evidential particles preceded by the complementizer that. Second, esą, neva, and girdi might be analysed as complementizers. In this case the structural duality could be explained by an assumption that such lexical items, permanently positioned next to the complementizer kad/jog, over the time will supersede it (Būda, 1986, 57). In a language, the process of lexical items changing their grammatical status and function is very slow, uneven and gradual; therefore, the co-existence in a language of older and newer structures at the same time is an inseparable feature of the syntactic development (Ambrazas, 2006, 25). Another fact signalling that these lexical items still have not completely acquired the function of the complementizer is that they are occasionally used in sentences with the matrix clause separated from the complement by the colon, which happens only in asyndetic sentences, as illustrated in example (9):

(9)  
  a. Pryn kartą su juo susiginčijo: esą barai jai patinka <...>.  
      (fiction)  
      [Once Pryn argued with him (that) she likes bars <...>.]  
  b. <...> buhalteriai, galbūt ims mums priekaištauti: girdi, yra premijos, kategorijos <...>.  
      (journalism)  
      [<...> accountants may expostulate (that) there are rewards, categories <...>.]  

In regard to English complement that-clauses, it can be observed that there are no directly corresponding structures. Therefore, it is difficult to compare them in the two languages. On the whole, considering the tendency of esą, neva and girdi to adopt the function of the complementizer, it would be logical to assume that these structures should rather be regarded as that-complement clauses with the overt form of the complementizer and the Lithuanian instances like those in example (7) should be translated with that in the place of esą, neva and girdi. Such a translation, however,
would not convey the evidential meaning of these lexical items, thus the translator should incorporate some compensational strategies in order to preserve it.

Another structural peculiarity of Lithuanian complement clauses after the verbs of saying which seems to favour the deletion of that is the complexity of the complement clause itself. The complementizer that is often absent when the complement clause is a complex sentence of condition joined to the matrix clause by the conjunction jei/jeigu “if”:

(10) a. Jis užsiminė, *jei situacija Lietuvos rinkoje blogės, DINAMIKA tauru* *koncertuoti už Atlanto. (journalism)*
   [He mentioned (that) *if the situation on the Lithuanian market deteriorates, DINAMIKA will head to concert across the Atlantic.*]
   b. Vaivada pats pažadėjo – *jeigu geras pasirodysite, važiuosite į Briuselį. (fiction)*
   [Vaivada promised himself – *if (you) perform well, (you)’ll go to Brussels.*]
   c. Dar pagrasisino: *jei jaunuolis neatveš dolerų, kreipsis į policiją. (journalism)*
   [(He) also threatened: *if the youngster doesn’t bring some dollars, (he) will go to the police.*]

It must be mentioned that such structures typically occur in the journalism section of the LLC, and the most common punctuation mark in them is the colon, as in (10c). In the BNC, the corresponding complement clauses are typically used with the overt form of that10, as illustrated in (11):

(11) a. He threatened *that if the Moldavian authorities did not notify him within 10 days <...> he would take “necessary measures” <...>. (non-academic)*
   b. Mr Kinnock promised *that if Labour come to power, all that will change. (spoken)*

Another exclusive feature of zero–that complement clauses after the verbs of saying in Lithuanian suggested by the corpus evidence is that the colon as a punctuation mark for the separation of the main clause from the complement clause is by far more typical than the comma in written language as the examples of such structures were found in LLC for 46 of the 75 studied matrix verbs. Unlike Lithuanian, no punctuation marks are used to separate zero–that complement clauses from main clauses in English.

---

10 The clauses where the complementizer that precedes the conditional are also characteristic of Lithuanian: Dar kiti sakys, *kad jei nepatinka, kreipkis į privatininkus <...>. (journalism)* [Still others will say that if (you) do not like (it), address private traders <...>.]
Conclusions

The thorough examination of the corpus data collected for the verbs of saying followed by complement clauses with the complementizer that omitted in the English and Lithuanian languages led to the following conclusions.

The zero-that complements are by far more frequent in English than in Lithuanian as a greater variety of English verbs of saying select zero-that complements in all the three investigated registers.

Surprisingly, few cases of zero-that complements occur with Lithuanian verbs in the spoken register, which, on the one hand, is opposite to the English language, in which the spoken mode of communication is considered a strong factor influencing the omission of that, and, on the other hand, contradicts the theoretical claims about asyndetic sentences being more characteristic of Lithuanian spoken language. However, this might be a result of the small size and content insufficiency of the spoken section of the LLC.

In Lithuanian, the zero-that is likely if the complement clause is preceded by evidential markers, such as esq, neva “supposedly”, girdi “(you) hear”. Such clauses cannot be regarded as instances of that deletion since in contemporary Lithuanian linguistics these items tend to function as conjunctions. However, the view needs further investigation. The English zero-that complement clauses do not have lexical elements alternative to the complementizer that.

Lithuanian zero-that complement clauses show greater complexity as they can be combined with conditional clauses, whereas in English the combination of complement that-clauses with conditionals is possible only if that is overt.

A major similarity between the English and the Lithuanian zero-that complement clauses is the structural peculiarity known as closeness of clause juncture, which is manifested by the absence of intervening elements between the matrix verb and the subject of the complement clause as well as by the subjects having definite and usually anaphoric reference. This finding is in accord with previous corpus-based studies on the English zero-that complement clauses.

Another major similarity between zero-that complements in the two languages is that the most frequent matrix verbs in both languages, such as say favour the deletion of the complementizer, while the rare matrix verbs tend to retain that. This finding agrees with the previous corpus-based studies on English zero-that complement clauses.

The current study was concerned with qualitative research, which enables one to detect certain tendencies for the deletion of the complementizer that in English and Lithuanian but does not allow to account for the significance of factors influencing the choice of the zero or overt form of the complementizer. Therefore, taking into account that no corpus-based studies on these structures have been conducted so far in Lithuanian, further corpus-based studies on the issue would be essential. First,
a corpus-based study on Lithuanian verbs taking *that*-complements would result in a complete list of Lithuanian verbs of this lexico-grammatical pattern. Second, quantitative corpus-based studies would provide more information upon how influential the factors favouring the choice of zero-*that* are in Lithuanian. Finally, parallel corpus-based studies of English–Lithuanian and Lithuanian–English texts would give a more detailed account for similarities and differences between *that*- and zero-*that* complement clauses in the two languages, which would be of substantial value in translation theory as well as in language teaching.

**Sources**


**Dictionaries**


**References**


Tekstynų duomenimis pagrįsta jungtukų *kad* ir *that* praleidimo atvejų analizė kalbėjimo veiksmažodžių reikalaujamuose šalutiniuose sakiniuose anglų ir lietuvių kalbose

Judita Giparaitė

Lietuvos edukologijos universitetas, Studentų g. 39, LT-2034 Vilnius, judita.giparaite@leu.lt

Julija Ritčik

Šiuolaikinės mokyklos centras, Žygimantų g. 4, LT-01102 Vilnius, j.ritcik@gmail.com

Santrauka

Straipsnyje analizuojami jungtukų *kad* / *jog* ir *that* praleidimo atvejai anglų ir lietuvių kalbų šalutiniuose prijungiamuosiuose (complement) sakiniuose, kurie yra valdomi kalbėjimo veiksmažodžių. Šiuo tekstyno medžiaga pagrįsta tyrimu siekiama išsiaiškinti, kokie kontekstiniai veiksmai turi įtakos jungtukų *kad* / *jog* ir *that* praleidimui, ypatingas dėmesys skiriamas registrui, šalutinius sakinius valdančių veiksmažodžių savybėms, sakinių sandaros ypatumams. Visa tyrimui būtina medžiaga surinkta iš Britų nacionalinio tekstyno ir Dabartinės lietuvių kalbos tekstyno.

Tyrimas parodė, kad, nors jungtukų praleidimas šalutiniuose sakiniuose yra būdingesnis anglų nei lietuvių kalbai, tačiau abiejose kalbose esama panašumų. Abiem kalboms būdinga tai, kad jungtukai *kad* / *jog* ir *that* dažniau praleidžiami sakiniuose su dažniausiai vartojamais kalbėjimo veiksmažodžiais. Be to, abiejose kalbose panaši ir šalutinių sakinių su praleistu jungtuku sandara: nėra antrininkių sakinio dalių, įterpiamų tarp valdančiojo veiksmažodžio ir šalutinio dėmens veiksnio. Tačiau yra ir skirtumų. Jungtuko *kad* / *jog* nebuvo lietuvių kalboje tikėtinas, jei šalutinis dėmuo pradedamas evidencialumo raiškos priemonėms priskiriamas žodžiais *esq, neva* ir *girdi*, kurie šiuolaikinėje lietuvių kalboje neretai atlieka jungtukų funkciją, o anglų kalbai tokios struktūros nebūdingos. Tik lietuvių kalbai yra būdingi atvejai, kai jungtukas *kad* / *jog* praleidžiamas pirmojo laipsnio šalutiniame sakinyme, kai prieš jį įterpiamas prijungiamasis antrojo laipsnio šalutinis sąlygos sakinys.

**Esminiai žodžiai:** jungtukų *kad* / *jog* ir *that* praleidimas, prijungiamieji sakiniai, kalbėjimo veiksmažodžiai, tekstyno duomenimis pagrįstas tyrimas, greitinamasis tyrimas.

Gauta 2017 04 26 / Received 26 04 2017
Priimta 2017 11 27 / Accepted 27 11 2017