THE THAT/ZERO-THAT ALTERNATION WITH 'I THINK' ACROSS FORMAL AND INFORMAL REGISTERS

RESMİ VE GAYRİ RESMİ SÖYLEMDE 'I THINK' İLE THAT/BOŞ THAT DEĞİŞİMİ

Songül TAŞ, Fatma SOLMAZ

65

Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi
Yıl: 2026, Sayfa: 1-16
https://dergipark.org.tr/tr/pub/cusosbil

DOI: 10.35379/cusosbil.1654315

Geliş Received: **20.03.2025**
Kabul Accepted **02.12.2025**

THE THAT/ZERO-THAT ALTERNATION WITH 'I THINK' ACROSS FORMAL AND INFORMAL REGISTERS [1]

RESMİ VE GAYRİ RESMİ SÖYLEMDE 'I THINK' İLE THAT/BOŞ THAT DEĞİŞİMİ

Songül TAŞ [2], Fatma SOLMAZ [3]

Keywords:
Comment clause,
That complementation,
Zero complementation,
That/zero-that alternation,
Corpus analysis

Anahtar Kelimeler:
Yorum cümlesi,
That tamlaması,
Boş tamlama,
That/boş that değişimi,
Derlem analizi.

ABSTRACT

The flourishing development of corpus linguistics has significantly advanced research on discourse markers; with I think being one of the most extensively studied examples. Recognizing the scarcity in the current literature regarding the alternation between that- complement and zero-that complement, this study aimed to investigate this variation in the verb think through a corpus-based analysis involving both spoken and written corpus. The overall results showed that I think with a zero-complementizer appears more frequently than I think with a that-complementizer, regardless of the genre, even though I think appears more frequently in the spoken corpus. Further results also showed that the that-complementizer is less frequent, irrespective of the formality of the genres. In addition, I think with a zero-that complement was found to occur more often in the sentence-initial position across genres, while it was rare in medial and final positions. Lastly, the findings demonstrated that instances of intervening adverbials before the that-complementizer are not frequent, whereas the zero-complement alternation with intervening adverbials is more frequent.

ÖZ

Derlem bilim dilbiliminin gelişen yükselişi, söylem belirleyicileri üzerine yapılan araştırmaları önemli ölçüde ilerletmiştir; I think ise en çok incelenen örneklerden biri olmuştur. Mevcut literatürde that ve boş that tamamlaması arasındaki değişimle ilgili sınırlı çalışma olduğunu dikkate alarak, bu çalışma hem sözlü hem de yazılı derlem verilerini içeren derlem tabanlı bir analiz yoluyla think fiiliyle kullanılan that/boş that değişimini incelemeyi amaçlamıştır. Genel sonuçlar, I think ifadesinin boş tamamlayıcı ile, that tamamlayıcı ile olduğundan daha sık kullanıldığını ve bunun türden bağımsız olarak geçerli olduğunu göstermiştir; ancak, I think ifadesi sözlü derlemde daha yaygın olarak görülmüştür. Daha ileri bulgular, that tamamlayıcısının, türlerin resmiyet düzeyinden bağımsız olarak, daha az sıklıkla kullanıldığını ortaya koymuştur. Ayrıca, I think ifadesinin boş that tamamlayıcısı ile türler arasında daha çok cümle başında yer aldığı, ancak cümlenin orta ve son konumlarında nadiren görüldüğü tespit edilmiştir. Son olarak, bulgular, that tamamlayıcısından önce gelen bağlamsal zarf kullanımlarının oldukça sınırlı olduğunu, buna karşın boş değişiminin bağlamsal zarflarla daha sık görüldüğünü göstermiştir.

[1] Preliminary findings from this study were presented as an abstract at the INCSOS 8th International Congress on Social Sciences.

[2] Öğretim Görevlisi, Kırşehir Ahi Evran Üniversitesi, Yabancı Diller Yüksekokulu, songultas@ahievran.edu.tr, Orcid: 0000-0003-2590-6063

[3] Öğretim Görevlisi, Kırşehir Ahi Evran Üniversitesi, Fen Edebiyat Fakültesi, fatmasolmaz@ahievran.edu.tr, Orcid: 0000-0001-8706-5728

Alıntılamak için/Cite as: Taş S. ve Solmaz F. (2026). The That/Zero-That Alternation With 'I Think' Across Formal And Informal Registers, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, s. 1-16

INTRODUCTION

Comment Clauses (CCs), which are called as "*parenthetical verbs*", "*parentheticals*", "*disjuncts*" (Espinal, 1991), or "*discourse markers*" (Biber et al., 1999), and as a feature of spoken language (Heine & Kaltenböck, 2021), are defined by Kaltenböck (2007) as main clause additions to another construction, namely, the Host Construction (HC) which occurs initial, middle, or final position in a HC and appears as a main clause. In that, they are stated to be related to the HC by linear adjacency though not syntactically, so as not a constituent of the host construction. They are rather related to such categories as reporting clauses, matrix clauses and discourse markers (Kaltenböck, 2009).

Most of the CCs, when used in face-to-face conversation, are indicated to show a tendency to be listener or speaker-oriented. In that, *I*-oriented comment clauses such as '*I suspect'* and *'I'm frustrated to admit'* might show what is said in a conversation at a given moment reflects the speaker's suggestion or opinion (Povolná, 2003). Those expressions are required to be followed by a complementizer, mostly linked to a *that* complementizer, such as '*I believe'*, '*I understand*', and '*I realize'* (Van Bogaert, 2009).

Moreover, the most common complement-taking verbs such as *think, believe, realize, say,* etc. are stated to be utilized mostly with *zero-that* and they are often used parenthetically (Thompson & Mulac, 1991). They have also related the syntactic mobility of a CTP (complement taking predicate) to its incidence rate and to the incidence of complementizer *that* omission.

Shank et al. (2014) examines *I think* as a discourse marker, highlighting its function in signaling the speaker's stance or attitude, similar to a modal particle, as discussed by Aijmer (1997). It also emphasizes that the alternation between *I think that* and *I think Ø* is tied to its discourse role in expressing mental states, underscoring its pragmatic function in conversation. Shank et al. (2016) also explains that common subject-verb combinations such as *I think* have evolved into fixed "epistemic phrases" or "discourse formulas," frequently functioning as parenthetical expressions in spoken discourse. Their frequent use without the *complementizer that* is viewed as evidence of their grammaticalization and growing independence as discourse markers.

I think functions as a discourse marker to convey speaker attitudes, beliefs, and stance, and to enhance coherence in spoken interaction (Li & Pang, 2022). Drawing on prior research, the researchers emphasize the increasing scholarly interest in the discoursal role of *I think*. They point out a strong correlation between the sentence position of *I think* and its communicative role. Juszczyk and Konat (2024) also discuss *I think* as a discourse marker that primarily serves as a hedging function, allowing speakers to express uncertainty and soften the force of their statements. Zidan and Yagi (2025) underscores that *I think* functions primarily as a pragmatic marker, contributing to discourse structure and modality rather than conveying concrete propositional meaning. They also highlight that *I think* serves multiple interpersonal functions, such as signaling agreement, disagreement, topic continuity, or shifts in stance.

It is a crucial choice for the speaker between two constructional variants in grammatical analysis, which is even more challenging when it is hard to differentiate between the meanings of them, as is often asserted for *that* clauses and *zero-that* clauses (Biber et al., 1999). The selection between retaining or omitting *that* complementizer has long been regarded as an instance of free variation in linguistic studies (Dor, 2005; Kaltenböck, 2006)

The omission of the *that* complementizer in noun complement clause, holding two contradictory views on the possibility of the omission, has been an important issue in grammar studies since there have been some corpus-based researchers (Biber et al., 1999; Carter & McCarthy, 2006) and generative syntax-based researchers (Pesetsky, 1992; Bošković & Lasnik, 2003) who contradict the use of *zero-that* in such clauses; whereas, there have also been some others (Berlage, 2014; Downing & Locke, 2006; Eastwood, 2005; Greenbaum, 1996; Huddleston & Pullum, 2002) who supported the idea of *that* omission in a noun complement clause (Hidarto & Andrianto, 2015). After conducting a semantic analysis of *that*-omission, Dor (2005) suggested

that only the predicates which entail a truth claim. In that, such deletion of *that* complementizer in complement structures is influenced by the semantic feature of "truth-claim" (Kaatari, 2018).

Though *that* complementation, as a formal syntactic phenomenon, has been worked on since the foundation of generative grammar by many researchers (Bresnan, 1972; Kayne, 1981; Rosenbaum, 1967; Stowell, 1981) in recent years, there has been a growing number of publications focusing on the alternation between *that* and *zero-that* complementation (Cacoullos & Walker, 2009).

Recognizing that scarcity in the field after reviewing the existing relevant literature, this study was set to investigate on the use of *that/zero-that* alternation in complement clauses with the verb '*think*' which has been pointed as the verb most frequently followed by *that* clauses (Cacoullos & Walker, 2009; Jaeger, 2010; Kaatari, 2018; Thompson & Mulac, 1991) by using the existing American corpora, namely COCA (The Corpus of Contemporary American English) (Davies, 2008). The research included spoken and written corpus to differentiate between the formal and informal register of language factors in the preference of *that/zero-that* alternation as well as the mode of language use as spoken and written. To this end, the present study aims to answer the following research questions:

RQ1: How does the use of the complementizer *that* after "*I think*" vary across register and mode of language modes (spoken, blog, and academic) in terms of frequency, syntactic structure, and clause position?

RQ2: How does intervening adverbials influence the presence or omission of the *that*-complementizer following "*I think*"?

As highlighted in the introduction, although prior studies (e.g., Thompson & Mulac, 1991; Cacoullos & Walker, 2009) have explored the *that/zero-that* alternation with the verb think, there remains a lack of comprehensive corpus-based research examining this phenomenon across different modes and registers. Addressing this gap, the present study uses updated corpus data to investigate the syntactic optionality of the complementizer that, examining the factors that influence its inclusion or omission. By comparing spoken and written corpora, the study reveals how mode (spoken vs. written) and register (formal vs. informal) shape grammatical choices. Since complement clause variation poses challenges for EFL and ESL learners, the findings have pedagogical implications, helping inform grammar instruction and improve learners' pragmatic awareness. Moreover, the study contributes to usage-based grammar by offering empirical insights drawn from large-scale natural language data (COCA).

In this context, the following sections of this paper will provide the existing relevant literature in detail under three main subheadings as the comment clauses, *that* and *zero-that* complementation, and the employment of the *that* and *zero-that* complementizers with the verb '*think*'; the procedure and the results of the current study; the discussion and conclusion parts.

EMPIRICAL STUDIES ON THAT/ZERO-THAT ALTERNATION

In this section, the existing literature on the comment clauses, *that* and *zero-that* complementation, and the use of the *that* and *zero-that* complementizers with the verb '*think'* will be discussed in detail by providing evidence from the empirical studies conducted on the relevant issue.

Comment Clauses

Comment clauses, with three different types at least, are stated to be both style and content disjuncts which form a heterogeneous syntactic class by Quirk et al. (1985). Ranger (2011) identified three distinct types of comment clauses in his review of Laurel J. Brinton's work, which aimed to define clausal pragmatic markers:

- Comment clauses of the *I think* type,
- Adverbial comment clauses of the *as you say* type,
- Relative comment clauses of the *what is more* type.

Another quite similar list of three syntactic types of finite CCs was previously presented by Povolná (2005) in her article which she examined *I*-oriented comment clauses as response elicitors in spoken English, as follows:

- the main clause (e.g. *I suppose*)
- an adverbial finite clause introduced by "as" (e.g. *as I've said so*)
- a nominal relative clause (e.g. *what I mean is*)

Quirk et al. (1985) pointed the comment clauses like the main clause as listed as the number one on all the lists above as the most important one since comment clauses that resemble matrix clauses typically contain transitive verbs or adjectives that, in other contexts, require a *that*-clause as a nominal object. Some common examples of that type are *I think, I believe, I expect, I feel, I would say, I assume,* etc. (Heine & Kaltenböck, 2021). Whereas the subject complement clauses, as in the third type are stated to be rare and low in predictability (Walter & Jaeger, 2005).

In her inquiry on the analysis of three conversational texts which were taken from a Corpus of English Conversation, Povolná (2003) observed that comment clauses are frequently used in spontaneous face-to-face conversations, where they operate distinctly from those in written discourse. Hence, she suggested for further research to consider the functional role of CCs to be able to contribute to the understanding of their functions in spoken discourse. Referring to the large number of issues that must be resolved to achieve a fully operational Functional Discourse Grammar, Harder (2008) argued that a critical aspect of linguistic interface models is their ability to connect with an interpersonal level by representing discourse elements fully, enhancing the model's overall integration potential.

Kaltenböck (2009) examined 830 instances taken from the spoken corpus of the British component of the International Corpus of English (ICE-GB), analyzing the distribution and communicative functions of comment clauses such as *I think* and *it seems*. His findings underscored the correlation between position, scope, and prosody, revealing that comment clauses can influence entire host clauses or highlight specific phrasal elements. Pöldvere et al. (2016), with the goal of identifying key contextual factors and their dialogic effects, examined how first-person epistemic and evidential Complement-Taking Predicate (CTP) constructions, like *I think, I know,* and *I suppose*, function in terms of expansion and contraction in spoken discourse by conducting an exploratory analysis in the London–Lund Corpus (LLC) of spoken British English and indicated that CTP constructions do not merely expand the dialogic context in which they appear, but play a role in limiting alternative perspectives as well.

That and *Zero-that* Complementation

That complementation holds significance in linguistic literature, particularly in discussions of extraposition (Chalker & Weiner, 1998), a process in which the notional subject shifts to a position following the predicate and is replaced by it (Collins, 1994), a structure commonly found in academic writing (Biber et al., 1999; Herriman, 2000; Hyland & Tse, 2005). Hyland and Tse (2005) explored the frequencies, forms, and functions of evaluative *that* in two corpora comprising of 465 abstracts from published research articles, as well as master's and doctoral dissertations written by L2 students and demonstrated that the evaluative *that* usage is prevalent in the published scholar and graduate student writings.

In the literature, it is agreed that spoken English, especially conversations, is found to be richer in terms of CTPs (complement taking predicates), compared to written texts (Biber et al., 1999; Bogaert, 2011; Kaltenböck 2005; Tottie & Paradis 1982). Biber (2000) found that *that* clauses are used with a limited set of verbs, particularly 'think, say, and know', which appear more frequently in conversation than in written texts. Rosefeldt (2008) argued the matter from a different perspective stating in his paper that *that clauses* are not singular terms. A formal analysis of non-nominal quantification revealed that *that* clause positions function as sites of non-nominal quantification. Man and Chau (2009) conducted a longitudinal corpus study of evaluative *that* clauses in the argumentative writings of Chinese EFL learners, using a common task. The results of their analysis showed a complex development in the use of evaluative *that* clauses in the long term. Notably, while the rate of verb-based *that* clauses showed a decrease, the rate of noun-based and adjective-based *that*-clauses grew in the student writings.

An earlier study on the persistence of optional *that* was conducted by Ferreira (2003) who carried out three experimental studies examined whether any prior exposure to a lexically or syntactically similar instance of *that* affected the utilization of an optional *that* in a sentence-complement structure. The results demonstrated that the inclusion of just a single word, namely, the optional complementizer *that*, is syntactic rather than a lexical phenomenon.

The phenomenon of *that*-deletion dates to early English texts, where the complementizer was often deleted when the subject of the main clause matched that of the subordinate clause or when reporting direct speech (Martín & Barranco, 2014). The use of this phenomenon has also been looked at in the field of ELT. Taşçı and Öztürk (2021) aiming to search the distribution of controlling verbs in post-predicate *that*-clauses (VTHAT) in academic texts written by Turkish EFL learners, comparing findings to that of NS (native speaker) English-students and experienced writers. To this end, the researchers compiled a specialized research corpus consisting of three sub-corpora: master and doctoral theses written by Turkish learners of English and NS students, as well as published research articles authored by NS experienced writers. Findings on VTHAT constructions revealed that eight out of the ten most common verbs in Turkish students' dissertations have also been frequently used in the three sub-corpora. The analysis showed an overreliance on VTHAT structures by both NS and NNS students, while Turkish students use them more often compared to the NS students. Similarly, Ishikawa (2016) examined VTHAT usage comparing Asian learners of English (ALE) and native English speakers using data from the ICNALE corpus. The findings revealed that NS students used VTHAT more frequently than ALE, while ALE omitted the complementizer more frequently in both spoken and written corpora. Additionally, ALE exhibited a preference for reporting verbs such as *think, believe, agree*, and *know*.

Shank et al. (2016) conducted a corpus-based study to track the historical trajectory of *that/zero-that* alternation across nine cognitive verbs including *think* employing stepwise logistic regression analysis which revealed that, unlike the general assumption that the *zero* form has increased over time, the study found a constant decline in *zero* usage, although the rate of this decrease varied across different verbs. Regarding the origin of the source complementizer, a more contentious issue, Cacoullos and Walker (2009) examined the complementizer *that/zero-that alternation* in introducing subordinate clauses from both synchronic and diachronic perspectives. Their multivariate analysis indicated that the complementizer *that* serves to determine the lines between two clauses with lexical content, although the alternation is largely limited by lexical factors. In contrast, the *zero* complementizer tends to appear when the clauses operate as a single unit.

When *that* and *zero-that* clauses regarding their selection of matrix predicate were compared by Kaltenböck (2006) by analyzing various potential conditioning factors using corpus data from ICE-GB, the findings showed that considerably high rate of NPs and VPs with *zero-that* clauses (16.3 % and 14.6 %), and low proportions of PPs and AdjPs (0% and 5.5 % respectively). Furthermore, the study identified key factors influencing the choice of *zero-that*, including text informality, matrix predicate type, and the informational weight of the complement clause. The likelihood of using an optional 'that' is influenced by processing complexity (Ferreira & Dell, 2000; Hawkins, 2004; Walter & Jaeger, 2005).

Furthermore, examining the impact of phonological optimization on the inclusion of *that*, Walter and Jaeger (2005, p.7) found that speakers tend to omit *that* when its presence would result in a shared place or manner of articulation with adjacent segments, providing the following example: *I believe (that) that drug makes you sleepy*. Conversely, they may retain it when neighboring segments exhibit identical articulation features.

More recently, it was reported that *that* and *zero-that* can function interchangeably without altering meaning. Several factors influence this alternation, including: *(i)* formality, distinguishing between formal and informal styles; *(ii)* matrix predicate frequency; *(iii)* pronominal or nominal variation of lower-clause subject; and *(iv)* the occurrence of any intervening item between the predicate and subordinate clause (Martín & Barranco, 2014). Many studies have identified these conditioning factors for the interchangeability of *that* and *zero complementizers* as stated in Shank and Plevoets (2018) discusses the interchangeability of *that* and *zero complementizers* and the factors influencing this alternation noting that some have argued the choice is due to stylistic preference not because of those factors. For the first conditional factor *formality* Kaltenböck (2006) states that *that* tends to be regarded as the more formal option, while zero is associated with informal registers.

For the second factor, researchers state that the subject of the matrix clause is often cited as influencing whether the

complementizer is used or omitted. Numerous studies (Bolinger, 1972; Elsness, 1984; Thompson & Mulac, 1991; Tagliamonte & Smith, 2005) argue that pronouns, especially *I* and *you*, tend to favor *zero usage* (omission of that). However, Torres Cacoullos and Walker (2009) found that the distinction between pronouns and full noun phrases (NPs) has a greater impact on *that-omission* than the difference between *I/you* and other subject types. They suggest that the previously reported strong effect of *I* and *you* likely stem from the high frequency of discourse formulas like *I think* and *I guess*, which they analyze separately. Hyland (2001, 2002) examined the functions of first-person pronouns (*I* and *we*) in students' academic writing in English as a second language and found that L2 writers often avoid asserting an individual authorial presence, a tendency that may reduce their credibility and limit engagement in academic discourse. He emphasized the importance of self-mentioning through first-person pronouns, arguing that such usage plays a crucial role in mediating the relationship between writers' arguments and their disciplinary communities. In another paper, Hyland (2005) discusses pronouns in the broader framework of engagement strategies in academic writing, introducing a model for analyzing stance that highlights how pronoun use contributes to the ways writers present themselves, position their arguments, and engage with readers. Later, Hyland and Jiang (2017) extended this line of inquiry by analyzing pronouns as markers of informality in academic writing, focusing on first-person pronouns (*I*, *we*), second-person pronouns (*you*, *your*), and unattended anaphoric pronouns (*this*, *these*, *that*, *those*, *it*). Their findings showed a notable increase in the use of first-person pronouns, underscoring their significance as indicators of informality in contemporary academic discourse.

About the third factor, several studies (Warner, 1982; Elsness, 1984; Finegan & Biber, 1985; Rissanen, 1991; Thompson & Mulac, 1991; Rohdenburg, 1996; Tagliamonte & Smith, 2005; Torres Cacoullos & Walker, 2009) have argued that complement clauses with pronominal subjects (e.g., he, she, they) are more likely to omit the *complementizer that* compared to clauses with full noun phrase (NP) subjects. In other words, *zero usage* is favored when the subject is a pronoun rather than a full NP. Research has consistently shown that the presence of intervening item between the matrix and complement clause increases the likelihood of using the *complementizer that* (Bolinger, 1972; Rohdenburg, 1996). This is because such material adds cognitive processing load, making an explicit marker of subordination (like that) more helpful for comprehension. In contrast, when the matrix and complement clauses are adjacent, the structure is cognitively and syntactically simpler, which favors *zero usage* (i.e., omission of that). Rohdenburg (1996) emphasizes that delayed processing due to intervening elements supports the use of *that*, while Torres Cacoullos and Walker (2009) found that intervening material is one of the strongest predictors of *complementizer presence*, especially with common complement-taking verbs.

Moreover, several other studies indicated that a set of subject-verb collocations with high frequency are closely linked to the use of *zero complementizers* (Shank & Plevoets, 2018). Cacoullos and Walker (2009), in that vein, suggested that the factors influencing complementizer selection may differ for frequently occurring "discourse formulas" like *I think* compared to more flexible complement-taking constructions. When it comes to the factors promoting the use of zero complementizer, various studies have explored language-internal and structural elements that contribute to the alternation between *that* and *zero-that*, with some employing advanced statistical methods to determine the significance of these factors (Shank et al., 2016). Dor (2005) suggested that "truth claim" concept plays an essential role. While *that* clauses indicate "propositions," *zero-that* clauses convey "asserted propositions.".

Numerous prior studies have analyzed *that/zero* variability in relation to register variation, highlighting that as the more formal choice, whereas the omission of *that* is typically associated with informal contexts (Kaltenböck, 2006). For instance, Hidarto and Andrianto (2015), suggesting that the complementizer *that* should be considered optional rather than strictly required, carried out a corpus-based study and found that *that* omission mostly appears with the preceding head nouns from the category signs of evidence, such as sign, clue and proof. It has also been revealed that most of the *that* omissions occur in spoken register which is generally informal. When those two factors were compared, it has been found that register factor had more significant role than the noun category in

influencing the use of *that* omission.

That-deletion (the process that transforms *I think that* into *I think* Ø) has commonly been regarded as a historical phenomenon, suggesting that the evolution of *that*-deletion follows a non-linear trajectory with no straightforward line from *that* to the *zero* form (Ranger, 2011).

One of the most widely discussed pathways has been provided by Thompson and Mulac (1991) suggesting that epistemic parentheticals (e.g. *I think, I guess,* or *I suppose)* are formed as matrix clauses which are followed by dependent *that*-complement clauses. Therefore, omitting *that* complementizer results in an ambiguous structure, as the original matrix clause is reinterpreted as a parenthetical and subsequently shifted to middle or final position within the clause (Heine & Kaltenböck, 2021). The various stages of the relevant hypothesis, which was called as "*the matrix clause hypothesis*" by Brinton (2006, p.35), are listed in the study of Heine and Kaltenböck (2021, p.3) as follows:

Stage 1: I think that it is possible.

Stage 2: I think Ø it is possible.

Stage 3: It is possible, I think.

In the first stage, *I think* is a straightforward matrix clause with a complement clause. In the second stage, it is ambiguous; it is either a matrix clause or a parenthetical disjunct. On the other hand, in the last stage, it is clearly parenthetical disjunct, not as a matrix clause Brinton (2006, p.35).

The Use of That and Zero-That Complementizers with the Verb 'Think'

I think, I believe, I suspect and similar complement-taking predicates as clause-initials have been an important reason of questioning since their syntactic status carries some ambiguities, as such*;* whether *"they are main clauses governing a complement clause or comment clauses, like parentheticals, which have a syntactically supplementary relationship with the following clause"* (Kaltenböck, 2009, p.49). These questions are even more difficult to answer when *that*-complementizer has been omitted.

In his study on the relevant issue, Kaltenböck (2009) analyzed 200 instances of sentence initial *I think* in a spoken English corpus to determine their syntactic role as either main or comment clauses. The corpus-based study revealed that structural variation (*that* vs. *zero*) does not correspond to differences in prosodic patterns. Furthermore, it was suggested that the *that*-complementizer is not widely used as an indicator of subordination; instead, it has been used as a filler to give weight to *I think* or for prosodic purposes in spoken language. Dehé and Wichmann (2010) investigated this issue from a prosodic perspective, suggesting that sentence initial pronoun-verb constructions such as *I think* and *I believe* may operate as main clauses, adverbials, or discourse markers. Their corpus analysis of spoken British English indicated that prosody plays a role in distinguishing these functions: main clauses are marked by accentuation on the pronoun, comment clauses by accentuation on the verb, and discourse markers by a lack of stress.

Bogaert (2011), on the other hand, provided a comprehensive analysis on the prior hypothesis that CTMPs (complement taking mental predicates) like *I think,* etc. The researcher assessed whether the framework could be applied to *I think* and eight other CTMPs (namely, *I suppose, I believe, I reckon, I expect, I understand, I guess, I imagine, and I realize*) critically concerning with their use in British English. The detailed examination of corpus data from ICE-GB and BNC showed that all CTMPs predominantly appear in the initial position, where they show a strong preference for the zero complementizer over *that*.

METHODOLOGY

This corpus-driven study adopts a bottom-up approach to investigate the *that/zero-that* alternation following *I think* across spoken, blog, and academic modes in the COCA corpus. Rather than applying a pre-existing theoretical framework, the bottom-up approach employed for the analysis emerges from patterns observed in the data since the present study aims to uncover how the *that/zero-that* alternation varies across different registers and modes of language use (spoken vs. written). In this study, the verb think with first-person subject was chosen on purpose since Quirk et al. (1985) and Povolná (2005) pointed that

the comment clauses taking the main clause form are the most frequent type of the comment clauses, and there is an agreement that the verb *think* is a common CTP (complement taking predicate), and it occurs frequently especially in spoken texts (Biber et al., 1999; Biber, 2000; Bogaert, 2011; Heine & Kaltenböck, 2021; Kaltenböck, 2005; Tottie & Paradis, 1982). The present study also focused on three different genres with different levels of formality and modes as Kaltenböck (2006) stated that formality of style was a significant factor in using *that*-complementizer and it was an indicator of formal register.

The analysis was based on tokens from blog, spoken and academic subcorpora retrieved from COCA (the Corpus of Contemporary American English) in line with the purpose of the study. The features of corpus in general have been summarized in Table 1 below.

Table 1: The Features of Subcorpora in the COCA

Genres	Texts (n)	Words (n)	*I think (n)*	Explanation
spoken	44803	127396932	277277	The transcription of more than 150 various TV and radio programs
academic	26137	120988361	8326	More than 200 different peer-reviewed journal articles from hard and soft disciplines, are included
web (blog)	98748	125496216	95044	Blog texts from the U.S. section of GloWbE were further classified into various web genres

taken from English-Corpora.org

Table 1 summarizes that *I think* occurs mostly in the spoken corpus (n=277,277), followed by blog (95,044), and academic corpora (277,277). Given the large number of instances found across the corpus (a total of 380,647), manual annotation of all tokens impractical. Therefore, in line with Brezina's (2018) recommendation, a random sample of 500 tokens was selected from each genre for analysis.

Data analysis was manually coded and analyzed using Excel. This study adopts a linguistic feature design, which enables us to examine linguistic features individually with the contextual factors affecting its usage (Brezina, 2018). Considering the enormous size of COCA, and feasibility of the research, a total of 1500 tokens, the first relevant 500 occurrences from each corpus (spoken, blog, and academic) were extracted and categorized into six groups: blog-zero, blog-*that*, spoken-zero, spoken-*that*, academic-zero, and academic-*that*. Only the tokens where *I think* was immediately followed by either a zero complementizer or the overt *that* were retained; tokens not containing the target structure (*I think* followed by a *that*-clause or *zero-that* clause) were discarded. In addition, each genre was examined for the distribution of zero-*that* constructions based on the syntactic structure of the clause that follows *I think*. These structures were categorized as:

I think (that) + NOUN

I think (that) + DETERMINER

I think (that) + ADJECTIVE

This classification aimed to explore whether syntactic environment influences the presence or omission of the complementizer that across different registers.

The corpus annotation was carried out by two independent raters, expertise in EFL and have prior experience in corpus data annotation. Both annotators received detailed guidelines prior to annotation and worked independently. In cases of disagreement, for instance, there were examples creating ambiguity caused by adverbials appearing before the subject of the complement clause in *zero-those* constructions, making it difficult to determine where the *that*-complementizer would have been placed in the original sentence, a consensus was reached through discussion. The comparison of the two analyses showed that the percentage agreement was met at the rate of 91%. Thus, the interrater reliability of the study was high.

RESULTS

The results are discussed in accordance with the research questions. The first research question asks how the use of the complementizer *that* after "*I think*" varies across registers and language modes (spoken, blog, and academic) in terms of frequency of occurrence, syntactic structure, and clause position and the second research question aims to put forward how does intervening adverbials influence the presence or omission of the *that*-complementizer following "*I think*".

To answer the first research question, the first five hundred tokens of *I think* in each genre were examined to determine the use of *that* complementizer with *I think*; the tokens not containing *I think* with *zero that* or *I think that* are discarded (see APPENDIX I). Numbers of the total instances containing *I think zero-that,* and *I think that* are presented for each genre below in Table 2.

Table 2: The Distribution of Instances with I Think across three Subcorpora of COCA

	Zero that		I think that	
Genre	**N**	**%**	**n**	**%**
blog	414	92.20	38	8.46
spoken	394	82.77	82	17.22
academic	395	87.97	54	12.03
total	*1203*	*87.36*	*174*	*12.63*

As shown in Table 2, the analysis of the tokens reveals a total of 174 instances of *I think* with *that*-complementizer (12.63%) in total; *I think* with *zero-that* complementizer (87.36%) is used more often across three corpora. It was also found that the frequency of *I think* with *zero that* is much more common than the frequency of *I think that* regardless of the genre. However, the use of *I think that* is higher in spoken corpus (17.22%) than academic genre (12.03%) and blogs (8.46%). Following extracts are taken from three subcorpora:

- **I think** it was pretty accurate in how we depicted it. (COCA: BLOG)
- This is sad, but **I think that** it is just an effect of growing older. (COCA: BLOG)
- **I think** the Modern Interface is a great tablet interface (COCA: ACAD)
- They were there in Miranda, but **I think that** the textual connection became more important for him. (COCA: ACAD)
- No, **I think** this is not going to end well. I don't think so. (COCA: SPOK)
- And **I think that** people feel very anxious and they're not dealing with the sense of loss. (COCA: SPOK)

Secondly, *e*ach genre is examined according to *zero that* distribution regarding the following syntactic structure*: I think (that) + NOUN+, I think (that) + DET* and *I think (that) + ADJ.* For each structure, the number of instances and the percentages of them in the subcorpus in question are provided in Table 3.

Table 3: The Distribution of the I Think (+That/-That) in Different Syntactic Structures across three Genres

		BLOG		SPOKEN		ACAD		Total	
I think	**Structure**	**n**	**%**	**n**	**%**	**n**	**%**	***n***	***%***
zero	**NOUN+**	8461	42	15813	88	395	30	*24669*	*34.39*
	DET	10709	53	33287	66	817	63	*44813*	*62.47*
	ADJ	881	4	1272	3	92	7	*2245*	*3.12*
TOTAL		*20051*	*92*	*50372*	*87*	*1304*	*84*	*71727*	*100*
that	**NOUN+**	735	44	2639	36	89	35	*3463*	*37.53*
	DET	748	45	4330	59	136	53	*5214*	*56.52*
	ADJ	193	12	325	5	30	12	*548*	*5.94*
TOTAL		*1676*	*8*	*7294*	*13*	*255*	*16*	*9225*	*100*

The syntactic structures of the examined instances show that in all the genres, *I think (that)* mostly co-occurs with a determiner except the academic subcorpus, in which *I think* with zero-*that* complementizer mostly co-occurs a common or proper noun (88%) *(Table 3). I think (that) + ADJ* structure is the least frequent pattern across all three subcorpora (2.77%), occurring less frequently in *I think zero that* condition (3.12%). The following excerpts are taken from each corpus for each structure.

- **I think** the conditions are no longer too hopeless because development in Internet technology has made it (COCA: BLOG) (zero+det)
- I wanted to talk about two options that **I think** far too few college ministries consider. (COCA: BLOG) (zero+adj)
- Indeed, **I think that** critics are correct in saying that Conrad fails to provide African individuals in Heart. (COCA: ACAD) (that+noun)
- **I think** technological change begins to move that dynamic in a different way. (COCA: ACAD) (zero+adj)
- **I think that** the idea that you take a fictional character… (COCA: SPOK) (that+det)
- …well, Julie, **I think that** Democrats absolutely believe in border security. (COCA: SPOK) (that+noun+)

In line with the first research question, the genres are also examined in terms of the position of the *I think* in the first 500 tokens of each genre *(Table 4).*

Table 4: The Position of I Think in three Genres

	Zero that							
	Blog		**Academic**		**Spoken**		***total***	
	n	**%**	**n**	**%**	**n**	**%**	***n***	***%***
Initial Position	400	96.61	349	88.35	377	91.50	*1126*	*92.21*
Medial Position	8	1.93	44	11.14	29	7.03	*81*	*6.63*
Final Position	6	1.44	2	0.51	6	1.46	*14*	*1.14*
total	*414*	*100*	*395*	*100*	*412*	*100*	*1221*	*100*

Table 4 shows that *I think* is most frequent in initial position (92.21%) across three subcorpora. On the other hand, middle position *I think* is more frequent in academic corpus (11.14%), and there are the least number of instances of *I think* in middle position in blogs (1.93). Furthermore, end-position *I think* is even rare in all corpora; blogs included eight examples of final position *I think* (1.44%), while spoken corpus includes six occurrences of end position *I think* (1.46%)*, and* academic corpus includes only two instances (0.51%). In the blog writings, there are similar numbers of middle position (n= 8) and end position *I think* (n= 6). Some examples taken from the corpora are presented as follows:

- It's a simple point of view, but a very telling one, **I think**. (COCA: SPOK)
- The investigation is, **I think**, close to being completed. (COCA: SPOK)
- This is implicitly aimed against the notion of culture, **I think**. (COCA: BLOG)
- Most of us, **I think**, feel uncomfortable with hearing our own voices outside of our head… (COCA: BLOG)
- "They're Russians, **I think**."(COCA: ACAD)
- We said nothing more and both, **I think**, enjoyed our long silence. (COCA: ACAD)

Finally, to answer the second research question, further analysis regarding the position of *I think* was carried out. The findings demonstrated that even though there are examples of intervening adverbial before *that* complementizer, it is scarce (only two instances) in the first five-hundred tokens of three genres. While there is only one example of intervening with *that* complementizer in academic corpus, a similar sentence structure is formed with *zero-that* complementizer. The instances are provided below:

- **I think** then, **that** Boy Scouts had better start testing their potential members at the age of 6… (COCA: BLOG)
- **I think**, therefore, **that** it is quite appropriate to "objectify" treatment decisions for… (COCA: ACAD)
- Even if they exist, there is only one, **I think** therefore they are not used, because there is no adequate equipment. (COCA: ACAD)

On the other hand, there are plenty of examples creating ambiguity via adverbials occurring before the subject of the complement clause with *zero-that* since it is hard to identify the position of *that* complementizer in the original sentence. In that, it is unclear what exactly the adverbial modifier is: the matrix clause or the complement clause. Some of the problematic clauses are presented below:

- And **I think**, frankly, when you look at things like the decision not to release the photos… (COCA: SPOK)
- **I think** for some of the time I look like a homeless girl or a prostitute... (COCA: BLOG)
- **I think**, however, it is important to say that if you are going to "talk the talk" of empowerment, creativity, and risk taking, you have to "walk that walk" as well. (COCA: ACAD)

DISCUSSION

The findings of the present study demonstrated that tokens that including *I think* occurred more frequently in the spoken while in the written (in blogs and academic writing), it is used relatively less often, which is parallel to the existing literature. There is an agreement that CTPs (complement taking predicates) are more common in spoken English (Biber et al., 1999; Bogaert, 2011; Kaltenböck, 2005; Tottie & Paradis, 1982). Moreover, the *zero-that* complementizer use was more often. Shank and Plevoets (2018) and Cacoullos and Walker (2009) argued that the use of zero-that complementizer is more frequent in formulaic sentences such as *I think, I guess etc.*

The present study indicated that the use of *I think* with zero-complementizer is more frequent than the use of *I think* with *that*-complementizer regardless of the genre. This finding matches with the statement of Thompson and Mulac (1991) that the most commonly used complement-taking verbs such as *think, believe, realize, say,* etc. mostly co-occur with *zero-that*. Even though Dor (2005) and Kaltenböck (2006) asserted that the use of *that*-complementizer was a typical example of free variation, the findings of the present study demonstrated that there is a high tendency in using zero-complementizer with *I think* without being affected by the genre, register or position.

Furthermore, Dor (2005) categorized the verbs into three types according to the possibility of *that* deletion by the semantic property, namely, truth-claim. He stated that "*think*" is a predicate of belief and knowledge, that is, it allows *that* deletion. In that line, the present study also demonstrated that *that-deletion* with the verb "*think*" is optional.

Elsness (1984), Kaltenböck (2006) and Hidarto and Andrianto (2015) stated that register format is a significant factor in using *that* complementizer and it was an indicator of formal register. The findings indicated that regardless of the formality of the genres, *that* complementizer is low in frequency. Indeed, in the corpus-based analysis of the first five-hundred tokens, the use of *that* complementizer with *I think* was more frequent in spoken corpus, which is regarded more informal in register. When the distribution of the complementizer was examined by syntactic structure, its use was found to be more frequent in the academic genre, which is considered a more formal register. Similarly, the study of Taşçı and Öztürk (2021) showed that *that* complementizer use is frequent in academic writings of both Turkish EFL learners and native speaker English students.

The findings of the present study indicated that along with the high frequency of the *zero-that* complementizer use, *I think* with *zero-that* is used more often in initial position in each genre when the first five-hundred tokens are examined. While the use of final position *I think* is rare regardless of genre, *I think* in medial position occurs mostly in academic corpus. These findings are in line with the study of Bogaert (2011), which concluded *that* CTMPs (complement taking mental predicates) like *I think* mostly appear in initial position with *zero-that* complementizer. Furthermore, Thompson and Mulac (1991), and Heine and Kaltenböck (2021) argued that epistemic parentheticals like *I think* are followed by a dependent *that* complementizer when they occur as matrix clauses. On the other hand, when they are used with *zero-that* complementizer, they are reanalyzed and moved to a medial or final position in the clause as parentheticals. Their assertion requires more instances of the medial or final position of *I think*. In that sense, it might be claimed that the findings of the present study are supported by existing literature to some extent.

Elsness (1984) stated that in the clauses which include an intervening adverbial between the matrix verb and the subject of the complement clause, *that* complementizer is more frequent to avoid potential ambiguity. The present study found that instances of intervening adverbials preceding the *that-complementizer* are rare, with only two cases identified across the first 500 tokens in each of the three genres. In the academic corpus, just one such instance was found, however, similar syntactic structures are more common with the *zero-that* variant.

On the other hand, numerous instances involving adverbials before the subject of the complement clause in *zero-that* constructions create ambiguity, as it becomes difficult to determine where the *that-complementizer* would appear in the original sentence. In such cases, it is often unclear whether the adverbial modifies the matrix clause or the complement clause.

CONCLUSION

This study aimed to investigate *that/zero-that* alternation with *I think* in spoken, blog and academic corpus in COCA, specifically, to dwell on and differentiate between the formal-informal uses of language factor in the alternation of *that/zero-that* as well as the mode of language use as spoken and written.

Conducting a corpus-based analysis, which was based on tokens from blog, spoken and academic genres retrieved from COCA (the Corpus of Contemporary American English), the current study provided the following overall findings.

First, the tokens including *I think* occurred more frequently in the spoken mode, while it is used relatively less often in the written mode (in blogs and academic writing). The current study also revealed that *I think* with zero-complementizer appears more frequently than *I think* with *that* complementizer notwithstanding the genre. It was also indicated in the present study that *I think* with *zero-that* complementizer occurs more often in the initial position in each genre, whereas the medial and final positions are rarely preferred. Further results showed that *that* complementizer is less frequent irrespective of the formality of the genres. Lastly, the findings demonstrated that the instances of intervening adverbials before *that* complementizer are quite scant, while zero alternation with intervening adverbials are more frequent, though they create ambiguity.

This study highlights how the choice to include or omit *that* can affect clarity, emphasis, and formality. For instance, including *that* might make sentences clearer or more formal, which is important in academic writing, whereas omitting *that* is more suitable for a casual conversation. Through the findings of the current study, teachers are suggested to help students understand that the use of *that* in complement clauses is not fixed but varies depending on the mode (spoken vs. written) and formality. For example, omitting *that* is more common in informal spoken English, while including it is more frequent in formal writing (Hidarto & Andrianto, 2015; Kaltenböck, 2006). Furthermore, instructors should incorporate examples from authentic spoken and written sources (like COCA) to illustrate when *that* is typically included or omitted instead of teaching grammar rules (Cobb & Boulton, 2015). This approach helps students grasp the pragmatic and stylistic reasons behind the variation. The introduction of corpus-derived sentences to expose students to natural usage patterns is highly advised for EFL classes (Chang, 2014; Muftah, 2023). This real-world evidence strengthens students' intuition about when *that* is optional, improving both comprehension and production skills (Fang, et al., 2021; Muftah, 2023). Teachers can also design activities where students practice using complement clauses both with and without *that* in different registers. Role-plays, debates, or writing tasks can simulate formal and informal contexts, helping learners adapt their language appropriately. Besides, teachers should explicitly discuss typical learner errors related to *that* omission/inclusion, and offer corrective feedback rooted in actual language use rather than prescriptive norms since complement clause variation can confuse EFL/ESL learners.

Despite the findings of the current study shed some light on the crucial issues in literature, there are some limitations that need to be considered. Firstly, in-depth analysis of the tokens taken from three genres were restricted with the number of five hundred for each in COCA. Thus, further studies might include more genres with a wider number of

examples from different corpora.

As the second limitation, it could be pointed that this study was conducted by considering only one verb with first person singular in the form of present simple tense. Hence, more studies focusing on other forms of verbs or other cognitive verbs on similar issues are required. Moreover, further study comparing how learners of English and native speakers of English use parentheticals such as *I think* under the same task conditions is suggested to provide insights on the effects of cross-linguistics differences on parenthetical use.

ACKNOWLEDGEMENTS

Preliminary findings from this study were presented as an abstract at the INCSOS 8th International Congress on Social Sciences.

DECLARATION OF INTERESTS

We have nothing to declare.

REFERENCES

Aijmer, K. (1997). I think – An English modal particle. In T. Swan & O. Jansen Westvik (Eds.), Modality in Germanic languages: Historical and comparative perspectives (pp. 1–47). Mouton de Gruyter.

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

Bogaert, J. V. (2011). I think and other complement-taking mental predicates: A case of and for constructional grammaticalization. *Linguistics, 49*(2), 295–332. https://doi.org/10.1515/LING.2011.009

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

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide.* Cambridge University Press.

Brinton, L. J. (1996). Pragmatic markers in English: Grammaticalization and discourse functions. Mouton de Gruyter.

Bryant, M. M. (Ed.) (1962*). Current American usage*. Funk & Wagnalls.

Cacoullos, R. T., & Walker, J. A. (2009). On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics, 47*(1), 1–43. https://doi.org/10.1515/LING.2009.001

Chang, J. Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL, 26*(2), 243–259. doi:10.1017/S0958344014000056

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), The Cambridge handbook of English corpus linguistics (pp. 478–497). Cambridge University Press.

Dehé, N., & Wichmann, A. (2010). Sentence-initial *I think (that)* and *I believe (that)*. *Studies in Language, 34*(1), 36–74. doi:10.1075/sl.34.1.02deh

Elsness, J. (1984). *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies, 65*(6), 519–533. https://doi.org/10.1080/00138388408598357

Fang, L., Ma, Q., & Yan, J. (2021). The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students. *Journal of China Computer-Assisted Language Learning*, *1*(1), 80-109.

Ferreira, V. (2003). The persistence of optional complementizer production: Why saying *that* is not saying *that* at all. *Journal of Memory and Language, 43*, 379–398.

Finegan, E., & Biber, D. (1985). That and zero complementizers in Late Modern English: Exploring ARCHER from 1650–1990. In B. Aarts & C. F. Meyer (Eds.), *The verb in contemporary English* (pp. 241–257). Cambridge University Press.

Harder, P. (2008). Comment clauses, functional discourse grammar and the grammar–discourse interface. In J. Mackenzie & M. Gómez-González (Eds.), A new architecture for functional grammar (pp. 197–226). John Benjamins. https://doi.org/10.1515/9783110197112.197

Heine, B., & Kaltenböck, G. (2021). From clause to discourse marker: On the development of comment clauses. *Language Sciences, 87*, 1–16.

Hidarto, A., & Andrianto, S. (2015). The omissibility of that in noun complement clauses: Corpus evidence. In Proceedings of the 3rd Atma Jaya Conference on Corpus Studies (pp. 1–21). (pp. 1-21).

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*(3), 207–226.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*(8), 1091–1112. https://doi.org/10.1016/S0378-2166(02)00035-8

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes, 24*, 123–139. https://doi.org/10.1016/j.esp.2004.02.002

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes, 45*, 40–51. https://doi.org/10.1016/j.esp.2016.09.001

Ishikawa, S. (2016). Use of that-clauses after reporting verbs in Asian learners' speech and writing: Frequency, verb type, and that-omission. *EPiC Series in Language and Linguistics, 1*, 202–215.

Kaatari, H. (2018). On the syntactic status of *I'm sure*. *Corpora, 13*(1), 1-25. https://doi.org/10.3366/cor.2018.0134

Kaltenböck, G. (2006). "… That is the question": Complementizer omission in extraposed that-clauses. *English Language and Linguistics, 10*, 371-396. https://doi.org/10.1017/S1360674306001961

Kaltenböck, G. (2009a). English comment clauses: Position, prosody, and scope. *Arbeiten aus Anglistik und Amerikanistik, 34*(1), 51–78.

Kaltenböck, G. (2009b). Initial "I think"*:* Main or comment clause? *Discourse and Interaction, 2*(1), 49–70.

Li, L. P., & Pang, Y. (2022). A corpus-based pragmatic analysis of discourse marker "I think". *Open Access Library Journal, 9*, e9301. https://doi.org/10.4236/oalib.1109301

Man, D., & Chau, M. H. (2009). Learning to evaluate through *that*-clauses: Evidence from a longitudinal learner corpus. *Journal of English for Academic Purposes, 37*, 22–33. https://doi.org/10.1016/j.jeap.2018.11.007

Martín, J. C., & Barranco, J. R. (2014). On the use of *that/zero* as object clause links in early English medical writing. *Studia Neophilologica, 86*(1), 1–16. https://doi.org/10.1080/00393274.2014.900595

Muftah, M. (2023). Data-driven learning (DDL) activities: do they truly promote EFL students' writing skills development?. *Education and Information Technologies*, *28*(10), 13179–13205.

Povolná, R. (2003). Comment clauses in English face-to-face conversation. *Theory and Practice in English Studies, 1,* 71-82.

Povolná, R. (2005). On some I-oriented comment clauses as response elicitors in spoken English. *Discourse and Interaction, 1*, 86–97.

Pöldvere, N., Fuoli, M., & Paradis, C. (2016). A study of dialogic expansion and contraction in spoken discourse using corpus and experimental techniques. *Corpora, 11*(2), 191–225. https://doi.org/10.3366/cor.2016.0092

Ranger, G. (2011). Review of The Comment Clause in English: Syntactic Origins and Pragmatic Development, by Laurel J. Brinton. *Lexis*, Book reviews. https://doi.org/10.4000/lexis.1951

Rissanen, M. (1991). On the history of that/zero as object clause links in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 272–289). Longman.

Ritčik, J. (2014). A corpus-based analysis of that-deletion in complement clauses after the verbs of saying, thinking, and discovering in English and Lithuanian. [Master's thesis, Lietuvos Edukologijos Universitetas]. Lituanistika Database

Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics, 7*(2), 149–182. https://doi.org/10.1515/cogl.1996.7.2.149

Rosefeldt, T. (2008). *That*-clauses and non-nominal quantification. *Philosophical Studies, 137*, 301–333. https://doi.org/10.1007/s11098-006-0002-8

Shank, C., Plevoets, K., & Cuyckens, H. (2014). A diachronic corpus-based multivariate analysis of "I think that" vs. "I think zero". In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics. Quantitative studies in polysemy and synonymy. John Benjamins Publishing Company.* https://doi.org/10.1075/hcp.43.11sha

Shank, C., Bogaert, J. V., & Plevoets, K. (2016). The diachronic development of zero complementation: A multifactorial analysis of the *that/zero* alternation with *think*, *suppose*, and *believe*. *Corpus Linguistics and Linguistic Theory, 12*(1), 31–72. https://doi.org/10.1515/cllt-2015-0074

Shank, C., & Plevoets, K. (2018). Investigating the impact of structural factors upon *that/zero* complementizer alternation patterns in verbs of cognition: A diachronic corpus-based multifactorial analysis. *Research in Corpus Linguistics, 6*, 83–112. https://doi.org/10.32714/ricl.06.07

Tagliamonte, S., & Smith, J. (2005). No momentary fancy! The zero 'complementizer' in English dialects. *English Language and Linguistics, 9*(2), 289–309. https://doi.org/10.1017/S1360674305001711

Taşçı, S., & Öztürk, Y. (2021). Post-predicate *that*-clauses controlled by verbs in native and non-native English. *Australian Journal of Applied Linguistics, 4*(1), 18–33. https://doi.org/10.29140/ajal.v4n1.486

Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics, 15*(3), 237–251. https://doi.org/10.1016/0378-2166(91)90107-Z

Walter, M., & Jaeger, T. (2005). Constraints on optional *that*: A strong word form OCP effect. In *Proceedings from the annual meeting of the Chicago Linguistic Society (Vol. 41, No. 1, pp. 505-519).*

Warner, A. R. (1982). *Complementation in Middle English and the methodology of historical syntax: A study of the Wyclifite Sermons.* Croom Helm.

Zidan, B. A., & Yagi, S. M. (2025). A parallel corpus analysis of the pragmatic marker "I think". *Theory and Practice in Language Studies, 15(*3), 706–715. https://doi.org/10.17507/tpls.1503.05

APPENDIX

https://docs.google.com/spreadsheets/d/1TXv_d8eH4aSrUK99-v2FVop2Gr4Ch6Vuedit?usp=drivesdk&ouid=105209749840716482695&rtpof=true&sd=true

Author Contributions

All the authors have equal contributions.