

Discourse Markers (DMs) in Turkish as a Foreign Language Coursebooks: A Corpus-Driven Approach

Doğan BAYDAL¹

Öz

This study examines the representation of discourse markers (DMs) in Turkish as a Foreign Language (TFL) coursebooks, comparing their distribution with spoken Turkish as documented in the Turkish National Corpus (TNC) (Spoken). The analysis focuses on two widely used coursebook series, Yeni Hitit (levels 1-3) and Yediiklim Türkçe (A1-C2), examining dialogue-based and interactional sections. The most frequent DMs identified in the TNC were manually analyzed and quantified in the coursebooks. Normalized frequency values per million words were calculated, and the data were compared through statistical and visual analyses, including ratio comparisons and corpus-based benchmarks. The findings reveal that discourse markers are consistently underrepresented and unevenly distributed across both series, particularly those central to interactional fluency and pragmatic use. Core markers such as *yani*, *şey*, and *işte* occur at substantially lower frequencies than in spoken Turkish, indicating a gap between pedagogical input and authentic language use. These results support previous research suggesting that language teaching materials tend to prioritize structural aspects of language over discourse-level competence.

Keywords: *Turkish as a foreign language, discourse markers, foreign language coursebooks.*

¹ Öğr. Gör. Dr., Ondokuz Mayıs Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, Mütercim ve Tercümanlık Bölümü, doganbaydal@omu.edu.tr, ORCID: 0000-0001-9392-4779

Kaynak gösterme: Baydal, D. (2026). Discourse markers (dms) in Turkish as a foreign language coursebooks: a corpus-driven approach. *Aydın TÖMER Dil Dergisi*, 11(2), 355- 383.

Geliş tarihi: 20.02.2025 – Kabul tarihi: 22.07.2026

Doi: 10.17932/IAU.TOMER.2016.019/tomer_v01i1i2002

Yabancı Dil Olarak Türkçe Ders Kitaplarında Söylem Belirleyicileri: Derlem Tabanlı Bir Yaklaşım

Öz

Bu çalışma, Türkçenin yabancı dil olarak öğretiminde kullanılan ders kitaplarında söylem belirleyicilerinin temsiliyetini incelemekte ve bu kullanımın Türk Ulusal Derlemi'nin (TUD) Sözlü bölümüyle karşılaştırmasını yapmaktadır. Çalışma, yaygın olarak kullanılan iki ders kitabı serisine odaklanmaktadır: *Yeni Hitit* (1-3 seviyeleri) ve *Yediiklim Türkçe* (A1-C2). Çözümleme kapsamında, bu serilerin diyalog ve etkileşime dayalı bölümleri incelenmiştir. TUD Sözlü konuşma verisinde en sık kullanılan söylem belirleyicileri belirlenmiş ve bu belirleyiciler ders kitaplarında manuel olarak tespit edilip sayısallaştırılmıştır. Milyon kelime başına normalize edilmiş sıklık değerleri hesaplanmış ve veriler oran karşılaştırmaları ve derlem temelli ölçütler kullanılarak istatistiksel ve görsel çözümlemelerle karşılaştırılmıştır. Bulgular, söylem belirleyicilerinin her iki seride de sistematik olarak az temsil edildiğini ve dengesiz dağıldığını göstermektedir. Özellikle etkileşimsel akıcılık ve pragmatik kullanım açısından önemli olan *yani*, *şey* ve *işte* gibi temel belirleyicilerin, konuşma diline kıyasla oldukça düşük sıklıkta yer aldığı görülmektedir. Bu durum, öğretim materyalleri ile doğal dil kullanımı arasında bir boşluk olduğunu ortaya koymaktadır. Elde edilen sonuçlar, dil öğretim materyallerinin söylem düzeyindeki yeterlilikten ziyade yapısal dil özelliklerine öncelik verme eğiliminde olduğunu ortaya koyan önceki araştırmaları desteklemektedir.

Anahtar kelimeler: *Yabancı dil olarak Türkçe, söylem belirleyicileri, yabancı dil ders kitapları.*

Introduction

Discourse markers, from now on DMs, are linguistic elements that play a crucial role in structuring spoken and written communication by signaling relationships between ideas, organizing discourse, and managing conversational flow. They serve various functions, such as indicating contrast, elaboration, cause-and-effect relationships, and topic shifts. Schiffrin (1987) defines DMs as expressions that provide contextual coherence by linking segments of discourse, while Fraser (1999) categorizes them based on their pragmatic functions, such as elaborative, contrastive, and inferential markers. These linguistic elements contribute to the coherence and cohesion of the discourse, making communication more effective and natural.

The importance of DMs in learning a foreign language cannot be overstated. They are essential for achieving communicative competence, as they enable learners to produce and comprehend discourse in a way that mirrors native speaker usage. Studies by Müller (2005) and Fung and Carter (2007) emphasize that foreign language learners who acquire DMs can better understand the pragmatic aspects of communication, thereby improving their fluency and interactional skills. Without adequate knowledge of DMs, learners may struggle with discourse organization, leading to unnatural or fragmented speech and writing.

Given their significance, DMs should be systematically incorporated into language syllabi and coursebooks to support learners in developing natural and cohesive communication skills. The Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) highlights the importance of DMs in achieving coherence and cohesion in communication, emphasizing their role in enabling learners to express ideas logically and fluently. Researchers such as Fung and Carter (2007) and Gilmore (2007) argue that course materials should include explicit instruction on DMs to help learners recognize and use them appropriately in various communicative contexts. A well-structured syllabus that integrates DMs can enhance language proficiency by providing learners with the tools to manage discourse effectively and align with the communicative competencies outlined in the CEFR.

This study aims to examine Turkish as a Foreign Language (TFL) coursebooks to determine how DMs are presented and whether their inclusion

aligns with the communicative needs of learners. The study adopts a corpus-driven approach, analyzing the treatment of DMs by referring to the most frequently used ones identified in TNC (Spoken) by Aksan et al. (2012). By evaluating their presence and usage within the course materials, the study aims to provide insights into their adequacy and suggest potential improvements for more effective foreign language learning. In line with this aim, the study attempts to answer the following research question:

How are the most frequently used DMs in Turkish, as identified in TNC (Spoken) by Aksan et al. (2012), presented in TFL coursebooks?

Theoretical Background

Discourse Markers

Spoken discourse is characterized by various verbal and non-verbal elements (Norrick, 2012). These elements play a role in collaboratively constructing meaning or actions and signaling the interpretation of utterances, including practices like turn-taking, backchannel responses, pragmatic markers, and gestures (Caffi & Janney, 1994). DMs are a type of pragmatic marker that helps facilitate the flow of conversation, organize contributions in sequence to maintain coherence, and express epistemic and affective perspectives (Fraser, 1996). The term “DM,” as defined by Roulet (1980), highlights the role of social interaction in shaping linguistic structures. He argues that, in addition to linguistic elements expressing epistemic and evaluative stances, illocutionary force, and persuasive intentions, DMs also serve important interactional and organizational functions in discourse. DMs include: (i) pragmatic markers, defined as words or phrases that signal the possible communicative intent of an utterance; (ii) both lexical and non-lexical elements like laughter and backchannels, which convey emotional and cognitive states and may serve as indicators of acknowledgment or lack thereof; and (iii) gestures and other non-verbal strategies, including prosodic features, which can convey a range of social meanings (Norrick, 2012). Example 1 below shows instances of each category from English.

- (1) a. *Well*, I think we should get started.
b. Speaker 1: I got the promotion!
Speaker 2: *Oh, really? That's great!*
c. Speaker 1: Do you agree?
Speaker 2: (nods) *Mm-hmm.*

In example 1a, the word “*well*” is a discourse marker and signals a transition or hesitation, preparing the listener for what comes next. It acts as a discourse-organizing device rather than adding semantic meaning. In example 1b, “*Oh, really?*” expresses surprise and acknowledgment. “*That’s great*” is a verbal reinforcement of positive affect. The phrase “*oh*” adds an emotional response, while “*really*” questions or verifies the information. In 1c, the head nod is a non-verbal cue signaling agreement. Combined with “*mm-hmm*”, it confirms acknowledgment without disrupting the speaker’s turn.

Norrick (2012) adopts a broad view of discourse markers, including lexical items, non-lexical vocal signals, and non-verbal resources that contribute to interactional meaning. In the present study, however, this broader framework is narrowed to lexical and pragmatic DMs that can be systematically identified in dialogue-based and interactional sections of the textbooks and compared with corpus-based spoken data. In this sense, the study does not attempt to analyse the full multimodal range of discourse markers discussed by Norrick but rather focuses on the subset most relevant to textbook representation and foreign language input. This narrower operationalization is particularly suitable for corpus-informed textbook analysis, as it allows the study to examine how frequently occurring spoken discourse markers are represented in pedagogical materials (Furkó, 2020).

DMs in Turkish

DMs in Turkish range from simple conjunctions to more nuanced expressions that reflect contextual subtleties. The most frequently used DMs in Turkish, as identified in the TNC (Spoken) (Aksan et al., 2012), along with their functional relationships and illustrative examples, are presented below.

“*Yani*” is a multifunctional DM that helps manage the flow of conversation, clarify meaning, emphasize points, or soften statements. Its closest English equivalents include “*I mean*”, “*well*”, “*you know*” or “*so*” depending on the context. Its flexibility makes it one of the most frequently used markers in everyday Turkish speech. Example 2 below shows the usage of “*yani*” as a DM in Turkish.

- (2) *Yani* neden ol-ma-sın?
 well why become-NEG-OPT.3SG

“Well, why not?” (S-BEABXw-0391-686)¹

In Example 2, the DM “*yani*” functions as a reinforcing and emphasis marker that strengthens the speaker’s stance. It signals reasoning or justification for the statement that follows, subtly implying that the idea being proposed is reasonable or logical.

The Turkish DM “*şey*” is one of the most frequently used and versatile elements in spoken language. While its literal meaning is “*thing*,” as a DM, “*şey*” serves various pragmatic functions that go beyond its dictionary definition. It plays a crucial role in managing conversations, structuring speech, and expressing cognitive or emotional states. Example 3 shows the usage of “*şey*”.

(3) *Şey* hani o zehirle-n -di-m de-di-m ya.
well you know this poison-PASS-PST-1SG say-PST-1SG DM
“Well, you know, I said I was poisoned” (S-BEABXO-0392-881)

In example (3), “*şey*” introduces a reminder of something that was mentioned earlier. It signals that the speaker is about to refer back to a previous conversation.

“*İşte*” serves as a multifunctional DM that helps organize conversation, highlight information, and express emotions like resignation or emphasis. Depending on the context, its English equivalents could be “*here*”, “*well*”, “*there you go*”, “*this is it*” or “*you know*”. Its flexibility makes it a fundamental part of natural, spoken Turkish. Example 4 below shows the usage of “*işte*” as a DM in Turkish.

(4) *İşte* birçok öğrenci gel-iyor ev tut-uyor para harca-ıyor.
you know many student come-PROG house rent-PROG money
spend-PROG
“You know, many students come, rent a house, spend money.”
(S-ADABO-0366-469)

In example (4), “*işte*” acts as a framing device that introduces a familiar scenario, suggesting that the actions of the students are typical or illustrative of a broader trend. It adds a conversational tone, making the statement feel more casual and relatable.

¹ The tags in the parentheses were given by Turkish National Corpus (TNC) (Aksan et al., 2012). The letter “S” stands for the spoken data.

“*Falan*” is a versatile DM that adds vagueness, generalization, or informality to speech. It can soften statements, convey indifference, or simply indicate an incomplete list. Depending on the context, its English equivalents may include “*and so on*”, “*etc.*” or “*stuff like that*”. Example 5 shows the usage of “*falan*”.

- (5) Hani erkek-ler-de *falan* çok fazla hisset-me-yiz bu-nu.
 you know man-PL-LOC and.stuff very much feel-NEG-AOR.1PL this-ACC
 “You know, we don't feel this much in men.” (S-AEABUZ-0021-1442)

In example (5), “*falan*” suggests that men are an example of the group being discussed, but the statement may also apply to similar groups or contexts. It broadens the scope without specifying exact limits.

Like “*falan*”, “*filan*” is another DM in Turkish that is used to indicate approximation, vagueness, or generalization. It functions similarly to “*etc.*”, “*and so on*”, or “*or something like that*” in English. Sometimes “*falan*” and “*filan*” are used interchangeably. Example 6 shows the usage of “*falan*” in Turkish.

- (6) Şimdi hiç kimse oğl-u papaz *filan* ol-sun iste-me-z.
 now, nobody son-3SG.POSS priest and.stuff be-OPT.3SG want-NEG-AOR
 “Now, nobody wants their son to be a priest or something like that.”
 (S-ACABUZ-0331-103)

In example (6), the speaker suggests that it's not just about becoming a priest but possibly any religious figure or something in that realm. It allows for an open-ended interpretation, implying “a priest or any similar role” without being too specific.

The DM “*zaten*” frequently reinforces statements, indicates inevitability, or provides additional justification for a point. It can be translated as “*anyway*”, “*already*” or “*as it is*” in English, depending on the context. Example (7) below shows the usage of this DM.

- (7) Şöyle yap-alım, ben fazla dur-mu-ca-m *zaten*.
 this way do-OPT.1PL I long stay-NEG-FUT-1SG anyway
 “Let's do it this way, I won't stay for a long time.” (S-BEABXX-0411-1832)

In example (7), the speaker emphasizes that their decision to not stay long was already the case, regardless of what happened. “*Zaten*” suggests that

the decision is not influenced by the current situation but was already determined beforehand.

“*Tamam*” as a DM in Turkish is a widely used marker that serves various functions in conversation, depending on tone, context, and emphasis. It can function as agreement or acceptance, ending or wrapping up a topic, seeking confirmation, or soothing or calming. Example 8 shows the usage of “*tamam*” in Turkish.

- (8) Şu-nu unut-tuğ-u-nuz an, soru git-ti *tamam*?
 this-ACC forget-PTCP-3SG.POSS-2PL.POSS moment question go-PST okay
 “The moment you forget that, the question is gone, okay?”
 (S-BEABXO-0147-1406)

In example (8), the speaker explains a rule or an important point and wants to ensure the listener understands it. By adding “*tamam*” at the end, the speaker checks if the listener follows along.

The DM “*tabi*” serves various functions, depending on the context and tone of speech. It generally conveys agreement, confirmation, emphasis, or expectation, similar to “*of course*”, “*naturally*”, or “*sure*” in English. Example 9 below shows the usage of “*tabi*”.

- (9) *Tabi* canım ism-in orda dur-uyor.
 of.course my.dear name-2SG.POSS there stay-PROG
 “Of course my dear, your name is there.” (S-BEABXO-0006-600)

In example (9), “*tabi*” reinforces certainty, meaning something like “*of course!*” or “*obviously*” in English. The speaker assures the listener that their name is indeed there.

The DM “*peki*” serves multiple functions depending on the context and tone. It can be translated as “*okay*”, “*well*”, “*alright*”, “*but*” or “*so*” in English. It is often used to signal agreement, introduce a follow-up question, or express hesitation. A related example is given in (10) below.

- (10) *Peki, bu yeterli mi?*
 so this enough Q
 “So, is that enough?” (S-AEABPZ-0331-36)

In example (10), “*peki*” is used to transition into a question that follows from the previous discussion. Here, it signals that the speaker checks whether the current situation meets expectations.

The DM “*haydi*” (also spelled “*hadi*”) is commonly used in spoken language to express encouragement, urgency, invitation, or impatience. It is similar to “*come on*”, “*let’s*”, “*go on*”, or “*hurry up*” in English. Its function depends on the tone and context of the conversation. A related example is given in (11) below.

- (11) *Haydi* *çabuk* *toparlan-ın*.
 let’s quickly get.ready-IMP.2PL
 “Let’s get together quickly.” (S-VI19E1A-4001-859)

In example (11), “*haydi*” signals that the speaker wants the listeners to act quickly. The word “*çabuk*” (quickly) reinforces this urgency. The DM “*neyse*” serves various functions depending on the context and tone. It is commonly used to change the subject, conclude a conversation, downplay something, or express resignation. It can be translated as “*any-way*”, “*never mind*”, “*oh well*” or “*let’s move on*” in English. Example 12 shows the usage of “*neyse*” in Turkish.

- (12) *Neyse* sen yaz yeni numara-yı.
 anyway you write new number-ACC
 “Anyway, write the new number.” (S-SA-16B2A-0311-240)

In example (12), the DM “*neyse*” is used to shift the focus and move on to the main point. The speaker likely mentioned something before but decides not to dwell on it and moves to the main action (*writing down the number*).

In Turkish, the DM “*aynen*” is widely used in informal conversations and serves multiple functions such as expressing agreement or confirmation, emphasizing that something happens in the same manner or indicating that something is the same for the speaker. It can be translated as “*exactly*”, “*totally*”, “*yeah, same*”, “*that’s right*”, or “*just like that*” depending on the context. Example 13 below shows the usage of “*neyse*” in Turkish.

- (13) Evet *aynen* hep-si var.
 Yes exactly all-3SG.POSS exist
 “Yes, that's all there is.” (S-BEABXs-0130-178)

In example (13), the DM “*aynen*” is used to confirm and emphasize agreement with the previous statement. By adding “*aynen*”, the speaker stresses that the information is completely accurate, leaving no room for doubt.

The individual analysis of DMs in Turkish show that these markers are highly context-dependent, often carrying multiple pragmatic functions that extend beyond their lexical meanings. It is undeniable that they serve diverse functions in structuring conversation, managing interaction, and shaping meaning. Next section presents the roles of DMs in foreign language materials.

DMs in Foreign Language Teaching Materials

Discourse markers, which occur frequently in spoken language, are a defining feature of speech. Their significance in everyday communication is considerable, as they support both interpretation and social engagement in interaction (Lam, 2010). Despite their significance in natural speech, research has consistently shown that foreign language teaching materials often fail to adequately represent DMs (Gilmore, 2007; Fung & Carter, 2007). Many textbooks tend to prioritize grammatical structures and lexical knowledge over the discourse-level features that facilitate authentic communication (Müller, 2005). This discrepancy raises concerns about whether learners are sufficiently exposed to the pragmatic and interactional strategies necessary for real-world conversations.

The Common European Framework of Reference for Languages (CEFR) outlines three types of competence - linguistic, sociolinguistic, and pragmatic - each described in terms of knowledge and skills. Among these, pragmatic competence involves the functional use of linguistic resources (Furkó, 2020). In the CEFR, discourse competence is considered as a component of pragmatic competence and is defined as the ability of language users or learners to organize sentences in a sequence in order to produce coherent stretches of language (Council of Europe, 2001). Furthermore, coherence and cohesion are identified as key criteria that must be fulfilled to ensure clear and effective communication. For the construction of coherent discourse, CEFR emphasizes the importance of discourse competence in language learning, underscoring the need for instructional

materials to reflect natural spoken patterns, including the appropriate use of DMs (Carlsen, 2010). The scale of coherence and cohesion reflects a fundamental distinction in the CEFR's description of language development, namely that between quantity and quality. The use of DMs across proficiency levels is described, on the one hand, in terms of the range of markers employed. At lower levels, learners are expected to use only very basic DMs; at the A2 level, simple markers are introduced; and at the A2+ level, the most frequently occurring DMs are expected. At higher levels, the range of DMs is assumed to expand progressively.

Although DMs have been examined in relation to language teaching materials, their representation and treatment in Turkish as a foreign language coursebooks remain underexplored. Thus, this study fills this gap in the literature by investigating the extent to which DMs are incorporated into teaching materials.

Previous Studies on DMs in Foreign Language Teaching Materials

Although different areas have been examined in foreign language materials before, including studies such as epistemic modality (Holmes, 1988), future time expressions (Mindt, 1997), dialogues (Wong, 2002) and multi-word lexical phrases (Koprowski, 2005). what seems to be limited is the study of the role of DMs.

Lam (2010) examines the discrepancies between the use of the discourse particle “well” by expert English speakers and its representation in secondary-level textbooks in Hong Kong. The findings of the study show that there is a significant misalignment between real-world usage and pedagogical presentation. Textbooks often present “well” in a highly skewed manner, over-emphasizing its use in discussions while virtually ignoring it in presentations. Furthermore, textbooks fail to reflect the diversity of the particle's positioning (especially medial use) and its broad functional range (including textual framing and cognitive processing). Alavi-Nia and Mozaffari (2014) conducts a corpus-based study evaluating the use of discourse particles (DPs) in English as a Foreign Language (EFL) and Persian as a Foreign Language (PFL) textbooks. The results of the study show that EFL course-books utilize DPs at a significantly higher rate (8.8% of total word count) than PFL course-books (6.5%), suggesting that EFL materials more closely approximate naturally occurring language. Regarding the presentation of DMs in English as a Foreign Language (EFL) textbooks,

Shimada (2012) finds significant discrepancies between textbook dialogues and natural speech. The results of the study show that while DMs are essential for communicative competence and indicative of authenticity, textbooks often fail to replicate the frequency and functional variety found in natural conversation. Kawashima (2023) analyses the representation of pragmatic markers - specifically “oh”, “okay”, “well”, and “yeah” - in beginner-level international and Japanese high school ELT textbooks and finds that there are significant discrepancies between textbook dialogues and naturally occurring discourse as represented in the Corpus of Contemporary American English (COCA). Results of the study show that textbooks often limit these markers to a single function and restrict their placement to the beginning of utterances. In contrast, authentic discourse employs these markers for a wider variety of functions and places them in medial or final positions to signal politeness or floor-sharing.

Taken together, these studies reveal a consistent gap in the representation and treatment of DMs in foreign language teaching materials. Next section presents the methodology of the current study.

Methodology

This section explores the study’s methodology in detail. It begins by presenting the DMs examined in the research. Next, the process of data collection is outlined, including an overview of the sample size and the raw frequency of occurrences. Following this, the section provides an in-depth discussion of the analysis of the gathered data.

DMs Selected for Analysis from the TNC (Spoken)

The DMs that are analyzed in this study are taken from TNC (Spoken) (Aksan et al., 2012). The reason for choosing DMs from the spoken section is that they play a crucial role in spoken communication, helping to structure conversations, manage turn-taking, and express interpersonal meanings. Analyzing DMs from the TNC (Spoken) provides valuable insights into how Turkish speakers naturally use these elements in real-life interactions. Moreover, since DMs are highly context-dependent, analyzing them in spontaneous speech rather than controlled settings ensures a more accurate representation of their usage. Table 1 below shows the most frequently used DMs in Spoken Turkish National Corpus.

Table 1. *The most frequently used DMs in Spoken Turkish National Corpus*

Rank	(DMs)	Number of Texts	<i>Of</i>	<i>Nf</i>
1	yani	342	8561	8442.61
2	şey	311	5714	5634.98
3	işte	310	4296	4236.59
4	falan	215	2207	2176.48
5	zaten	275	2156	2126.18
6	tamam	244	2107	2077.86
7	tabi	248	1785	1760.32
8	ya (yaa)	133	1450	1429.95
9	peki	182	986	972.36
10	haydi (hadi)	168	847	835.29
11	filan	100	433	427.01
12	neyse	121	318	313.6
13	aynen	101	262	258.38

Note. *Of* = observed frequency; *Nf* = normalized frequency (per million words). Number of texts refers to the number of texts in which each discourse marker occurs (out of 456 texts). Observed frequency indicates the raw frequency of each discourse marker in the corpus (N = 1014023 words). Normalized frequency represents occurrences per million words.

Table 1 shows that “*yani*”, “*şey*”, “*işte*”, “*falan*”, “*zaten*”, “*tamam*”, “*tabi*”, “*ya (yaa)*”, “*peki*”, “*haydi (hadi)*”, “*filan*”, “*neyse*”, “*aynen*” are the most frequently used DMs in spoken Turkish (as represented in the TNC) according to normalized frequency values. McEnery et al. (2006) emphasize the importance of careful interpretation when analyzing frequency data, as raw counts from a corpus often need to be standardized for meaningful comparisons. Therefore, this study used normalized values of the examined DMs as a reference point in selecting which DMs would be analyzed. Thus, only DMs with normalized frequency values exceeding 250 were selected for analysis in this study.

Data collection

The selection of textbooks for this study is based on their widespread use, proficiency coverage, and relevance to TFL instruction. The research analyses two widely used TFL coursebook series: *Yeni Hitit Yabancılar İçin Türkçe Ders Kitabı* and *Yediiklim Türkçe*. The *Yeni Hitit* series (levels 1-2-3), developed by Ankara University’s TÖMER (Turkish and Foreign Languages Research and Application Centre) (2019), is designed to cover CEFR levels A1 to C1 and is recognized for its structured and pedagogical-

ly controlled approach. In contrast, the *Yediiklim Türkçe* series, developed by the Yunus Emre Institute (2016), covers a broader proficiency range from A1 to C2 and is also widely used in teaching Turkish to non-native speakers, offering a more extensive progression across levels within the CEFR framework.

The books include thematic units, dialogues, reading passages, and exercises, making them ideal for analyzing how DMs are integrated into structured language learning. The selection ensures that the study captures how DMs are introduced across different proficiency levels, providing insights into their representation in both beginner and advanced stages of language learning. Additionally, both coursebook series are frequently used in university TFL programs and language institutions, making their analysis relevant for understanding standard TFL instruction. Table 2 below gives detailed information about the coursebooks analyzed in this study.

Table 2. *The coursebooks analyzed in this study*

No.	Coursebook	Publisher	Publication Year	Intended Learners
1	<i>Yeni Hitit 1</i>	Ankara University TÖMER	2019	A1-A2
2	<i>Yeni Hitit 2</i>	Ankara University TÖMER	2019	B1
3	<i>Yeni Hitit 3</i>	Ankara University TÖMER	2019	B2-C1
4	<i>Yediiklim Türkçe A1</i>	Yunus Emre Institute	2016	A1
5	<i>Yediiklim Türkçe A2</i>	Yunus Emre Institute	2016	A2
6	<i>Yediiklim Türkçe B1</i>	Yunus Emre Institute	2016	B1
7	<i>Yediiklim Türkçe B2</i>	Yunus Emre Institute	2016	B2
8	<i>Yediiklim Türkçe C1</i>	Yunus Emre Institute	2016	C1
9	<i>Yediiklim Türkçe C2</i>	Yunus Emre Institute	2016	C2

Table 2 shows Whereas the *Yeni Hitit* series comprises three coursebooks covering CEFR levels A1 to C1, the *Yediiklim Türkçe* series consists of six coursebooks and extends from A1 to C2.

Data analysis

The data analysis in this study adopts a quantitative approach to examine the representation of DMs in both the *Yeni Hitit* and *Yediiklim Türkçe* series and to compare their usage with the TNC (Spoken). Since the reference corpus consists of spoken Turkish, only dialogue-based and interactional sections of the textbooks - including conversations, listening transcripts, speaking activities, as well as workbooks and transcriptions of audio ma-

materials used as supplementary resources - were included in the analysis to ensure comparability.

First, a manual search was conducted to identify occurrences of selected DMs in the textbooks. The total instances of each DM were counted by two researchers. The inter-coder reliability between the researchers is 0.995 (Cohen's Kappa). McHugh (2012) explains that a Cohen's Kappa value ranging from 0.81 to 1.00 indicates an almost perfect level of agreement between raters. Therefore, it can be confidently concluded that the coding process yields reliable results.

The analysis of DMs in *Yeni Hitit* and *Yediiklim Türkçe* was conducted using normalized frequency per million words and ratio comparison to evaluate their distribution against the TNC (Spoken) section. Since the TNC (Spoken) data and the textbook data differ significantly in size, raw frequency counts alone are insufficient for meaningful comparison. To account for this discrepancy, normalized frequencies (i.e., occurrences per one million words) were calculated to standardize the data and enable a proportional comparison. Additionally, ratio analysis was employed to assess the relative representation of DMs in coursebooks compared to their natural usage in the TNC Spoken Section. A ratio value of *1.0* indicates that a discourse marker appears at an equal rate in both corpora, while values above *1.0* indicate overrepresentation, and values below *1.0* indicate underrepresentation or omission. This approach allows for a quantifiable, objective comparison, ensuring that DM usage in coursebooks is evaluated in relation to authentic Turkish discourse patterns.

Findings and Discussion

Findings

Table 3 presents the observed and normalized frequencies of DMs in the *Yeni Hitit-1* (A1-A2) coursebook and the *Yediiklim* A1-A2 coursebooks, as well as in the TNC (spoken), along with their ratio values.

Table 3. Comparison of DMs in *Yeni Hitit-1* (A1-A2), *Yediiklim* (A1-A2), and TNC (Spoken)

Rank	DM	<i>Of</i> (Yeni Hitit)	<i>Nf</i> (Yeni Hitit)	<i>Of</i> (Yediiklim)	<i>Nf</i> (Yediiklim)	<i>Nf</i> TNC	<i>Ratio</i> (Yeni Hitit)	<i>Ratio</i> (Yediiklim)
1	yani	4	117.5	10	704.4	8442.61	0.01	0.08
2	şey	5	146.8	14	985.7	5634.98	0.03	0.17

3	işte	3	88.1	6	352.2	4236.59	0.02	0.08
4	fa lan	0	0.0	0	0.0	2176.48	0.00	0.00
5	za ten	2	58.7	4	281.7	2126.18	0.03	0.13
6	ta mam	17	499.1	19	985.7	2077.86	0.24	0.47
7	ta bi	17	499.1	15	1056.1	1760.32	0.28	0.60
8	ya	0	0.0	4	281.7	1429.95	0.00	0.20
9	pe ki	19	440.5	15	774.7	972.36	0.45	0.79
10	ha di	5	146.8	5	352.2	835.29	0.18	0.42
11	fi lan	0	0.0	0	0.0	427.01	0.00	0.00

Note. *Of*= observed frequency; *Nf*= normalized frequency (per million words). *Yeni Hitit* values are based on a total of 8742 words (A1-A2 levels), and *Yediiklim* values are based on a total of 14199 words (A1-A2 levels). TNC values are normalized per million words (N = 1014023 words). Ratios represent the proportion of textbook frequencies relative to the spoken corpus (TNC).

As shown in Table 3, all DMs are considerably less frequent in both textbook series than in the TNC (spoken). Among the markers, *yani*, *şey*, and *işte* show particularly low normalized frequencies in both *Yeni Hitit* and *Yediiklim* compared to the corpus. In contrast, *peki*, *tabi*, and *tamam* display relatively higher frequencies within the textbooks, although they are still below the TNC values. The marker *ya* is entirely absent in *Yeni Hitit* but appears at a low frequency in *Yediiklim*. Additionally, markers such as *fa lan* and *fi lan* are not attested in either series. Overall, while all DMs are underrepresented in comparison to spoken Turkish, *Yediiklim* generally exhibits higher frequencies across most markers than *Yeni Hitit*.

Table 4 presents the observed and normalized frequencies of DMs in the *Yeni Hitit-2 (B1)* coursebook and the *Yediiklim B1-B2* coursebooks, in comparison with the TNC (spoken), along with their ratio values.

Table 4. Comparison of DMs in *Yeni Hitit-2 (B1)*, *Yediiklim (B1-B2)*, and TNC (Spoken)

Rank	DM	<i>Of</i> (<i>Yeni Hitit</i>)	<i>Nf</i> (<i>Yeni Hitit</i>)	<i>Of</i> (<i>Yediiklim</i>)	<i>Nf</i> (<i>Yediiklim</i>)	<i>Nf</i> TNC	Ratio (<i>Yeni Hitit</i>)	Ratio (<i>Yediiklim</i>)
1	yani	18	528.6	13	525.2	8442.61	0.06	0.06
2	şey	10	293.7	14	525.2	5634.98	0.05	0.09
3	işte	5	146.8	14	613.1	4236.59	0.03	0.14
4	fa lan	2	58.7	0	0.0	2176.48	0.03	0.00
5	za ten	5	146.8	6	350.1	2126.18	0.07	0.16
6	ta mam	5	146.8	9	262.6	2077.86	0.07	0.13

7	tabi	8	234.9	14	613.1	1760.32	0.13	0.35
8	ya	2	58.7	11	613.1	1429.95	0.04	0.43
9	peki	13	381.8	22	963.3	972.36	0.39	0.99
10	hadi	5	146.8	11	613.1	835.29	0.18	0.73
11	filan	0	0.0	0	0.0	427.01	0.00	0.00

Note. *Of* = observed frequency; *Nf* = normalized frequency (per million words). *Yeni Hitit* values are based on a total of 11386 words (B1 level), and *Yediiklim* values are based on a total of 21109 words (B1-B2 levels). TNC values are normalized per million words (N = 1014023 words). Ratios represent the proportion of textbook frequencies relative to the spoken corpus (TNC).

As shown in Table 4, all DMs remain less frequent in both textbook series than in the TNC (Spoken). The markers *yani*, *şey*, and *işte* continue to display relatively low normalized frequencies in both *Yeni Hitit* and *Yediiklim* compared to the corpus. However, *Yediiklim* shows noticeably higher frequencies for several markers, particularly *işte*, *tabi*, *ya*, and *hadi*. The marker *ya* is especially more prominent in *Yediiklim*, whereas it appears only minimally in *Yeni Hitit*. Among the markers, *peki* demonstrates the highest frequency in both series, approaching the TNC value in *Yediiklim*. In contrast, *falan* and *filan* are either absent or minimally represented, particularly in *Yediiklim*. Overall, while DMs remain underrepresented relative to spoken Turkish, *Yediiklim* exhibits a broader and more frequent use of DMs compared to *Yeni Hitit* at this level.

Table 5 presents the observed and normalized frequencies of DMs in the *Yeni Hitit-3 (B2-C1)* coursebook and the *Yediiklim C1-C2* coursebooks, in comparison with the TNC (spoken), along with their ratio values.

Table 5. Comparison of DMs in *Yeni Hitit-3 (B2-C1)*, *Yediiklim (C1-C2)*, and TNC (Spoken)

Rank	DM	<i>Of</i> (Yeni Hitit)	<i>Nf</i> (Yeni Hitit)	<i>Of</i> (Yediiklim)	<i>Nf</i> (Yediiklim)	<i>Nf</i> TNC	Ratio (Yeni Hitit)	Ratio (Yediiklim)
1	yani	8	234.9	19	737.6	8442.61	0.03	0.09
2	şey	9	264.3	17	737.6	5634.98	0.05	0.13
3	işte	7	205.6	15	614.6	4236.59	0.05	0.15
4	falan	3	88.1	1	40.9	2176.48	0.04	0.02
5	zaten	4	117.5	8	327.8	2126.18	0.06	0.15
6	tamam	3	88.1	1	40.9	2077.86	0.04	0.02
7	tabi	6	176.2	4	163.7	1760.32	0.10	0.09

8	ya	1	29.4	15	614.6	1429.95	0.02	0.43
9	peki	5	146.8	13	532.7	972.36	0.15	0.55
10	hadi	7	205.6	5	204.6	835.29	0.25	0.24
11	filan	2	58.7	2	81.8	427.01	0.14	0.19

Note. *Of*= observed frequency; *Nf*= normalized frequency (per million words). *Yeni Hitit* values are based on a total of **13928 words** (B2-C1 levels), and *Yediiklim* values are based on a total of **27066 words** (C1-C2 levels). TNC values are normalized per million words (N = 1014023 words). Ratios represent the proportion of textbook frequencies relative to the spoken corpus (TNC).

As shown in Table 5, DMs remain less frequent in both textbook series than in the TNC (spoken). The markers *yani*, *şey*, and *işte* continue to occur at relatively low frequencies in both *Yeni Hitit* and *Yediiklim* compared to the corpus. However, *Yediiklim* again shows higher frequencies for most markers, particularly *yani*, *şey*, and *işte*. The marker *ya* remains extremely limited in *Yeni Hitit* but is considerably more frequent in *Yediiklim*, indicating a clearer presence of interactional features. Among the markers, *peki* continues to be one of the most frequent in both series, with *Yediiklim* showing a higher value. In contrast, *filan* and *tamam* occur at low frequencies in both series, while *filan* remains relatively infrequent. The markers *tabi* and *hadi* display similar frequencies across the two series. Overall, although DM use increases at this level, both series still fall short of reflecting spoken Turkish, with *Yediiklim* providing a relatively richer and more balanced distribution than *Yeni Hitit*.

Discussion

This section examines the results outlined in the previous section, with support from Figures 1, 2, and 3. Figure 1 below specifically compares DMs in the *Yeni Hitit* 1 (A1-A2) coursebook and the *Yediiklim* A1-A2 coursebooks with the TNC spoken section.

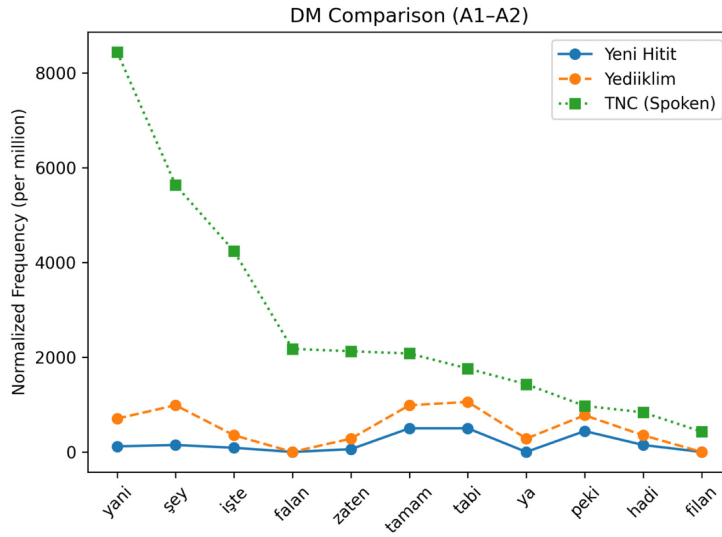


Figure 1. Representation of DMs in Yeni Hitit 1 (A1-A2) and Yediiklim (A1-A2) Compared to TNC (Spoken)

The patterns observed in Figure 1 can be interpreted in light of Carlsen’s (2010) discussion of coherence and cohesion in the CEFR. Although the CEFR predicts an increase in the range of DMs across proficiency levels, it does not assume a corresponding increase in their frequency or density. The findings at the A1-A2 level appear to align with this expectation, as both textbook series display a limited range and low frequency of DMs, reflecting the CEFR’s emphasis on “very basic” and “simple” linking devices at lower levels.

However, when compared to the TNC (spoken), the markedly lower frequencies of core DMs such as *yani*, *şey*, and *işte* suggest that textbook input may not adequately reflect the role these items play in naturally occurring spoken discourse. In this respect, while the textbooks seem to follow the CEFR’s focus on controlled and limited use of cohesive devices, they may simultaneously restrict learners’ exposure to the interactional functions of DMs. Furthermore, the relatively higher presence of markers such as *peki*, *tamam*, and *tabi* indicates a tendency toward more structured and pedagogically oriented discourse, which contrasts with the more varied and spontaneous use of DMs in spoken Turkish. This supports Carlsen’s (2010) observation that the CEFR prioritizes control and efficiency over

frequency but also highlights a potential gap between CEFR-informed materials and authentic language use.

Figure 2 below presents a comparison of DMs in the *Yeni Hitit 2* (B1) and *Yediiklim* B1-B2 coursebooks with the TNC spoken section.

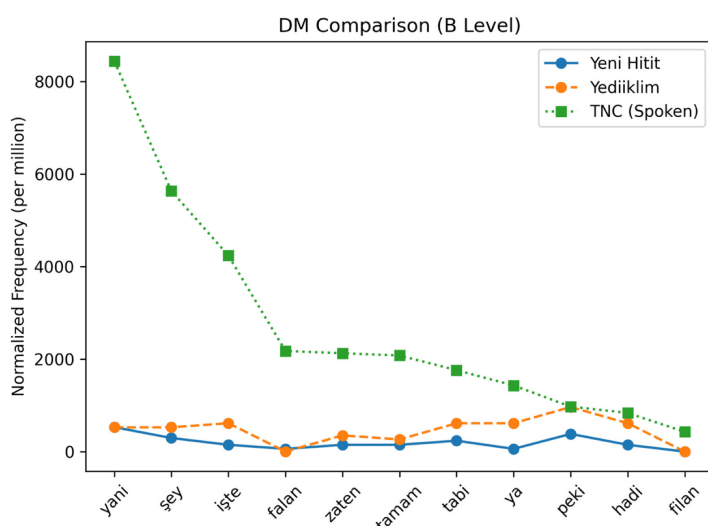


Figure 2. Representation of DMs in *Yeni Hitit 2* (B1), *Yediiklim* (B1-B2), and TNC (Spoken)

The patterns observed in Figure 2 align with Carlsen’s (2010) discussion of coherence and cohesion in the CEFR, which emphasizes limited use of cohesive devices at intermediate levels. Although some increase in DM use is evident, the frequencies in both textbook series remain substantially lower than in the TNC, particularly for the most frequently used marker *yani*, as well as other high-frequency markers such as *şey* and *işte*, suggesting restricted exposure to authentic spoken discourse. This finding is consistent with studies reporting underuse or uneven distribution of discourse features in learner language (Hinkel, 2001; Müller, 2005), as well as research showing variability across proficiency levels (Castro, 2004; Johnson, 1992). Furthermore, the relatively higher presence of certain markers in *Yediiklim* compared to *Yeni Hitit* reflects variation in how teaching materials operationalize CEFR principles, reinforcing Carlsen’s point that the framework prioritizes control and range rather than frequency.

Figure 3 below presents a comparison of DMs in the *Yeni Hitit 3 (B2-C1)* and *Yediiklim C1-C2* coursebooks with the TNC spoken section.

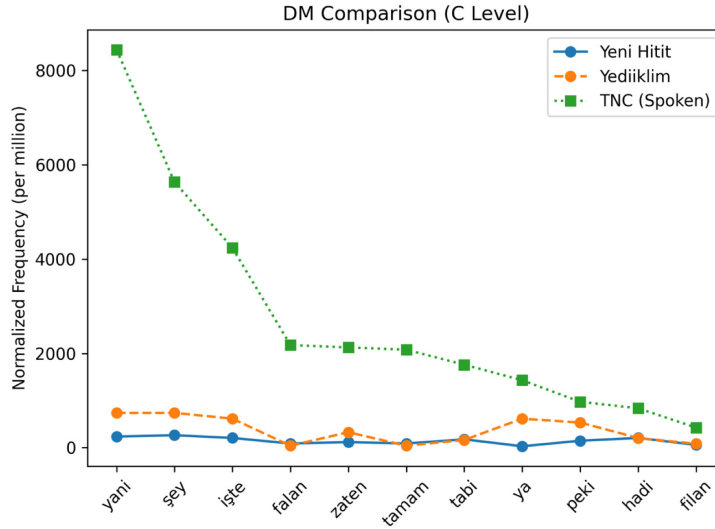


Figure 3. Representation of DMs in Yeni Hitit 3 (B2-C1), Yediiklim (C1-C2), and TNC (Spoken)

The patterns observed in Figure 3 highlight important characteristics of DM use in advanced-level teaching materials. Despite the expectation that higher proficiency levels should involve a wider range and more appropriate use of cohesive devices, the distribution of DMs in both textbook series still differs markedly from the TNC, particularly for high-frequency markers such as *yani*, *sey*, and *işte*. This suggests that advanced learners are not fully exposed to the discourse-pragmatic richness of spoken Turkish, especially in terms of interactional markers.

This pattern reflects the CEFR's emphasis on control, efficiency, and appropriateness rather than increased frequency of use (Council of Europe, 2001). However, the limited presence of key DMs in the textbooks indicates that advanced-level materials may prioritize structurally coherent and formally controlled discourse at the expense of more natural, spontaneous interaction. This observation aligns with previous research showing that learner language - even at advanced levels - may diverge from native usage in terms of the frequency and distribution of cohesive devices (Hinkel, 2001; Müller, 2005), and that increased proficiency does not necessarily

lead to greater density of such features (Castro, 2004; Johnson, 1992). At the same time, the relatively richer and more varied use of certain markers in *Yediiklim* suggests that some materials move closer to reflecting authentic discourse practices, although a noticeable gap with spoken Turkish remains.

Taken together, the findings across all three proficiency levels reveal a consistent gap between the representation of DMs in textbook materials and their use in naturally occurring spoken Turkish. Although some increase in variety and frequency can be observed as proficiency levels rise, both *Yeni Hitit* and *Yediiklim* continue to underrepresent core DMs such as *yani*, *şey*, and *işte* in comparison to the TNC. This suggests that learners are exposed to a limited and pedagogically controlled version of discourse, which does not fully reflect the interactional richness of authentic language use.

These findings are in line with previous research highlighting discrepancies between textbook input and real-world language use. For instance, Lam (2010) demonstrates that textbooks may present discourse particles in a skewed and functionally restricted manner, while Shimada (2012) similarly reports that textbook dialogues fail to capture the frequency and functional diversity of DMs found in natural conversation. Likewise, Kawashima (2023) shows that pragmatic markers in textbooks are often limited to a narrow range of functions and positions, unlike their more flexible and multifunctional use in authentic discourse.

Furthermore, the differences observed between *Yeni Hitit* and *Yediiklim* support the claim that teaching materials vary in how closely they approximate natural language. In this respect, the relatively higher frequency of DMs in *Yediiklim* aligns with Alavi-Nia and Mozaffari's (2014) findings that some instructional materials incorporate discourse features more extensively than others. Overall, the results of the present study reinforce the view that foreign language textbooks tend to simplify and constrain the use of DMs, potentially limiting learners' exposure to the pragmatic and interactional features essential for communicative competence.

Conclusion

This study aimed to answer the research question: “How are the most frequently used DMs in Turkish, as identified in the TNC (Spoken) by Aksan et al. (2012), presented in TFL coursebooks?”

The findings reveal a clear mismatch between the distribution of DMs in naturally spoken Turkish and their representation in TFL coursebooks, raising concerns about the adequacy of these materials for developing communicative competence. Across both *Yeni Hitit* and *Yediiklim* series, DMs are generally underrepresented and unevenly distributed, particularly those central to interactional fluency and pragmatic use. Core markers such as *yani*, *şey*, and *işte* remain considerably limited compared to their high frequency in natural discourse, while several markers occur only minimally or inconsistently. This pattern supports previous research indicating that language teaching materials tend to prioritize structural aspects of language over pragmatic and discourse-level features (e.g. Gilmore 2007; Fung & Carter 2007; Schmitt, 2023; Curry & McEnery, 2025).

From a pedagogical perspective, this lack of authentic DMs conflicts with the CEFR, which emphasizes interactional competence, pragmatic skills, and discourse coherence as key components of communicative language teaching (Council of Europe, 2001). The CEFR explicitly states that learners should be equipped with the discourse strategies necessary for maintaining conversations, structuring speech, and expressing stance - all of which are heavily dependent on the appropriate use of DMs. The absence of such elements in TFL textbooks suggests that learners may struggle with spontaneous communication, interactional coherence, and pragmatic flexibility, leading to speech that feels unnatural and less effective in real-world interactions.

These findings underscore the urgent need for a more corpus-informed approach to TFL material development. Coursebook designers should integrate authentic spoken data, ensuring that DMs are not only included but also presented in meaningful contexts that reflect their natural usage. As previous studies in foreign language teaching have suggested (e.g. Timmis, 2015; Müller, 2005), textbook dialogues must align more closely with real-life communication patterns to enable learners to acquire both linguistic accuracy and pragmatic fluency. Without such improvements, TFL learners will remain at a disadvantage in real-world interactions, unable to fully grasp the nuances of conversational Turkish.

This study has some limitations related to the nature of the data. While the TNC (Spoken) reflects naturally occurring spoken language, the coursebooks represent pedagogically designed and more controlled input. Al-

though only dialogue-based and interactional sections were analyzed to enhance comparability, these sections may still not fully capture the spontaneity and variability of authentic speech. As a result, the lower frequency of DMs in the textbooks may reflect not only underrepresentation but also the effects of pedagogical simplification and material design.

Future research should extend this analysis by examining other TFL textbooks and instructional materials, exploring the explicit teaching of DMs in classroom settings, and assessing the impact of corpus-informed textbook design on learner proficiency. A pedagogical shift toward discourse-focused language instruction, incorporating frequent and functional DMs, would better equip learners for authentic communication and bridge the gap between textbook language and actual spoken Turkish.

Research and Publication Ethics

In this study, all rules specified in the Higher Education Institutions Scientific Research and Publication Ethics Directive were followed. None of the actions described under the heading of Actions Contrary to Research and Publication Ethics have been carried out.

Ethics Committee Permission

The research does not require ethics committee approval.

Conflict of Interest

There is no issue in the research that could lead to a conflict of interest.

Author Contribution

The author solely conceived and designed the study, collected and analyzed the data, interpreted the findings, and wrote, revised, and approved the final version of the manuscript.

Use of Artificial Intelligence Tools

No artificial intelligence (AI) tools were used in the conception, design, data collection, data analysis, interpretation of the findings, or preparation of this manuscript.

References

[1] Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H., Atasoy, G., Öz, S., Yıldız, İ., & Kurtoğlu, Ö. (2012). Construction of the Turkish National Corpus (TNC). In *Proceedings of the*

Eighth International Conference on Language Resources and Evaluation (LREC'12) (pp. 3223–3227). European Language Resources Association.

[2] Alavi-Nia, M., & Mozaffari, M. (2014). Discourse particles in EFL and PFL course-books: Tracing text authenticity. *Procedia—Social and Behavioral Sciences*, 98, 161–170. <https://doi.org/10.1016/j.sbspro.2014.03.402>

[3] Ankara Üniversitesi Türkçe Öğretim Araştırma ve Uygulama Merkezi. (2019). *Yeni Hitit yabancılar için Türkçe ders kitapları 1-3*. Ankara Üniversitesi Türkçe Öğretim Araştırma ve Uygulama Merkezi.

[4] Caffi, C., & Janney, R. W. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics*, 22(3–4), 325–373. [https://doi.org/10.1016/0378-2166\(94\)90115-5](https://doi.org/10.1016/0378-2166(94)90115-5)

[5] Carlsen, C. (2010). Discourse connectives across CEFR levels: A corpus-based study. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 191–210). European Second Language Association.

[6] Castro, C. D. (2004). Cohesion and the social construction of meaning in the essays of Filipino college students writing in L2 English. *Asia Pacific Education Review*, 5(2), 215–225. <https://doi.org/10.1007/BF03024959>

[7] Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

[8] Curry, N., & McEnery, T. (2025). Corpus linguistics for language teaching and learning: A research agenda. *Language Teaching*, 58(2), 232–251. <https://doi.org/10.1017/S0261444824000430>

[9] Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167–190. <https://doi.org/10.1075/prag.6.2.03fra>

[10] Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31(7), 931–952. [https://doi.org/10.1016/S0378-2166\(98\)00101-5](https://doi.org/10.1016/S0378-2166(98)00101-5)

[11] Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410–439. <https://doi.org/10.1093/applin/amm030>

[12] Furkó, P. B. (2020). *Discourse markers and beyond: Descriptive and critical perspectives on discourse-pragmatic devices across genres*

and languages. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-37763-2>

[13] Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118. <https://doi.org/10.1017/S0261444807004144>

[14] Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, 12(2), 111–132.

[15] Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9(1), 21–44. <https://doi.org/10.1093/applin/9.1.21>

[16] Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RELC Journal*, 23(2), 1–17. <https://doi.org/10.1177/003368829202300201>

[17] Kawashima, C. (2023). An analysis of pragmatic markers presented in ELT textbooks. In B. Ferguson, B. Lacy, & R. Derrah (Eds.), *Learning from students, educating teachers: Research and practice* (pp. 162–169). Japan Association for Language Teaching. <https://doi.org/10.37546/JALT-PCP2022-19>

[18] Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59(4), 322–332. <https://doi.org/10.1093/elt/cci061>

[19] Lam, P. W. Y. (2010). Discourse particles in corpus data and textbooks: The case of *well*. *Applied Linguistics*, 31(2), 260–281. <https://doi.org/10.1093/applin/amp026>

[20] McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

[21] McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>

[22] Mindt, D. (1997). Corpora and the teaching of English in Germany. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 40–50). Longman.

[23] Müller, S. (2005). *Discourse markers in native and non-native English discourse*. John Benjamins. <https://doi.org/10.1075/pbns.138>

[24] Norrick, N. R. (2012). Listening practices in English conversation:

The responses responses elicit. *Journal of Pragmatics*, 44(5), 566–576. <https://doi.org/10.1016/j.pragma.2011.08.007>

[25] Roulet, E. (1980). Interactional markers in dialogue. *Applied Linguistics*, 1(3), 224–233.

[26] Schifffrin, D. (1987). *Discourse markers*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511611841>

[27] Schmitt, N. (2023). Norbert Schmitt’s essential bookshelf: Formulaic language. *Language Teaching*, 56(3), 420–431. <https://doi.org/10.1017/S0261444822000039>

[28] Shimada, K. (2012). Discourse markers in EFL textbooks and spoken corpora: Materials design and authenticity. *Language Education & Technology*, 49, 215–244.

[29] Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.

[30] Wong, J. (2002). Applying conversation analysis in applied linguistics: Evaluating dialogue in English as a second language textbooks. *International Review of Applied Linguistics in Language Teaching*, 40(1), 37–60. <https://doi.org/10.1515/iral.2002.003>

[31] Yunus Emre Enstitüsü. (2016). *Yedi İklim Türkçe ders kitapları A1-C2*. Yunus Emre Enstitüsü.

Genişletilmiş Özet

Bu çalışma, Türkçede söylem belirleyicilerin yabancı dil olarak Türkçe (YDT) öğretiminde kullanılan ders kitaplarında nasıl temsil edildiğini incelemekte ve bu kullanımın Türk Ulusal Derlemi (TUD) Sözlü verileriyle karşılaştırmasını yapmaktadır. Söylem belirleyiciler, konuşma sırasında akıcılığı sağlama, söylemi yapılandırma ve konuşmacılar arasındaki etkileşimi düzenleme gibi işlevlere sahip önemli dil unsurlarıdır. Bu öğeler, konuşmanın doğal akışını destekleyerek anlamın yorumlanmasını kolaylaştırır ve iletişimde pragmatik yeterliliğin gelişimine katkıda bulunur (Schifffrin, 1987). Bu nedenle, söylem belirleyicilerinin yabancı dil öğretiminde nasıl sunulduğu, öğrencilerin hedef dilde etkili, akıcı ve bağlamsal olarak uygun iletişim kurabilmeleri açısından büyük önem taşımaktadır. Ancak literatürde, yabancı dil ders kitaplarının söylem belirleyicileri yeterince yansıtmadığı ve bu tür öğelerin çoğunlukla ihmal edildiği sıklıkla vurgulanmaktadır (Gilmore, 2007; Fung ve Carter, 2007).

Bu çalışmada, TUD (Sözlü) verilerine dayalı olarak en sık kullanılan söylem belirleyicilerinin, Yeni Hitit (1-3 seviyeleri) ve Yediiklim Türkçe (A1-C2) ders kitabı serilerinde nasıl temsil edildiği incelenmiştir. Çalışmanın temel amacı, ders kitaplarında yer alan söylem belirleyicilerinin kullanım sıklığını ve çeşitliliğini ortaya koyarak bu materyallerin doğal konuşma dilini ne ölçüde yansıttığını değerlendirmektir. Çözümleme kapsamında yalnızca diyalog temelli ve etkileşimsel bölümler incelenmiş; ayrıca çalışma kitapları ve ses kayıtlarının yazılı dökümleri de çözümleme kapsamına dahil edilmiştir. Bu yaklaşım, ders kitaplarının özellikle sözlü etkileşim boyutunu daha doğru ve geçerli biçimde değerlendirmeye olanak sağlamaktadır. Türkçede en sık kullanılan söylem belirleyicileri *yani*, *şey*, *işte*, *falan*, *zaten*, *tamam*, *tabi*, *ya*, *peki*, *haydi* ve *filan* olarak belirlenmiş ve bu öğelerin ders kitaplarındaki kullanım sıklıkları manuel olarak tespit edilmiştir. Kodlama sürecinde iki araştırmacı tarafından gerçekleştirilen çözümleme sonucunda Cohen Kappa katsayısı 0.995 olarak hesaplanmış ve bu değer kodlayıcılar arasında oldukça yüksek düzeyde güvenilirlik olduğunu göstermiştir (McHugh, 2012).

Elde edilen veriler doğrultusunda, her bir söylem belirleyicisinin toplam kullanım sıklığı hesaplanmış ve milyon kelime başına normalize edilmiş sıklık değerleri belirlenmiştir. Bunun yanı sıra, ders kitaplarındaki kullanım ile TUD Sözlü verilerindeki kullanım arasındaki farkı ortaya koymak amacıyla oran çözümlemesi yapılmıştır. Oran değeri 1.0'a yaklaştıkça ders kitaplarının doğal konuşma diline daha yakın olduğu, 1.0'ın altındaki değerlerin ise eksik temsili gösterdiği kabul edilmiştir. Bu yöntem, söylem belirleyicilerinin ders kitaplarında ne ölçüde yer aldığını nesnel, sistematik ve karşılaştırmalı biçimde değerlendirmeye olanak sağlamaktadır. Ayrıca, üç farklı yeterlik düzeyinde yapılan karşılaştırmalar, söylem belirleyicilerin dağılımının seviyeye göre nasıl değiştiğini ortaya koymak açısından önemli veriler sunmaktadır ve bu durum öğretim materyallerinin gelişimsel yapısına dair önemli ipuçları sağlamaktadır.

Bulgular hem Yeni Hitit hem de Yediiklim serilerinde söylem belirleyicilerinin genel olarak düşük sıklıkta ve dengesiz biçimde temsil edildiğini ortaya koymaktadır. Tüm seviyelerde, özellikle Türkçede en sık kullanılan *yani*, *şey* ve *işte* gibi temel söylem belirleyicilerinin TUD verilerine kıyasla oldukça düşük sıklıkta yer aldığı görülmektedir. Alt seviyede her iki seride de söylem belirleyicilerinin sınırlı sayıda ve düşük yoğunlukta kullanıldığı, ancak Yediiklim serisinin bazı belirleyicilerde görece daha yük-

sek değerlere ulaştığı tespit edilmiştir. Orta seviyede söylem belirleyici kullanımında belirli bir artış gözlenmekle birlikte, bu artışın doğal konuşma diline yaklaşmak açısından yeterli olmadığı görülmektedir. Bu seviyede özellikle *ya* belirleyicisinin Yediiklim serisinde daha belirgin şekilde yer aldığı, buna karşın Yeni Hitit serisinde oldukça sınırlı kaldığı dikkat çekmektedir. Üst seviyede ise söylem belirleyicilerin çeşitliliğinde ve kullanımında artış gözlenmekle birlikte, doğal konuşma dilindeki dağılımla hâlâ belirgin bir fark bulunmaktadır. Bu durum, ileri düzey materyallerde bile söylem belirleyicilerin kullanımının sınırlı kaldığını göstermektedir. Buna ek olarak, iki kitap serisi arası karşılaştırmalar, Yedi İklim serisinin bazı söylem belirleyicilerde daha yüksek oran değerlerine ulaştığını ve bu yönüyle doğal konuşmaya görece daha yakın bir dağılım sergilediğini göstermektedir. Bununla birlikte, her iki seride de söylem belirleyicilerin kullanımının genel olarak sınırlı kaldığı ve özellikle etkileşimsel işlevi güçlü belirleyicilerin yeterince temsil edilmediği görülmektedir. Bu durum, ders kitaplarının söylem belirleyicilerini daha çok yapılandırılmış ve pedagojik olarak sadeleştirilmiş bağlamlarda sunduğunu düşündürmektedir ve bu bağlamların doğal konuşma örüntülerini yeterince yansıtmadığını ortaya koymaktadır.

Bu sonuçlar, ders kitaplarının söylem belirleyicilerini pedagojik olarak kontrollü, sınırlı ve çoğu zaman yapay bir biçimde sunduğunu göstermektedir. Bu durum, öğrencilerin doğal konuşma diline özgü etkileşimsel ve pragmatik özellikleri yeterince edinmesini engelleyebilir. Avrupa Ortak Başvuru Metni söylem yeterliliğini iletişimsel yeterliliğin önemli bir bileşeni olarak tanımlamakta ve öğrencilerin söylemi organize etme, akıcılığı sağlama ve bağlamsal uygunluk geliştirme becerilerini kazanmalarını hedeflemektedir. Ancak elde edilen bulgular, mevcut ders kitaplarının bu hedefleri tam anlamıyla karşılayamadığını ortaya koymaktadır. Sonuç olarak, bu çalışma yabancı dil olarak Türkçe öğretiminde kullanılan ders kitaplarının söylem belirleyicileri açısından doğal konuşma dilini yeterince yansıtmadığını göstermektedir. Her ne kadar Yedi İklim serisi bazı açılardan daha yüksek sıklık değerleri sunsa da her iki seride de belirgin bir eksiklik söz konusudur. Bu durum, ders materyallerinin sözlü derlem verilerine dayalı olarak yeniden düzenlenmesi ve söylem belirleyicilerinin daha doğal, çeşitli ve işlevsel bağlamlarda sunulması gerektiğini ortaya koymaktadır. Gelecekte yapılacak çalışmaların, farklı ders kitaplarını ve öğretim uygulamalarını inceleyerek söylem belirleyicilerinin öğretimini daha kapsamlı biçimde ele alması önerilmektedir.