

The Use of English Definite Article “the” with Generic Reference: A Comparative Corpus-Based Study of Turkish EFL Learners and Native English Speakers

Songül TAŞ
Kırşehir Ahi Evran Üniversitesi
songultas@ahievran.edu.tr
ORCID ID: 0000-0003-2590-6063

Araştırma Makalesi	DOI: 10.31592/aeusbed.1681606
Geliş Tarihi: 22.04.2025	Revize Tarihi: 21.10.2025
	Kabul Tarihi: 30.11.2025

Atıf Bilgisi

Taş, S. (2025). The use of English definite article “the” with generic reference: A comparative corpus-based study of Turkish EFL learners and native English speakers. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(3), 1009-1028.

ABSTRACT

The English article system, particularly the definite article *the*, continues to pose challenges for both native and non-native speakers of English. Although extensive research has focused on the non-generic uses of *the*, there remains a notable gap concerning its use with generic reference. Addressing this gap, the present study investigates the use of the definite article *the* with generic reference by analyzing and comparing its occurrence in blog writings of native English speakers and academic writings of Turkish EFL learners on an identical topic. A comparative corpus-based design was employed, and instances of *the* with generic reference were identified, categorized, and quantified using two concordance programs. Each occurrence of *the* was classified as generic or non-generic and compared quantitatively across groups. The results revealed that both groups employed *the* generically at very low frequencies. The findings were then discussed in relation to the pre-established research hypotheses. Finally, the study concludes by acknowledging its limitations and offering recommendations for future research, particularly the need to expand participant samples and incorporate more diverse genres and proficiency levels to enhance generalizability.

Keywords: comparative analysis, corpus linguistics, definite article, EFL learners, generic reference

İngilizce Belirli Tanımlık “the”nin Türsel Anlamda Kullanımı: Türk EFL Öğrencileri ve Anadili İngilizce Olan Konuşmacılar Üzerine Karşılaştırmalı Bir Derlem Çalışması

ÖZET

İngilizce tanımlık sistemi, özellikle belirli tanımlık *the*’nin doğru kullanımı, hem anadili İngilizce olan hem de yabancı dil olarak İngilizce öğrenen bireyler için önemli bir güçlük oluşturmaktadır. Önceki araştırmalar genellikle *the*’nin türsel olmayan (non-generic) kullanımlarına odaklanmış, ancak *the*’nin türsel (generic) anlamdaki kullanımı üzerine yapılan çalışmalar sınırlı kalmıştır. Bu çalışmada, belirli tanımlık *the*’nin türsel anlamdaki kullanımı, anadili İngilizce olan bireylerin blog yazıları ile İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerin aynı konuda yazdıkları akademik metinler karşılaştırılarak incelenmiştir. Karşılaştırmalı derlem temelli bir araştırma deseni kullanılmış; *the*’nin türsel anlamda kullanıldığı örnekler iki farklı eşdizim programı aracılığıyla belirlenmiş, kategorize edilmiş ve nicel olarak analiz edilmiştir. Her bir *the* kullanımı, türsel veya türsel olmayan olarak sınıflandırılmış ve gruplar arası nicel karşılaştırmalar yapılmıştır. Bulgular, her iki grubun da *the*’yi generic anlamda oldukça düşük sıklıkta kullandığını göstermiştir. Elde edilen sonuçlar, önceden belirlenen araştırma hipotezleri çerçevesinde tartışılmıştır. Son olarak, çalışmanın sınırlılıkları belirtilmiş ve gelecek araştırmalara yönelik olarak, örneklem çeşitliliğinin artırılması ve farklı türlerde yazıların dâhil edilmesi önerilmiştir.

Anahtar Kelimeler: karşılaştırmalı analiz, derlem dilbilimi, belirli tanımlık, EFL öğrencileri, türsel anlam

Introduction

The English article system represents one of the most complex and challenging aspects of language acquisition for both second and foreign language learners. This is mostly because of the dissimilarities between the article system of the learners’ mother tongue and English article system. Articles serve a central role in expressing definiteness and specificity in English, but their forms and meanings rarely align across languages. Consequently, learners often face persistent difficulties in selecting the appropriate article form, particularly the definite article *the*. Among the most frequently used free morphemes, articles have the first place in English which has three articles, namely, *the*, *a*,

and $\emptyset$ (zero). The articles “the” and “a” in the English language present the most basic interpretation for definiteness and indefiniteness (Marinas, 2011).

On the relevant issue, Bickerton (2016) indicated that NP falls into one of four types; generic, definite, indefinite, and others involving *a/an* and *zero article*. Although the articles in English language seem to occur in different contexts, they have overlapping distributions. According to Ojeda (1991), a definite noun phrase might be considered either as a definite description or as a definite generic. In that, the use of the article ‘the’ is commonly perceived as having two main categories as generic and non-generic uses (Celce-Murcia and Larsen- Freeman, 1999). Furthermore, it is used before plural and singular nouns when the noun is specific or particular (Marinas, 2011). Lyons (1977) classified the uses and expressions of generics in English as; definite singulars as in the example ‘*The bird is a warm-blooded animal*’, indefinite singulars as in the example ‘*A cat has nine lives*’, bare plurals as in the example ‘*Dinosaurs are extinct*’, or definite article plus adjective as in the example ‘*The elderly need better health care*’ (Gelman and Raman, 2003). The definite article *the* primarily precedes non-generic nouns and is commonly interpreted in a non-generic sense initially. However, this situation might change according to the context where *the* is used generically as such; ‘*The mobile phone is on the table. /The mobile phone is a means of communication*’, which is quite rare except in the scientific register (Liu and Gleason, 2002; Master, 1997). As illustrated in the examples above, the definite article conveys specificity and individuality in the first sentence, whereas it expresses a generic meaning in the second.

When *the* is used generically, it denotes all or majority of a group, such as a species, an ethnic group, or the people of a nation. It also tends to render the noun more formal or abstract, particularly when used with singular nouns (Radden, 2009; Yıldırım, 2015). In its generic context, *the* precedes a singular noun to represent an entire class of entities rather than individual members of that class (e.g., *The internet has affected our life in many ways*). Conversely, when utilized before a plural count noun, it conveys a sense of collective reference (e.g., *The Italians wear fashionably*.) (Master, 1997; Radden, 2009). All uses of the that fall outside the aforementioned generic contexts are considered non-generic, which is significantly broader and more frequently encountered in English (Liu and Gleason, 2002; Yıldırım, 2015). The debate proposed that the forms employed for generic reference closely resemble those used for non-generic or individuating reference. In generic usage, the typically precedes a singular noun (Alenizi, 2013), whereas non-generic usage is often marked by its occurrence before plural nouns (Radden, 2009). Liu and Gleason (2002) categorized the use definite article *the* with non-generic into four distinct types: cultural, situational, textual, and structural. This classification has since been widely referenced in subsequent research, as their taxonomy provides a foundational framework for the analysis of non-generic uses of the English definite article.

Every language possesses quite complex linguistic forms, which are difficult to understand or acquire by L2 learners. One example of this in English could be stated as definite article *the* (Dikilitaş and Altay, 2011) since the English article system is perceived as one of the most challenging structures to learn for both EFL and ESL learners. The main reason relying behind this complexity in English article system is both its frequency of use and its complexity (Alenizi, 2013; Garcia Mayo, 2008; Lee, Park and Heinz, 2018; Liu and Gleason, 2002; Master, 1997; Miller, 2005; Prior, Fujise and Fenwick, 2014; Thomas, 1989; Yıldırım, 2015). It is suggested by the existing research that all kinds of article use are not equally difficult for all English learners, and that the L1 of those learners has an important role in the difficulty they experienced while learning the article system (Lee, Park, and Heinz, 2018; Giles, 2019). Furthermore, studies comparing article choice of L2 learners from various L1 backgrounds revealed that the definite article is acquired earlier than indefinite article in L2 (Chrabaszcz and Jiang, 2014). Master (1997), one of the pioneering researchers to highlight the L1 impact on the acquisition of articles, suggests that full mastery of the English language article system is typically not achieved until the higher levels of interlanguage development. He further argues that even advanced learners of English may continue to have trouble with article usage.

Studies conducted on second language acquisition indicated well that non-native speakers of English commit errors in article use although they mastered on all other elements of the language since

they have difficulty in acquiring articles (Geng, 2010). It is more complicated when learners' L1 and English differ in expressing definiteness or when a specific article is not used for definiteness in their mother tongue (Alenizi, 2013; Koç, 2015). In this regard, learners whose first language (L1) possesses an article system like that of English tend to experience fewer difficulties in acquiring English articles compared to those whose L1 lacks such a system (Giles, 2019). The acquisition becomes particularly more challenging when learners' native languages do not include articles at all (Ekiert, 2004; Master, 1997; Miller, 2005; Thomas, 1989). Turkish is one of those languages, in that its native speakers mostly have difficulty in deciding when or which article is required to be used.

While definiteness is overtly marked through articles in English, in several other languages such as Japanese and Russian, it is typically conveyed through word order or contextual cues (Göksel and Kerslake, 2005). Turkish, as an agglutinative language which has a rich system of case morphology, also does not utilize separate articles to express definiteness. Instead, definiteness in Turkish is typically indicated through syntactic structures, the presence of indefinite determiners, case marking, tense-aspect-modality, stress, and determiners (Dede, 1986; Dikilitaş and Altay, 2011; Göksel and Kerslake, 2005; Koç, 2015; Yıldırım, 2015). Therefore, 'the' causes problems for Turkish learners of English as Turkish context is one of those EFL settings (Dede, 1986; Yıldırım, 2015). Since there are structural differences between the English and Turkish article systems, native Turkish speakers often encounter challenges in appropriately acquiring and using the English definite article (Koç, 2015; Yıldırım, 2015). Hence, this problematic issue needs to be addressed more in the studies to be able to comprehend the matter and search for remedies accordingly.

Recognizing the challenges faced by second and foreign language learners in acquiring and appropriately using the English definite article, and the scarcity of research focusing specifically on the use of the definite article *the* with generic reference, the current study seeks to contribute to the field by analyzing Turkish EFL learners' writings implementing a concordance program, namely AntConc 3.5.9, and analyzing the utilization of English definite article by comparing them with the existing written database of blog writings of natives of English on one of the most common and widely used concordance programs, namely COCA (Corpus of Contemporary American). In line with these, both Turkish EFL learners' and native English speakers' definite article use in their writings is classified and categorized based on whether it refers to generic or non-generic contexts, with particular emphasis on the primary usage. Existing literature of studies on the relevant topic in the field provided in the second chapter by considering their focus and findings.

Literature Review

The existing literature on the English article system, in addition to the EFL and ESL learners' use of articles, reviewed in terms of its chronological progression and the recurring themes identified throughout the studies. Early investigations in this field (e.g., Burton-Roberts, 1976; Lyons, 1977; Maratsos, 1974) established the semantic foundations of definiteness and genericity, identifying key distinctions between referential and non-referential uses. These foundational works demonstrated that the definite article *the* can encode both specific and non-specific meanings depending on the context, but they offered limited empirical evidence on how these forms are acquired by non-native speakers. Carlson (1982) also addressed the semantics of generic terms and sentences, proposing that their interpretation has implications for the structure of other grammatical components. His analysis offered a comprehensive account of generics while critically assessing alternative theoretical perspectives. Similarly, Mason and Pimm (1984) investigated the conceptual ambiguities surrounding generality and genericity, highlighting that the construction of a generic example involves both the emphasis and omission of certain salient features.

The 1990s and early 2000s marked a shift toward second language acquisition (SLA) research on article use. Highlighting the relative neglect of definite generics within both linguistic and philosophical discourse, Ojeda (1991) grounded his study in the influential theory of definite descriptions, offering a formally rigorous and empirically grounded semantic account of definite generics. Scholars such as Master (1997), Epstein (2002), and Liu and Gleason (2002) analyzed L2

learners’ use of *the* across different semantic categories, leading to taxonomies of article usage such as cultural, situational, textual, and structural reference. These frameworks have since become foundational in analyzing article errors and developmental sequences among learners from different L1 backgrounds. Master (1997), a key early figure in the field, aimed to link the acquisition and functional application of the English article system to a well-structured pedagogical sequence, spanning three proficiency levels: beginner, intermediate, and advanced. He argued that, due to the system’s inherent complexity and multi-faceted nature, English articles should be taught progressively over an extended period. Epstein (2002) introduced an alternative perspective on the issue of definiteness by incorporating referential interpretations within a discourse-based framework. He contended that the definite article fulfills various functions beyond simple uniqueness or familiarity, such as indicating discourse prominence, signaling role or value status, and marking shifts in perspective. In the same year, Liu and Gleason (2002) explored various non-generic uses of the definite article *the*, demonstrating that these uses correspond to differing levels of difficulty. Their findings revealed that learners’ accuracy in supplying all four identified non-generic uses improved significantly with increased proficiency. This study became foundational for subsequent research in the field, particularly due to the taxonomy it introduced for the English definite article use with non-generic reference.

Building on this line of inquiry, Gelman and Raman (2003) conducted a series of five studies investigating how preschool children utilize linguistic form classes and pragmatic cues to interpret generic expressions. Their results indicated that children begin distinguishing between generic and non-generic noun phrases at an early age, suggesting that young learners rely on multiple linguistic and conceptual indicators in acquiring an understanding of generic constructions. Snape (2008) conducted two experiments investigating how Japanese and Spanish ESL learners acquire the nominal domain, to determine whether they can reset the relevant parameter to align with the English system. The judgment task resulted that the Japanese participants tended to reject grammatical mass noun contexts, while simultaneously accepting ungrammatical mass plural constructions, suggesting difficulties in fully acquiring the target parameter settings. While the Spanish group exhibited similar behavior to the Japanese group in mass plural contexts, they generally encountered fewer difficulties with count nouns but struggled more with mass nouns. That said, Spanish speakers appeared to experience fewer issues with mass nouns compared to the Japanese learners. Ionin and Montrul (2010), on the other hand, worked on L1 interference and its influence on the interpretation of definite plurals by EFL learners. The study suggested that while L1 transfer has an impact on the interpretation of definite plurals in English, it is possible to recover from such a transfer. Another study with Spanish context was presented by Marinas (2011) focusing on the marker of definiteness to comprehend the processes which could have an impact on its occurrence in the EFL learners’ interlanguage. The results of three exercises carried out to analyze article production has revealed that a significant proportion of both intermediate and upper-intermediate students presented some examples of interference when the generic reference in the plural is concerned.

A growing body of corpus-based studies has explored cross-linguistic effects on article acquisition. For instance, Ekiert (2004) compared learners from article-based and article-less languages, showing that L1 transfer strongly affects the accuracy of article use, particularly in contexts requiring definiteness. Learners whose native languages lack articles, such as Turkish, Chinese, or Japanese, tend to underuse *the*, whereas learners from languages with definite markers (e.g., Spanish, Arabic) exhibit overuse. Chrabaszcz and Jiang (2014) also investigated the non-generic use of the English definite article, focusing on highly proficient L1 speakers of Spanish and Russian. Their study aimed to examine the influence of L1 on the use of the English non-generic definite article, as well as to test the difficulty hierarchy proposed by Liu and Gleason (2002). The results indicated that the native language had a significant role on the participants’ utilization of the definite article in non-generic reference. Furthermore, the two L1 groups displayed distinct patterns of difficulty depending on the specific linguistic context.

In the Turkish EFL context, several studies have focused on learners’ non-generic use of *the*. In a study involving Turkish EFL learners across varying proficiency levels, Dikilitaş and Altay

(2011) explored the non-generic use of the definite article *the* across four categories, utilizing a questionnaire which was developed by Liu and Gleason (2002). Their findings revealed that article use accuracy did not consistently improve with increasing proficiency. Similarly, Koç (2015) worked with Turkish EFL learners to investigate the use of *the* in non-generic contexts, comparing the performance of two groups of different level of proficiency students (50 intermediate and 50 low-advanced). The findings demonstrated that participants were more prone to omitting the definite article than overusing it across proficiency levels. The researcher observed that Turkish learners frequently omitted *the* in obligatory contexts, suggesting persistent interlanguage fossilization. In the same year, within the Turkish context, Yıldırım (2015) conducted an analysis of four non-generic uses of *the* by administering a grammaticality judgment test which was adapted from Liu and Gleason (2002) and found that even advanced Turkish learners struggled with contextual uses, particularly in cultural and structural categories. The results revealed that participants encountered difficulties in accurately using the non-generic *the*, particularly in contexts involving cultural reference. In a related study, Alenizi (2013) investigated the non-generic use of the English definite article among Arabic-speaking learners, employing a quantitative methodology that included both descriptive and statistical analyses. Unlike the findings in the Turkish context, this study identified proficiency as a significant factor influencing accurate article usage. Building on these investigations, Prior, Fujise, and Fenwick (2014) explored the connection between Japanese learners' use of the non-generic definite article and their cognitive style, specifically field dependence versus field independence. The results of a non-generic definite article test, completed by twenty-seven Japanese students studying in Canada, revealed that the situation-based use of the non-generic definite article was the easiest for ESL learners.

However, despite this substantial body of research, studies addressing the generic use of the definite article remain scarce. Radden (2009) and Barton, Kolb, and Kupisch (2015) emphasized that native speakers themselves rarely use *the* in generic contexts, favoring bare plurals or indefinite singulars. Radden (2009) examined article *the* with generic reference by analyzing various types of generic expressions and their underlying conceptual motivations. He highlighted four key factors that shape the meaning and usage of generic reference: the concepts of exclusiveness and inclusiveness, the metonymic relationship between instance and type, the metonymy of type for subtype, and the conceptual blending of instance and type. Expanding on this research, Geng (2010) carried out an experimental study with Chinese learners to analyze the misuse of the definite article in English from a semantic perspective. By evaluating learner performance on an article usage test and interpreting the results through semantic analysis, the study emphasized the importance of strengthening instruction on the generic use of English articles. Barton, Kolb, and Kupisch (2015) researched about the generic use of *the* by examining factors that influence article usage with plural subjects that have a generic reference. Their study considered both linguistic variables, such as kind-level versus individual-level predicates, and sociolinguistic variables, including age and regional background. Their findings indicated that while the use of definite articles is optional with generic plural subjects, bare plural forms are generally preferred.

Recent research has revisited these issues with a stronger corpus-linguistic and psycholinguistic orientation. Giles (2019), for instance, investigated the hierarchy of difficulty in the written use of the English non-generic definite article among L1 Azerbaijani speakers. The study built upon the categorical framework proposed by Liu and Gleason (2002). The examination of the written production of Azerbaijani learners of English a thesis project revealed that article overuse was most frequent in cultural and structural contexts, while generic reference remained underrepresented. The findings indicated that learners demonstrated equal levels of difficulty in omitting obligatory uses of the definite article within both the cultural and structural categories. However, the cultural category posed a significantly greater challenge in terms of article overuse, indicating that learners were more prone to incorrectly inserting the definite article in these contexts. Park (2021) analyzed Korean EFL learners' writing and observed that generic *the* was virtually absent across proficiency levels, reinforcing the assumption that it is rarely acquired through input alone. The challenge of acquisition of English article system compounds by the fact that article systems are not a common grammatical feature globally. A study by Yang (2024) on 126 university students in Korea found that learners "tended not to understand the English articles officially and clearly." Students were often only taught

about definite and indefinite articles and were "embarrassed that there were three articles instead of two," being unaware of the concept of the zero article.

A learner's L1 has a profound effect on their acquisition of English articles. This influence is evident both for speakers of languages without articles and for those whose native languages have different article systems. A study by Derkach and Alexopoulou (2023) analyzing a learner corpus confirmed the lower overall accuracy of learners with no-article L1s, specifically citing speakers of Chinese and Russian. Chan's research (2022) on 57 advanced Cantonese ESL learners in Hong Kong found that the participants had confusion about the structures of noun phrases for representing generic reference. Cantonese, which lacks articles, led learners to misuse the + plural or bare singular nouns for generic reference, while the definite generic (the dodo) was largely unnoticed. Even when an L1 has articles, differences in how they function create significant hurdles for learners. The study by Thanh (2022) notes there is no exact equivalence between Vietnamese and English generic sentences. Vietnamese has no single article equivalent to the, and its zero article is definite ([+definite]), whereas the English zero article is always indefinite ([+indefinite]). In her paper, Kovács (2023) also highlights that while both languages (Hungarian and English) have articles, a crucial difference lies in generic usage. English uses the zero article for a generic sense (e.g., "Milk is good for you"), whereas Hungarian uses the definite article in the same context ("A tej jó neked"). It is recognized that there is a scarcity of studies on the generic use of English definite article *the*. In that, most of the studies conducted on either the general aspects of English article system or the non-generic use of the definite article among learners and speakers of English. Based on this scarcity, this study aims to investigate on the generic use of English definite article *the* by text mining both the existing written database of natives' on their blog writings and the writings of Turkish EFL learners on similar topics of the previous data.

The current study particularly aims to analyze and categorize the patterns of generic reference realized through the definite article *the* in native English speakers' written discourse, utilizing text-mining techniques applied to a corpus of blog writings. In addition, it aims to investigate the use of the definite article *the* with generic reference in the writings of Turkish EFL learners on parallel topics, through corpus-based analysis aimed at identifying both frequency and structural tendencies. Finally, it aims to compare and contrast the generic uses of the definite article *the* between native English speakers and Turkish EFL learners, in order to reveal possible L1 transfer effects and to delineate the similarities and differences in article usage across the two groups.

By comparing both natives' and non-natives' uses of definite article *the*, this study attempts to demonstrate the differences or commonalities between the uses of *the* between the two groups and to see the effect of L1 interference in the use of definite article *the*. The current study aims to investigate on the use of English definite article *the* with generic reference and to reveal the L1 interference on the use of those articles in Turkish EFL learners' writings. Based on the ultimate purpose of the study abovementioned, three research questions and hypotheses generated before conducting the analysis were as follows:

RQ1: How do native English speakers use the definite article *the* with generic reference in their blog writings?

Hypothesis 1: Native English speakers consistently employ the definite article *the* with generic reference according to established grammatical norms, showing clear and systematic usage patterns across different contexts.

RQ2: How do Turkish EFL learners use the definite article *the* with generic reference in their English writings?

Hypothesis 2: Turkish EFL learners do not generally prefer to use English definite article *the* with generic reference for particularly the kind generic as there is not any separate word for definiteness in Turkish and generic meaning is mostly referred by plurals.

RQ3: What are the similarities and differences between native speakers and Turkish EFL learners in their use of the definite article *the* with generic reference, and to what extent can these differences be attributed to L1 (Turkish) interference?

Hypothesis 3: Use of English definite article *the* with generic reference is not common among both the non-native and the native speakers of English. Differences between native and Turkish EFL learners' use of the definite article *the* with generic reference are significantly influenced by L1 transfer effects.

Methodology

Research Model and Design

This study adopts a comparative corpus-based research design, which allows for the systematic comparison of linguistic features across different corpora representing distinct speaker groups. The comparative corpus approach has been widely used in second language acquisition and applied linguistics to uncover usage patterns and transfer effects (Biber, Conrad, and Reppen, 1998; Granger, 2015). The present study compares the use of the English definite article *the* with generic reference in two corpora: (1) blog posts written by native English speakers, and (2) academic essays written by Turkish EFL learners. The design facilitates both quantitative frequency analysis and qualitative interpretation of contextualized examples. This design employed because of the need to identify potential L1 influences and genre effects on the use of *the* in generic contexts. Since native and non-native speakers operate within different discourse conventions, a corpus-based comparison provides an empirically grounded way to examine how meaning and form interact in authentic language use.

Participants and Sampling

The study involved 35 Turkish EFL learners enrolled in a compulsory English preparatory program at a state university in Turkey during the 2020–2021 academic year. The participants were between 18 and 21 years old and had studied English for at least seven years in formal education settings. Their proficiency levels ranged from B1 to B2 according to the Common European Framework of Reference for Languages (CEFR). The native speaker corpus compiled from publicly available English blog posts collected from the *Corpus of Contemporary American English* (COCA). To ensure comparability, five blog entries written between 2012 and 2017 were selected. All blog posts centered on the topic of *board games*, which aligns with the Turkish learners' essay task. A key methodological consideration was the genre difference between the corpora. This discrepancy was identified as a limitation and explicitly discussed in the "Discussion" section. To mitigate its impact, both datasets were analyzed in terms of comparable semantic and grammatical criteria rather than stylistic ones.

Data Collection Procedure and Tools

The data for the present study was collected via using two different concordance programs, namely COCA (Corpus of Contemporary American English) and AntConc 3.5.9 (A freeware corpus analysis tool designed for concordancing and textual analysis.). While the existing database of native speakers' blog writings was included, 35 writings collected from the prep school students of a Turkish state university were uploaded and a database was formed in the second one. The writings of both groups were selected on similar topics, namely on board games.

The existing data from COCA was limited to five chosen blog sites on the board games between the years 2012 and 2017 to be able manage the analysis and see the exact match when compared to the students’ writings since they are also on the same topic and about the same number. The native speaker data consisted of blog entries downloaded from the COCA database. These entries were chosen based on topic similarity, recency, and readability to ensure comparability with learner data. The total word count across the five blog posts was approximately 9,000 words.

The Turkish learner corpus was constructed from the participants’ written assignments. Each student was asked to write a 100–120-word essay on the topic “*My Favorite Board Game*” as part of their midterm writing assessment. The instructions required the use of modals (*must, have to, mustn’t, don’t have to*) but placed no explicit focus on article usage, thus allowing natural language production. The essays were typed by the students and collected digitally through the university’s learning management system. All participants were informed about the study’s purpose, procedures, and the use of their writings for research purposes. Since COCA data is publicly available, no additional ethical clearance was required for the native speaker corpus. The collected texts were uploaded into AntConc 3.5.9, a corpus analysis and concordance tool developed by Laurence Anthony (2020). This software enables keyword searches, frequency counts, and contextual concordance analysis. A checklist was developed by the researcher to categorize each occurrence of the definite article *the* according to its reference type. The checklist, adapted from Liu and Gleason (2002) and refined with insights from Radden (2009), included two main categories - generic and non-generic -and subcategories for *kind-generic* and *delimited-generic* usage. A second expert who is a colleague from the same institution reviewed the instrument to ensure construct validity and inter-rater reliability. Percentage agreement between coders was calculated as 92%, indicating a high level of consistency.

Data Analysis

The data collected from both existing and created databases were analyzed via the checklists (see Appendix II and Appendix III) for both databases prepared by the researcher for the present study. First, each instance of *the* was identified and tagged for its reference type for quantitative frequency analysis. Frequency counts were calculated for both corpora and normalized per 1,000 words to allow direct comparison. Second, representative concordance lines were examined to determine whether the functioned as a generic or non-generic marker for the qualitative contextual analysis. Then, the generic uses were also divided into two sub-categories as generic *the* plus singular (kind generic) and generic *the* plus plural noun (delimited generic). With the help of this checklist, all the uses of definite article *the* could be listed, counted, and grouped according to the meanings they conveyed. Instances that were ambiguous or contextually dependent were cross-checked by both raters. As this study concentrates on the use of definite article *the* with generic reference, non-generic instances of definite article were counted but not divided and grouped into subcategories. Finally, the two corpora were compared in terms of distributional patterns of the usage. Special attention was given to occurrences of kind-generic (e.g., The tiger is a carnivore) and delimited-generic (e.g., The Italians are known for their cuisine). Statistical testing was not applied due to the small dataset size, as the study primarily aimed to provide descriptive insights rather than inferential generalizations. The findings thus represent exploratory evidence of article use tendencies rather than definitive cross-population comparisons.

Research Ethics

This study is a corpus based study, so ethical permission is not a requirement.

Findings

Four distinct types of generic reference were identified by Radden (2009) in English and categorized according to two scopes whether the article used is definite or indefinite, and whether the noun is singular or plural. Each type is assumed to convey a distinct form of generic meaning.

In the current study, both natives' and non-natives' use of generic reference with definite article *the* are discussed by text mining the two different databases and concordance programs.

Totally, 1,103 instances of the definite article *the* were identified across the two corpora, 577 occurrences in the native speaker corpus and 526 in the Turkish EFL learners' corpus. Each instance was categorized according to reference type. Generic uses accounted for only 3.6% of all tokens in the native corpus and 3.2% in the learner corpus, confirming the initial assumption that the generic *the* is infrequent in both native and non-native discourse. The generic occurrences were further divided into two subtypes:

- Kind-generic: definite article + singular noun (e.g., *The whale is a mammal.*)
- Delimited-generic: definite article + plural noun referring to a delimited group (e.g., *The Italians are creative.*)

In the native corpus, 10 kind-generic and 11 delimited-generic instances were identified while 15 kind-generic and only four delimited-generic instances occurred in the learner corpus. These results suggest that both groups use *the* with generic reference rarely and unevenly across subtypes. Further results are presented in detail under three subtitles below.

Use of Definite Article *the* with Generic Reference by Native Speakers

When the created corpus on COCA (Corpus of Contemporary American English) by involving five blogs considering their topics and dates was analyzed, it was found out that a total number of 577 definite article was used in those blog entries of native speakers. It was not surprising since the definite article *the* is one of the most used words of English with the number of 50066305 according to COCA (English-Corpora.org). All the 577 uses of definite article were listed and categorized first into two as generic and non-generic, then the generic uses of native speakers were categorized into two as the definite singular (kind generic) and definite plural (delimited generic). Some extractions from the completed checklist were provided in Table 1 below.

Table 1

Use of Definite Article the of the Native Speakers of English

GENERIC USE OF <i>THE</i>		NON-GENERIC USE OF <i>THE</i>
The + singular noun (kind generic)	The + plural noun (delimited generic)	-
1. ...still needing the cash....	1. The rules are guidelines.	1. ... through the whole process...
2. Whenever the opportunity presents itself.	2. The games are already playable.	2. ...from the maker of Middens...
3. The house feels good when it feels a little bit empty.	3. The descriptions are generalizations only.	3. ...the clever sorts behind Fallen...
4. The bride is in the goal of perfecting her figure for her wedding day.	4. ...paring down to the basics...	4. ...at the top of this article...
5. ...in the feminine hygiene products section of the grocery store....	5. ...any that can rob space from the books does not have to...	5. ...links in the comments is...

Total number of 577 definite article use of *the* in 5 blogs by native speakers

The results of the analysis revealed that a number of 21 generic use of definite article was detected in the blog entries of native speakers of English. In that, the number of kind generic use (n=10) was counted as almost equal to the number of delimited generic use (n=11), whereas the uses of delimited generic were not denoted to the exact and correct use of definite article with plural generic since none of them were used for humans. For instance, in the example '*The Turks are hospitable people.*' '*The Turks*' refer to the all the group of people who are related to the country of Turkey. Whereas, in the example depicted in the Table 1 above '*The descriptions are generalizations*' does not refer to delimited generic since '*the descriptions*' do not denote to humans. In that, the delimited generic form in English is limited to human referents and non-human subjects with definite plural are not considered with a generic sense, instead they describe individuated instances (Radden,

2009). Therefore, although the uses of definite article with plural nouns in a sense of generic were listed in Table 1 above, they do not convey the meaning with generic reference exactly since they are not denoted to the humans.

These findings confirm that native speakers seldom employ *the* generically, preferring bare plural or indefinite forms to express generality. This aligns with Radden’s (2009) argument that the definite article is semantically constrained in generic contexts, especially outside academic registers. About this limitation of delimited generic, the delimited generic is sometimes excluded from certain analyses of English generics, as noted by Radden (2009) because of its constraints. The influence of those constraints could be observed in the use of delimited generic of natives in this study since there is no such use with humans in all the texts and uses of definite article examined. The case presented not much difference with the kind generic use of natives since only ten examples were detected among a total of 577 uses of definite article *the*. Five of those usages were involved in Table 1 above and the whole table of uses of definite article *the* was shared in the Appendix II. As could be seen from the examples in Table 1 above, the definite singular often denotes a specific type or category in generic reference. To exemplify, ‘*the bride, the house, the cash, the opportunity, and the grocery*’ used in those contexts lead to the hearer establishing mental contact to the type a single instance with its inclusive totality (Radden, 2009). The limited presence of generic *the* even in native writing underscores its rarity as a productive grammatical form in English. This finding reinforces the notion that generic reference is more often expressed through bare plurals (e.g., *Dogs are loyal*) or indefinite singulars (e.g., *A dog is a loyal animal*). As it could be clearly deduced from the findings above, the hypothesis one which anticipates that the limited number of generic uses of definite article *the* exists even in the writings of native speakers of English was supported by the data examined.

Use of definite article *the* with generic reference by Turkish EFL learners

Turkish EFL learners displayed a similar pattern but with slightly different tendencies. When the writings of EFL learners’ (n=35) were included in the concordance program and mined for the definite article *the* uses, it was found out that a total number of 526 uses of English definite article occurred in the students’ writings on board games. Among those 526 uses, only 19 were detected as uses in generic reference, most of which (n= 15) were pointed as kind generic while only four were delimited generic. In the Table 2 below, some examples of all three kinds of definite article uses were listed from the uses of Turkish EFL learners.

Table 2
Use of definite article *the* of Turkish EFL learners’

GENERIC USE OF <i>THE</i>		NON-GENERIC USE OF <i>THE</i>
The + singular noun (Kind generic)	The + plural noun (Delimited generic)	-
1. The chessboard consists of black and white	1. The elephants always have to...	1. the adjacent squares
2. the backgammon game	2. ...the international Olympics to contest.	2. the aim is to
3. the ludo has six phase	3. The rules	3. the aim is to
4. The chess player has to	4. the chess pieces.	4. the aim of the
5. the aristocracy long		5. the aim of the

Total number of 526 definite article use *the* in 35 writings of Turkish EFL learners’

The examples presented for kind generic uses like; ‘*The chessboard consists of black and white*’ demonstrate that although ‘*the chessboard*’ is used by definite article *the* denotes to a generic reference since it does not refer to one specific chessboard, instead it refers to all kinds without any exception. Similarly, ‘*the aristocracy*’ is also used with generic reference by the Turkish EFL learners denoting to all the members of the group by using the definite article and a singular (non-count) noun. This kind of usage was interesting for the Turkish learners since the generic meaning of this word in Turkish is conveyed by using plural as ‘*asiller, soylular, etc.*’ Hence, this was one of the unforeseeable uses of definite article by Turkish EFL learners.

The first example of delimited generic uses demonstrates an example of reinforcement of generic reference by using some lexical devices as frequency adverbs (Barton, Kolb, and Kupisch, 2015). In that, ‘*the elephants*’ referring to a group of elephants are reinforced by the frequency adverb ‘*always*’ to establish genericity.

For the kind generic use of definite article, Turkish students mostly preferred to use definite article *the* when they intended to use genericity for the names of the games such as ‘*the backgammon*’ and ‘*the chess*’ referring to a delimited set within a given pragmatic context (Radden, 2009), as they were demonstrated above in the Table 2. Other than the names of the games, they used definite article with the tools or players of the games by allowing exceptions. The findings from the text mining of Turkish EFL learners’ writings supported the hypothesis II to some extent since it was found that the number of definite article *the* use with generic reference is very limited (n=19) among Turkish EFL learners. However, since the number of kind generic uses (n=15) is more than the number of uses of definite article *the* with delimited generic reference (n=4), reverse of which was anticipated before in the hypothesis II, the hypothesis two was not completely approved by the findings of the analysis.

Comparison of Use of Definite Article *the* with Generic Reference by Natives and Turkish EFL Learners

When compared quantitatively, both groups demonstrate similar proportions of generic *the*, confirming Hypothesis 1 that generic use of *the* is uncommon among both native and non-native speakers. However, the distributional difference between kind-generic and delimited-generic forms supports Hypothesis 2 only partially. Turkish learners indeed produced more kind-generic forms, but this pattern likely stems from L1 transfer effects from Turkish, where generality is conveyed via singular bare nouns or plural forms without determiners (Göksel and Kerslake, 2005). This might be because of the constraints of using definite article *the* for especially delimited generic suggested by Radden (2009). In that, both groups prefer other tools to convey meaning of genericity such as indefinite article or bare plural, which was presumed before for the Turkish learners of English because of the article system in English is not similar to the features used in Turkish for conveying the meaning of definiteness or indefiniteness.

The limited use of delimited-generic *the* in the learner corpus suggests that Turkish learners have difficulty associating the article with collective human referents. This is consistent with findings from Koç (2015) and Yıldırım (2015), who observed that Turkish EFL learners underuse *the* in both generic and non-generic human-referential contexts. Furthermore, the low overall frequency in both corpora confirms earlier reports (Barton, Kolb, and Kupisch, 2015; Giles, 2019) that native speakers themselves rarely employ *the* generically, restricting its use to formal or academic genres. Hence, the observed learner performance might reflect not only L1 transfer but also limited input frequency, as Turkish learners are rarely exposed to such structures even in authentic materials.

From a theoretical perspective, these findings support to Usage-Based and Cognitive Grammar frameworks (Langacker, 1991; Krifka, 2012), which propose that grammatical knowledge develops through exposure frequency and conceptual salience. The similarity between native and learner corpora also aligns with Radden’s (2009) claim that the semantic scope of generic *the* is inherently restricted. Therefore, even advanced learners may produce native-like frequencies not because they have acquired the form, but because its use is naturally rare. In addition, the findings highlight a crucial genre effect. The native corpus consisted of informal blogs, while the learner corpus consisted of semi-formal essays. As Barton, Kolb, and Kupisch (2015) argue, generic expressions are more common in scientific and expository genres than in personal or narrative writing. Thus, the contextual and stylistic nature of both corpora likely contributed to the scarcity of generic *the* instances.

To sum up, the results indicate that both native and Turkish EFL writers rarely use the definite article *the* in generic contexts. In that, kind-generic forms are slightly more common than delimited-generic forms in both groups. In addition, Turkish learners’ preference for kind-generic forms may

reflect L1 transfer from Turkish, which encodes genericity differently. Finally, the rarity of generic *the* in both corpora suggests that limited exposure, genre effects, and semantic constraints influence the acquisition of the structure jointly.

Conclusion, Discussion and Recommendation

The article system -especially the definite article *the*- in English has a great importance since the articles are the most commonly used semantic units of the language. However, it carries upon considerable difficulties and challenges for both natives and learners of English language, particularly for the ones whose L1 has a different article system than English such as Turkish, Japanese, Chinese, Arabic, Spanish, and Russian (Alenizi, 2013; Chrabaszcz and Jiang, 2014; Geng, 2010; Marinas, 2011; Prior, Fujise, and Fenwick, 2014; Yıldırım, 2015). Geng (2010) clarified that though the generic use of articles present significant challenges for learners, it has often been overlooked in instructional texts. Recognizing this importance and the lack of studies on the field especially on the use of definite article *the* with generic reference, the present study set out to examine the use of the English definite article *the* with generic reference among Turkish EFL learners compared with native English speakers. Adopting a comparative corpus-based approach, the analysis focused on two comparable datasets. The findings confirmed that both groups rarely employed *the* generically, with only minimal variation across subtypes and contexts. Low frequency of generic *the* was the leading result of the current study. It was revealed that the generic use of *the* is exceptionally rare for both native and non-native speakers. This supports previous research (Barton, Kolb, and Kupisch, 2015; Giles, 2019; Radden, 2009) suggesting that definite singular generics occur infrequently in English discourse.

Another finding was about the preference for kind-generic forms rather than the delimited-generic forms. In that, Turkish EFL learners showed a clear preference for kind-generic structures (*The tiger is...*), while delimited-generic forms (*The Italians are...*) were nearly absent. This may be because of the L1 transfer effects from Turkish, which expresses genericity through unmarked plurals rather than determiners. Influence of genre and input is another crucial finding of the current study on the frequency of generic *the* usage. The rarity of generic *the* in both corpora reflects the influence of genre and limited input frequency. As noted in previous corpus research (Park, 2021; Chen, 2023), generic constructions appear more frequently in scientific or expository writing than in personal or narrative genres like blogs or short essays. About the L1 transfer effect on the use of generic *the*, the findings revealed that though some traces of Turkish structure were observed, the similarity between native and learner frequencies indicates that the difficulty arises less from L1 interference and more from the inherent rarity of generic *the* in English.

These results supports the view that grammatical acquisition is shaped by input frequency and discourse function, as proposed in usage-based and cognitive linguistic models (Krifka, 2012; Langacker, 1991). Because generic *the* occurs rarely in natural input, learners are not able to acquire it automatically without explicit focus. Moreover, the findings support Radden’s (2009) claim that generic *the* occupies a peripheral position within the article system, semantically marked, context-dependent, and pragmatically constrained. From a pedagogical standpoint, the results highlight several key considerations. First, explicit instruction on the semantic distinction between generic and non-generic reference is necessary since standard grammar materials rarely address this difference. Next, corpus-based learning activities can help EFL learners observe authentic usage patterns of *the*, fostering greater form–meaning awareness (Chen, 2023). Moreover, genre-sensitive writing tasks such as argumentative essays or scientific summaries should be integrated into instruction to provide contexts where generic reference naturally arises. Lastly, teachers should emphasize that correct article choice depends not only on grammar but also on discourse function, register, and communicative purpose.

Although this study attempted to address an important gap in the field and provided some crucial findings for further practices and research, it had some limitations that need to be considered. First, the methodological limitations of this study primarily concern the corpus scope and genre variation. Although efforts were made to align topics and text lengths, the native speakers’ blogs and

learners' essays differ in communicative intent and stylistic conventions. This difference may influence article use frequency and function. Although this study attempted to control for topic similarity, such differences may influence linguistic choices, including article usage. Consequently, the results should be interpreted as illustrative rather than generalizable. Additionally, the relatively small sample size ($n = 35$) limits the generalizability of the findings. Nevertheless, the results provide a valuable case study illustrating how article usage patterns emerge across distinct linguistic and cultural backgrounds. However, the study's strength lies in its focus on a neglected area -the generic use of *the*- and its combination of quantitative and qualitative corpus techniques. By combining quantitative and qualitative corpus techniques, this study offers an example of exploratory model for analyzing low-frequency grammatical phenomena like generic *the*. Future research can expand the dataset to include other genres like academic essays by native speakers and employ statistical testing to validate observed patterns.

Considering the limitations mentioned above, it is suggested for further research to employ mixed-methods approaches, combining corpus analysis with judgment tasks or interviews to examine learner awareness. In that, other instruments like an interview or questionnaire could support the data with the participants. Besides, a grammaticality test on articles might be implied such as AJT (acceptability judgment task), Truth-Value Judgment Task (TVJT) which was used in the study of Ionin and Montrul (2010) to get more concrete results on the differences between native and Turkish EFL learners' use of definite article *the* with generic reference. Furthermore, future research may explore pedagogical interventions, testing whether explicit instruction or data-driven learning improves accuracy in generic reference. Lastly, future studies might investigate cross-linguistic variation, comparing learners from other article-less languages (e.g., Korean, Japanese, Russian) to identify universal versus language-specific trends. To sum up, this study emphasizes the need for targeted reinforcement of generic article usage in English, based on its findings. In conclusion, this study contributes to the growing body of research on English article acquisition by addressing an underexplored area, the generic use of the definite article *the*. Although limited in scope, the results demonstrate that both native and non-native speakers rarely use generic *the*, reflecting its scanty role in English discourse. For Turkish EFL learners, the difficulty lies not only in L1 transfer but also in the inherent infrequency and semantic obscurity of the structure. By combining corpus-based description with pedagogical reflection, this research bridges linguistic theory and classroom practice. The findings underscore the importance of raising learners' metalinguistic awareness and incorporating authentic data into instruction to promote deeper understanding of how English articles function beyond surface grammar.

Author Contributions

The author contributed 100% to this article.

Interest of Conflict

There is no conflict of interest in this article.

Bibliography

- Alenizi, A. (2013). *The Acquisition of the Nongeneric Uses of the English Definite Article the by Arabic Speakers of English*. Yüksek lisans tezi. Southern Illinois University Carbondale, Illinois. ProQuest Dissertations and Theses veri tabanından erişildi. (UMI No. 1541913).
- Anthony, L. (2020). *AntConc* (Version 3.5.9) [Computer software]. Waseda University. Available from <https://www.laurenceanthony.net/software>
- Barton, D., Kolb, N., and Kupisch, T. (2015). Definite article use with generic reference in German: An empirical study. *Zeitschrift für Sprachwissenschaft*, 34(2), 147–173.

- Biber, D., Conrad, S., and Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Bickerton, D. (2016). *Roots of language*. Berlin: Language Science Press.
- Burton-Roberts, N. (1976). On the generic indefinite article. *Language*, 52, 427–448.
- Butt, J., and Benjamin, C. (1988). *A new reference grammar of modern Spanish*. London: Arnold.
- Can, C. (2017). A learner corpus-based study on verb errors of Turkish EFL learners. *Journal of Education and Training Studies*, 5(9), 167-175.
- Carlson, G. (1982). Generic Terms and Generic Sentences. *Journal of Philosophical Logic*, 11(1), 145-181.
- Celce-Murcia, M., and Larsen-Freeman, D. (1999). *The grammar book: An ESL teacher's course* (2nd Ed). Boston: Heinle and Heinle.
- Chan, A. Y. W. (2023). Hong Kong Cantonese ESL Learners' Interpretation of the Use of Articles for Generic Reference: a Bilingual Perspective. *English Teaching and Learning*, 47(4), 469-492.
- Chrabaszcz, A., and Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research*, 30(3), 351–379.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. Retrieved from: <https://www.english-corpora.org/coca/>.
- Dede, M. (1986). *Studies in Turkish Linguistics*. Amsterdam: John Benjamins.
- Derkach, K., and Alexopoulou, T. (2024). Definite and indefinite article accuracy in learner English: A multifactorial analysis. *Studies in Second Language Acquisition*, 46(3), 710-740.
- Dikilitaş, K., and Altay, M. (2011). Acquisition sequence of four categories of non-generic use of the English definite article the by Turkish speakers. *Novitas-ROYAL*, 5(2), 183–198.
- Ekiert, M. (2004). Acquisition of the English article system by speakers of Polish in ESL and EFL settings. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics*, 4, 1-23.
- English-Corpora.org. (n.d.). *Corpus of Contemporary American English*. Retrieved November 25, 2020, from <https://www.english-corpora.org/corpora.asp>
- Epstein, R. (2002). The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics*, 12(4), 333–378.
- Erguvanlı, E.E. (1984). *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Garcia Mayo, M. D. P. (2008). The acquisition of four non-generic uses of the article *the* by Spanish EFL learners. *System*, 36, 550-565. doi: 10.1016/j.system.2008.08.001
- Gelman, S., and Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, 74(1), 308–325.

- Geng, J. (2010). The semantic analysis of the definite article misuse by Chinese learners of English. *Asian Social Science*, 6(7), 180–184.
- Giles, D. (2019). *The L2 English non-generic definite article use of L1 Azerbaijani speakers*. Yüksek lisans tezi. St. Cloud State University, St. Cloud, Minnesota.
- Göksel, A. and Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7–24. doi: 10.1075/ijlcr.1.1.01gra
- Ionin, T., and Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60(4), 877–925.
- Koç, D. K. (2015). The non-generic use of the definite article the in writing by Turkish learners of English. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(2), 56-68.
- Kovács, É. (2023). On the use of articles in English and in Hungarian. *Acta Academiae Beregsasiensis, Philologica*, 2(2), 81-91.
- Krifka, M. (2012). Definitional generics. In *Genericity*, (pp. 372-389).
- Langacker, R. W. (1991). *Foundations of cognitive grammar: Vol. II. Descriptive application*. Stanford CA: Stanford University Press.
- Lee, J., Park, S., and Heinz, M. (2018). Exploring patterns of article use by advanced Korean learners of English and Spanish. *International Review of Applied Linguistics in Language Teaching*, 41(1), 79-101.
- Liu, D., and Gleason, J. (2002). Acquisition of the Article the by Non-native Speakers of English: An Analysis of Four Non-generic Uses. *Studies in Second Language Acquisition*, 24(1), 1-26.
- Lyons, J. (1977). *Semantics*: Vol. 1. New York: Cambridge University Press.
- Maratsos, M. (1974). Preschool Children's Use of Definite and Indefinite Articles. *Child Development*, 45(2), 446-455.
- Marinas, L. T. (2011). Definite Article use in the L1 of Spanish Speakers: A multi-dimensional problem. *Miscelánea: A Journal of English and American studies*, 43, 87-105.
- Mason, J., and Pimm, D. (1984). Generic Examples: Seeing the General in the Particular. *Educational Studies in Mathematics*, 15, 277-289.
- Master, P. (1997). The English Article System: Acquisition, Function, and Pedagogy. *System*, 25(2), 215-232.
- Miller, J. (2005). Most ESL students have trouble with articles. *International Education Journal*, 5, 80-88.
- Nickel, B. (2005). On semantics for characterizing sentences. Retrieved from <https://philpapers.org/rec/NICOSF>.
- Ojeda, A. (1991). Definite descriptions and definite generics. *Linguistics and Philosophy*, 14, 367-397.

- Park, S. (2021). The necessity of minimal representation of genericity in a newly developed language, Unish. *Journal of Universal Language*, 22(2), 87-104.
doi: 10.22425/jul.2021.22.2.87
- Prior, S., Fujise, K., and Fenwick, K. (2014). Acquisition of the non-generic definite article in English: The influence of cognitive style. *International Journal of Linguistics*, 6(3), 217-230.
- Radden, G. (2009). Generic reference in English: A metonymic and conceptual blending analysis. In *Metonymy and metaphor in grammar* (pp. 199-228). John Benjamins Publishing Company.
- Sinclair, J. M. (Ed.) (1991) *Corpus, concordance, collocation*. Oxford University Press, Oxford.
- Snape, N. (2008). Resetting the nominal mapping parameter in L2 English: definite article use and the count-mass distinction. *Bilingualism: Language and Cognition*, 11(1), 63-79.
- Taylor, J. R. (1995). *Linguistic categorization: Prototypes in linguistic theory*. In *Cognitive grammar* Oxford: Oxford University Press.
- Thanh, T. M. (2022). Generic sentences in Vietnamese and English: Similarities and Differences. *International Journal of Language and Linguistics*, 10(2), 166-175.
doi: 10.11648/j.ijll.20221002.25
- Thomas, M. (1989). The acquisition of English articles by first-and second-language learners. *Applied psycholinguistics*, 10(3), 335-355. doi: 10.1017/S0142716400008663
- Yang, Y. J. (2024). A study on the understanding of English articles among university students in an EFL. *The Journal of Mirae English Language and Literature*, 29(1), 223-243. doi: 10.46449/MJELL.2024.02.29.1.223
- Yıldırım, B. (2015). Turkish ELT students' acquisition of the definite article: An analysis of four non-generic uses. *Procedia - Social and Behavioral Sciences*, 748-755.

Genişletilmiş Özet

Giriş

İngilizce tanımlık sistemi hem ikinci dil hem de yabancı dil öğrenen bireyler için dil ediniminin en karmaşık ve zorlayıcı alanlarından birini temsil etmektedir. Bu durumun temel nedeni, öğrencilerin anadilindeki tanımlık sistemi ile İngilizce tanımlık sistemi arasındaki yapısal farklılıklardır. İngilizcede yer alan "the", "a" ve Ø (sıfır tanımlık) olmak üzere üç adet tanımlık, dilde en sık kullanılan serbest morfemler arasında ilk sırada yer almaktadır. "The" ve "a" tanımlıkları, belirli ve belirsiz anlamları ifade etme konusunda en temel araçlardır (Marinas, 2011). İkinci dil edinimi üzerine yapılan çalışmalar, anadili İngilizce olmayan bireylerin, dilin diğer öğelerinde yüksek yeterlilik sergilemelerine rağmen tanımlık kullanımında hatalar yaptıklarını ortaya koymuştur. Bunun nedeni, tanımlık sisteminin ediniminin görece daha zor olmasıdır (Geng, 2010). Özellikle öğrencinin anadili ile İngilizce arasında belirgin bir fark bulunuyorsa ya da anadilinde belirli anlamı ifade eden ayrı bir tanımlık bulunmuyorsa, bu durum daha da karmaşık hale gelmektedir (Alenizi, 2013; Koç, 2015). Anadili İngilizce gibi bir tanımlık sistemine sahip olan bireyler, İngilizce tanımlıkları edinmede daha az zorluk yaşarken; anadilinde tanımlık bulunmayan bireyler, İngilizce tanımlıkları kullanma konusunda ciddi zorluklarla karşılaşmaktadır (Giles, 2019). Bu bağlamda, Türkçenin tanımlık sisteminden yoksun olması, Türk öğrencilerin "the" kullanımında karar verme süreçlerinde sıkıntı yaşamalarına neden olmaktadır (Ekiert, 2004; Master, 1997; Miller, 2005; Thomas, 1989). İkinci ve yabancı dil olarak İngilizce öğrenen bireylerin İngilizce tanımlık sistemini özellikle de belirli tanımlık olan "the"nin doğru kullanımıyla ilgili yaşadığı güçlükler göz önüne alındığında ve "the"nin cinsel/generik referansla kullanımı üzerine yapılan çalışmaların sınırlılığı dikkate alındığında, bu çalışma, alana katkı sağlamayı amaçlamaktadır. Bu doğrultuda, Türk EFL öğrencilerinin yazılı metinleri, AntConc 3.5.9 adlı eşdizim analiz programı kullanılarak analiz edilmiş ve bu veriler, İngilizce anadil konuşurlarının blog yazılarından oluşan ve en yaygın derlem araçlarından biri olan COCA (Corpus of Contemporary American English) veri tabanı ile karşılaştırılmıştır. Hem Türk EFL öğrencilerinin hem de İngilizce anadil konuşurlarının yazılı ürünlerinde yer alan belirli tanımlık kullanımları, cinsel (generik) ve cinsel olmayan (non-generik) bağlamlara göre sınıflandırılmış ve özellikle generik kullanım üzerine odaklanılmıştır.

Alan yazında "the"nin generik anlamla kullanımına yönelik araştırmaların sınırlı olması, çalışmanın önemini artırmaktadır. Nitekim önceki araştırmalar çoğunlukla genel tanımlık sistemi ya da "the"nin generik olmayan kullanımlarına odaklanmıştır. Bu çalışma ise, hem İngilizce anadil konuşurlarının blog yazılarındaki mevcut yazılı veri tabanlarını hem de Türk EFL öğrencilerinin benzer konular üzerine yazılmış metinlerini metin madenciliği yoluyla analiz ederek, "the"nin generik kullanımlarını ortaya koymayı ve iki grubun kullanım farklılıklarını ya da benzerliklerini belirlemeyi amaçlamaktadır. Ayrıca, birinci dilin (L1) etkisinin bu kullanımlar üzerindeki rolü de incelenmiştir. Bu çalışma, anadili İngilizce olan bireyler ile İngilizceyi yabancı dil olarak öğrenen bireylerin belirli tanımlık *the*'yi kullanım biçimlerini karşılaştırarak, iki grup arasındaki benzerlikleri ya da farklılıkları ortaya koymayı ve Türkçenin birinci dil (L1) etkisinin *the* tanımlığının kullanımına olan yansımalarını incelemeyi amaçlamaktadır. Mevcut çalışma, İngilizce belirli tanımlık *the*'nin türsel (generik) anlamda kullanımını araştırmayı ve Türk EFL öğrencilerinin yazılarında bu tanımlıkların kullanımında görülen birinci dil etkisini ortaya koymayı hedeflemektedir. Yukarıda belirtilen genel amaç doğrultusunda, analizden önce oluşturulan üç araştırma sorusu ve hipotez aşağıdaki gibidir:

AS1: Anadili İngilizce olan konuşmacılar, blog yazılarında belirli tanımlık *the*'yi türsel anlamda nasıl kullanmaktadır?

Hipotez 1: Anadili İngilizce olan konuşmacılar, belirli tanımlık *the*'yi türsel anlamda yerleşik dilbilgisel normlara uygun biçimde, farklı bağlamlarda açık ve sistematik kullanım örüntüleri sergileyerek kullanmaktadır.

AS2: Türk EFL öğrencileri, İngilizce yazılarında belirli tanımlık *the*’yi türsel anlamda nasıl kullanmaktadır?

Hipotez 2: Türk EFL öğrencileri, özellikle *kind-generic* türünde olmak üzere, İngilizce belirli tanımlık *the*’yi türsel anlamda kullanmayı genellikle tercih etmemektedir. Bunun nedeni, Türkçede belirli anlamı karşılayan ayrı bir sözcüğün bulunmaması ve türsel anlamların çoğunlukla çoğul yapılarla ifade edilmesidir.

AS3: Anadili İngilizce olan konuşmacılar ile Türk EFL öğrencilerinin belirli tanımlık *the*’yi türsel anlamda kullanımlarında hangi benzerlikler ve farklılıklar bulunmaktadır ve bu farklılıklar hangi ölçüde birinci dil (Türkçe) etkisine bağlanabilir?

Hipotez 3: İngilizce belirli tanımlık *the*’nin türsel anlamda kullanımı, hem anadili İngilizce olanlar hem de olmayanlar arasında yaygın değildir. Anadili İngilizce olan konuşmacılar ile Türk EFL öğrencilerinin *the*’nin türsel kullanımı konusundaki farklılıkları, büyük ölçüde birinci dil aktarımından kaynaklanmaktadır.

Yöntem

Bu çalışma, farklı konuşur gruplarını temsil eden iki ayrı derlem arasındaki dilsel özelliklerin sistematik biçimde karşılaştırılmasına olanak tanıyan karşılaştırmalı derlem temelli bir araştırma deseni benimsemektedir. Karşılaştırmalı derlem yaklaşımı, ikinci dil edinimi ve uygulamalı dilbilim alanlarında, kullanım örüntülerini ve dil aktarımı etkilerini ortaya çıkarmak amacıyla yaygın biçimde kullanılmaktadır (Biber, Conrad ve Reppen, 1998; Granger, 2015). Çalışma kapsamında, İngilizce anadil konuşurlarının mevcut derlem verileri ile Türk EFL öğrencilerinin yazılarından oluşturulan derlem verileri karşılaştırılmıştır. Veriler, iki farklı eşdizim programı kullanılarak toplanmıştır: COCA (Corpus of Contemporary American English) ve AntConc 3.5.9 (ücretsiz bir metin çözümleme ve eş dizim analiz aracıdır). İngilizce anadil konuşurlarına ait blog yazıları COCA’dan alınmış, Türkiye’deki bir devlet üniversitesinin hazırlık okuluna devam eden öğrencilerden toplanan 35 yazı ise AntConc’a yüklenerek veri tabanı oluşturulmuştur. Her iki grup için de yazı konuları benzer şekilde belirlenmiştir: kutu/masa oyunları. COCA’dan alınan veriler, 2012–2017 yılları arasında yayımlanmış ve “board games” konulu beş blog sitesiyle sınırlandırılmıştır. Böylece, öğrencilerin yazılarıyla daha uyumlu ve doğrudan karşılaştırılabilir veriler elde edilmiştir. Toplanan veriler, araştırmacı tarafından geliştirilen kontrol listeleri (Ek II ve Ek III) yardımıyla analiz edilmiştir. Bu listelerde, tüm “the” kullanımları önce generik ve non-generik olarak ikiye ayrılmış; ardından generik kullanımlar kendi içinde “kind generic” (tekil isimle kullanılan) ve “delimited generic” (çoğul isimle kullanılan) olmak üzere iki alt kategoriye ayrılmıştır. Böylece, “the”nin tüm kullanımları sayılmış, anlamlarına göre gruplandırılmış ve listelenmiştir. Bu çalışmanın odak noktası generik kullanımlar olduğu için, non-generik kullanımlar yalnızca sayılmış, alt kategorilere ayrılmamıştır. Bu desen, *the*’nin türsel bağlamlardaki kullanımında ortaya çıkabilecek birinci dil (L1) etkilerini ve tür (genre) farklılıklarının olası etkilerini belirleme gereksinimi nedeniyle tercih edilmiştir. Anadili İngilizce olanlar ve olmayanlar farklı söylem gelenekleri içinde üretim yaptıklarından, derlem temelli bir karşılaştırma, anlam ve biçim arasındaki etkileşimi özgün dil kullanımı verileri üzerinden deneysel inceleme olanağı sunmaktadır.

Bulgular

Çalışmada, iki farklı eş dizim veri tabanı analiz edilerek elde edilen sonuçlar üç başlık altında sunulmuştur.

Anadili İngilizce Olan Konuşmacılarda “the” nin Generik Kullanımı

Birinci hipotez doğrultusunda beklenildiği üzere, analiz sonuçları İngilizce anadil konuşurlarının blog yazılarında yalnızca 21 adet generik *the* kullanımı tespit edildiğini göstermiştir.

Bu kullanımların 10'u "kind generic", 11'i ise "delimited generic" kategorisinde yer almıştır. Ancak, "delimited generic" kullanımların hiçbiri insanlara yönelik değildi. Bu da söz konusu kullanımların tam anlamıyla doğru bir generik tanımlık kullanımı olmadığını göstermektedir. "Kind generic" kullanımlar ise toplamda sadece 10 örnekle sınırlı kalmıştır. 577 adet "the" kullanımı arasında bu kadar az generik örnek bulunması, hipotez 1'i desteklemiştir.

Türk EFL Öğrencilerinde "the"nin Generik Kullanımı

Türk EFL öğrencilerinin yazılarında yapılan metin madenciliği sonuçları hipotez 2'yi kısmen desteklemiştir. Generik anlamda kullanılan "the" sayısı oldukça sınırlı kalmıştır (n=19). Bununla birlikte, bu kullanımların 15'i "kind generic", sadece dördü "delimited generic" olarak sınıflandırılmıştır. Hipotez 2'de beklenti, "delimited generic" kullanımın daha yaygın olacağı yönündeydi. Bu durum gerçekleşmediği için hipotez yalnızca kısmen doğrulanmıştır.

İki Grup Arasında Karşılaştırma

Her iki grubun generik "the" kullanımları, hipotez 3'ü destekler şekilde oldukça düşük çıkmıştır. Özellikle Türk öğrencilerde "delimited generic" kullanımın beklenenden az olması, Türkçede genellik anlamının çoğunlukla yalın çoğul yapılarla verilmesine bağlanabilir. Bu nedenle, L1 etkisinden söz etmek bu bağlamda zordur.

Sonuç, Tartışma ve Öneriler

Çalışmanın başında ortaya konulan üç hipotez büyük oranda desteklenmiştir. İstisnai olarak, Türk öğrencilerde "kind generic" kullanımların, delimited generik kullanımlardan fazla çıkması, hipotez 2 ile çelişmiştir. Bunun dışında, hem anadil konuşurlar hem de EFL öğrencileri, generik anlamda "the" kullanmaktan kaçınmış, çoğunlukla non-generik bağlamlarda tercih etmişlerdir. Generik kullanımlar arasında ise her iki grup da kind generik yapıları, delimited olanlara göre daha sık kullanmıştır. Bu durum, Radden (2009) tarafından dile getirilen delimited generik yapının sınırlılıklarıyla açıklanabilir. Her iki grubun generik *the* kullanımlarının sayıca benzer çıkması ve Türk EFL öğrencilerinin beklenenin aksine delimited generik'i fazla kullanmamaları, L1 etkisinin bu bağlamda belirgin olmadığını göstermektedir. Bu bulgular ışığında, İngilizce tanımlık sisteminin özellikle generik kullanımlarına yönelik öğretimsel pekiştirmenin önem taşıdığı sonucuna varılmıştır. Bu çalışma, alandaki önemli bir boşluğu ele almaya çalışmış ve ileriye dönük uygulama ve araştırmalar için bazı önemli bulgular sunmuş olsa da, göz önünde bulundurulması gereken bazı sınırlılıklara sahiptir. İlk olarak, hem katılımcı sayısının (n=35) hem de seçilen derlem büyüklüğünün (tüm girdileriyle birlikte yalnızca 5 blog) oldukça sınırlı olması, çalışmanın sonuçlarının genellenmesini zorlaştırmaktadır. Seçilen derlem sayısının sınırlı tutulmasının nedeni, benzer konularda ve benzer miktarda verinin karşılaştırılmasıyla daha güvenilir ve karşılaştırılabilir sonuçlara ulaşılmak istenmesidir. Bu nedenle, gelecekte yapılacak çalışmalar daha geniş katılımcı gruplarıyla ve daha büyük çaplı derlem verileriyle yürütülebilir.

Çalışmanın ikinci sınırlılığı olarak, yazma görevinde belirlenen konunun etkisi de değerlendirilebilir. Bu konu, özellikle delimited generik referans kullanımını teşvik etmeyebileceği için bir kısıt oluşturmuş olabilir. Gerçi bu konu, öğrencilerin ders içerikleriyle ilişkili olması ve dil yeterlilik düzeylerine uygunluğu nedeniyle seçilmiştir. Ancak ileride gerçekleştirilecek çalışmalar, belgeseller ya da haber içerikleri gibi farklı konularla yürütülerek özellikle generik *the* kullanımına dair daha fazla veri elde edebilir. Bu sınırlamaların yanı sıra, çalışmanın verileri farklı veri toplama araçlarıyla desteklenebilirdi. Örneğin, katılımcılarla yapılacak bir görüşme ya da anket uygulaması, ya da tanımlıklarla ilgili bir gramatikalite testi (örneğin Ionin ve Montrul'un (2010) çalışmasında kullanılan Kabul Edilebilirlik Yargısı Görevi [AJT] ya da Doğruluk-Değer Yargısı Görevi [TVJT]) kullanılabilirdi. Bu tür ölçme araçları, ana dili İngilizce olan bireylerle Türk EFL öğrencileri arasındaki "the" nin generik kullanımı farklarını daha somut biçimde ortaya koymaya yardımcı olabilir. Son olarak, gelecekteki çalışmalar, diller arası değişkenliği inceleyebilir; bu kapsamda, diğer tanımlık sistemine sahip olmayan dillerin (örneğin Korece, Japonca, Rusça) öğrencileri

karşılaştırılarak evrensel ve dile özgü eğilimlerin belirlenmesi sağlanabilir. Özetle, bu çalışma, elde edilen bulgular doğrultusunda İngilizcede türsel tanımlık kullanımının hedefe yönelik olarak pekiştirilmesi gerekliliğine dikkat çekmektedir. Sonuç olarak, bu araştırma, İngilizce tanımlık edinimine ilişkin artan çalışma grubuna, belirli tanımlık *the*'nin türsel kullanımını ele alarak önemli bir katkı sağlamaktadır. Kapsamı sınırlı olmakla birlikte, elde edilen sonuçlar hem anadili İngilizce olanların hem de olmayanların *the*'yi türsel anlamda oldukça seyrek kullandıklarını ve bu yapının İngilizce söylemde oldukça sınırlı bir role sahip olduğunu göstermektedir. Türk EFL öğrencileri açısından bu güçlük, yalnızca birinci dil (L1) aktarımından değil, aynı zamanda bu yapının doğası gereği seyrek görülmesinden ve anlamsal açıdan belirsiz olmasından kaynaklanmaktadır.