EMPIRICAL INSIGHTS INTO LANGUAGE AND COGNITION

Editor: Bilal Kırkıcı

nobel bilimsel

Empirical Insights into Language and Cognition

Editor: Bilal Kırkıcı

Publishing No.: 2678
ISBN: 978-625-376-633-7
E-ISBN: 978-625-376-634-4
Edition: 1st Edition, September 2025

Executive Editor: Nevzat Argun -nargun@nobelyayin.com-
Publishing Coordinator: Gülfem Dursun -gulfem@nobelyayin.com-

Interior Designer: Samet Tekin -samet@nobelyayin.com-
Proofreader: İpek Oral -ipek@nobelyayin.com-
Cover Designer: Sezai Özden -sezai@nobelyayin.com-

Library Information Card

Kırkıcı, Bilal.
Empirical Insights into Language and Cognition / Editor: Bilal Kırkıcı
1st Edition. X + 240 p. 16,5x24 cm. Includes references and no index.
ISBN: 978-625-376-633-7
E-ISBN: 978-625-376-634-4
1. Experimental Linguistics 2. Psycholinguistics 3. Cognition

General Distribution

ATLAS AKADEMİK BASIM YAYIN DAĞITIM TİC. LTD. ŞTİ.
Adress: Bahçekapı Mh. 2465 Sk. Oto Sanayi Sitesi No:7 Bodrum Kat, Şaşmaz/ANKARA
Phone: +90 312 278 50 77 - **Sales:** siparis@nobelyayin.com
E-Sales: www.nobelkitap.com - esatis@nobelkitap.com - www.nobelbilimsel.com
Points of Delivery and Salesı: Alfa, Kırmızı Kedi, Arkadaş, D&R, Dost, Kika, Kitapsan, Nezih, Odak, Pandora, Prefix, Remzi Kitabevleri, Yeryüzü Dağıtım
Print and Binding: Ada Matbaacılık Yayın San. Tic. Ltd. Şti. Sertifika No.: 44093
ASO 1. Organize Sanayi Bölgesi Anadolu Cad. No:4 Sincan/ANKARA

List of Authors

Editor: Bilal Kırkıcı

Prof. Dr., METU, bkirkici@metu.edu.tr

Seçkin Arslan

Dr., CNRS, seckin1984@gmail.com

Özgür Aydın

Prof. Dr., Ankara University, ozguraydin66@gmail.com

Doğan Baydal

Dr., OMÜ, doganbaydal@omu.edu.tr

Hakan Cangır

Assoc. Prof. Dr., Ankara University, hcangir@ankara.edu.tr

İpek Çelik Gençer

MA, METU, ipekcelik959@gmail.com

Orhan Demir

Asst. Prof. Dr., Başkent University, orhandemir@baskent.edu.tr

Nazik Dinçtopal Deniz

Asst. Prof. Dr., Boğaziçi University, nazik.dinctopal@bogazici.edu.tr

Onur Evcen

MA Student, METU, onur.evcen@metu.edu.tr

Onur Keleş

MA, Boğaziçi University, onur.keles1@bogazici.edu.tr

Hazel Zeynep Kurada

Asst. Prof. Dr., Ankara Medipol University, hazelkurada@gmail.com

Enis Oğuz

Dr., METU, enisoguz@metu.edu.tr

Çiğdem Sağın Şimşek

Prof. Dr., METU, sagin@metu.edu.tr

Serkan Uygun

Asst. Prof. Dr., Bahçeşehir University, serkan.uygun@bau.edu.tr

İpek Pınar Uzun

Assoc. Prof. Dr., Ankara University, bekar@ankara.edu.tr

Contents

Preface

This book contains data-driven empirical studies by linguists who work in different subfields of linguistics, approach the discipline from diverse perspectives, and adopt various methodological frameworks. The scholars contributing to this volume aim to provide linguists, as well as graduate and undergraduate students, with a broad perspective of the field by presenting studies that showcase the methodological and theoretical richness of linguistic research. Engaging with different approaches and data sources, the book encourages readers to critically evaluate various linguistic phenomena and explore new directions in linguistic inquiry.

Beyond its empirical contributions, this book also serves as a bridge between theoretical discussions and real-world linguistic data. By encompassing a diverse array of studies, it offers readers an opportunity to explore how different methodological approaches can be applied to the study of language and cognition. The volume not only enriches the reader's understanding of linguistic phenomena but also encourages critical thinking and interdisciplinary dialogue, making it a valuable resource for both students and researchers.

This volume brings together ten empirical studies that investigate various aspects of language and cognition, using diverse methodologies and theoretical perspectives. Each chapter offers insights into linguistic processing, acquisition, and structure, shedding light on the interplay between language, thought, and experience.

In *Exploring Lexical Associations in the Bilingual Mind: A Study of Turkish-English Speakers*, Hakan Cangır examines how bilingual individuals organize and retrieve words, shedding light on lexical connections in the bilingual lexicon. Similarly, *The Effect of the Turkish Evidential Morpheme on Scalar Implicature Computation* by Onur Evcen explores how Turkish evidential markers influence scalar implicature interpretation, highlighting the interplay between morphology and pragmatics in real-time language processing.

Another central focus is the acquisition and processing of morphological structures, particularly in second-language learning. In *Morphological Generalization in L2 Turkish: The Case of Turkish Aorist*, Serkan Uygun investigates how learners of Turkish generalize aorist tense morphology.

Complementing this, *Morphological Complexity Patterns Across L2 English Essay Scores: The Case of a Morphologically Rich and Productive L1 Background* by Enis Oğuz explores how various indices of morphological complexity relate to L2 English essay quality among Turkish-English bilinguals in order to offer new insights into the influence of L1 morphology and contribute to ongoing debates on how to measure complexity in L2 writing.

Questions that go beyond morphology, syntax and sentence processing are also explored through various methodologies. In *Motor Activation in Action Language: Task-Dependent but Non-Specific Effects*, Hazel Zeynep Kurada explores how concurrent motor activity influences the processing of action verbs in Turkish, revealing general facilitation effects that highlight the task-dependent nature of embodiment in language comprehension. *Subject Control in Temporal Converb Constructions in Turkish: Evidence from the Turkish National Corpus (TNC)* by Doğan Baydal investigates subject control mechanisms in Turkish converbs using corpus data. Besides, *Good-Enough Parsing in Turkish: Task Effects in Online Processing of Role-Reversals* by Onur Keleş and Nazik Dinçtopal Deniz examines how task demands influence sentence processing in Turkish, showing that readers sometimes rely on plausibility-based heuristics over syntactic parsing. Extending this discussion, *What Eye Movements Reveal about Pre-Verbal Focus Processing in Turkish: An Eye-Tracking Study on Sentence Comprehension During Reading* by İpek Pınar Uzun, Seçkin Arslan, and Özgür Aydın investigates how focus position and argument type affect sentence processing in Turkish, reflecting a processing advantage tied to focus structure and word order.

In its final chapters, this volume delves into bilingual processing and metalinguistic awareness. *The Role of Syntactic and Semantic Information in Cross-Linguistic Processing: A Translation Priming Study* by Orhan Demir investigates how syntactic and semantic congruence influence translation priming, shedding light on the complex interactions between languages in the bilingual mind. Finally, *The Effect of Cognateness on Tip-of-the-Tongue (TOT) States and Their Resolution in Turkish EFL Learners* by İpek Çelik Gençer and Çiğdem Sağın Şimşek explores how cognate words impact the occurrence and resolution of TOT states in Turkish learners of English, emphasizing the connection between linguistic experience and memory.

By integrating psycholinguistic experiments, corpus analyses, and eye-tracking studies, this collection presents a comprehensive exploration of language processing and acquisition. Each study offers empirical insights that not only deepen our theoretical understanding of linguistic phenomena but also provide valuable implications for language learning, bilingualism, and cognitive science.

Together, these studies offer a rich and multifaceted perspective on language and cognition, combining empirical findings with theoretical discussions to deepen our understanding of linguistic structures and processes.

Bilal Kırkıcı (ed.)

Chapter 1:
Exploring Lexical Associations in the Bilingual Mind: A Study of Turkish-English Speakers

Hakan Cangır

ABSTRACT

This chapter investigates word associations in the bilingual mental lexicon of Turkish-English speakers, comparing them with L1 English users and exploring the influence of word type, lexical frequency, and word level (i.e., lexical proficiency according to CEFR[1] levels) on these associations. While existing literature suggests L1 users favour paradigmatic relations and L2 users, especially at lower proficiency levels, tend towards syntagmatic relations, this study examines these assumptions within the understudied context of Turkish as an L1 and English as an L2. A word association task is employed to investigate the responses from L1 Turkish-L2 English bilinguals. To provide a comparative baseline, data from the "Small World of Words" semantic association database is also analysed, representing L1 English association strengths and frequencies. The experimental stimuli consist of 20 high and 20 lower-frequency nouns and adjectives, carefully controlled for concreteness and cognateness, while verbs are excluded due to Turkish verb conjugation complexities. By examining the potential influence of word type (nouns – adjectives), cue word frequency (high – low according to British National Corpus), and word CEFR level on lexical associations, this research aims to contribute to psycholinguistic models of bilingual mental lexicon organisation from a different angle and offer insights for language practitioners regarding vocabulary acquisition and development in L2 English.

Key Words: Word Associations, Bilingual, Mental Lexicon, Frequency

[1] Common European Framework of Reference for Languages (Council of Europe, 2020)

1. INTRODUCTION

The bilingual mental lexicon, the cognitive system responsible for storing and processing words in two or more languages, gives us a complex and a comprehensive area of psycholinguistic inquiry. Investigating how bilinguals organise and access their lexical knowledge has the potential to guide us in drawing conclusions for theories of language processing, as well as for practical applications in language education. There have been various attempts to illustrate the organisation of bilingual mental lexicon and each with different approaches. In some of the earlier ones, Collins and Quillian (1969) and Collins and Loftus (1975) argue that word knowledge is stored as nodes in an interconnected semantic network. When we encounter a word or when a lexical node is activated in the mind, the stimulus triggers some intra-lexical links and the activation spreads onto other related lexical nodes. The degree of spreading activation depends on the strength of the associations between the nodes, which is the effect of personal experience and the frequency of exposure. As a consequence of this spreading activation, the presentation of a specific word (e.g., dog) is likely to increase the possibility of producing an associated word (e.g., cat).

Having their spark in the basic notion of spreading activation, several models of bilingual mental lexicon have been proposed to explain how the two (or more) languages are represented and accessed in the bilingual mind. Some claim bilinguals have separate lexicons, whereas others assert that bilinguals share an integrated (single) mental dictionary. The Revised Hierarchical Model (RHM) (Kroll & Stewart, 1994) suggests that L1 lexical representations are strongly attached to conceptual representations, while L2 lexical representations are closely connected to L1 representations. As the language user gains proficiency, direct links and associations between L2 lexical and conceptual representations emerge. This model also asserts that L2 learners, particularly at lower proficiency levels, are more likely to depend on L1 mediation, which can eventually affect the lexical associations in their lexicon. Another important framework attempting to explain how the bilingual mental lexicon is structured is the Distributed Feature Model (DFM) (de Groot, 1992). This model foregrounds the role of shared and language-specific semantic features in lexical representation. To be more precise, bilinguals represent concepts through distributed networks of features, with some features being shared across languages and others being language specific. The DFM assumes that the degree of overlap between L1 and L2 semantic representations affects the duration of lexical access and what lexical associations are formed in the mental lexicon.

Using these mental lexicon models as their foundation, research on bilingual lexical representation and processing has explored the way L2 lexicon differs from or how much it resembles L1 using the word association paradigm (e.g., Fitzpatrick, 2006; 2007). Previous research studies have suggested that L1 speakers have a tendency towards paradigmatic (meaning-based) associations in their lexical choices, which are based more on semantic resemblance (e.g., *chicken* – egg), while L2 learners, especially those below a certain proficiency level, tend to stick with syntagmatic (collocational/position-based) associations, which are mainly based on sequential or contextual relationships (e.g., *cat* – *meow*) (Fitzpatrick, 2007). This pattern is often attributed to the L2 learner's overreliance on contextual cues and formulaic language in addition to their restricted access to the full range of semantic relations available in the L2 and can possibly be due to L1 effect.

In earlier studies exploring word associations, Meara (1983) concludes that L2 learners particularly in high proficiency levels produce more paradigmatic associations in a word association task. This may suggest that the lexical representations in their mental lexicon become more semantically organised with more exposure to the L2. In another study, Schmitt and Meara (1997) suggest that the word association responses of L2 English learners resemble those of L1 speakers more as they gain proficiency. A recent and comprehensive review of word association studies in L2 acquisition has been provided by Fitzpatrick and Thwaites (2020), highlighting the difference between paradigmatic and syntagmatic associations. They have listed several patterns in word association behaviour, underlining the fact that paradigmatic or meaning-based responses are generally the most common in word association research, whereas form-based (or clang) responses appear to be the least (e.g., *bear* – *beer*). As Wolter (2001) suggests, it is more likely for the participants who have a deeper and stronger understanding of the cue words or who are more proficient (Zareva & Wolter, 2012) to come up with a meaning-based response in a word association task. Similarly, if the cue word is a noun, the chances are higher for the participants to give semantically associated responses (Nissen & Henriksen, 2006). On the other hand, form related (clang) responses are more prevalent among younger participants (especially those in the last year of preschool) (Namei, 2004), lower level L2 users (Wolter, 2001) and those learning English as a foreign language rather than a second language (Håkansson & Norrby, 2010). As for position based (or syntagmatic / collocational) responses (e.g., *black – coffee*), the patterns observed by the researchers are rather complicated. Some studies claim that position-based responses are more ubiquitous in L2 users (Håkansson & Norrby, 2010), whereas some others assert the opposite stating that they are more common in L1 users' responses (Fitzpatrick & Izura, 2011). The collocational responses in L2 are particularly triggered by adjective cue words and are given by lower-level language users. Furthermore, responses which are synonymous with the cue words are more likely to be given by either L1 users or higher proficiency language users (Khazaeenezhad & Alibabaee, 2013).

In addition, according to Xiaosa and Wenyu (2016), the grammatical category of a word may influence the type of associations it may trigger in both L1 and L2 users. To give an example, studies show that verbs and adjectives tend to elicit more syntagmatic responses (e.g., collocations) compared to nouns. According to Nissen and Henriksen (2006), nouns are predominantly organised in paradigmatic associations, while verbs and adjectives are more likely to be associated by syntagmatic relations. This difference may indicate that the mental lexicon of bilinguals varies in its structure depending on the type of the word represented, which could also differ in L2 or L3. Investigating lexical associations in L3 Italian, Kuzu and Cangır (2025) concludes that the collocational associations in L1 Turkish - L3 Italian users' mental lexicon is strong when the cue word is an adjective. On the other hand, when the same cues are considered, stronger semantic relations are observed in the L1 Italian speakers' lexicon. For noun cues, semantic associations are more robust in both groups, but L1 speakers tend to have more nuanced and sophisticated lexical choices, indicating deeper knowledge of vocabulary.

Research studies primarily exploring word association patterns from the lens of an understudied language, Turkish (though employing different methodologies from the current study), offer valuable insights into the mental lexicons of L1 Turkish users learning English as their L2. Studies to date have investigated cross-linguistic collocational networks between L1 Turkish and L2 English, unearthing complex ways in which the two languages may interact (Cangır & Durrant, 2021). Findings reveal patterns of cross-linguistic collocational priming; particularly congruent collocations are processed faster in L1-L2 direction (e.g., *cold war – soğuk savaş*). There is also research (Cangır, 2021a) indicating the relationship between the subjective judgements of L1

Turkish – L2 English users' collocational frequency intuitions and corpus-driven objective frequency measures, which reveal a partial overlap between the two, particularly for high frequency words. Investigating the issue from the perspective of L1 Turkish users and focusing on their native speaker intuitions for word frequency, another study (Cangır, 2021b) examines the potential correlation between the lexical (collocational) associations in the L1 Turkish mental lexicon and the association strength measures gathered from the Turkish National Corpus (Aksan et al., 2012). The results indicate a robust relationship between L1 intuitive and corpus-based lexical association strengths, which the writer claims could provide fresh insights into both word association research and language teaching practices.

Although previous studies scrutinising L1 Turkish and L2 English users and exploiting various experimental techniques have sought to shed light on the way words are associated in their lexicons and have provided new insights into the word associations in the bilingual mind, no study, to the writers' knowledge, has attempted to utilise a word association task trying to illuminate how L1 Turkish – L2 English users may differ from or resemble L1 English users and if word frequency, type and level play a role in the process. As an understudied phenomenon, the lexical associations in the Turkish-English bilingual mental lexicon are likely to offer a novel perspective on the way words are organised in the mind. Turkish is an agglutinative language with a rich morphology, which differs significantly from Indo-European languages like English. For instance, being more adept to decomposing words into constituent morphemes in his/her native language, a Turkish-English bilingual could be more inclined to associate the word "break" with "unbreakable" or "broken". It is likely that these typological difference affects how Turkish-English bilinguals represent and process lexical information in L2 English, which could reveal distinct patterns of lexical association compared to bilinguals of other language pairs.

The current study builds on the two bilingual mental lexicon models explained above (i.e., Spreading Activation Model and Revised Hierarchical Model) and considers the main categories of Fitzpatrick's (2007) classification for lexical associations in L1 and L2 (i.e., meaning based [semantic] (e.g., *cat – dog*), position-based [collocation] (e.g., *cold – war*), and form-based [clang] (e.g., *hat – fat*). It aims to examine the validity of the earlier assumptions of paradigmatic and syntagmatic relations within the specific context of Turkish as an L1 and English as an L2. Furthermore, it seeks to explore how/if factors such as word frequency, word type and level can modulate these association patterns. As mentioned earlier, no study has yet attempted to investigate this phenomenon from this angle and considering an understudied L1 background, Turkish.

To achieve these aims, a word association task is employed, presenting L1 Turkish-L2 English bilinguals with English stimuli and collecting their intuitive responses for lexical association. The data obtained from this task is compared with data from the "Small World of Words" (SWOW) (De Deyne et al., 2019) a semantic association database, which provides a large-scale representation of L1 English semantic associations to establish a baseline for L1 English association strengths and frequencies. By analysing the interplay of these variables, the researcher aims to contribute to a more nuanced understanding of the bilingual mental lexicon and to provide valuable insights for language practitioners regarding vocabulary acquisition and development.

Based on the existing literature, this study aims to address the following research questions:

1. How do the word association patterns of L1 Turkish-L2 English bilinguals compare to those of L1 English speakers (as represented in the SWOW database)?
2. How does (a) word type [nouns vs. adjectives], (b) lexical frequency [high vs. low], and (c) word level [A1-B2 according to CEFR] influence the patterns of word associations produced by L1 Turkish-L2 English bilinguals?

2. METHODOLOGY

2.1. Item Development

There were some methodological considerations while choosing the lexical items in the word association task. The starting point of item selection was the frequency profiles of the items in the British National Corpus (BNC) because L2 English learners at the institution where this study was conducted are trained using textbooks dominant in British English. The initial filtering was based on the most frequent 500 words in BNC. To observe the possible influence of lexical frequency and considering the possible language levels of the students, out of 500 words, top 100 and bottom 100 items of the frequency list was extracted. Earlier studies have detected certain patterns in terms of the effect of frequency. For instance, Meara (2009) finds that cue words with low frequency tend to get more inconsistent and even blank responses (Higginbotham, 2010) than more frequent words. This may suggest that word associations in an L2 is more sensitive towards frequency effects than L1 association. Another way to name frequency effect is the word familiarity effect. For example, Wolter (2001) suggests that a decline in word familiarity leads to increasing number of form-based responses, while some researchers like Riegel and Zivian (1972) indicate the otherwise (see also Wilks, 2009).

The list was further filtered in terms of word type because the current research was only interested in nouns and adjectives. A rationale for this could be that association categories depend partly on word class. That is to say, as earlier research suggests, nouns appear to be mainly structured in paradigmatic relations, while adjectives are characterised by syntagmatic relations (Nissen & Henriksen, 2006). The researcher attempted to make sure the selected items were not synonyms or antonyms as this could mislead the participants. To avoid the items with the same noun and verb forms, most of the selected nouns on the list had only noun forms. However, this rule had to be softened for some items as there were also other concerns, such as frequency, potential CEFR level, and concreteness of the items.

The degree of concreteness of the chosen items was also a matter of importance for the current research and thus was determined based on the data available in Brysbaert, Warriner, and Kuperman (2014). The writer wanted to ensure that half of the items in the chosen word categories were below the average concreteness rating of the whole items in one word category, and the other half was above the average concreteness rating and the degree of concreteness was used as a confounding variable in this research. This decision can be attributed to the fact that participants in earlier research studies responded more quickly to the concrete cue words in word association tasks (Brown, 1971; de Groot, 1989; van Hell & de Groot, 1998), left fewer blank responses (de Groot, 1989; Bøyum, 2016), the data output exhibited tendency towards increased homogeneity in its dispersion (Brown, 1971; de Groot, 1989), and participants showed a higher propensity for responding (de Groot, 1989) when they were faced with more concrete words rather than less concrete ones.

Another concern during item filtering and later during data analysis was the CEFR level of the selected lexical items since the participants in the study were

L2 English learners and the writer wanted to make sure they knew all the words presented. To this end, only items in A1, A2 and B1 levels were given priority based on the website https://englishprofile.org. This rule was violated for only three B2 level items due to other methodological concerns summarised earlier. Table 1 provides an overview of normalised frequency values and concreteness scores of the selected cue words (20 nouns and 20 adjectives).

Table 1. Distribution of relative frequency and concreteness across word levels

					Skewness	
	CEFR	**Mean**	**SD**	**IQR**	**Skewness**	**SE**
Relative frequency	A1	524.85	366.808	181.230	2.4579	0.616
	A2	192.23	160.175	116.630	1.7825	0.913
	B1	141.90	173.767	152.790	1.6439	0.580
	B2	84.20	69.588	81.070	1.0941	0.794
Concreteness	A1	3.05	0.967	1.230	0.2954	0.616
	A2	3.35	0.995	0.760	0.4718	0.913
	B1	2.45	0.710	1.130	0.4625	0.580
	B2	3.06	0.975	1.270	0.0484	0.794

As one would expect, as the CEFR level of the chosen cue words increase, the relative frequency values decrease (from 524.85 at A1 to 84.20 at B2). The concreteness values (ranging from 2.45 at B1 to 3.35 at A2), on the other hand, seem to be reasonably symmetrical across the data as indicated by lower standard deviations and inter quartile ranges (IQR). Table 2 considers the same distribution based on the different word types investigated in this study.

Table 2. Distribution of relative frequency and concreteness across word types

					Skewness	
	POS	**Mean**	**SD**	**IQR**	**Skewness**	**SE**
Relative frequency	adj	205.25	238.115	295.05	1.7045	0.512
	noun	319.85	349.472	310.47	2.7340	0.512
Concreteness	adj	2.54	0.699	1.01	0.2603	0.512
	noun	3.19	0.999	1.43	0.0824	0.512

The mean values suggest that the noun cues in the study have higher frequency values (M = 319.85) and naturally have higher concreteness ratings (M = 3.19) than adjective cues (M = 205.25; 2.54, respectively). Adjectives, on the other hand, have lower standard deviations from the mean both for relative frequency and concreteness ratings. Standard deviations are 238.115 for the frequency of adjective cue words and 349.472 for noun cue words, whereas standard deviations for concreteness are 0.699 for adjectives and 0.999 for nouns, again indicating the variability within each word type.

2.2. Participants

Table 3. Participants' age and self-rated proficiency

	Age	Reading*	Writing	Listening	Speaking
N	53	56	56	56	55
Missing	3	0	0	0	1
Mean	21.0	4.11	3.86	3.84	3.55
Standard deviation	4.85	0.755	0.841	0.848	1.03
Minimum	18.0	2.00	2.00	2.00	2.00
Maximum	46.0	5.00	5.00	5.00	5.00
Skewness	3.37	-0.444	-0.292	-0.238	-0.126
Std. error skewness	0.327	0.319	0.319	0.319	0.322

* For proficiency in each language skill, 5 is the highest score and 1 is the lowest score (weak proficiency).

The research study included 56 participants; however, the number varies slightly across language skills due to missing data (N = 53 for age; N = 56 for reading, writing, and listening; N = 55 for speaking). The mean age of the participants was 21.0 years, with a standard deviation of 4.85, indicating a range between 18 and 46 years. The age distribution was positively skewed (skewness = 3.37, standard error of skewness = 0.327), indicating a concentration of younger participants. Self-rated proficiency of the participants in L2 English, on a scale of 1 to 5 (where 5 represents the highest proficiency), revealed the following means and standard deviations: reading (M = 4.11, SD = 0.755), writing (M = 3.86, SD = 0.841), listening (M = 3.84, SD = 0.848), and speaking (M = 3.55, SD = 1.03). Skewness values for the proficiency ratings ranged from -0.444 to -0.126, with standard errors between 0.319 and 0.322, indicating relatively symmetrical distributions for these measures.

The subjects included 39 females (70%), 12 males (21%), and 5 individuals who did not specify their gender (9%). The majority (95%) of the participants were pursuing BA degrees, while a small proportion (5%) were MA students. These demographic characteristics are important to consider when interpreting the word association data, as factors such as age, gender distribution, and educational level may influence lexical organisation and processing in the bilingual mental lexicon (Fitzpatrick & Thwaites, 2020).

2.3. Tools

2.3.1. Word Association Tasks (WATs) in Bilingual Research

WATs have been broadly utilised to explore the structure of monolingual and bilingual mental lexicons. Subjects doing this task are given a stimulus word (i.e., a cue word) and they are asked to respond with the first word(s) that come to their mind. It is claimed by researchers investigating word associations in the internal lexicon that these responses have the potential to reveal insights into the associative relationships (e.g., semantic, collocational etc.) between lexical nodes in the lexicon. WATs are conducted either in simple questionnaire formats (e.g., Cangır, 2021a) or using more sophisticated tools, such as PsychoPy (Peirce et al., 2019) to record response times (e.g., Fitzpatrick & Izura, 2011).

One of the pioneers in the field, Fitzpatrick (2007) comprehensively reviews the studies focusing on word associations in L2 acquisition. She particularly highlights the distinction between paradigmatic (e.g., synonymy, antonymy, or category membership) and syntagmatic associations (e.g., collocations or typical contexts of use) between the cue words and responses. The overall observation by Fitzpatrick (2007) is that L1 speakers are inclined to produce more paradigmatic associations, while L2 learners, particularly at lower proficiency levels, tend to produce more syntagmatic associations, which could be attributed to learners' overreliance on the use of formulaic language as well as their limited semantic knowledge in their L2. It is also claimed by Schmitt and Meara (1997) that as L2 leaners' proficiency increases, their responses to the cue words in a word association task resemble L1 users' responses.

2.3.2. English Vocabulary Profile Online[2]

During the item development phase and to be used in data analysis, "The English Vocabulary Profile Online" is utilised to assign CEFR levels to the cue

[2] The website can be accessed at https://englishprofile.org.

words in the experiment. The database is a resource developed by Cambridge University providing comprehensive information about the English vocabulary knowledge of learners from different L1 backgrounds and at different levels of proficiency according to CEFR. Researchers and practitioners can access the platform to explore the words and phrases L2 English learners typically know, understand and use at each CEFR level (A1-C2). The platform provides vocabulary lists with certain filters, lexical frequency details, and grammatical patterns, giving a detailed insight into vocabulary acquisition in L2 English. It is also used for material development, language testing and practical applications in classroom settings (Capel, 2012). The current study made use of the website to label the target words with their CEFR levels and used them as an independent variable in data analysis.

2.3.3. The Small World of Words Database[3]

In an attempt to provide a more comprehensive analysis and present the data in a comparative manner, the current study employed the "Small World of Words" (SWOW) database (De Deyne et al., 2019). SWOW is a database of word associations gathered from thousands of volunteer participants, which provides a reliable representation of semantic relationships in different languages including English. For this research, the database serves a baseline for comparing the association patterns of L1 Turkish-L2 English participants with those of L1 English users.

2.4. Word Type, Lexical Frequency and Association Strength

As has been documented in earlier research, word type and lexical frequency (or familiarity) are considered important factors that may influence the lexical associations in the bilingual mental lexicon. Fitzpatrick and Izura (2011) state that nouns and adjectives, for example, often show distinct patterns of association. Participants in word association studies tend to associate nouns with other nouns from the same semantic category; however, they are inclined to associate adjectives with some nouns that act as modifiers (e.g., *green – tea*). Additionally, both certain distinctive patterns and some similarities have been observed between L1 and L2 users' responses, the reasons of which are attributed to proficiency in L2, cue word properties and individual differences but the phenomenon need further explanation.

[3] The website can be accessed at https://smallworldofwords.org/en/project.

Lexical frequency, or how often a word occurs in a language (or its familiarity), can also play a role in how words are organised in the lexicon. As Ellis (2002) states, high-frequency words tend to be accessed more quickly and efficiently as they are processed more easily than low-frequency words. Additionally, high frequency words generally have more robust and well-defined semantic representations in the mind. Studies exploring word associations have suggested that high-frequency cue words tend to stimulate more consistent and predictable association responses and lexical patterns than low-frequency cue words (Nelson & Schreiber, 1992). As far as L2 acquisition is concerned, L2 users may find it more challenging to acquire and integrate low-frequency words into their mental lexicon, which in turn could result in less consistent and more variable association patterns. Therefore, the comparison of distinct word association patterns between first and second language users presents an intriguing avenue for further research, particularly when considering understudied languages such as Turkish. For this study, relative lexical frequency was computed using the BNC (British National Corpus) through SketchEngine (Kilgarriff et al., 2014) interface. The numerical output was used both for item selection and data analysis.

Studies exploring word associations in the bilingual lexicon tend to use reference corpora both for item extraction and calculate the association strength measures of the responses to compare them with the frequency of participants' responses. There are some studies (e.g., Wettler et al., 2005) claiming that a corpus-driven model of word association can resemble the intuitive human responses in a word association task. However, others with more compelling evidence suggest that the co-occurrences in a corpus cannot correlate with human experiences (e.g., Mollin, 2009; Kang, 2018), emphasising the fact that the processes underlying the learning of lexical association may not overlap with the word associations participants generate in word association tasks (McRae, Khalkhali & Hare, 2012). Most studies comparing word association responses to corpus co-occurrences report evident differences between the two (e.g., Kang, 2018), particularly highlighting the lack of figurative uses of words in these tasks (e.g., erupt-violence rather than erupt-volcano) (Thwaites, 2019). Given this evidence in the literature and the warning against using corpus-driven association strength measures, the current study employs lexical frequency of the cue words only for item extraction and to categorise the words into frequency groups (high-lower) for data analysis.

To calculate a reliable association strength measure which could be comparable with other means (i.e., with other association measures employed in earlier studies), the frequency of a particular response was divided by the total number of responses given to that cue word so that the proportion of the participants producing the same word could be considered and the value can be normalised (Playfoot et al., 2018). For instance, if 35 participants out of 60 respond with "book" when presented with the cue word LIBRARY, the strength of association is calculated as 35/60, which equals to 0.58. There is also empirical evidence in psycholinguistics studies (e.g., Cañas, 1990) suggesting that strength of association could be a good indicator of processing speed in priming experiments with higher associative strength eliciting faster reaction times.

2.5. Data Analysis

In an attempt to address the research questions, descriptive statistics were used and the L2 English responses with the highest frequency were reported in comparison to the most frequent L1 English responses as evidenced in SWOW. In addition, the type of associations between the cue words and responses were also reported comparatively. To help readers visualise the detected differences between L1 and L2 English responses and the possible effect of association type, the effect of cue word frequency and the CEFR level of the lexical items, bar charts were utilised. Furthermore, to reveal the relationship between association strength of the L2 English users' lexical choices and the frequency of the L1 English responses, a partial correlation analysis was conducted. Finally, two separate ANOVA tests were computed to explore the possible effect of cue word frequency, level and type on the association scores.

3. RESULTS

This section presents the findings of the current study in a step-by-step manner. To provide clarity and facilitate interpretation, the results are organised under four distinct subheadings, each directly addressing one of the research questions and its corresponding sub-questions. Within each subsection, explanatory tables and visual representations are employed to illustrate the key patterns and statistical outcomes.

3.1. Differences and Similarities Between The Responses in L1 English and L2 English

Table 4 illustrates the order of the responses by L1 and L2 English users (based on frequency and association score) and reports the types of association between the adjective cues and the responses comparatively.

Table 4. Responses with adjective cues

Corpus Freq.	Cue words	L2 English Responses	Assoc. Score	Assoc. Type	L1 English Responses	SWOW Freq.	Assoc. Type
950.1	*New*	Old	0.346	Coll.	Old*	37	Coll.
465.01	*High*	Low	0.166	Sem.	Low*	42	Sem.
423.69	*Different*	Same	0.129	Sem.	Same*	25	Sem.
397.75	*Local*	People	0.094	Coll.	Nearby	18	Sem.
344.67	*Important*	Thing	0.075	Coll.	Necessary	13	Sem.
311.43	*Early*	Morning	0.173	Coll.	Morning*	59	Coll.
296.94	*Possible*	Impossible	0.283	Form	Impossible*	32	Form
253.12	*Public*	Transportation	0.148	Coll.	Open	25	Sem.
214.49	*Major*	Minor	0.296	Sem.	Minor*	48	Sem.
208.25	*Clear*	Mind	0.132	Coll.	Transparent	50	Sem.
27.17	*Proud*	Pride	0.075	Form	Happy	12	Sem.
26.4	*Remote*	Control	0.500	Coll.	Control*	64	Coll.
25.71	*Accurate*	Correct	0.120	Sem.	Correct*	44	Sem.
25.45	*Honest*	Lie	0.094	Sem.	True	26	Sem.
24.86	*Smooth*	Criminal	0.113	Coll.	Soft	16	Sem.
24.2	*Violent*	Blood	0.056	Sem.	Anger	13	Sem.
23.05	*Sensible*	Reasonable	0.094	Sem.	Smart	23	Sem.
21.01	*Flexible*	Body	0.057	Coll.	Bends	26	Coll.
21.0	*Frequent*	Often	0.260	Sem.	Often*	61	Sem.
20.71	*Bitter*	Chocolate	0.470	Coll.	Sweet	36	Coll.
		- 10 collocational, - 8 semantic, and - 2 form associations			- 5 collocational, - 14 semantic, and - 1 form associations		

As is seen in the table, nine lexical items (signalled by an asterisk*) match in terms of the first words that are triggered the subjects' lexicon when they see the cue word. When the second lexical choices they provided in the word association test are analysed, the overlap is more than half (12 items). However, the second responses provided by the participants are not reported in this table. The syntagmatic vs. paradigmatic distinction put forward in earlier sections of the paper is confirmed given the results in Table 4. To be more precise, L1 English users produce paradigmatically related (i.e., meaning-based) words more than syntagmatically related (i.e., position-based) words as opposed to L2 English users whose results show the opposite trend. Another striking pattern is that more responses overlap between L1 and L2 items when the cue words are highly frequent (6 matching items for frequent items, and 3 matching items for relatively low frequent items). Table 5 presents the same comparison with the noun cue words.

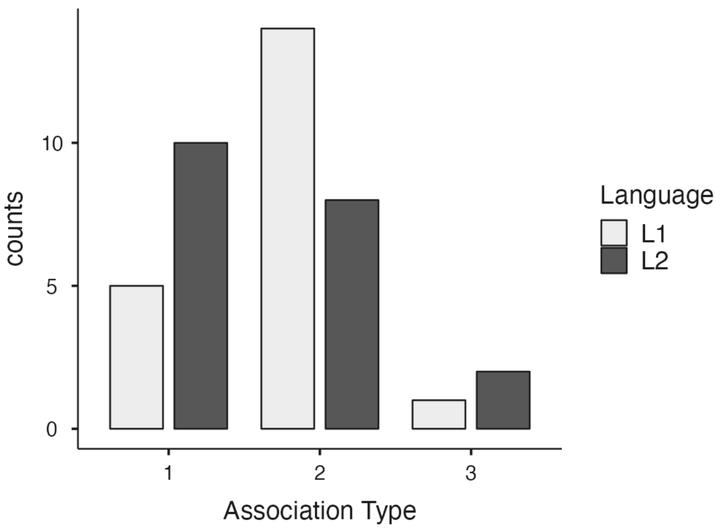
Table 5. Responses with noun cues

Corpus Freq.	Cue words	L2 English Responses	Assoc. Score	Assoc. Type	L1 English Responses	SWOW Freq.	Assoc. Type
1592.14	*Time*	Clock	0.200	Sem	Clock*	43	Sem
254.32	*Person*	Human	0.204	Sem	Human*	100	Sem
592.62	*Government*	State	0.102	Sem	Politics	19	Sem
560.52	*Life*	Short	0.106	Coll	Death	34	Sem
520.8	*World*	Earth	0.200	Sem	Earth*	91	Sem
471.5	*Place*	Home	0.155	Sem	Home*	32	Sem
383.62	*Example*	Instance	0.159	Sem	Sample	19	Sem
373.71	*Family*	Mother	0.152	Sem	Love	47	Sem
339.57	*Question*	Answer	0.422	Sem	Answer*	121	Sem
331.72	*Interest*	Hobby	0.232	Sem	Hobby*	44	Sem
114.18	*Concern*	Worry	0.116	Sem	Worry*	125	Sem
106.2	*Science*	Fiction	0.133	Coll	Study	20	Sem
102.77	*Discussion*	Debate	0.266	Sem	Talk	47	Sem
98.56	*Client*	Customer	0.175	Sem	Customer*	86	Sem
95.25	*Disease*	Illness	0.181	Sem	Sick	25	Sem
94.72	*Attitude*	Behaviour	0.255	Sem	Bad	49	Coll
92.57	*Ability*	Talent	0.227	Sem	Skill	31	Sem
91.62	*Library*	Book	0.659	Sem	Books*	72	Sem
90.59	*Drug*	Medicine	0.116	Sem	Medicine*	17	Sem
90.08	*Variety*	Various	0.048	Form	Show	24	Coll
		- 17 semantic, - 2 collocational, and - 1 form associations			- 18 semantic and - 2 collocational associations		

As the table indicates, ten lexical items (signalled by an asterisk*) overlap between L1 and L2 responses. Six items overlap in the high frequency category and 4 matching in the low frequency category. In other words, the more frequent the cue words were, the more consistent the responses of the L1 and L2 English users. The syntagmatic vs. paradigmatic distinction put forward in earlier sections of the paper cannot be confirmed given the results for the noun cues in Table 5.

Considering the responses which are semantically related with the cue words in both the tables, the analysis of adjective associations revealed a similar pattern in the generation of antonyms, with both L2 and L1 English speakers identifying 5 antonyms. A notable difference emerged in synonym generation, where L1 English speakers provided more than twice as many synonyms or words with similar senses (7) as their L2 counterparts (3). In contrast, the noun association data indicated that L2 English users generated a slightly larger pool of semantically related words (12 or 13) compared to the 10 such words produced by L1 English speakers.

In order to visualise the difference between the association types in L1 and L2 English responses, Figure 1 compares the frequency of the association categories for both the adjective and noun cues in the dataset.



*1 (collocational), 2 (semantic), 3 (form)

Figure 1. Comparison of Association Type in L1 vs. L2 English

As is revealed in the figure, when the responses in both word types are combined, L1 English users tend to have more semantically associated responses, whereas L2 English users are inclined to have more collocational responses, which is in line with earlier research findings. Both groups have few responses with form associations (e.g., various – variety).

To further investigate the strength of associations in L1 and L2 English responses and their potential negative or positive correlation, a partial correlation analysis is computed with concreteness as the controlling variable.

Table 6 shows that as the mean frequency of responses in SWOW increases, the mean frequency (or association strength scores) of the participants' lexical responses also increases. This relationship also indicates the similar trends in the predictability of the associated words when a cue is given regardless of the participants L1 or L2. The controlling variable 'concreteness" value seems to play a moderate role in the equation. A separate correlation (though not reported in Table 6) between association strength scores and concreteness ratings reveals that as the concreteness of the cue words increases, the association strength scores also appear to increase ($r(78) = .21, p = .050$), which means participants are slightly more consistent with their responses for more concrete cue words. However, the correlation coefficient indicates a weak positive relationship, suggesting that the finding may require further scrutiny and should be interpreted with caution.

Table 6. Relationship between association strength and SWOW

		Association Strength	
Association Strength	Pearson's r	—	
	p-value	—	
	Spearman's rho	—	
	p-value	—	
SWOW (Small World of Words)	Pearson's r	0.607	***
	p-value	< .001	
	Spearman's rho	0.690	***
	p-value	< .001	

Note. controlling for 'Concreteness'

Note. * p < .05, ** p < .01, *** p < .001

3.2. The Effect of Word Type on Word Associations

As an alternative way to approach the data, Figure 2 pays attention to the potential effect of word type on the types of associations revealed as a result of the participants' responses.

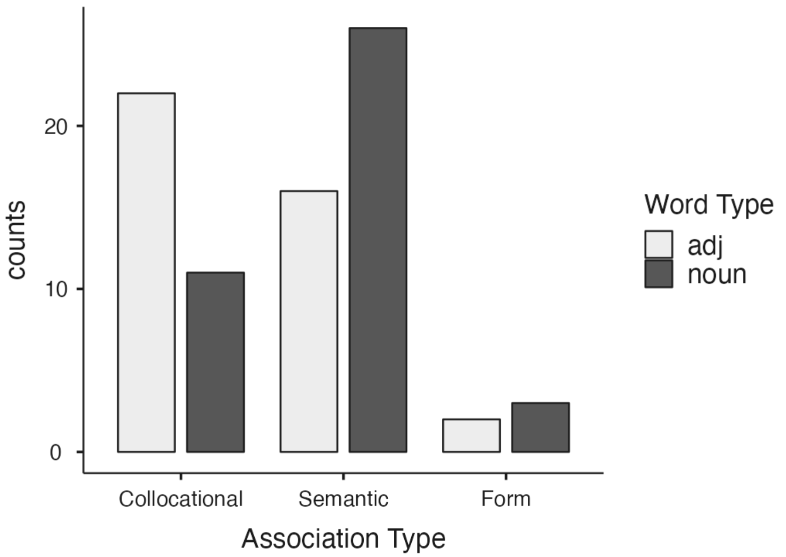


Figure 2. Types of associations across word forms

As has partly been discussed in Section 3.1., Figure 2 illustrates the types of associations detected for two different word types. More syntagmatic (position-based OR collocational) associations are observed with the adjective cues, whereas slightly more paradigmatic (semantic) associations are seen with the noun cues. As for form associations (i.e., lexical-syntactic), slightly more noun cues trigger these associations than adjectives.

To further analyse the effect of word frequency, level and type on the association scores and report the possible difference statistically, two ANOVA tests were conducted controlling for (a) word frequency and type, (b) word level and type, respectively. The results indicated no main effect of the independent variables on the association scores. The numerical output is provided in Table 8 and 9 under Sections 3.3 and 3.4.

3.3. The Effect of Lexical Frequency on Word Associations

As is highlighted in Section 3.1., more matching words are observed between L1 and L2 English responses when the cue words are highly frequent (6 matching items for frequent items, and 3 matching items for relatively low frequent items. Table 7 summarises the descriptive details for the association strength scores across two frequency categories.

Table 7. Descriptive details of the association scores across frequency and word types

						Skewness	
	Frequency	**Type**	**Mean**	**SD**	**IQR**	**Skewness**	**SE**
Association Score	high	adj	0.200	0.0958	0.137	1.06	0.512
		noun	0.213	0.1058	0.107	1.67	0.512
	low	adj	0.215	0.1583	0.115	1.95	0.512
		noun	0.244	0.1424	0.147	2.02	0.512

The descriptive values indicate variations in association strength scores across different frequency and word types. For high-frequency cue words, the mean association score of the responses is 0.200 (SD = 0.0958, IQR = 0.137) for adjectives and 0.213 (SD = 0.1058, IQR = 0.107) for nouns. For cue words with lower frequency, the mean association score of the responses is 0.215 (SD = 0.1583, IQR = 0.115) for adjectives and 0.244 (SD = 0.1424, IQR = 0.147) for nouns. What is striking is that noun cues triggered responses with higher association scores than adjective cues. Additionally, low-frequency cue words, particularly nouns, elicited responses with slightly higher association measures and greater variability (as evidenced by the standard deviations and the interquartile ranges) when compared to the high frequency cue words. To validate if the detected difference is statistically significant, an ANOVA test was computed with the association strength score as the dependent variable, and frequency and word type as the fixed factors.

Table 8. The effect of word frequency and type on association measures

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	**η^2**
Frequency	0.01062	1	0.01062	0.6463	0.424	0.008
Type	0.00876	1	0.00876	0.5335	0.467	0.007
Frequency ✻ Type	0.00129	1	0.00129	0.0785	0.780	0.001
Residuals	1.24824	76	0.01642			

The results indicate that lexical frequency ($F(1, 76) = 0.6463$, $p = 0.424$, $\eta^2 = 0.008$) or word type ($F(1, 76) = 0.5335$, $p = 0.467$, $\eta^2 = 0.007$) have no significant main effects on the association strength scores. In addition, the interaction between lexical frequency and word type ($F(1, 76) = 0.0785$, $p = 0.780$, $\eta^2 = 0.001$) has no significant effect on the association scores of the responses. To be more precise, although certain patterns are observed in the participants' responses, the results of the ANOVA test suggest that neither word frequency nor word type, nor their combination, have a statistically significant effect on the association scores of the responses in the word association task.

Even if no significant differences between word types and frequency categories were detected, the overall trends across the variables were illustrated in Figure 3. The figure indicates the possible effect of frequency class on the associative strength scores of the responses for two different word forms.

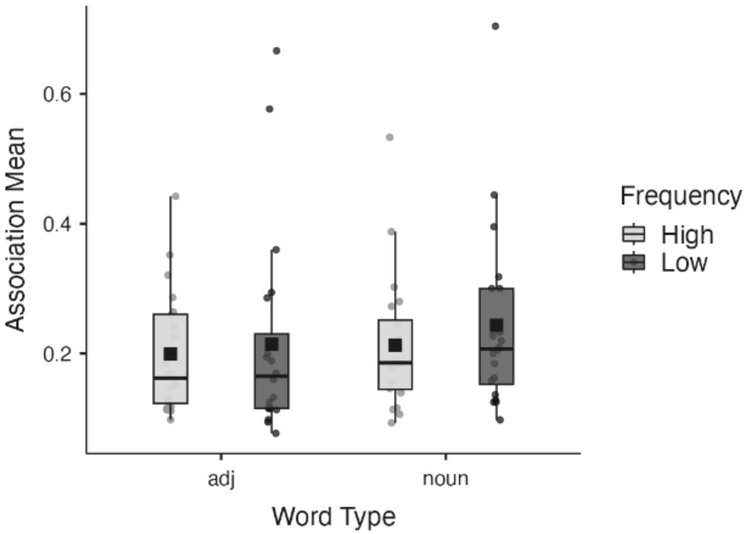


Figure 3. Association scores across frequency and word types

The figure depicts that nouns reveal slightly higher association scores than adjectives. The difference between the association strength scores of adjectives is smaller than the scores of nouns. More interestingly, the nouns in the lower frequency category indicate a higher association strength score, which means the participants agreed more on their responses in the noun cue category, particularly when the cue words have lower frequency (e.g., attitude, variety, etc.).

3.4. The Effect of Word CEFR Level on Word Associations

Building on the analysis in the previous sections, Figure 4 illustrates the overall tendencies across word types considering the lexical items in different CEFR levels and Table 9 displays the results of an ANOVA test suggesting the effect of word level, word type, and their interaction on word association scores.

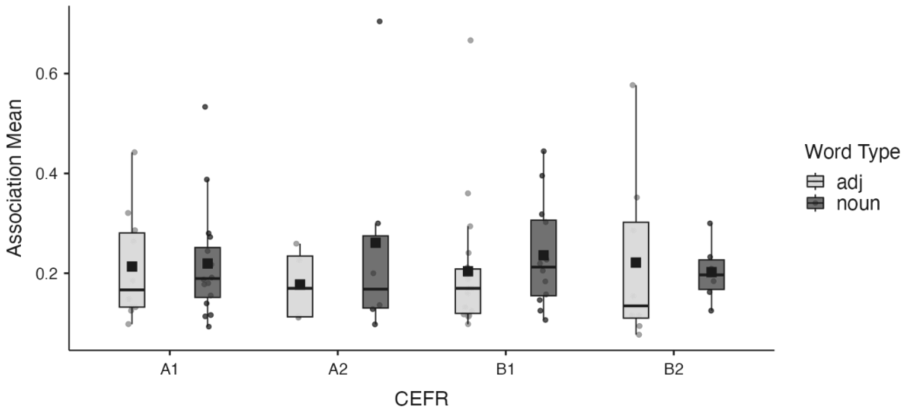


Figure 4. The distribution of association scores across CEFR levels

The figure illustrates how word association scores are scattered across CEFR levels of the words (A1 to B2) for both adjectives and noun cues. Overall, one can see no strong, consistent increase or decrease in association mean scores as the CEFR level of the cue words progresses from A1 to B2 for either word type. Additionally, both adjective and noun cue words indicate considerable variability within each level, signalled by the spread of the data points and the length of the box plots. The fluctuations in the median and mean across the levels are visible; however, these variations do not indicate a clear linear trend. Outliers also exist across all the levels represented for both the word types, suggesting that some individual cue words triggered lexical items that have notably higher or lower association strength scores compared to the general distribution within their respective categories. As an overall observation, the mean association strength scores appear to remain relatively stable across all levels, which could postulate that the strength of word associations does not dramatically differ as the level of the cue words vary between A1 and B2.

Although an overall trend of nouns having higher association scores than adjectives in each CEFR level is observed in the visual output, the ANOVA in Table 9 suggests no significant main effect of word level ($F(3, 72) = 0.0133$, $p = 0.998$, $\eta^2 = 0.001$) or word type ($F(1, 72) = 0.6044$, $p = 0.439$, $\eta^2 = 0.008$) on association strength scores. Additionally, according to the ANOVA, no significant interaction between word level and word type ($F(3, 72) = 0.3312$, $p = 0.803$, $\eta^2 = 0.013$) is observed.

Table 9. The effect of word level and type on association measures

	Sum of Squares	df	Mean Square	F	p	η^2
Word CEFR Level	6.86e-4	3	2.29e-4	0.0133	0.998	0.001
Word Type	0.0104	1	0.01042	0.6044	0.439	0.008
Level ✻ Type	0.0171	3	0.00571	0.3312	0.803	0.013
Residuals	1.2420	72	0.01725			

4. OVERALL DISCUSSION

As Collins and Quillian (1996), Collins and Loftus (1975), and Stevyers and Tenenbaum (2005) suggest, word knowledge in the mental lexicon is stored as nodes in an interconnected semantic network. In Steyvers and Tenenbaum's (2005) model, when we encounter a word (or when it is activated), some activation spreads onto other lexical nodes and the degree of spreading activation vary in strength, which could stem from varying individual experiences. The more a node is activated and the more its activation spreads onto the other related nodes, the more robust the interconnected links seem to become. Thus, this repeated spreading activation could increase the possibility of retrieving a related word when a stimulus is presented to language users. The word association task (Fitzpatrick, 2007) builds on the idea that the words a person knows are stored in a network of interconnected nodes in the lexicon. As De Deyne et al. (2019) and Playfoot et al. (2018) state the "first word that comes to mind" in a word association task should be the most strongly associated concept in the lexicon since the strongest activation ought to be passed from the stimuli (i.e., activated cue) along the strongest intra-lexical networks. The precise associations triggered by the cue words could vary from person to person; however, some commonalities and patterns can still be observed in the responses and these responses are primarily employed to create word association norms (e.g., De Deyne et al., 2019; Nelson, McEvoy & Schreiber, 2004). De Deyne et al.'s (2019) "Small World of Words" database is one of these attempts to investigate word association norms in different languages and the dataset in L1 English was used as a reference to compare with L2 English users' responses in the current study. Results from this type of mega studies and publicly available databases have revealed certain correlations between association strength scores computed via word associations tasks and human judgements of semantic similarity. These results also appear to correlate with reaction times in lexical decision tasks (Steyvers, Shiffrin & Nelson, 2004; De Deyne, Navarro, & Storms, 2012). To put it in a nutshell, these results indicate that the association networks emerging from word association tasks go beyond rudimentary measures of semantic similarity for psycholinguistic experiments, offering useful insights into lexical processing and bilingual mental lexicon.

Having its spark in Collins and Loftus' Spreading Activation Model (1975), using the main association categories of Fitzpatrick's (2007) taxonomy, the current study has employed a word association task and De Deyne et al.'s (2019) database to explore the difference and similarities between the L1 English and L1 Turkish-L2 English users' lexical choices in a word association task, scrutinise the possible effect of lexical frequency, word type and cue word level on the association strength scores.

4.1. Differences and Similarities Between the Responses in L1 English and L2 English

When the data from both groups are analysed (Table 4 and 5), both similarities and differences in word associations between L1 and L2 English speakers can be observed. The patterns reveal that a certain set of lexical items seem to trigger identical first-word associations in both groups, indicating potential shared associations in their lexicon. Though not reported in the current study, the lexical overlap expands when considering second lexical choices in the association task. To be more precise, half of the items the L2 English participants produced either as their first or second choice appears to intersect with the L1 English users' responses. On the other hand, a key distinction is prevalent in the syntagmatic (collocational/position-based) and paradigmatic (meaning-based) organisation of the lexicon. As the SWOW database indicates, L1 English users tend to produce words with paradigmatic associations (semantic relationships) with the cue words, whereas L1 Turkish-L2 English users tend towards syntagmatic associations (collocational relationships) in their responses. This finding aligns with most of the earlier research (e.g., Håkansson & Norrby, 2010; Zareva & Wolter, 2012), suggesting the effect of language proficiency on associative patterns and highlighting the fact that L2 learners often rely more on context and collocations. There are also studies stating that if participants are L1 users or advanced L2 users, they are more inclined to produce paradigmatic responses (Jiang & Zhang, 2019; Zareva, 2007). Furthermore, if the cue words are participants' L1, they are more likely to respond with synonyms, which is also a meaning-based response (Fitzpatrick, 2006; Fitzpatrick & Izura, 2011). The current study supports this idea, showing some similarities and the different tendencies between the groups and how L2 English users' associations reflect their learning process and reliance on language input.

4.2. The Effect of Word Type on Word Associations

As is evident in Tables 4-5 and Figures 1-2, there are different types of associations based on the two different word types under scrutiny. Adjective cues in the word association task elicited responses with more syntagmatic (collocational) associations, while noun cues tended to prompt more paradigmatic (semantic) associations. This finding is in line with some earlier research (e.g., Nissen & Henriksen, 2006) stating that if the cue words are nouns, the chances are higher for the participants to respond with a meaning-related lexical item. Furthermore, earlier research also suggests that if the cue words are adjectives, the participants tend to produce a position-based (collocational) items (Nissen & Henriksen, 2006). This could eventually indicate that different word types play distinct roles in language, with adjectives more closely connected with the nouns which play a central role in semantic networks. Furthermore, the associative patterns of L1 English users demonstrate a greater tendency towards generating synonyms for adjective cues as opposed to L2 English users who produced slightly more synonyms for the noun cues. This finding also goes hand in hand with some earlier research which conclude that L1 users are likely to produce more synonymous responses than L2 users (Fitzpatrick & Izura, 2011). Despite the distinct patterns detected, the results of the ANOVA test reveal that the effect of word types on the association strength scores is not statistically significant. This could tentatively mean that while word type affects the type of associations based on the participants' responses, the strength of association scores is not influenced to a great degree, which might encourage us to ask further questions and conduct more fine-grained research to better understand the complex structure of (bilingual) mental lexicon (see Fitzpatrick & Thwaites, 2020 for a comprehensive review).

4.3. The Effect of Lexical Frequency on Word Associations

The study also explored the potential effect of cue word frequency on L2 users' responses and the detected word associations. The overall results indicate that higher frequency cue words trigger greater consistency in responses between L1 and L2 English users. This may suggest that more frequently used lexical items due to frequent exposure strengthen the connections in the mental lexicon of both the groups under investigation. There are studies in the literature stating that word frequency and age of acquisition (AoA) exhibit a strong relationship with the centrality measures observed in networks derived from word association data in the lexicon (e.g., De Deyne & Storms, 2008; Meara, 2007). Several other studies also find that low-frequency cue words result in less consistent responses (e.g., Meara, 2009), more blank responses (Higginbotham, 2010) than higher-frequency cue words. These studies investigating the phenomenon in L2 English also suggest that word associations in L2 could be more sensitive towards lexical frequency effects than L1. This has encouraged some researchers to embrace Stolz and Tiffany's (1972) approach that these results should be interpreted as the effect of word familiarity rather than word frequency. Following this approach, Wolter (2001) finds that both L1 and L2 users produce more form-based responses as the cue word familiarity decrease. Research exploring the effect of AoA has more conclusive results, suggesting that the latency of word association responses is inversely correlated with the age of acquisition of the cue word (e.g., Brysbaert, Van Wijnendaele, & De Deyne, 2000). Overall, earlier studies posit that the distributional properties of cue words (i.e., lexical frequency, contextual diversity and AoA) may have a partial role in word associations, but they are poor predictors of the association strength scores (Van Rensbergen, Storms, & De Deyne, 2015).

Although there are observable trends in the descriptive values of the current study, with lower-frequency nouns triggering responses with higher association scores and variability, the output from the ANOVA test reveals that cue word frequency does not affect association strength scores, just like earlier studies (e.g., Van Rensbergen et al., 2015). This finding on the surface seems to contradict with the studies emphasising the robust impact of frequency on the organisation of mental lexicon (e.g., Ellis, 2002). However, studies employing word association tasks have so far claimed that cue word frequency has a weak or a counter effect on word associations. For instance, de Groot (1989) states that higher frequency cues elicit slightly slower response times and more diversity. On the contrary, Brown (1971) asserts that higher frequency cue words yield marginally faster responses. As a result, we may say that the findings are rather controversial and more rigorous research methodologies and further studies with more cue words may be needed for more conclusive results.

4.4. The Effect of Word CEFR Level on Word Associations

The current study also examined the relationship between word level (according to CEFR) and association strength scores and found no strong or consistent correlations between the two variables. Association scores vary to a great extent across the word levels for both word types and the fluctuations do not signal a clear linear trend. This could suggest that the strength of word associations does not particularly increase as the level of the cue words decrease or the level dispersion of cue words in this study are not normally distributed. The lack of effect could also be attributed to the limited number of cue words used in this study. More cue words in different word levels could have yielded a different picture. It may also be the case that the language proficiency of the participants in the study was above a standard level and resembled one another (as evidenced by low standard deviations). Thus, their similar responses due to their similar proficiencies (though speculative due to self-ratings) could misguide the interpretations.

Earlier studies so far have not explicitly categorised cue words by their CEFR level, but they have used partly related concepts like frequency, AoA, or familiarity as discussed in the earlier parts of this section (see an extensive overview in Fitzpatrick & Thwaites, 2020). It can be broadly stated that lower CEFR level may correspond to higher frequency and words acquired earlier in life. For example, Zareva (2007) suggests that higher frequency (i.e., lower CEFR level) cue words are likely to be more entrenched in L2 lexicon, which may lead to more predictable and paradigmatic responses among higher proficiency learners. Fitzpatrick (2007) also states that low frequency (i.e., higher CEFR) cue words tend to elicit responses with weaker association strength. Additionally, more variation among the participant responses is expected, which signals less established connections in the lexicon. Because the current study cannot detect a direct effect of CEFR word level on the association scores and the related literature lacks explicit reference to the potential effect of word level, further research may be required to observe this possible influence.

What is unique about this study and its humble contribution to word association research could be summarised as follows: we have had no previous knowledge of L1 Turkish – L2 English users' word associations tendencies and if (or to what extent) their preferences intersect with L1 English users and if the earlier findings indicating syntagmatic associations for L2 English users also apply to the Turkish context. The findings reveal a similar pattern and partly support earlier remarks.

5. LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Some of the findings in the current research (e.g., the effect of word frequency and word CEFR level) have led us to ask further questions rather than providing a clear answer to the initial research questions. This may be attributed to methodological challenges, the limited number of cue words employed, and the restricted number of participants across different proficiency levels. As an alternative variable, *dispersion,* which indicate a more nuanced measure of word's salience, could have been added to the statistical analysis as a distributional measure. As Adelman, Brown, and Quesada (2006) suggests, the number of different documents in a corpus in which a word appears (i.e., word dispersion) have the potential to provide a better insight into a word's degree of salience than frequency. In addition, though not explored in this study, the potential influence of AoA on word associations could be considered in future research. If words learnt in earlier periods of life help develop lexical knowledge in L1, it may be logical to ask if the same effect is observed in L2 acquisition. Future research should also consider the possible impact of language typology on word association patterns in L1 and L2. The structural differences between a person's L1 and their s L2 might influence how words are linked in their mind. Since Turkish is an agglutinative language, L1 Turkish users may be more inclined to break down words into morphemes, which could in turn affect how they make form-based or meaning-based associations in English. As Fitzpatrick and Thwaites (2020) rightly mention, typological differences pose a challenge for the studies comparing bilinguals' responses across two languages. They also note that not many studies exist trying to incorporate the effect of typology in their analysis and future research should consider investigating that.

Finally, it should be noted that word association responses may vary based on the respondents' strategies, the demands of the association task, and many other individual factors, such as gender and age. Therefore, the patterns reported here, which lack the analysis of all those independent variables, may not be fully accurate models of the human mind, but maybe just a rough sketch of the structure of the mental lexicon.

REFERENCES

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823.

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yilmazer, H., Atasoy, G., Öz, S., Yildiz, İ., & Kurtoğlu, Ö. (2012). Construction of the Turkish National Corpus (TNC*). Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. (pp. 3223-3227) İstanbul. Türkiye.

Bøyum, M. (2016). *A new word association test for Norwegian* (M.A. dissertation). University of Oslo.

Brown, W. P. (1971). A retrospective study of stimulus variables in word association. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 355–366.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*(2), 215–226.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cañas, J. J. (1990). Associative Strength Effects in the Lexical Decision Task. *The Quarterly Journal of Experimental Psychology Section A*, *42*(1), 121-145. https://doi.org/10.1080/14640749008401211

Cangır, H. (2021a). Objective and subjective collocational frequency. Association strength measures and EFL teacher intuitions. *Pedagogical Linguistics*, *2*(1), 64–91.

Cangır, H. (2021b). Lexical Associations in the L1 Turkish Mental Lexicon: Can L1 Lexical Intuition and a Representative Corpus Guide Teaching of Turkish as a Foreign Language Materials? *Dilbilim Dergisi / The Journal of Linguistics*, *0*(37), 45–66. https://doi.org/10.26650/jol.2021.972166

Cangır, H., & Durrant, P. (2021). Cross-linguistic collocational networks in the L1 Turkish--L2 English mental lexicon. *Lingua*. https://doi.org/10.1016/j.lingua.2021.103057

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, *3*, e1. doi:10.1017/S2041536212000013

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment—Companion Volume*. Council of Europe Publishing.

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231.

De Deyne, S., Navarro, D. J., & Storms, G. (2012). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, *15*(5), 824–845.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1001-1018.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. In S. H. Foster-Cohen, M. Medved Krajnovic, & J. Mihaljevic Djigunovic (Eds.), *EUROSLA Yearbook* (Vol. 6, pp. 121–45). Amsterdam, Netherlands: John Benjamins.

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*(3), 319–331.

Fitzpatrick, T., & Izura, C. (2011). Word Association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition*, *33*(3), 373–398.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237–274. https://doi.org/10.1017/S0261444820000105

Håkansson, G., & Norrby, C. E. (2010). Environmental influence on language acquisition: Comparing second and foreign language acquisition of Swedish. *Language Learning*, *60*(3), 628–650.

Higginbotham, G. M. (2010). Individual learner profiles from word association tests: The effect of word frequency. *System*, *38*(3), 379–390.

Jiang, N., & Zhang, J. (2019). Form prominence in the L2 lexicon: Further evidence from word association. *Second Language Research*, 1–22. https://doi.org/10.1177/0267658319827053

Kang, B. (2018). Collocation and word association: Comparing collocation measuring methods. *International Journal of Corpus Linguistics*, *23*(1), 85–113.

Khazaeenezhad, B., & Alibabaee, A. (2013). Investigating the role of L2 language proficiency in word association behavior of L2 learners: a case of Iranian EFL learners. *Theory and Practice in Language Studies*, *3*(1), 108–115.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174.

Kuzu, C. & Cangır, H. (2025). Investigating the Lexical Associations in the L3 Italian Mental Lexicon. In H. Cangır, K. Uzun, T. Can, & E. Oğuz (Eds.), *Exploration of the Intersection of Corpus Linguistics and Language Science* (pp. 191-222). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-8146-5.ch009

Lam, B. P. W., & Sheng, L. (2020). The Nativelikeness Problem in L2 Word-association Tasks: Examining Word Class and Trials. *English Language Teaching*, *13*(5), 125. https://doi.org/10.5539/elt.v13n5p125

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in adolescents and young adults: examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The adolescent brain: Learning, reasoning, and decision making* (pp. 39–66). Washington, DC: American Psychological Association.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistic Circular*, *11*, 28–38.

Meara, P. (2007). Simulating word associations in an L2: The effects of structural complexity. *Language Forum*, *33*(2), 13–31.

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam, Netherlands: John Benjamins.

Mollin, S. (2009). Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory*, *5*(2), 175–200.

Namei, S. (2004). Bilingual lexical development: A Persian–Swedish word association study. *International Journal of Applied Linguistics*, *14*(3), 363–388.

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language, 31*(2), 237–260. https://doi.org/10.1016/0749-596X(92)90013-N

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods*, *36*(3), 402–407.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, *16*(3), 389–408.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Playfoot, D., Balint, T., Pandya, V., Parkes, A., Peters, M., & Richards, S. (2018). Are word association responses really the first words that come to mind? *Applied Linguistics*, *39*(5), 607–624.

Riegel, K. F., & Zivian, I. W. M. (1972). A study of inter- and intralingual associations in English and German. *Language Learning*, *22*(1), 51–63.

Schmitt, N., & Meara, P. (1997). RESEARCHING VOCABULARY THROUGH A WORD KNOWLEDGE FRAMEWORK: Word Associations and Verbal Suffixes. *Studies in Second Language Acquisition*, *19*(1), 17–36. doi:10.1017/S0272263197001022

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. Healy (Ed.), *Experimental cognitive psychology and its applications*: *Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (pp. 237–249). American Psychological Association.

Stolz, S., & Tiffany, J. (1972). The production of'child-like' word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Behaviour*, *11*(1), 38–46.

Thwaites, P. (2019). *Lexical and distributional influences on word association response generation* (Ph.D. Dissertation). Cardiff University.

van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211.

Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic Bulletin & Review*, *22*(6), 1717–1724.

Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, *12*(2), 111–122.

Wilks, C. (2009). Tangled webs…: Complications in the exploration of L2 lexical networks. In T. Fitzpatrick, & A. Barfield (Eds.), *Lexical processing in second language learners: Studies in honour of Paul Meara* (pp. 25–37). Bristol, UK: Multilingual Matters.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69.

Xiaosa, L. & Wenyu, W. (2016). Word class influence upon L1 and L2 English word association. *Chinese Journal of Applied Linguistics*, *39*(4), 440-458. https://doi.org/10.1515/cjal-2016-0028

Zareva, A. (2007). Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization? *Second Language Research*, *23*(2), 123–153.

Zareva, A., & Wolter, B. (2012). The'promise' of three methods of word association analysis to L2 lexical research. *Second Language Research*, *28*(1), 41–67.

Chapter 2: **The Effect of the Turkish Evidential Morpheme on Scalar Implicature Computation**

Onur Evcen

ABSTRACT

Propositions containing scalar quantifiers such as “some” and “all” often allow for two interpretations. For example, the statement “Some of the students passed the exam” can be understood semantically as “Some, and possibly all, of the students passed,” or pragmatically as “Some, but not all, of the students passed”—an inference known as a scalar implicature. Prior research has demonstrated that scalar implicature computation can be influenced by the speaker’s knowledge state, epistemic commitment, and broader contextual factors. Building on this, the present study examines whether the Turkish indirect evidential morpheme -*mIş*, which typically signals indirect experience or inference, impacts the likelihood of scalar implicature generation by suggesting limited speaker knowledge. Specifically, we tested whether -*mIş* reduces pragmatic interpretations by conveying epistemic uncertainty. One hundred and one native Turkish speakers participated in a controlled experiment where they were presented with visual scenarios accompanied by statements that either included the evidential morpheme -*mIş* or a bare (direct) form. Participants answered follow-up questions designed to assess whether they interpreted the scalar term “some” pragmatically or logically. The results support prior findings that the epistemic status of the speaker plays a crucial role in determining whether listeners derive scalar implicatures. Furthermore, our data provides novel empirical evidence that the evidential -*mIş* reliably conveys speaker ignorance or lack of direct knowledge to the hearer, thereby reducing the rate of scalar implicature computation. These findings contribute to our understanding of how grammatical evidentiality interacts with pragmatic inference, and they expand the cross-linguistic literature on scalar implicature and evidential markers.

Keywords: Scalar Implicatures, Speaker Ignorance, Evidentiality, Experimental Pragmatics, Contextual Processing

1. INTRODUCTION

1.1. Scalar Implicatures

When we communicate, we often interpret meanings that extend beyond the literal content of sentences. For example, upon hearing the statement "Some of the students passed the exam," a listener may infer that not all of the students passed, a phenomenon known as scalar implicature. Scalar implicatures arise when speakers use terms that have stronger alternatives along a linguistic scale, such as some versus all. This inference is motivated by the cooperative nature of communication: listeners assume that the speaker deliberately chose the weaker term (some) because the stronger alternative (all) would have been false (Grice, 1975; Horn, 1972). In other words, when listeners encounter the sentence in (1a), they typically infer that the alternative in (1b) is false, thereby favouring the pragmatically enriched interpretation in (1c) over the purely semantic reading in (1d).

(1) a. Some of the students passed the exam.

b. All of the students passed the exam.

c. Some, but not all, students passed the exam.

d. Some, and perhaps all, students passed the exam.

Scalar implicatures are not limited to quantifier words alone. In a language, many concepts can include scalar structures, sometimes without speakers even realizing it. For example, the verb "love," can be viewed as a stronger alternative to the verb "like." Thus, when a speaker says, "I liked the movie," and not "I loved the movie," it can lead the listener to infer that the speaker enjoyed the movie but did not love it, producing a scalar implicature. In addition to verbs of ranking, some degree adjectives can also function as scalar expressions. Indeed, in a study conducted by Sedivy, Tanenhaus, Chambers & Carlson (1999), it was shown that when listeners heard an expression like "Pick up the tall glass" in a context where a short glass was also present, they processed the information more quickly. This finding suggests that when individuals hear an adjective, they immediately generate a comparison set, typically structured along a short-to-tall continuum within a category. Furthermore, some theories propose that modal concepts are similarly structured along a scale. For example, “certainty” asymmetrically entails both “likelihood” and “possibility” (Santorio & Romoli, 2017). In other words, if a proposition is “certain,” then it is both “likely” and “possible” but not the other way around, since the scalar entailment is asymmetrical. An example scale for quantifier expressions representing gradable asymmetry is provided in Figure 5 (adapted from Barner & Bachrach, 2010).

Example Quantifier Scale Between "Some" and "All"

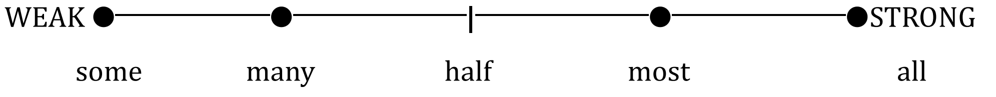


Figure 5

1.1.1. Access Models to Scalar Implicatures

1.1.1.1. Direct Access Model to Scalar Implicatures

According to the Direct Access Models, scalar implicatures arise effortlessly as a natural part of language processing and do not require additional cognitive effort. However, when an implicature needs to be cancelled, this demands extra time and cognitive resources during language comprehension. The rationale behind this is that the pragmatic meaning of scalar expressions is accepted by default, whereas their purely logical interpretation must be derived through an implicature mechanism (Levinson, 2000). In other words, this model proposes that, unless the context signals otherwise, language users initially interpret expressions like "some" as "some but not all," and only later consider broader interpretations such as "at least some." However, it is important to note that communication usually occurs within a broader context. Conversational context and the Question Under Discussion (QUD) are significant factors that can influence scalar implicature processes (Breheny et al., 2006; Skordos et al., 2022). In such cases, the precompiled and readily available nature of scalar alternative pairs, as predicted by the direct access model, plays a crucial role in enhancing efficiency, speed, and accuracy in communication.

1.1.1.2. Indirect Access Model to Scalar Implicatures

Another theory in the literature, the Indirect Access Models, suggests that the lower-bound semantic interpretation is generated quickly and automatically as a byproduct of basic sentence processing. In contrast, all types of inferences, including scalar implicatures, require additional time and cognitive resources. In this view, often referred to as the Literal-First model, it is assumed that in every context, the lower-bound interpretation is established first, and the upper-bound interpretation is considered subsequently (Huang & Snedeker, 2009). Researchers supporting this hypothesis find it consistent with traditional linguistic observations (Horn, 2004). For example, the syntactic and semantic interpretation of a simple proposition containing the word "some" is seen as more basic and primary compared to its upper-bound pragmatic interpretation. That is, the enriched upper-bound meaning ("some but not all") always presupposes the prior construction of the lower-bound meaning ("at least one"). Therefore, the upper-bound interpretation can only arise after the lower-bound interpretation has been processed, implying a two-stage processing model: first, the semantic interpretation of "some" is established, and then, depending on the context, the pragmatic enrichment is activated. Accounts proposing that scalar implicatures arise through four core cognitive stages emphasize the crucial role of mentally representing alternatives. These stages can be outlined as follows (Barner et al., 2011):

(i) *Computation of the basic meaning:* The listener automatically interprets the literal meaning of the utterance.

(ii) *Generation of alternative expressions:* Other possible utterances the speaker could have produced, namely the alternatives that differ in informativeness and are mutually exclusive on a scale, are mentally generated.

(iii) *Elimination of less informative alternatives:* The listener discards alternatives that convey less information, typically those that are more general or less specific, retaining only the more informative ones within the context.

(iv) *Strengthening of the interpretation:* The remaining alternatives are pragmatically strengthened according to the context. At this stage, a sentence like "Some of the students passed the exam" is interpreted as "Some, but not all, of the students passed," yielding a pragmatically enriched meaning.

1.1.2. Scalar Implicature Computation in Neurotypical Adults

There is a large body of literature showing the pragmatic competence of adults, particularly in their ability to compute scalar implicatures across a wide range of languages and contexts. For example, Papafragou and Musolino (2003) showed that Greek-speaking adults overwhelmingly (with an accuracy of 92.5%) rejected under-informative statements like *"Some of the horses* *jumped"* when in fact all had jumped, which indicates pragmatic thinking. Chierchia, Crain, Guasti, Gualmini, and Meroni (2001) investigated English-speaking adults and children and reported similar findings. Another early study by Noveck (2001) used universal factual statements rather than a controlled experimental setup and found that French-speaking adults rejected under-informative statements such as *"Some elephants are mammals"* in 59% of cases. A replication of Noveck (2001) by Guasti et al. (2005) examined scalar implicature computation in Italian-speaking adults and found that they rejected under-informative factual statements like *"Some giraffes have long necks"* with a rejection rate of 50%. While still more pragmatically sensitive than children, who accepted such statements 87% of the time, adult responses in this study were less robust than in Papafragou and Musolino's. As Slabakova (2010) notes, when materials rely on factual generalizations rather than contextually grounded scenarios, adults may interpret sentences through the lens of world knowledge or imagine exceptions. For example, when evaluating *"Some giraffes have long necks,"* listeners might think of baby giraffes that do not, making the statement seem acceptable. A study with Turkish adults by Renans et al. (2018) showed the participants a scene where a lion carried 4/4 apples within the experimental context and asked about the felicity of a *bazı* ("some") sentence uttered by a puppet: *"The lion carried some of the apples"*. Turkish-speaking adults consistently stated that they thought the puppet did not answer well (approximately 85% of the time). The results from crosslinguistic studies show that adults typically compute scalar implicatures in the necessary contexts, especially when there is a carefully controlled experimental environment with rich context, which helps prevent a potential underestimation of the pragmatic performance of adults.

1.1.3. Factors that Influence Scalar Implicature Computation

1.1.3.1 Working Memory

Working memory can be described as the cognitive system that is responsible for temporarily storing and managing information (Alloway & Copello, 2013; Cowan, 2008). Since scalar implicature computation involves considering literal meanings, scalar alternatives, and pragmatic meanings, it is reasonable to expect that working memory capacity might play a role in alleviating this additional cognitive load. Antoniou, Cummins, and Katsos (2016), for instance, examined this link by combining scalar implicature tasks with two working memory assessments: a backward digit span task and a reading span task.

Their results indicated that participants with higher working memory scores were more likely to reject under-informative sentences such as *"There are hearts on some of the cards"* when all the cards visibly contained hearts. However, the link between working memory and scalar implicature computation remains inconclusive. For example, Banga, Heutinck, Berends, and Hendriks (2009) reported no significant difference between individuals with high and low working memory capacity when participants evaluated under-informative factual statements such as *"Some elephants have trunks."*. The dichotomy between the two lines of research might be because of the fact that the experimental designs that require additional effort, such as processing complex visual displays, may accentuate the role of memory capacity. In other words, deriving scalar implicatures may not always be effortful enough to tax memory resources, thereby diminishing observable differences (Heyman & Schaeken, 2015; Antoniou et al., 2016).

For instance, Dieussaert, Verkerk, Gillard, and Schaeken (2011) showed that participants with lower working memory capacity only struggled to make pragmatic judgments when they were also asked to remember complex visual patterns at the same time. When the additional task was easier or absent, even those with lower working memory were just as successful at deriving scalar implicatures as those with higher capacity. This suggests that working memory may only affect implicature computation when the task places extra demands on cognitive resources.

1.1.3.2. Theory of Mind

Theory of Mind (ToM) refers to the ability to attribute mental and epistemic states such as beliefs, desires, intentions, and knowledge to oneself and others and this capacity may help individuals to interpret and predict others' epistemic states and communicative intentions, which is an important component in pragmatic processing (Wellman, 2018; Noveck, 2018). Most studies investigate the relationship between ToM skills and scalar implicature computations by focusing on children since they are still in the process of developing these skills. A study done by Mazzaggio, Foppolo, Job, and Surian (2021) compared typically developing children with children who had autism spectrum disorder (ASD). They found that in children younger than 6;0, ToM skills were positively correlated with scalar implicature computation. Notably, this relationship appears to be independent of lexical knowledge, morphosyntactic skills, and general intelligence. Eventually, understanding others' thoughts, beliefs, and intentions helps children in their pragmatic processing.

In a recent study, Fairchild and Papafragou (2021) recruited English-speaking adults and investigated their ToM skills using the *"Mind in the Eyes Task"* and *"Strange Stories Task"* by Baron-Cohen et al. (1997) and Happé (1994), respectively. Then, they measured the participants' scalar implicature computation by asking them to rate the felicitousness of informative and under-informative sentences. Eventually, they found a positive correlation between scalar implicature computation and ToM skills of the adult participants.

1.1.3.3. Speaker Knowledge States

The ability to reason about what the speaker knows and what he/she intends to communicate is highly significant in cooperative conversations, which requires reasoning about the speaker's knowledge states (Sauerland, 2004). For example, in the sentence "Some of the students passed," the listener makes an inference about the speaker's knowledge state by thinking that the speaker does not believe the proposition "All of the students passed" because, if they did, they would have said so. The speaker's knowledge state significantly shapes this interpretation. For example, if the speaker is a professor who has seen all of the students' grades, the statement implies "some, but not all, of the students passed," which leads to a scalar implicature computation and the formation of the pragmatic meaning. However, if the speaker is a person who has observed only a small subset of the students' grades, the listener might cancel the upper-bounded "not all" meaning by recognizing the limitations in the speaker's knowledge state and can reason accordingly, which will lead to the logical meaning, some, and possibly all, students passed (Goodman & Stuhlmüller, 2013). This interaction between the speaker and the listener shows that the computation of scalar implicatures is highly influenced by context and speaker knowledge.

Empirical data suggests that context and the speaker's perceptual access influence the generation of scalar implicatures. A self-paced reading task by Bergen & Grodner (2012) tested adult native speakers of English. They created stimulus passages that consist of three sentences written from a first-person perspective. These passages were designed by combining two factors: knowledge context (either full or partial) and trigger type (scalar or focused). In the full-knowledge passages, the opening sentence described the speaker as an expert in the topic, indicating they would know whether the stronger alternative was true or false. In contrast, the partial-knowledge passages left the extent of the speaker's knowledge unclear. In the critical passages, the second sentence began with either a scalar quantifier (Some of) or a focused quantifier (Only some of), followed by a noun phrase and predicate. In addition, a final complement sentence was added that started with an anaphor like "the rest" or "the others", implying that the stronger alternative was false (some, but not all). The scalar trigger should have been costlier in processing (Bott & Noveck, 2004; De Neys & Schaeken, 2007) since the scalar implicature had been generated by the readers. They hypothesized that there should be no significant difference between full knowledge and partial knowledge conditions in terms of processing cost. The results of the self-paced reading experiment indicated that the full-knowledge condition caused significantly higher RTs when a scalar quantifier was presented afterwards compared to the partial knowledge condition, possibly because the readers generated a scalar implicature in that condition. Additionally, in the final complement sentence, the full-knowledge condition helped reduce processing costs as it referenced the complement set triggered by the scalar implicature. Crucially, speaker knowledge did not have any effect in the focused quantifier condition. The online data showed that speaker knowledge significantly affected the likelihood of readers computing a scalar implicature.

Another study by Hochstein, Bale, and Barner (2018) investigated the role of speaker knowledge in the scalar implicature computations of neurotypical adults and adolescents with ASD. They introduced the participants to a puppet named Farmer Brown and displayed three boxes on the table. The puppet always opened the first two boxes to see what was inside; in the full-knowledge condition, it also peeked into the third box to see the contents and in the partial-knowledge condition, it did not check what was inside the third box. In the end, the puppet produced some statements about the contents of the boxes as in "all/some/two of the boxes have x". Then, the participants were asked to answer the question of whether there was an x in the third unopened box using categorical responses such as "yes", "no", and "I do not know". Hochstein et al. (2018) predicted that the participants would produce a pragmatic meaning- some, but not all- when the puppet had the full knowledge about the contents of the boxes and would produce literal meaning- some, and possibly all- in conditions where the puppet had not peeked into the third box, partial-knowledge condition. The results indicated that both neurotypical and ASD subjects successfully generated the scalar implicature in full-knowledge trials. In partial-knowledge trials, while neurotypical adults were able to cancel the computation of the pragmatic meaning, ASD subjects were not able to perform well and opted for the pragmatic meaning, overgenerating the scalar implicature. This behaviour of individuals with ASD could not be explained by ToM abilities since they could easily evaluate the knowledge states of their interlocutors when explicitly asked.

1.2 Speaker Knowledge States and Evidentiality

The conveyance of speaker ignorance is bound to a created experimental context in the aforementioned literature where speakers acquire information about the world through direct experience, or through salient contexts provided. Speakers can also update their knowledge states by receiving information from a third party, or by making inferences based on evidence (Karaca, 2018). Some languages overtly mark these instances in their grammar with evidentiality including some South and North American Indian languages, some Caucasian and Balkan languages and Turkish (Aksu-Koç, Ögel-Balaban, & Alp, 2009; Aksu-Koç, 2016; Aikhelvald, 2004; Angelopoulos, Bagioka, & Terzi, 2023; Freidman, 2008).

Evidentiality refers to the "source of information" conveyed in a statement (Aikhelvald, 2004). While all languages have ways of expressing how the speaker knows something, not all languages include evidentiality as a formal grammatical category. In many languages, information source is conveyed lexically rather than grammatically. For instance, English allows optional reference to the origin of knowledge through expressions like *I guess*, *they say*, *I hear that*, or through verbs such as *allege*, and adverbials such as *reportedly* (Aikhenvald, 2004). In Russian, hearsay can be marked with particles such as *jakoby*, *mol*, and *deskatj* (Rakhilina, 1996, as cited in Aikhelvald, 2004). Similarly, Japanese employs adverbial expressions that reflect the speaker's stance and signal the evidential source of the information (Aoki, 1986, as cited in Aikhelvald, 2004). On the other hand, some languages (especially South & North American Indian, Caucasian, Balkan, and Turkic languages) have evidentiality systems embedded as a category within their morphology and syntax, using clitics (Shipibo-Conibo, Tamil, Ngiyambaa) affixes (Quechua, Tariana, Eastern Pomo, Aymara, Tucano), and additional tense-aspect paradigms (Turkic, Iranian, Finno-Ugric languages, Bulgarian) (Aikhelvald, 2004; Bashir, Butt, & King, 2006; Freidman, 2008; Fitneva, 2008; Johanson, 2008; Aksu-Koç, Ögel-Balaban, & Alp, 2009; Murray, 2017; Angelopoulos, Bagioka, & Terzi, 2023).

1.2.1 Evidentiality in Turkish

Turkish is an agglutinative language that has an extensive inflectional morphology. In this code, the past events are marked by two distinct suffixes: –DI and –mIş. While the suffix –DI indicates direct experience, –mIş conveys indirect experience (hearsay) or knowledge inferred from evidence (Aksu-Koç & Slobin, 1986; Aksu-Koç, 1988). Example sentences representing the verbal inflection paradigm for coding evidentiality in Turkish are presented in (2) and (3) (Arslan, 2016).

(2) Adam elma-y-ı ye-di

Man apple ACC eat DIRECT EVID

"The man ate the apple" [witnessed]

(3) Adam elma-y-ı ye-miş

Man apple ACC eat INDIRECT EVIDENTIAL

"The man ate the apple" [reported or inferred]

(Arslan, 2016, 116)

Arslan (2016) argues that while in (2), the speaker personally experienced or directly observed the apple being eaten; in (3), the speaker did not witness the event but has come to know about it through indirect means such as hearing it from someone else or by making a logical guess based on clues. For instance, if there are apple cores and peels left on the table, the speaker might deduce that someone ate an apple and use the evidential morpheme to describe that (also see, Aksu-Koç, 2016).

Johanson (2008) specifies three different uses of evidentials in Turkish: Reportive use, inferential use, and perceptive use, presented in (4), (5), and (6), respectively.

(4) Reportive Use: The knowledge is not firsthand but based on what others have said, such as hearsay or reported speech.

a. Bakan hasta-y-mış

minister sick-COP-EVID

"The minister is reportedly sick" (uttered by someone who knows about the sickness)

(5) Inferential Use: The speaker makes a statement based on internal reasoning or logical deduction.

b. Uyu-muş-um
sleep-EVID.PAST-1SG
"I have obviously slept" (uttered by someone who has just woken up)

(6) Perceptive Use: The speaker's statement is grounded in personal sensory experience, either through directly observing the event or noticing signs of it afterward.

c. İyi çal-ıyor-muş
well play-PROG-EVID
"She is, as I hear, playing well" (uttered by someone listening to the play)

(Johanson, 2008, 274)

The usages in (4), (5) and (6) depend on the context in which they are used in (Stott, Smith, Chang, & Bond, 2010). The same sentence in (4a), for example, may also refer to inferential and perceptive uses within the salient context (Johanson, 2008).

1.3 Current Study

Although various studies have documented evidence of the relationship between scalar implicature and speaker ignorance, no previous study has investigated the role of the evidential in conveying such mental states. If the Turkish evidential morpheme *-mIş* conveys speaker ignorance in a restricted experimental context, and Turkish native speakers opt for the so-called inferential use in such a context (Johanson, 2008), then the following research question comes to discussion:

RQ 1: Does the Turkish evidential morpheme affect the computation of scalar implicature in young adults by hinting them the contextual ignorance of the speaker?

To address the question, an experimental environment where there is only the participant, speaker, and the stimuli was created. The participants were explicitly told that this was a game that hosted only the participant and the cartoon character, in an attempt to rule out the reportive use of -mIş (Johanson, 2008). The reasoning was to compare predicates that had bare forms with predicates that had *–mIş*, the former implies a context where the speaker had the full-knowledge, and the latter implies that the speaker had the partial-knowledge, and speaking through inference based on evidence- using the inferential use of Turkish evidential morpheme (Johanson, 2008). If participants did not consistently generate a pragmatic meaning in the *-mIş* condition, then this could suggest that Turkish evidential had an effect on the computation of scalar implicatures because it conveyed the speaker's knowledge state to the participants. Thus, this study hypothesizes that neurotypical adult speakers of Turkish will compute scalar implicatures less frequently in contexts where the predicate contains the evidential morpheme *-mIş*, compared to contexts where the predicate is in bare form. This difference is expected due to *-mIş* signalling speaker ignorance, thereby weakening the implicature trigger.

2. METHODS

2.1 Participants

We recruited 101 native Turkish speakers (mean age = 20.6, *SD* = 2.5, *N* = 96 post exclusions). The majority of the participants were university students recruited in a quasi-experimental manner using convenience sampling. As per our exclusion criteria, we removed six participants failing to respond to attention checks for more than half of the time (2 incorrect responses out of 4 attention checks).

2.2 Materials & Procedure

The stimuli were programmed using PCIbex Farm experimental software (Zehr & Schwarz, 2018). The stimuli consisted of 4 attention and 6 critical trials that comprised of sentences that had the Turkish evidential morpheme *-mIş* (3 sentences) and bare forms with no morpheme (3 sentences). The participants were presented with text and pictures through the screens of their own devices. Prior to the experiment, the participants received preliminary instructions via text, explaining that their task in each trial was to guess the contents of a closed, third box, with assistance from a speaker in the form of a cartoon character, named Ali, that would share what he knew about the boxes via text using speech balloons. The participants were told that in the experiment, there were no third parties involved, and this was a game between the participant and the character.

Every trial introduced the speaker character and a row of three boxes. Two of the boxes were open and the contents were visible to both the character and the participants. There was a third, closed box in the scene. In all of the trials, the contents of the two open boxes were identical (a banana in both). A different fruit was used for each sentence. The closed box paradigm and experimental setting was adapted from Hochstein et al. (2018). Example trials from the experiment can be seen in (7a-d).

(7)

a. Full knowledge/all [attention]

Two of the boxes show fruit along with a third, closed box. Then, the speaker said,

'Kutuların hepsinde elma var'

kutu-lar-ın hepsi-nde elma var

box-PL-GEN all-LOC apple exist

'All of the boxes have apples'

b. Full knowledge/some: [critical]

Two of the boxes show fruit along with a third, closed box. Then, the speaker said,

Kutuların bazılarında elma var

kutu-lar-ın bazı-ları-nda elma var

box-PL-GEN some-PL-LOC apple exist

'Some of the boxes have apples.

c. Full knowledge/two: [attention]

Two of the boxes show fruit along with a third, closed box. Then, the speaker said,

Kutuların ikisinde elma var

kutu-lar-ın iki-si-nde elma var

box-PL-GEN two-POSS-LOC apple exist

'Two of the boxes have apples'

d. Partial knowledge/some: [critical]

Two of the boxes show fruit along with a third, closed box. Then, the speaker said, Kutuların bazılarında elma varmış

kutu-lar-ın bazı-ları-nda elma var-mış

box-PL-GEN some-PL-LOC apple exist-EVID

‘It appears that some of the boxes have apples’

In every trial, participants were presented with a multiple-choice question via text asking, ‘Do you think that there is an apple/banana/orange inside the third box?’ with the alternatives “Yes”, “No”, and “I don’t know”. If listeners considered the speaker's knowledge when interpreting implicatures, they would compute a scalar implicature and respond "No" when the speaker had the full knowledge about the contents of the third box. In contrast, if the speaker had partial knowledge and was just inferring based on the evidence, participants would not compute a scalar implicature and would instead respond with "I don’t know."

In addition to the critical trials, the stimuli also had 4 attention trials which included 2 ‘yes’ and 2 ‘no’ answers to help confine order effects. The first type of attention check trial was full knowledge/all condition where the speaker obviously gave a hint about the contents of the third box, which required a ‘yes’ answer for the question of whether there was a banana in the third box, where there were bananas in the open boxes. The second type was full knowledge/two condition, where the required answer was ‘no’, following from the logic that if 2 had bananas, then the third one should not have had bananas.

The order of the trials was pseudorandomized in such a way that it always started with an attention trial followed by two critical trials of the same type. There was no incident of bare form critical condition following the evidential critical condition or vice versa to control for carryover effects. The attention trials served as filler items between the critical conditions to distract participants from explicitly seeing the difference in trials. It should be noted that the ordering arrangement in the experiment is different than what was mentioned in the pre-registration. Randomly shuffling the order of the stimuli might have brought the risk that a direct trial, directly followed by an evidential trial, to be possibly present for some participants, which would create carryover effects. Therefore, a pseudorandomized order was preferred.

3. RESULTS

To ensure that participants were generally attentive to the task, the proportion of selections on the attention check trials was analysed across conditions. In the *attention all* condition, the proportion of "Yes" responses was high (*M* = 88.42, *SD* = 4.41), while the *attention two* condition showed a different pattern, with a higher proportion of "No" responses (*M* = 84.73, *SD* = 4.95). These results suggest that participants were attentive to the task, though not at ceiling levels, with some variability in their responses. For the critical trials, the proportion of selections on the critical trials varied across conditions. In the *direct* condition, the most common response was "No" (*M* = 60.00, *SD* = 8.27), followed by "I don't know" (*M* = 34.03, *SD* = 7.99), and a smaller proportion of "Yes" responses (*M* = 5.96, *SD* = 3.99). In the *evidential* condition, "I don't know" responses were most common (*M* = 60.83, *SD* = 8.25), followed by "No" responses (*M* = 35.31, *SD* = 8.08), and a small proportion of "Yes" responses (*M* = 3.84, *SD* = 3.25). We excluded trials where participants responded "Yes" to critical trials. While "Yes" responses were possible, they were not expected, as the listener did not have direct evidence of what was contained in the third button. These responses were rare, making up only 4.93% of the dataset (28 data points). Figure 6 below shows a distribution plot of the proportion of "no" responses (compared to "I don't know" responses) for each marker type (direct vs. evidential) in critical trials.

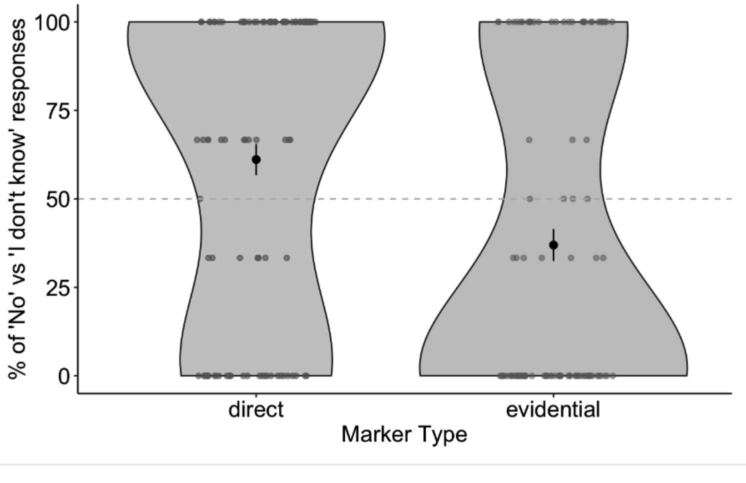


Figure 6. Proportion of "no" responses (vs "I don't know" responses) in each Marker Type (direct, evidential) in critical trials. Each dot represents a participant. Error bars represent standard error of the mean.

As shown in Figure 6, to investigate whether participants interpreted differently based on Marker Type (direct, evidential), we built generalized logistic mixed-effects models predicting the proportion of “No” responses (vs. “I don’t know” responses) based on fixed effects of Marker Type, with random intercepts for participant and item. Adding Marker Type to a null model significantly improved the model fit ($\chi^2(1) = 19.13$, $p < .001$), indicating that participants’ responses varied significantly based on the condition. Further, we added Marker Type as a random slope by participant, which further improved the model fit ($\chi^2(2) = 51.75$, $p < .001$). The best-fitting model revealed a main effect of Marker Type, with significantly more "No" responses in the Direct Marker condition, suggesting that participants derived the implicature more frequently in this condition than in the Evidential Marker condition ($\beta = -4.78$, $SE = 1.34$, $z = -3.56$, $p < .001$).

4. DISCUSSION

The present study investigated the role of evidentiality in the computation of scalar implicatures by focusing on the Turkish evidential morpheme *-mIş.* We aimed to determine whether the evidential marker, which can convey speaker uncertainty in certain contexts, would reduce scalar implicature computation compared to the direct bare forms. Since the findings revealed a significant difference between statements that included the evidential morpheme *-mIş* and a bare (direct) form, we found three main results. Firstly, we found that Turkish neurotypical adults could successfully compute scalar implicatures, judging by the common “No” responses in the *direct* condition. Secondly, we found that Turkish neurotypical adults could consistently cancel the implicature and opt for the literal meaning when the speaker was epistemically uncertain and talking based on inference, judging by the common “I don’t know” responses in the *evidential* condition. Following this, thirdly, Turkish evidential morpheme *-mIş* successfully conveyed the epistemic state of the speaker in a controlled experimental context, leading participants to give significantly more “I don’t know” responses to conditions where the evidential is present.

The results supported prior research on Turkish adults’ pragmatic competence by Renans et al. (2018). Our participants thought pragmatically in the presence of the existential quantifier “some” when the sentences were presented with bare predicates. These findings also add to the growing body of literature suggesting that the hearer’s interpretation of scalar terms is sensitive to the speaker's epistemic stance, as previously demonstrated in studies such as Hochstein et al. (2018) and Bergen & Grodner (2012). Unlike studies that manipulate speaker knowledge via broader contextual manipulations, our study shows that a grammatical evidential marker on its own—without additional cues—can significantly reduce the likelihood of scalar implicature computation. This supports claims that implicature computation is not purely a matter of semantics or pragmatics in isolation but is deeply influenced by how hearers assess the speaker's knowledge state. In particular, the results align with models of implicature that consider Theory of Mind and inferencing processes as central components (Fairchild & Papafragou, 2021).

Though our experiment successfully showed that *-mIş* can signal epistemic uncertainty, one alternative explanation is that the "I don't know" responses might not reflect the cancellation of implicatures per se, but rather a reluctance to commit to an answer under epistemic uncertainty. Participants might have interpreted *-mIş* as a signal to withhold judgment rather than as a cue to reinterpret scalar terms literally. Further research exploiting online methods, such as reaction time studies, could help us understand these possibilities.

There are also some limitations to our study. For instance, our participants were adult neurotypical speakers studying in a private university and the results may be different for the speakers from different socio-economic backgrounds. It is also unclear if the same benefits apply to people with atypical language or cognitive profiles or maybe to children, whose evidential and pragmatic skills are still developing. By studying larger populations, future research could answer these questions. The degree to which these mechanisms are universal or language-specific may also be revealed by cross-linguistic studies comparing other languages that overtly mark evidentiality in their grammar.

5. SUMMARY AND CONCLUSION

The study investigated whether the Turkish evidential morpheme *-mIş* can have an influence on scalar implicature computation by hinting information about the speaker's knowledge states. Building on prior research showing that pragmatic inferences are shaped by contextual factors such as speaker knowledge, our goal was to explore whether an overt grammatical marker of epistemic uncertainty could similarly modulate scalar implicature generation. We designed an experiment in which Turkish-speaking adults were asked to interpret sentences containing either bare forms or the evidential *-mIş* in a closed-box paradigm adapted from previous work (Hochstein et al. 2018). The results revealed three central findings. First, participants opted for the pragmatic "some, but not all" meaning in the bare predicate condition, as indicated by frequent "No" responses when the sentence contained "some." Second, in the *-mIş* condition, participants significantly reduced implicature computation and instead opted for the "I don't know" response, suggesting that the evidential marker successfully conveyed speaker uncertainty. Third, this difference in response patterns argues for the fact that Turkish speakers can interpret evidential morphemes pragmatically and adjust their interpretation of scalar quantifiers accordingly.

These findings can add to the broad literature on the interface between syntax, semantics, and pragmatics along with novel experimental evidence from Turkish which shows how language-specific grammatical features interact with general pragmatic mechanisms. The findings also suggest that Turkish evidentiality might be operating within an epistemic framework which can include the modal domain. This is in line with Palmer's (2001) discussion about the evidential being a distinct modal category and Aksu-Koç's (2016) discussion about Turkish evidentiality, signalling the need for future studies on the possible interface between Turkish evidentiality and epistemic modality.

REFERENCES

Aikhenvald, A. Y. (2004). *Evidentiality*. OUP Oxford.

Aksu-Koç, A. (1988). The acquisition of aspect and modality. *(No Title)*.

Aksu-Koç, A., & Slobin, D. I. (1986). A psychological account of the development and use of evidentials in Turkish.

Aksu-Koç, A., Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, *2009*(125), 13-28.

Aksu-Koc, A. (2016). The interface of evidentials and epistemics in Turkish. In M. Güven, D. Akar., B. Öztürk, & M. Kelepir (Eds.), Exploring the Turkish Landscape: Essays in honor of Eser Erguvanli-Taylan (pp. 143-156): John Benjamins, which is an important study on the topic.

Alloway, T. P., & Copello, E. (2013). Working memory: The what, the why, and the how. *The Educational and Developmental Psychologist*, *30*(2), 105-118.

Angelopoulos, N., Bagioka, D. V., & Terzi, A. (2023). Exploring syntactically encoded evidentiality. *Language Acquisition*, *30*(1), 50-75.

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics, 99*, 78–95.

Aoki,H.(1986).'Evidentials in Japanese',in Chafe and Nichols (eds.) (1986: 223–38).

Arslan, S. (2016). Neurolinguistic & psycholinguistic investigations on evidentiality in Turkish.

Banga, A., Heutinck, I., Berends, S. M., & Hendriks, P. (2009). Some implicatures reveal semantic differences. *Linguistics in the Netherlands, 26*, 1–13. https://doi.org/10.1075/avt.26.02ban

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive psychology*, *60*(1), 40-62.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84-93.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. Journal of Child Psychology and Psychiatry, 38(7), 813–822.

Bashir, E., Butt, M., & King, T. H. (2006). Evidentiality in South Asian Languages. In *Proceedings of the LFG06 Conference, Universität Konstanz* (pp. 30-50).

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, 51(3), 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434-463.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University conference on language development* (Vol. 25, pp. 157-168).

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, *169*, 323-338.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental psychology*, 54(2), 128-133.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *Quarterly Journal of Experimental Psychology, 64*(12), 2352–2367. https://doi.org/10.1080/17470218.2011.588799

Fairchild, S., & Papafragou, A. (2021). The role of executive function and theory of mind in pragmatic computations. *Cognitive Science*, *45*(2), e12938.

Fairchild, S., & Papafragou, A. (2021). The role of executive function and theory of mind in pragmatic computations. *Cognitive Science*, *45*(2), e12938.

Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of child language*, *35*(4), 845-868.

Friedman, V. A. (2008). Evidentiality in the Balkans with special attention to Macedonian and Albanian. In *Studies in evidentiality* (pp. 189-218). John Benjamins Publishing Company.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1), 173-184.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill.

Happe, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and ´ feelings by able autistic, mentally handicapped, and normal children and adults. Journal of Autism and Developmental Disorders, 24(2), 129–154.

Heyman, T., & Schaeken, W. (2015). Some differences in some: Examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica, 55*(1), 1–18. https://doi.org/10.5334/pb.bc

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar implicature in absence of epistemic reasoning? The case of autism spectrum disorder. *Language Learning and Development*, 14(3), 224-240.

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Horn, L. R., & Ward, G. L. (Eds.). (2004). *The handbook of pragmatics* (p. 3). Oxford: Blackwell.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376-415.

Johanson, L. (2008). Evidentiality in Turkic. In *Studies in evidentiality* (pp. 273-290). John Benjamins Publishing Company.

Karaca, F. (2018). Comprehension of evidentiality in spoken Turkish: Comparing monolingual and bilingual speakers.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of communication disorders*, *90*, 106089.

Murray, S. E. (2017). *The semantics of evidentials* (Vol. 9). Oxford University Press.

Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge University Press.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165-188.

Palmer, F. R. (2001). *Mood and modality*. Cambridge university press.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253-282.

Rakhilina, E. (1996).'Jakoby comme procédé de médiatisation en russe', in Guentchéva (ed.) (1996: 299–304).

Renans, A., Sag, Y., Ketrez, N., Tieu, L., Folli, R., Tsoulas, G., ... & Romoli, J. (2018). Plurality and cross-linguistic variation: An experimental investigation of the Turkish plural.

Santorio, P., & Romoli, J. (2017). Probability and implicatures: A unified account of the scalar effects of disjunction under modals. *Semantics and Pragmatics*, *10*, 13-1.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy*, *27*, 367-391.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109-147.

Skordos, D., Myers, A., & Barner, D. (2022). Quantifier spreading and the question under discussion. *Cognition*, *226*, 105059.

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua*, *120*(10), 2444-2462.

Stott, A., Smith, M., Chang, T., & Bond, A. (2010). Which-mis is MIS?: Turkish indirectivity and negative scope. *Evidence from Evidentials*, 263.

Teresa Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes*, *20*(5), 667-696.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728-755.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

Chapter 3:
Morphological Generalization in L2 Turkish: The Case of Turkish Aorist

Serkan Uygun

ABSTRACT

The question of which mechanisms speakers use to generalize inflectional patterns to novel words is controversial. While some researchers have argued that morphological generalizations are based on associative memory representations, others have claimed that they are rule-based. A third group posits both associative and rule-based mechanisms (i.e., dual-morphology models) for morphological generalization. In order to test these views, 30 L1 Turkish and 31 L2 Turkish speakers were tested via a nonce verb production experiment. The test case was Turkish aorist, which includes both regular and irregular forms. A written elicited-production experiment containing nonce verbs with varying degrees of similarity to existing verbs was administered to L1 and L2 Turkish speakers. The elicited-production responses were first coded as regular vs. irregular aorist forms, and then they were compared using linear mixed effect models. While both the L1 and the L2 groups seemed to employ rule-based mechanism for regular forms and associative mechanism for irregular forms, the L2 group had a tendency to use regular forms to a lesser extent than the L1 group. The L2 group also exhibited more variability in applying rule-based mechanism. These results indicate similarities in the generalization processes of regular and irregular aorist forms in the L1 and L2 speakers of Turkish, providing additional support for the dual-morphology models. However, the difficulty in applying the rule-based generalization in the L2 group may be taken to indicate that L2 learners rely less on rule-based processes for morphological generalization than L1 speakers.

Keywords: morphological generalization, Turkish aorist, variability, L1 and L2 Turkish

Words have been viewed as the core element of our linguistic knowledge, which enables us to communicate with other people. If a word is in its bare or root form, it is labelled as a *monomorphemic* word (e.g., hair). There are also *multimorphemic* words that are formed by the combination of a root and an affix (e.g., hairs, hairy) or two roots (e.g., hairbrush). The affixes that are attached to root forms are divided into two categories: derivational morphemes and inflectional morphemes. Derivational morphemes (e.g., hair + *y*) have a lexical function, may change the word class, and may cause some meaning change. In addition, while some derivational morphemes can be very productive, most of them are nonproductive. Conversely, inflectional morphemes (e.g., hair + *s*) have a grammatical function (i.e., they mark properties such as tense, number, person, and so forth), do not cause a change in the word class and meaning, are very productive, and are often required by the rules of grammar (Fromkin et al., 2011). Researchers have mostly been interested in the linguistic representation of inflected words for the last four decades and have conducted numerous studies because of a unique feature inflected forms have: they apply freely to nearly every root form except for the *irregular* forms.

The linguistic representation of regular vs. irregular inflection, which is also known as the English past tense debate (Pinker, 1999; Pinker & Ullman, 2002), has been the main source of the ongoing discussions. The productivity of the inflected forms has enabled the researchers to use new and novel word forms that the speaker has never encountered before, and different views have been proposed regarding the morphological generalization processes of inflected forms. Elicited production studies conducted with first-language (L1) English speakers revealed that they used two distinctive mechanisms in their morphological generalization processes. The rule-based generalization mechanism considers the rule (i.e., add *–ed* to any verb) as the default form and applies it easily and rapidly to any word form, known or novel, that the speaker encounters. On the other hand, the associative generalization mechanism takes the novel word form's phonological overlap with the previously known form and its frequency as the basis for morphological generalization. These two mechanisms have led the researchers to propose two distinct models for morphological generalization. According to the single-mechanism model (Rumelhart & McClelland, 1991), speakers use the phonological overlap between the known words and the novel words and conjugate the verb in its regular or irregular form based on this phonological overlap. This model claims that all forms, whether regular or irregular, are generalized by employing the associative generalization mechanism only. In contrast, the dual-mechanism model (Clahsen, 1999; Pinker & Prince, 1988) posits that two independent mechanisms are employed in morphological generalization processes. The first one is the rule-based mechanism, which is used in the generalization process of regular forms and applies the default rule rapidly. For the generalization of the irregular forms, the associative mechanism is employed, and this mechanism benefits from the phonological overlap between the novel and real forms. According to the dual-mechanism model, the mental lexicon represents and processes regular and irregular forms in different ways. The crucial difference between these two models is in the generalization processes of regular forms. The dual-mechanism model claims that regular forms have a different representation in the mental lexicon, and their generalization process employs the rule-based mechanism. On the contrary, the single-mechanism model does not differentiate the mental representation of regular and irregular forms because it uses only one mechanism, namely the associative mechanism.

Researchers have also investigated how bilinguals, particularly second-language (L2) speakers, represent regular and irregular inflectional forms in their minds and if they differ from L1 speakers in their morphological generalization processes. Previous research has shown that inflectional morphology is one of the most challenging areas for L2 speakers as a result of displaying high rates of inaccuracy and more variability, such as using inflectional morphemes optionally in required contexts (see Ionin & Montrul, 2023 for a review). In addition, Ullman (2001; 2005; 2020) proposed that there are two modules in the human mind: the declarative lexicon module, which stores all the irregularly inflected forms, and the procedural grammar module, which is responsible for the linguistic computations such as applying the rules for grammar. According to this proposal, late L2 learners, unlike L1 speakers of that language, fail to use the procedural grammar module due to maturational changes and rely more on the declarative lexicon module, even for regularly inflected forms. Previous elicited production studies have revealed mixed results (see Clahsen & Jessen, 2021), and more research is needed to unveil the morphological generalization processes in L2 speakers. The present study aims to investigate the morphological generalization in L2 Turkish speakers for the first time and compare them to L1 Turkish speakers with an attempt to explore if L1 and L2 speakers differ from each other in this domain. The study employs an elicited production task of novel word forms and explores the morphological generalization of Turkish aorist, which encodes habitual aspect or general present tense and involves 13 irregular verb forms. The results will yield further insight into the linguistic representation of regular vs. irregular inflection in L2 speakers and whether they can employ linguistic computations like L1 speakers.

1. MORPHOLOGICAL GENERALIZATION IN L1 AND L2 SPEAKERS

Many studies have been conducted with English L1 speakers to find out the morphological generalization of inflected forms. In one of these studies, Bybee and Moder (1983) used rhyme similarity for the construction of novel verbs. If the novel verb (e.g., *spling*) was highly similar to an existing irregular verb (e.g., *cling, sling, string*), participants produced irregular forms such as *splang* or *splung* with a rate of 80%. When the novel verb's similarity to an existing irregular verb decreased, the rate of the production of irregular forms also decreased. The researchers concluded that similarity played an important role in the generalization of irregular forms. In another study, Prasada and Pinker (1993) also used rhyme similarity and constructed novel verbs with different degrees of similarity to existing verbs (e.g., regular prototypical: *plip*; regular intermediate: *brilth*; regular distant: *ploamph*; irregular prototypical: *spling*; irregular intermediate: *cleef*; irregular distant: *goav*) to explore the morphological generalization of English past tense. They carried out two acceptability judgment tasks and one elicited production experiment. Participants preferred an equal number of *–ed* past tense forms for regular novel verbs, irrespective of their similarity to existing regular verbs. Conversely, they preferred more *–ed* past tense forms for irregular verbs when their similarity to existing irregular verbs decreased. These results indicated that, at least, two distinct mechanisms were employed in the morphological generalization of regular and irregular forms and provided support for the dual-mechanism model. Similar results were obtained for novel word production studies with L1 speakers of other languages (e.g., German: Clahsen, 1997; 1999; Italian: Say & Clahsen, 2002; Portuguese: Veríssimo & Clahsen, 2014; Russian: Romanova, 2008; Spanish: Brovetto & Ullman, 2005).

Morphological generalization processes have also been examined in Turkish, but only with L1 Turkish speakers. The linguistic structure that was investigated was Turkish aorist, which is one of the few structures with irregular forms. Kırkıcı and Kırkıcı (2009) used 28 real monomorphemic Turkish verbs (11 irregulars and 17 regulars) and created 84 novel verbs with three levels of similarity: high, medium, and low similarity. They tested 55 L1 Turkish speakers in an elicited production experiment. The results suggested different representations for the regular and irregular forms. Regular aorist was viewed as the default rule and was applied to novel verbs regardless of phonological similarity. In contrast, irregular aorist was only applied to novel verbs that had high phonological similarity to existing irregular verbs. Similar results were obtained by Uygun et al. (2023), who used rhyme similarity to create novel verbs with high similarity to existing irregular and regular verb forms together with novel verb forms whose rhyme does not exist in any Turkish verb. They tested 50 L1 Turkish speakers via an elicited production experiment and replicated the results of Kırkıcı and Kırkıcı (2009), which provide further support for the dual-mechanism model. In another study, by using 12 irregular verbs (ending with /r/ and /l/), Michon (2017) created 168 sonorant-ending novel verbs and tested 90 L1 Turkish speakers in an oral production experiment, where the stimuli were presented auditorily. In line with the previous studies, the results revealed that L1 Turkish speakers used the regular aorist form as the default with a rate of 86%. Michon (2017) further focused on the irregular aorist forms and investigated whether the production of the irregular aorist forms correlated with the type frequency of existing irregular verbs that share the same vowel, the same first consonant, or the same first and last consonants. She observed the crucial influence of the coda in the novel verbs; that is, irregular aorist was used more in novel verbs ending with /r/ than those ending with /l/. In addition, she also found a correlation between the participants' answers and the type and token frequencies of several trigrams and quadrigrams in Turkish. Similar findings were observed by Michon and Nakipoğlu (2019) and Nakipoğlu and Michon (2020) as well. These results indicate that L1 Turkish speakers use the regular aorist as the default rule with its wide application. They also employ the dual-route mechanism for morphological generalization of Turkish aorist and use two distinct mechanisms, namely rule-based mechanism for regulars and associative mechanism for irregulars. The crucial question is whether this generalization process is employed by L2 Turkish speakers as well.

A number of studies have examined morphological generalization in L2 speakers, and the results are mixed and less clear than those from L1 speakers. Some studies found no reliable differences between L1 and L2 speakers, suggesting that they do not differ from each other in their morphological generalization patterns (e.g., Dutch: Lalleman et al., 1997; English: Beck, 1997; Murphy, 2004; Kırkıcı, 2010; German: Hahne et al., 2006; Experiment 1A).

Conversely, there are also studies that found differences between these two groups (e.g., English: Cuskley et al., 2015; German: Neubauer & Clahsen, 2009; Clahsen & Jessen, 2021; Greek: Clahsen et al., 2010). These studies conclude that L2 speakers differed from L1 speakers in their morphological generalization processes of novel verbs that have phonological overlap with existing regular verbs because L2 speakers used significantly less regular tense forms than L1 speakers. A proposal to account for the observed differences between L1 and L2 speakers in their morphological generalization is that while L2 speakers mostly rely on lexical storage and associative generalizations, L1 speakers mainly benefit from rule-based generalization that focuses on the internal structure of the words. According to Ullman (2001; 2005; 2020), this contrast can be explained by the distinction between two separate memory systems, namely the declarative system, which is involved in learning and storing lexical items, and the procedural system, which is responsible for learning and processing grammatical rules. Because of maturational changes, adult L2 speakers are more dependent on their declarative memory for L2 learning and processing and less dependent on their procedural memory when compared to L1 speakers. As a result, L2 speakers, who are late learners of L2, are expected to rely less on rule-based mechanism in their morphological generalization process than L1 speakers. However, more research is needed to draw any generalizable conclusions on the morphological generalization in L2 speakers. Since no study has investigated morphological generalization in L2 Turkish, the present study will be the first to compare L1 and L2 Turkish speakers and provide a testing ground for the claims of Ullman (2001; 2005; 2020).

2. PURPOSE OF THE PRESENT STUDY

This study investigates the morphological generalization of Turkish aorist, which encodes habitual aspect or general present tense. Since irregular aorist forms only exist in monosyllabic verbs, the exponents regarding the monosyllabic verbs are as follows: Monosyllabic verbs ending with a vowel take the *–r* suffix (e.g., *ye* “eat”, *ye-r* “eats”). Most monosyllabic verbs ending with a consonant take the *–Ar* suffix (*–ar* or *–er*) depending on the vowel harmony (e.g., *kes* “cut”, *kes-er* “cuts”; *sor* “ask”, *sor-ar* “asks”). Only 13 monosyllabic verbs that end with a consonant take the *–Ir* suffix (*–ır*, *–ir*, *–ur*, or *–ür*) again depending on the vowel harmony and comprise the irregular verbs (e.g., *al* “take”, *al-ır* “takes; *gör* “see”, *gör-ür* “sees”). A common property of these 13 irregular verbs is that they all end in sonorants /l/, /r/, and /n/;

however, there are 70 other sonorant-ending monosyllabic verbs that are regular and take the *–Ar* suffix rather than the *–Ir* suffix (Nakipoğlu &Üntak, 2008). Therefore, it can be said that irregularity in Turkish aorist results from an existing suffix that is exceptionally applied to a limited number of monosyllabic verbs.

The present study explores morphological generalization in L2 Turkish speakers and compares their generalization processes with L1 Turkish speakers. If Ullman's proposal (2001; 2005; 2020) is on the right track, L2 Turkish speakers should rely more on their declarative memory system, resulting in significantly less usage of regular aorist forms and more reliance on irregular aorist forms when compared to L1 speakers. On the other hand, no significant group difference in the use of regular aorist forms will indicate that L1 and L2 speakers employ the same morphological generalization processes. These hypotheses were tested via a written-elicited novel verb production experiment.

3. Method

3.1. Sample

The L1 group consisted of 30 participants (22 women, between ages 20-50, *M* = 31.30 years, *SD* = 8.95), and they all spoke the standard dialect of Turkish. The L1 group was recruited and tested in Istanbul, Turkey. The L2 group, which was recruited from the Russian/Ukrainian community living in Istanbul, consisted of 31 participants, but one participant was excluded due to many illegible responses. The data from the remaining 30 L2 speakers (28 women, between ages 22-44, *M* = 30.60, *SD* = 7.39) were used for further analyses. The group can be considered late L2 learners as they learnt Turkish after age 8 (between ages 11-25, *M* = 18.60, *SD* = 3.98), and their age of arrival in Turkey was quite high (between ages 17-31, *M* = 22.80, *SD* = 4.60), with varying lengths of residence in Turkey (between years 3-14, *M* = 6.47, *SD* = 2.97). All L2 speakers had learned Turkish by attending Turkish language courses, which means that all of them must have been informed about the rule of Turkish aorist and the exceptions to the rule. As a proficiency measure, all L2 speakers completed the grammar section of the TELC test (https://www.telc.net/tr.html), and the results indicate high scores in the Turkish TELC test (*M* = 18.33, *SD* = 1.35, *Maximum Score* = 20). Self-ratings also revealed a high proficiency level of the L2 Turkish group (Speaking: *M* =

8.93, *SD* = 0.96, Listening: *M* = 9.07, *SD* = 1.10, Reading: *M* = 9.13, *SD* = 0.92, Writing: *M* = 8.40, *SD* = 1.12, Overall Turkish: *M* = 8.60, *SD* = 1.06, *Maximum Score* = 10). The L2 group also used Turkish regularly in their daily lives, according to self-ratings based on percentages (Speaking: *M* = 72.67, *SD* = 20.17, Listening: *M* = 75.33, *SD* = 19.22, Reading: *M* = 70.67, *SD* = 18.31, Writing: *M* = 64.67, *SD* = 20.31, *Maximum Score* = 100).

3.2. Instrument

The experimental items were taken from Uygun et al. (2023), which can be found at the Center for Open Science Framework website at https://osf.io/jnpt6. Because irregularity occurs only in monosyllabic Turkish verbs, 78 monosyllabic novel verbs were created with different levels of phonological similarity to existing monosyllabic verbs. These novel verbs were divided into three conditions: while the *Irregular* condition consisted of novel verbs similar to existing verbs with irregular aorist forms, the *Regular* condition involved novel verbs that were similar to existing verbs with regular aorist forms. Finally, the *No Similarity* condition had novel verbs, which were phonotactically legal but do not exist among monosyllabic Turkish verbs. There were 26 novel verbs in each condition. While all novel verbs in the Irregular condition were sonorant-ending, the novel verbs in the Regular condition involved both sonorant-ending and nonsonorant-ending novel verbs. However, the novel verbs in the No Similarity condition were nonsonorant-ending.

This is the first study that investigated the morphological generalization of all 13 irregular verbs and two novel verbs were created from each existing irregular verb. In order to ensure maximum phonological similarity to existing verbs, novel verbs for the Regular and Irregular conditions were created by changing the beginning phonemes of existing verbs. By following Yavuz (2011) for the Turkish phonology system, these changes were made according to the distinctive features of the beginning phonemes with an attempt to keep the rhyme of the novel verb as similar as possible to the existing verb. For instance, a novel verb with a distinct bilabial consonant (e.g., *pitmek* or *mitmek*) was created if the onset of an existing verb stem was a bilabial consonant (e.g., "b" as in *bitmek* "to finish"). For the labiodental, alveolar, palatal, and velar consonants, the same protocol was followed. Regarding the vowel-initial verb stems, novel verbs were generated based on the vowel's roundedness feature. If an existing verb starts with a rounded vowel (e.g., "ö" as in *ölmek* "to die"), the rounded vowel is replaced with another rounded vowel (e.g., *ülmek*) by ensuring maximal phonological similarity. Similarly, verb-initial unrounded vowels (e.g., "a" is in *aşmak* "to exceed) were replaced with other phonologically similar unrounded vowels (e.g., *ışmak*). For the No Similarity condition, phonotactically legal novel verbs with onsets, rhymes, and codas that are not found in the Turkish verb lexicon were created, and these novel verbs do not match any verbs that already exist in Turkish with regular or irregular aorist forms (e.g., *vöfmek*, no existing monosyllabic Turkish verb ends in *–öf*).

All novel verbs were created from 2- to 4-letter monosyllabic verbs, and they were matched in terms of mean length in letters across conditions (Regular: M = 2.96, SD = 0.45, Irregular: M = 2.77, SD = 0.43, No Similarity: M = 3.0, SD = 0.0). However, it was not possible to match the frequency of the existing verbs from which the novel verbs were created because irregular verbs have a higher frequency than regular verb forms based on 'TS Corpus' (https://tscorpus.com/, Sezer, 2017), which consists of 4,950,407 word types and 491,360,398 tokens taken from sources such as newspapers, social media, forums, blogs, and academic journals and books. According to this corpus, while the mean lemma and aorist-form frequencies (per million) for irregular verbs are 3.23 and 1.75, for regular verbs, the corresponding means are 2.01 and 0.28.

3.3. Design

Both L1 and L2 speakers completed the experiment online on their personal computers. Google® Forms was used in the preparation of the experiment. Each participant filled out a consent form and a detailed demographic background questionnaire prior to the experiment's start. Subsequently, the participants went over the experiment's instructions and became familiar with the process by doing two examples using novel verbs. In the main experiment, participants were presented with sentences one by one, and there was no time limitation. In each trial, there were three stimuli: firstly, the novel verb in its infinitive form, then a sentence with the novel verb in the present continuous tense, and finally a sentence with a blank space where participants had to fill in the aorist form. As can be seen in (1), which uses the novel verb *pöl*, the experimental sentences also included novel words as direct objects (e.g., *jopus*) so that the novel verbs' context did not evoke any evident semantic links with verbs that already existed (e.g., *pölmek* from *bölmek* "to divide).

After completing the experiment, the L2 group did the grammar section of the TELC test. A complete session lasted approximately 20 minutes.

(1) PÖLMEK

Deniz şimdi jopusunu pölüyor.

"*At the moment, Deniz is pöl+ing his/her jopus.*"

Deniz zaten düzenli olarak her gün jopusunu _______.

"*Every day, Deniz regularly _______ his/her jopus.*"

Regular responses were coded with "1" and irregular responses were coded with "0". Responses that were real words, not a Turkish aorist form, or not interpretable (e.g., *pöletir/pölütür* or *pölmer/pölmür* instead of *pöler/pölür*) were removed prior to the data analysis. A total of 0.86% of the L2 group's responses were removed, while there was no data removal in the L1 group. The remaining data were analyzed with R statistical software version 4.2.1 (R Core Team, 2021).

Two statistical analyses were carried out. The first analysis examined *between-group* differences, and a generalized linear mixed-effects regression model (binomial family, with the bobyqa optimizer), with crossed random effects by participant and item (Baayen et al., 2008) was fitted to the participants' responses. In the model, the participant-level variable *Group* (L1 and L2) and the item-level variable *Condition* (Regular, Irregular, and No Similarity) were used as fixed effects. The models were fitted using the package *lme4* (Bates et al., 2015). Initially, a model with random intercepts and slopes for all fixed effects and their interactions was constructed, and when this maximal model failed to converge, it was gradually simplified until convergence was reached (Barr et al., 2013). In the simplification process, random slopes by participant and item for each fixed effect in the model were only retained if they improved the model fit significantly. The Akaike Information Criterion (AIC) was used for model comparison because it provides a measure that penalizes complexity and leads to predictors being kept only when they substantially contribute to explaining variance in the data (Venables & Ripley, 2002). The model with the lower AIC value was selected, and this procedure was repeated until the simplification process did not produce a model with a lower AIC. For the main effects and overall interactions, sum-coded contrasts (-0,5, 0,5) were employed to the factor Group. For single comparisons in the factor Condition, treatment contrasts were applied. *No Similarity* was taken as the control condition, and potential similarity effects for novel verbs in the Irregular and Regular conditions were assessed based on differences between No Similarity and these two conditions. Similarly, the L1 group was used as the baseline to investigate the morphological generalization in the L2 group. The effect sizes are reported by using model coefficients in log odds (*ß*), standard errors (*SE*), *z*-statistics, and *p* values.

The second analysis examined the *within-group* differences to explore whether morphological generalization in the L2 group does indeed exhibit a larger degree of inter-individual variability than in the L1 group. For this analysis, each participant's mean scores of regular aorist responses for each condition, namely Regular, Irregular, and No Similarity conditions, were calculated, and then the obtained mean scores were subtracted from the group's mean score of regular aorist responses for each condition. The attained score for each participant indicates the extent to which that participant's score deviates within the group for each condition. The scores were then compared across the two participant groups for each condition by using Levene's test of Equality of Variances.

4. Results

The means and standard deviations, which are computed as proportions of regular responses of the total of regular and irregular responses, for the three conditions and the two participant groups are presented in Table 10, and Table 11 presents the results from the statistical model comparing Group and Condition differences of these proportions.

Table 10 Means and SDs (in parenthesis) of Regular Responses for the L1 and L2 Groups

	L1 group (2340 responses)	L2 group (2320 responses)
Irregular	0.47 (0.50)	0.53 (0.50)
Regular	0.80 (0.40)	0.65 (0.48)
No Similarity	0.92 (0.27)	0.66 (0.47)

Table 11 Fixed Effects from the Model of the Three Conditions

Fixed Effects	Estimate	Std. Error	z value	p value
(a) Overall Model				
Intercept	1.365	0.232	5.875	**.000**
Main effect: Condition (Irregular vs. No Similarity)	-2.524	0.395	-6.385	**.000**
Main effect: Condition (Regular vs. No Similarity)	-0.688	0.357	-1.928	.054
Group (L2 vs. L1) * Condition (Irregular vs. No Similarity)	-3.316	0.604	-5.488	**.000**
Group (L2 vs. L1) * Condition (Regular	-1.293	0.490	-2.640	**.001**

vs. No Similarity)				

Formula in R: glmer (answer ~ Condition * Group + (1 + Condition | participant) + (1 + Group | item)

(b) Relevelled by Condition				
Group (L2 vs. L1, Irregular)	0.550	0.525	1.047	.295
Group (L2 vs. L1, Regular)	-1.471	0.487	-3.024	**.003**
Group (L2 vs. L1, No Similarity)	-2.765	0.518	-5.334	**.001**

Formula in R: Same formula releveled for condition and checking for group differences

Model Split by Group				
L1 group				
(c) Intercept	2.004	0.331	6.059	**.000**
Main effect: Condition (Irregular vs. No Similarity)	-4.415	0.717	-6.161	**.000**
Main effect: Condition (Regular vs. No Similarity)	-1.501	0.610	-2.463	**.001**
L2 group				
(d) Intercept	0.745	0.297	2.506	**.000**
Main effect: Condition (Irregular vs. No Similarity)	-0.863	0.364	-2.372	**.002**
Main effect: Condition (Regular vs. No Similarity)	-0.064	0.322	-0.197	.844

Formula in R: glmer (answer ~ Condition + (1 + Condition | participant) + (1 | item)

The first analysis was conducted to unveil the *between-condition* and *between-group* differences. The results indicated more regular aorist forms for the No Similarity condition in general. While the No Similarity condition received significantly more regular aorist forms than the Irregular condition (*ß*: -2.524, *SE:* 0.395, z = -6.385, p < .001), this difference was marginally significant when compared to the Regular condition (*ß*: -0.688, *SE:* 0.357, z = -1.928, p = .054). More importantly, the model revealed significant two-way interactions of Group and Condition (see Table 2a). The Group and Condition interaction for the No Similarity and Irregular conditions (*ß*: -3.316, *SE:* 0.604, z = -5.488, p < .001) stems from the smaller contrast between the proportion of regular aorist forms in the L2 group (M = 0.66 vs. 0.53) than the L1 group (M = 0.92 vs. 0.47). Similarly, the smaller contrast in the L2 group between No Similarity and Regular conditions (M = 0.66 vs. 0.65) when compared to the L1 group (M = 0.92 vs. 0.80) is also the reason for the second significant Group and Condition interaction (*ß*: -1.293, *SE:* 0.490, z = -2.640, p < .002). Further *between-group* comparisons for the three condition types (Table 2b) showed no group difference in the Irregular condition (*ß*: 0.550, *SE:* 0.525, z = 1.047, p = .295), indicating the successful use of the associative generalization mechanism in both groups. However, the L2 group differed significantly from the L1 group in both Regular (*ß*: -1.471, *SE:* 0.487, z = -3.024, $p < .004$) and No Similarity (*ß*: -2.765, *SE:* 0.518, z = -5.334, $p < .002$) conditions due to their reduced proportion of regular aorist responses. This shows that the L2 group has problems using the symbolic generalization mechanism and is less successful in applying the rule-based mechanism. In addition, both groups produced the highest regular aorist forms for the No Similarity condition. Table 2c shows that the No similarity condition differed significantly from the other two conditions in the L1 group (Irregular: *ß*: -4.415, *SE:* 0.717, z = -6.161, $p < .001$; Regular: *ß*: -1.501, *SE:* 0.610, z = -2.463, $p < .002$). Conversely, as can be seen in Table 2d, in the L2 group, the No Similarity condition only differed significantly from the Irregular condition (Irregular: *ß*: -0.863, *SE:* 0.364, z = -2.372, $p < .003$; Regular: *ß*: -0.064, *SE:* 0.322, z = -0.197, $p = .844$).

The second analysis explored the *within-group* differences to see if the L2 group displayed more variability in their morphological generalization process. To test this, the difference between each individual's regular aorist responses and the corresponding group's mean for each of the three conditions was calculated. Then, each individual's mean score was subtracted from the corresponding group's mean score for each condition separately. Levene's test of Equality of Variances was used to make statistical comparisons. The results revealed significantly more individual variability in the L2 group's regular aorist responses than the L1 group only in the No Similarity condition (L2 group: M = 0.0035, SD = 0.25; L1 group: $M = -2.53\text{x}10^{-8}$, SD = 0.11; F = 22.21; $p < .002$). For the Regular (L2 group: M = 0.0013, SD = 0.24; L1 group: $M = 4.87\text{x}10^{-8}$, SD = 0.18; F = 1.35; $p = .249$) and Irregular (L2 group: M = 0.0037, SD = 0.27; L1 group: $M = 1.79\text{x}10^{-8}$, SD = 0.29; F = 0.16; $p = .692$) conditions, no significant difference was observed (see Figure 7). This shows that the L2 group displays more variability in morphological generalization for novel verbs that do not phonologically match with any existing Turkish verb.

Density Plots for Three Conditions (L1: Straight Line; L2: Dotted Line)

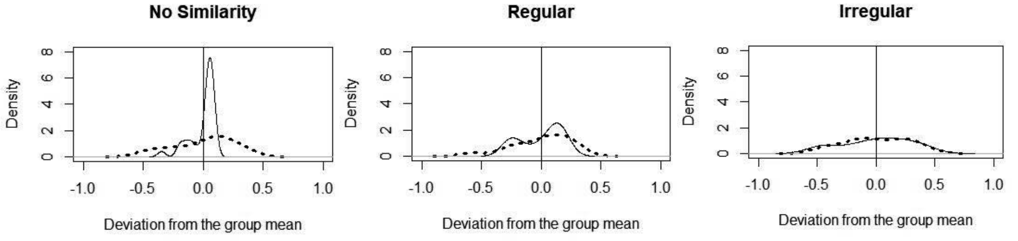


Figure 7

5. DISCUSSION

The main results of the study can be summarized as follows: In general, the regular aorist form was the most preferred form for all conditions. While both groups produced the least regular aorist forms for the Irregular condition, the most regular aorist forms were produced for the No Similarity condition. For the Irregular condition, both groups employed the associative generalization mechanism and did not differ from each other. For the Regular condition, rule-based generalization was used by both groups, but the L1 group produced regular aorist forms significantly more than the L2 group. A similar pattern was also observed in the No Similarity condition, with more variability in the L2 group's production of the regular aorist forms.

The findings for the L1 Turkish group can be explained by the dual-mechanism model (Clahsen, 1999; Pinker & Prince, 1988). According to this model, two separate mechanisms are employed for morphological generalization, the associative generalization for the limited number of irregular forms and the rule-based generalization for the wide number of verb forms, which makes the regular aorist form as the default form in generalizing novel verb forms. These findings are also in line with the findings of the previous L1 Turkish studies (Kırkıcı & Kırkıcı, 2009; Michon, 2017; Michon & Nakipoğlu, 2019; Nakipoğlu & Michon, 2020; Uygun et al., 2023) and other L1 studies conducted in different languages (e.g., German: Clahsen, 1997; 1999; Italian: Say & Clahsen, 2002; Portuguese: Veríssimo & Clahsen, 2014; Russian: Romanova, 2008; Spanish: Brovetto & Ullman, 2005), indicating that the dual-mechanism model can explain the morphological generalization processes in many languages other than English.

The dual-mechanism model can also explain the observed morphological generalization process in the L2 Turkish group since they produced regular aorist forms less for the Irregular condition because of using the associative generalization mechanism and more for the Regular and No Similarity conditions due to the wide application of the rule-based mechanism. Although they employed the associative generalization mechanism as effectively as the L1 group, significant group differences were observed in the application of the rule-based mechanism for the Regular and No Similarity conditions because the L2 group produced significantly fewer regular aorist forms when compared to the L1 group. Similar results were also observed in other L2 studies (e.g., Clahsen et al., 2010; Clahsen & Jessen, 2021; Cuskley et al., 2015), and these results indicate that while L1 speakers mainly benefit from rule-based generalization that focuses on the internal structure of the words, L2 speakers have difficulties using rule-based generalization as effectively as the L1 group and rely less on rule-based generalization. The L2 results provide further support for the account proposed by Ullman (2001; 2005; 2020) because the L2 speakers of the present study were not able to attain the degree of proceduralization of their L2 grammar as L1 speakers.

The L2 group also displayed more inter-individual variability than the L1 group only in the No Similarity condition. This condition was employed as a control condition to explore how participants will apply morphological generalization processes when they cannot rely on similarity-based associative generalization because the novel verbs in this condition were phonotactically legal but do not exist among the Turkish verbs. The L1 group mostly relied on the rule-based mechanism and performed more homogeneously in this condition, whereas the L2 group's performance was less homogeneous. This shows that the L2 group faced more difficulties in applying the rule-based mechanism when similarity-based associative generalization fails. This is also in line with Ullman's proposal (2001; 2005; 2020) and supports the claims of Ionin and Montrul (2023), who advocate that L2 speakers exhibit more variability in the use of inflectional morphemes.

The observed pattern in both groups cannot be explained by the single-mechanism model (Rumelhart & McClelland, 1991), which claims that all forms, whether regular or irregular, are generalized by employing the associative generalization mechanism only, and this mechanism benefits from phonological overlap between the novel verb and a previously learned verb. However, this is not the case in the present study because regular aorist forms were widely produced across different conditions, especially in the No Similarity condition, where both groups produced the highest rate of regular aorist forms. The associative generalization mechanism was only employed for the Irregular condition, where both groups produced the least regular aorist responses. Therefore, the obtained results can be best explained by the dual-mechanism model (Clahsen, 1999; Pinker & Prince, 1988), in which the L2 group relies less on the rule-based mechanism when compared to the L1 group.

6. CONCLUSION

The aim of the present study was to explore the morphological generalization process in L2 Turkish speakers for the first time and compare their generalization pattern to that of L1 Turkish speakers. The results suggested that both groups employed the dual-mechanism model, namely using the associative generalization mechanism for irregulars and the rule-based generalization mechanism for regulars. However, the L2 group seemed to rely less on the rule-based mechanism and produced regular aorist responses to a lesser extent in the Regular and No Similarity conditions, where the L1 group mostly relied on rule-based mechanism with very high proportions of regular responses. In addition, the L2 group also appeared to have a less homogeneous performance and exhibited more inter-individual variability in the No Similarity condition. These results are in line with Ullman's (2001; 2005; 2020) account, which claims that late L2 learners rely less on rule-based mechanism in morphological generalization.

Acknowledgment: I would like to thank Kateryna Iefremenko for her help with participant recruitment and data collection.

REFERENCES

Baayen R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi:10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler,M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beck, M. L. (1997). Regular verbs, past tense and frequency: Tracking down a potential source of NS/NNS competence differences. *Second Language Research*, *13*(2), 93-115. https://doi.org/10.1191/026765897670780840

Brovetto, C., & Ullman, M. (2005). The mental representation and processing of Spanish verbal morphology. In D. Eddington (Ed.), *Selected proceedings of the 7th Hispanic linguistics symposium* (pp. 98-105). Sommerville, MA; Cascadilla Press.

Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, *59*(2), 251-270. https://doi.org/10.2307/413574

Clahsen, H. (1997). The representation of participles in the German mental lexicon: Evidence for the dual-mechanism model. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 73-96). Springer: Dordrecht. https://doi.org/10.1007/978-94-017-3718-0_6

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences,* *22*(6), 991-1060. https://doi.org/10.1017/s0140525x99002228.

Clahsen, H., & Jessen, A. (2021). Morphological generalization in bilingual language production: Age of acquisition determines variability. *Language Acquisition, 28*(4), 370-386. https://doi.org/10.1080/10489223.2021.1910267

Clahsen, H., Martzoukou, M., & Stavrakaki, S. (2010). The perfective past tense in Greek as a second language. *Second Language Research,* *26*(4), 501-525. https://doi.org/10.1177/0267658310373880

Cuskley, C., Colaiori, F., Castellano, C., Loreto, V., Pugliese, M., & Tria, F. (2015). The adoption of linguistic rules in native and non-native speakers: Evidence from a Wug task. *Journal of Memory and Language, 84*, 205-223. https://doi.org/10.1016/j.jml.2015.06.005

Fromkin, V., Rodman, R., & Hyams, N. (2011). *An introduction to language*. Canada: Wadsworth Cengage Learning.

Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience,* *18*(1), 121-134. https://doi.org/10.1162/089892906775250067

Ionin, T., & Montrul, S. (2023). *Second language acquisition: Introducing intervention research*. Cambridge: Cambridge University Press.

Kırkıcı, B. (2010). Distinct mechanisms in the processing of English past tense morphology: A view from L2 Processing. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 67-84). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/celcr.13.07kir

Kırkıcı, B., & Kırkıcı, K. (2009). Türkçe geniş zamanda sözcükler ve kurallar [Words and rules in Turkish aorist]. Presented at the 23rd National Linguistics Conference, Famagusta, Turkish Republic of Northern Cyprus.

Lalleman, J. A., van Santen, A. J., & van Heuven, V. J. (1997). L2 processing of Dutch regular and irregular verbs. *International Journal of Applied Linguistics,* *115*(1), 1-26. https://doi.org/10.1075/itl.115-116.01lal

Michon, E. (2017). *Velir or veler? Exploring adults' implicit knowledge of an irregular verbal pattern: The Turkish aorist* [Unpublished master's thesis]. Boğaziçi University.

Michon, E., & Nakipoğlu, M. (2019). Rule-based vs. similarity-based generalization: An experimental study on the Turkish aorist. In J. Audring, N. Koutsoukos & C. Manouilidou (Eds.), *Proceedings of the 12th Mediterranean Morphology Meeting, MMM12* (pp. 54-63). Patras: University of Patras. https://doi.org/10.26220/mmm.3249

Murphy, V. (2004). Dissociable systems in second language inflectional morphology. *Studies in Second Language Acquisition, 26*(3), 433-459. https://doi.org/10.1017/S0272263104263033

Nakipoğlu, M., & Üntak, A. (2008). A complete verb lexicon of Turkish based on morphemic analysis. *Turkic Languages*, *12*(2), 221-280.

Nakipoğlu, M., & Michon, E. (2020). Abstraction vs analogy in the Turkish aorist. In A. Gürer, D. Uygun-Gökmen & B. Öztürk (Eds.), *Morphological complexity within and across boundaries: In honour of Aslı Göksel* (pp. 13-38). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/slcs.215

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of German participles. *Studies in Second Language Acquisition, 31*(3), 403-435. https://doi.org/10.1017/S0272263109090354

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1-2), 73-193. https://doi.org/10.1016/0010-0277(88)90032-7

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463. https://doi.org/10.1016/S1364-6613(02)01990-3

Prasada, S., & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*(1), 1-56. https://doi.org/10.1080/01690969308406948

R Core Team (2021). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available online at: https://www.rproject.org/

Romanova, N. (2008). Mechanisms of verbal morphology processing in heritage speakers of Russian. *Heritage Language Journal, 6*(1), 105-126. https://doi.org/10.46538/hlj.6.1.5

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2: Psychological and biological models* (pp. 216-271). Cambridge, MA: MIT Press.

Say, T. & Clahsen, H. (2002). Words, rules and stems in the Italian mental lexicon. In S. Nooteboom, F. Weerman & F. Wijnen (Eds.), *Storage and computation in the language faculty* (pp. 93-129). Springer: Dordrecht. https://doi.org/10.1007/978-94-010-0355-1_4

Sezer, T. (2017). TS corpus project: An online Turkish dictionary and TS DIY corpus. *European Journal of Language and Literature, 9*(1), 18-24. https://doi.org/10.26417/ejls.v9i1.p18-24

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37-69. https://doi.org/10.1023/A:1005204207369

Ullman, M. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory and practice* (pp. 141-178). Washington, DC: Georgetown University Press.

Ullman, M. (2020). The declarative/procedural model. A neurobiologically motivated theory of first and second language. In B. VanPatten, G. Keating & S. Wulff (Eds.), *Theories in second language acquisition* (pp. 128-161). New York: Routledge. https://doi.org/10.4324/9780203628942

Uygun, S., Schwarz, L., & Clahsen, H. (2023). Morphological generalization in heritage speakers: The Turkish aorist. *Second Language Research*, *39*(2), 519-538. https://doi.org/10.1177/02676583211059291

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer: New York.

Verissimo, J., & Clahsen, H. (2014). Variables and similarity in linguistic generalization: Evidence from inflectional classes in Portuguese. *Journal of Memory and Language, 76*, 61-79. https://doi.org/10.1016/j.jml.2014.06.001

Yavuz, H. (2011). Turkish consonants. In Z. Balpınar (Ed.), *Turkish phonology and morphology* (pp. 14-31). Eskişehir: Anadolu University.

Chapter 4: **Morphological Complexity Patterns Across L2 English Essay Scores: The Case of a Morphologically Rich and Productive L1 Background**

Enis Oğuz

ABSTRACT

Morphological complexity and its potential to explain language proficiency have received interest for many years, especially in second language acquisition research. Bilinguals generally exhibit more morphological complexity in their L2 essays as their L2 skills develop. However, the exact patterns of how morphological structures are used in L2 writing are difficult to determine, especially since bilinguals experience different levels of difficulty in acquiring L2 morphology. One essential factor modifying this challenge is L1 background, as bilinguals with a less complex L1 morphology might encounter an increased challenge in acquiring their L2 morphology. In the opposite scenario in which L1 morphology is more complex with its productivity and richness, it is unknown whether the challenge of L2 morphology acquisition will be modified. Furthermore, the controversy on how to measure morphological complexity has yet to be resolved. This study estimated eleven different morphological complexity indices for 692 L2 English essays written by Turkish-English bilinguals, whose L1 morphology was more complex than their L2 morphology. The indices were examined for their abilities in differentiating L2 essay scores. The results revealed important insights regarding the role of morphological complexity across different essay scores, as well as providing valuable data for further crosslinguistic comparisons with complexity studies focusing on certain L1 backgrounds.

Keywords: morphological complexity, L2 writing, bilingualism, derivation, inflection.

1. INTRODUCTION

Second language acquisition involves the construction of an additional language system, but this new system is open to the influence of the first language (L1). Due to this crucial difference, such constructs are also called interlanguages (Selinker, 1972), underscoring the involvement of L1 in both acquisition and production processes. Unlike first languages, interlanguages are characterized by their unstable and idiosyncratic nature, which results in complicated language usage and error patterns (Brezina & Pallotti, 2019). As an essential component of language acquisition, morphology acquisition in a second language (L2) likewise results in various usage and error patterns, some of which are similar to L1 acquisition and others are unique to L2 acquisition. In particular, morphological complexity differs across different L2 proficiency levels and L1 backgrounds. In an attempt to investigate morphological complexity patterns in bilinguals with a certain L1 background, this study examined essays written by Turkish-English bilinguals and calculated different morphological complexity measures to reveal potential differences across different score bands.

2. MORPHOLOGICAL COMPLEXITY

Morphological complexity indicates the relative difficulty and richness in creating complex words. In general, the way morphemes combine with words and other morphemes cannot be explained by a limited number of rules (e.g., Ackerman & Malouf, 2013), making morphology acquisition a challenging but important task for both native speakers and second language learners (Crossley, Tywoniw, & Choi, 2024; DeKeyser, 2005; Slabakova 2009). L2 morphology acquisition is more difficult than L1 morphology acquisition, especially for L2s with richer morphological systems (Schepens, van der Slik & van Hout, 2013). That is one of the main reasons why L2 acquisition as an adult usually ends up with a rather limited proficiency, despite the rare cases in which adult learners reach native-like skills in an L2 (e.g., Birdsong, 2005).

The exact reasons why L2 acquisition, and L2 morphology acquisition in particular, lags behind L1 acquisition are difficult to determine. Second language learners might depend more on their declarative memory (Ullman, 2006), acquire complex words as simple structures (Hopp, 2013), or use L2 morphological process only when these are parallel to their L1 (Hawkins & Casillas, 2008). Whatever the reasons are, the complexity of this task results in long-lasting errors even among advanced second language learners (Larsen-Freeman, 2010). In addition to errors, the occurrence of morphemes usually does not follow a straightforward pattern, as one morpheme appearing in a developmental stage might disappear in the next one (Long, 2003).

Finding significant predictors of L2 writing quality has been one of the main concerns in complexity studies (e.g., Egbert, 2017; Kyle & Crossley, 2017). In the lexical level, lexical diversity has been argued to predict higher essay scores (e.g., Kim, 2014; Yu, 2010), while lexical sophistication can reveal information regarding L2 proficiency level (Treffers-Daller, Parslow & Williams, 2018). For example, the average number of root morphemes is a significant indicator of L2 writing proficiency (Crossley et al., 2024). However, findings are inconclusive as to whether morphological complexity at the lexical level can differentiate L2 writing abilities and explain some of the variety in writing scores.

L2 learners are varied in their use of L2 morphological structures (Muroya, 2019), and several studies indicate that morphological complexity can help differentiate L2 proficiency levels (e.g., Verspoor et al., 2012). Morphological complexity measurements have also been found to have discriminate power in proficiency scores of L2 learners, suggesting that the more advanced an L2 learner becomes, the more complexity is shown in their L2 production (e.g., Březina & Pallotti, 2016; Schepens, van der Slik & van Hout, 2013; Van der Slik et al., 2019). In addition to complexity, more advanced L2 learners produced more morphemes (Crossley, Tywoniw & Choi, 2024), further supporting the difficulty of morpheme production in beginning-level L2 learners (e.g., Larsen-Freeman, 2010).

Until recently, the potential role of derivational morphemes had been neglected in morphological complexity research. Similar to how more variety in producing inflectional morphemes indicates a higher language proficiency and a better writing quality, more varied derivational morphemes also indicate higher L2 proficiency (e.g., Crossley, Tywoniw & Choi, 2024), confirming the existence of a complex lexicon in higher L2 proficiency levels (Kyle, Crossley & Berger, 2018). Such findings show the necessity of investigating both derivational and inflectional morphemes in morphological complexity research.

Although overall findings then suggest a developmental pattern in which morpheme complexity and usage increases along with L2 proficiency, there are also opposing claims regarding the important role of lexical and morphological features in L2 writing quality. Qian et al., (2021) found only two measurements of morphological complexity (V100 and T/F) to be significant indicators of L2 English writing quality, and García, Ainciburu and Buyse (2021) failed to find a significant effect of morphological complexity in differentiating L2 Spanish proficiency, supporting other studies that offered a limited role of morpheme knowledge in improving L2 writing proficiency (e.g., Malvern et al., 2004; Yu, 2010; Bulté & Housen, 2014). The findings of Qian et al. (2021) based on Chinese-English bilinguals also raise the question of whether L1 background has an effect on the role of morphological complexity in L2 writing development. The bilinguals in García, Ainciburu and Buyse (2021), on the other hand, had different L1 backgrounds, which might have acted as a confounding factor considering the small sample size (113 texts).

L1-L2 contrasts in morphology can also result in mistakes, determining the challenge presented by the target language (Muroya, 2019). In particular, when L1 has relatively a less complex morphology, the acquisition of an L2 with a more complex morphology might be negatively influenced (Schepens et al., 2013; Van der Slik et al., 2019). L2 learners can eventually reach a level of complexity that is indifferent from native speakers, but the differences in L1 background and the complexity of the target language morphology can make L2 learners experience different levels of difficulty (e.g., Housen & Simoens, 2016), which persist until advanced proficiency levels (Hopp, 2013).

The relative challenge in acquiring a particular L2 morphology has been confirmed in an experimental study. L2 English learners with a L1 Japanese background were found be more successful than L2 Japanese learners with an English L1 background in acquiring imperfective structures in the target language. This difference was associated with the complexity of L2 grammar and the transparency of input cues that help rule out L1-influenced interpretations (Gabriele, 2009). The findings show the importance of exploring the role of morphological complexity across different L1 backgrounds and target languages. However, this task is more challenging than it sounds, as different studies have offered different indices and methods to evaluate morphological complexity.

Measuring morphological complexity has been the center of controversy in complexity research for some time (Qian, Cheng & Zhao, 2021). The complexity of the highly abstract morphological system manifest itself in texts where affixes are abundant; especially derivational affixes seem to increase text difficulty (Crossley et al., 2024), probably due to their higher processing cost (Amenta & Crepaldi, 2012). On the other hand, inflectional affixes are argued to ease text comprehension, as they provide crucial grammatical information (Crossley, Tywoniw & Choi, 2024). However, the complexity of morphological structures involves more than a simple count of affixes in the text. For instance, considering word length in essays is essential in counting morphological structures since longer texts have higher chances of including a particular structure (Brezina & Pallotti, 2019). Although traditional complexity measures ignore the existence of particular structures in text (Kyle & Crossley, 2018), more fine-grained analyses have emerged over the years to capture the true extent of morphological complexity (e.g., Crossley, Tywoniw & Choi, 2024). There have been several attempts come up with an accurate index, including but not limited to Inflectional Diversity (Malvern et al., 2004), the Mean Size of Verbal Paradigms (Xanthos & Gillis, 2010), Morphological Complexity Index (Pallotti, 2015), and Kolmogorov Morphological Complexity (Wang, Wang, & Wang, 2022).

Developed by Malvern et al. (2004), inflection diversity measures morphological complexity by considering inflectional diversity. The measure is based on the inflection variation of words and considers both diversity and sample size. However, Xanthos and Gillis (2010) underline that this measurement can be inflated by increasing lexical diversity and propose *mean size of paradigm* (MSP) as an alternative method to measure inflectional diversity. This new approach considers inflectional morpheme counts within different word classes, which allows a more accurate computation for both small and large sample sizes. Brezina and Pallotti (2019) argue that this measurement would create a bias against word classes with smaller sample sizes, as the fewer the sample, the less likely they are to appear with different inflectional forms. As an alternative, they created The Morphological Complexity index following the initial work of Pallotti (2015), which also focused on calculating inflectional diversity. The measure also considers inflectional diversity within each different word class, but it is based on random samples of 10 words at a time in order to prevent the bias against shorter texts.

Wang, Wang and Wang (2022) points at that although the results of Brezina and Pallotti (2019) were informative, this previous study only focused on inflectional morphology and ignored the potential patterns in derivational complexity, similar to how derivational morphology is somewhat neglected in L2 research common (Clahsen et al., 2010). This is crucial since derived words are known to produce less priming for L2 learners compared to null priming effects for L2 inflectional affixes (Kirkici & Clahsen, 2013; Silva & Clahsen, 2008), showing different processing patterns than inflectional morphemes in L2. In order to fill this gap, Wang, Wang and Wang (2022) used the Kolmogorov Complexity to determine the morphological complexity of texts, in which 10% of characters are randomly removed from words and a comprehension technique is applied to see how successful the compression is despite the distortion. Morphologically richer texts are thought to be less affected by this distortion since they possess more word forms; therefore, the lesser success in the compression of the distorted text compared to the original text, the lesser morphological complexity it is associated with. The Kolmogorov morphological complexity index showed a similar pattern with Type-Token Ratio (TTR), as both successfully differentiated beginner and low-intermediate L2 learners, unlike MCI based on based on inflectional diversity. However, these measures failed to differentiate higher levels, revealing potential limitations.

One challenge in measuring morphological complexity was the necessity of computing skills, limiting the scope of this line of research (Wang, Wang, & Wang, 2022). Tywoniw and Crossley (2020) came up with their automatized tool, the Tool for Automatic Measurement of Morphological Information (TAMMI), which is a free software with the ability to process hundreds of texts and calculate different morphological complexity indices. Later, the second and more comprehensive version of this tool (TAMMI 2.0) was released by Crossley, Tywoniw & Choi (2024). A morphological complexity index (MCI) was developed for derivational morphemes, which was proven to be a good predictor of text readability in the analysis since greater variety in derivational affixes resulted in more difficult texts (Crossley, Tywoniw & Choi, 2024). In the study, the average number of morphemes per word was the strongest predictor of L2 proficiency, followed by inflectional variety, the average number of root morphemes (per content word), derivational variety (derivational TTR), and word frequency.

Even though the results Crossley, Tywoniw and Choi (2024) are promising, the data was based on the ELLIPSE corpus, which has 6500 essays written by L2 learners with different L1 backgrounds. Considering the potential role of L1 influence mentioned above (e.g., Gabriele, 2009), further studies on specific L1 backgrounds are needed. In one recent study, Uzun (2025) used TAMMI 2.0 to examine morphological complexity patterns in the Written Essays v2.6 of the International Corpus Network of Asian Learners of English (Ishikawa, 2013). This corpus includes argumentative essays written in L2 English, reflecting the writings of four different proficiency levels. The findings showed a considerable increase in inflectional and derivational MC measurements as L2 essay scores increases, with derivational MCI showing the biggest and inflectional MCI showing the smallest change. Despite its significant contribution to reveal how morphological complexity indices created by TAMMI 2.0 can differentiate L2 proficiency levels, the study of Uzun (2025) also used L2 essays written by bilinguals from 10 different Asian countries and regions, making it difficult to disentangle the potential influence of L1.

Utilizing the automatized TAMMI 2.0 tool for calculating morphological complexity indices and a leaner corpus written by bilinguals with a single L1 background for creating the necessary dataset, the current study attempts to reveal how morphological complexity patterns of Turkish-English bilinguals change across different essay scores. After further studies focusing on different L1 backgrounds are conducted, the current findings will also help reveal crosslinguistic influence on the morphological complexity development in L2 writing.

3. METHOD

The present study used the TELC corpus (Cangır et al., 2024) to examine morphological complexity indices in L2 writing. The corpus based on opinion essays written by 262 Turkish-English bilinguals in their L2, comprising a total number of 691 essays. Based on the LexTALE scores (Lemhöfer & Broersma, 2012), the overall proficiency level of the participants was calculated as intermediate. The scores in this corpus were estimated by taking the average of two scores given by independent raters, and in instances in which the score difference was more than two, a third rater also examined the essay and help determine the assigned score (Cronbach’s alpha: 0.74). The rubric used by the raters had 10 intervals, each referring to a single score, with half points being allowed.

Turkish, which was the L1 background of the bilinguals, has a shallow orthography that allows Turkish speaking first graders to acquire and improve impressive reading skills within a school year (e.g., Babayiğit & Stainthorp, 2007; Durgunoğlu & Öney, 1999). This rapid acquisition process is thought to contribute a faster development of certain skills, such as connective comprehension (e.g., Oğuz & Özge, 2020). That kind of an advantage can also help processing morphological information earlier, as Turkish second graders show patterns of automatized morphological processing even at the early stages of reading (Oğuz & Kırkıcı, 2023). Considering the richness and productiveness of Turkish morphology, this achievement is impressive for children who has received only 18 months of formal education. As discussed above, bilinguals with a less complex L1 might experience more challenge in acquiring an L2 with a more complex morphology (Schepens et al., 2013; Van der Slik et al., 2019). However, the bilinguals in the current study had a more complex L1 morphology compared to the morphology of their target language. This study will provide valuable information regarding how such a mismatch affects the morphological complexity patterns in L2.

Based on previous research (e.g., Brezina & Palotti, 2019; Crossley, Tywoniw & Choi, 2024), eleven indices were identified, which had strong potential to explain differences in L2 writing scores. Then, essays were analyzed using TAMMI 2.0 (Crossley, Tywoniw & Choi, 2024), and the indices in the output file were matched with essay scores. The following indices are investigated in the current study:

1. Number of affixes per content word
2. Number of roots per content word
3. Root log frequency
4. Affix log frequency per content word
5. Affix length per content word
6. Mean Subset Inflectional Variety
7. Mean Subset Derivational Variety
8. Inflectional type-token ratio (TTR)
9. Derivational type-token ratio (TTR)
10. Morphological complexity index (MCI) for derivations
11. Derivational Morphological complexity index (MCI) for inflections

All analyses were carried out in R (R Core Team, 2024), using the packages lme4 (Bates et al., 2015), mgcv (Wood & Wood, 2015), relaimpo (Grömping, 2007), and dominanceanalysis (Navarrete, Soares & Navarrete, 2020). The figures were created using the ggplot2 package (Wickham, 2011).

TAMMI 2.0 (Crossley, Tywoniw & Choi, 2024) calculates variety scores for inflection and derivation using a similar formula to the within-subset score in Brezina and Pallotti (2019). The software extracts content words and divides them into windows of 10 words. Each window is checked for inflectional morphemes, and each unique observation is reported (Words without inflectional morphemes are considered as null tokens). Finally, the sum of unique inflection morphemes (including the null token) is divided by the total number of windows. The same steps are applied for derivational morphemes as well. The variety scores calculated by this method allows controlling text length by avoiding longer texts to produce higher scores regardless of morphological variety they contain. The original formula of Brezina and Pallotti (2019), on the other hand, considers each word class separately (e.g., verbs, nouns) and draws random samples (without repetition within the sample) to determine morphological variety.

Type-token ratios (TTR) in TAMMI 2.0 are calculated by a similar method. Similar to variety calculations, content words are extracted and divided into windows of 10. The unique number of inflections within a window is then divided by the length of the window (as the window at the end might have fewer content words). The average of this score is then taken for all the windows in the text. The same formula is used for derivational morphemes as well, but rather than 10-content-word windows, 10-morpheme windows are used since words can have more than a single derivational morpheme.

Finally, TAMMI 2.0 again uses a modified version of the formula presented in Brezina and Pallotti (2019) for calculating inflectional and derivational MCI indices. This modified formula uses the variety score mentioned above by considering the unique number of affixes (inflectional or derivational) in each window and taking the average of these. In addition to this, a between-subset diversity score is estimated by comparing the unique number of inflectional morphemes (or derivational morphemes) across all samples. This, however, can be problematic since if you have a long enough text, this score can become 0, as no affixes will be unique in the text. The variety score and the diversity score are added together, divided by the number of windows, and finally decreased by 1.

4. RESULTS

A linear mixed effects model was created in which essays scores were the dependent variable and morphological indices were independent variables. Additionally, Participant ID and Task ID were considered as random effects since some bilinguals wrote multiple essays on different topics (i.e., tasks). Linearity was ensured through a visual inspection of the relationships between independent variables and the dependent variable. All variables were standardized using the scale function of R; the scaled values were used in the models, while raw values were used in the figures. Furthermore, the linear model provided a lower AIC value than the Generalized Additive Model (GAM), further supporting the advantage of using a linear model. Figure 8 shows the correlations between independent variables and essay scores.

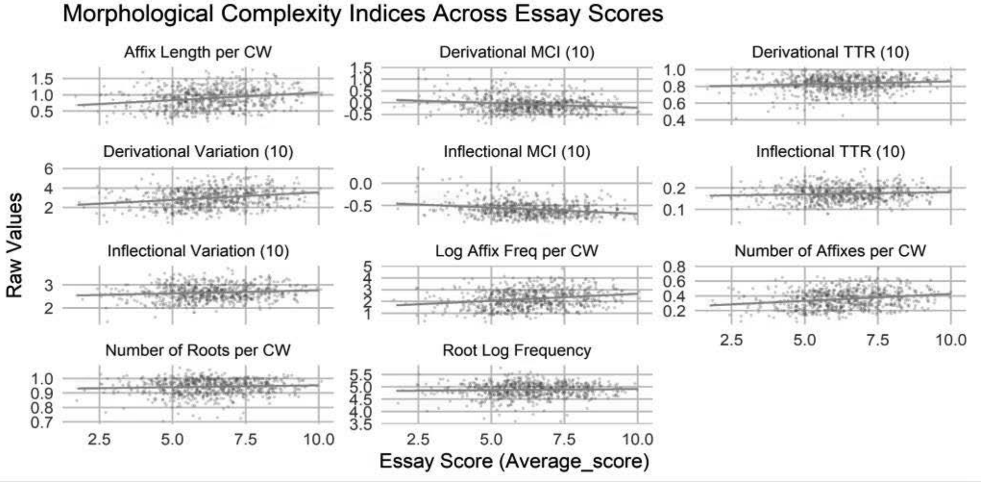


Figure 8. The relationships between morphological complexity indices and essays scores.

Three significant independent variables were found. The relative importance of each variable was calculated using relaimpo and dominance analyses packages. Both analyses showed that root log frequency and number of roots per content word had poor predictive values. These two variables were removed from the model along with two insignificant variables with the highest p-values, resulting in a reduced model (Model 2). In the final step, three more insignificant variables were removed, creating a third model (Model 3). The comparison of three models showed that Model 3 had the best fit, with the lowest AIC value (1743.6 compared to 1750.1 of Model 1 and 1746.3 of Model 2). Model 3 is presented in Table 12.

Table 12. The final model with the best predictive value and lowest AIC value.

	Average_score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.03	-0.20 – 0.15	0.749
Inflectional Variety	0.10	0.03 – 0.17	**0.003**
Morphological Complexity Index (Inflection)	-0.13	-0.20 – -0.06	**0.001**
Derivational Variety	0.30	0.22 – 0.38	**<0.001**
Morphological Complexity Index (Derivation)	-0.18	-0.26 – -0.11	**<0.001**
Random Effects			
σ^2	0.48		
$\tau_{00\ Participant}$	0.35		
$\tau_{00\ Task}$	0.02		
ICC	0.44		
$N_{Participant}$	262		
N_{Task}	4		
Observations	692		
Marginal R^2 / Conditional R^2	0.133 / 0.516		

The number of derivational and inflectional tokens were also calculated for further discussions. Figure 9 shows the calculated tokens. The figure must be examined with caution since TAMMI 2.0 counts morpheme usage regardless of its accuracy.

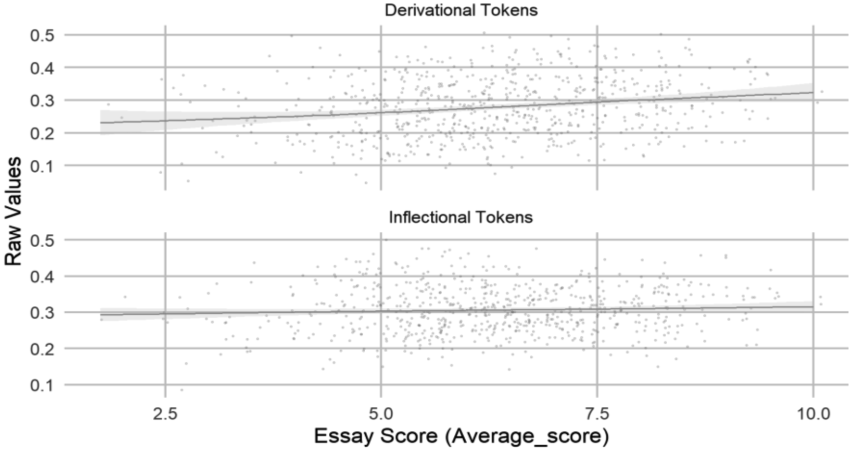


Figure 9. Tokens for derivational and inflections morphemes across different essay scores.

5. DISCUSSION

Morphology acquisition in L2 presents a difficult challenge for L2 learners. The complexity of morphological structures has been argued to increase as bilinguals improve their L2 skills, making morphological complexity indices potential tools in differentiating L2 proficiency levels. Furthermore, the relative complexity and similarity between L1 and L2 can influence the degree of the difficulty of L2 morphology acquisition. In order to examine the patterns of morphological complexity in a novel L2 population and provide valuable insights for further crosslinguistic comparisons, the current study analyzed the morphological complexity and usage patterns of L2 English essays in the TELC corpus. The analyses showed that four essential morphological complexity indices could successfully differentiate essay scores (and presumably L2 proficiency levels) in the L2 essays written by Turkish-English bilinguals.

Root frequency and number of roots per content word showed small change across essay scores and were insignificant in the analysis. The role of average root number was significant in Crossley, Tywoniw and Choi (2024), contradicting the current results. This can potentially point at a crosslinguistic difference for Turkish-English bilinguals; however, as the bilinguals in Crossley, Tywoniw and Choi (2024) had different L1 backgrounds, the conclusive remarks require further evidence with more studies focused on a single L1 background.

Some independent variables which were excluded in the final model also showed an increasing trend with increasing essay scores. Affix length, frequency, and length showed a steady increase, as well as Inflectional and Derivational Type/Token Ratios, mimicking the results of some previous studies (e.g., Crossley, Tywoniw & Choi, 2024; Uzun, 2025). The reason why such indices failed to make a meaningful contribution to the final model might be the existence of similar indices; the potential explanatory power of these specific indices might have been taken over by other complexity indices in the final model. Therefore, future studies should not readily exclude the insignificant measurements in the current study, as L1 influence is a strong possibility in the development of L2 morphology (e.g., Gabriele, 2009; Housen & Simoens, 2016; Schepens et al., 2013; Van der Slik et al., 2019).

Overall findings revealed four essential morphological complexity indices. Inflectional and Derivational Variety scores were significant predictors in explaining L2 essay scores, with Derivational Variety being the most essential variable with the highest estimate value (Table 1). This is in line with the overall token increase for derivational morphemes across essay scores; derivational tokens showed a sharper increase with increasing essay scores compared to inflectional tokens. The difference is expected since inflectional affixes are limited to conveying grammatical information with a limited number of forms in the English language. This finding once again underscores the crucial limitation of morphological complexity studies ignoring the contribution of derivational morphemes (e.g., Březina & Pallotti, 2016).

Finally, MCI estimates for both derivation and inflection decreased with increasing essay scores in the current study. Such a finding might suggest that although bilinguals use a different variety of morphemes in their writing, there is also this increasing affix repetition preventing higher diversity scores due to the lack of unique morphemes across different windows. However, even in such a scenario, MCI scores should have been a number equal to or bigger than zero, following the formula of Březina & Pallotti (2019). Although the lowest possible value for MCI is zero in that formula, TAMMI 2.0 produced negative numbers, suggesting essential changes to the original formula. The calculations for MCI in TAMMI 2.0, therefore, should be evaluated with caution.

Overall, the current results show Derivational and Inflectional Variety estimates calculated by TAMMI 2.0 can successfully help differentiate L2 English essays scores and thus had the potential to explain some variety across L2 proficiency levels as well. As for MCI estimates, the majority of values for derivational and inflectional morphemes were negative, and unlike all other estimates (Figure 1), they show a decreasing trend with increasing essay scores. This suggests the modified formula has potential limitations in calculating morphological diversity, possibly as a result of penalizing longer texts or coding mistakes.

REFERENCES

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, *89*(3), 429-464.

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in psychology*, *3*, 232.

Babayiğit, S., & Stainthorp, R. (2007). Preliterate phonological awareness and early literacy skills in Turkish. *Journal of Research in Reading, 30*(4), 394-413.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *convergence*, 12(1), 2.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *IRAL*, 43, 319–328.

Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research*, *35*(1), 99–119. https://doi.org/10.1177/0267658316643125

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of second language writing*, *26*, 42-65.

Cangır, H., Uzun, K., Can, T., Oğuz, E., & Kaya, Ö. F. (2024). The Development of an Error-tagged Learner Corpus: TELC (Turkish-English Learner Corpus) and its Web-interface. *Dilbilim Araştırmaları Dergisi*, *35*(2), 279-307.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language learning*, *60*(1), 21-43.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (2024). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods*, *56*(6), 5918–5929. https://doi.org/10.3758/s13428-023-02324-w

De La Torre García, N., Ainciburu, M. C., & Buyse, K. (2021). Morphological complexity and rated writing proficiency: The case of verbal inflectional diversity in L2 Spanish. ITL-International Journal of *Applied Linguistics, 172*(2), 290-318.

Durgunoğlu, A. Y., & Öney, B. (1999). A cross-linguistic comparison of phonological awareness and word recognition. *Reading and Writing, 11*, 281-299.

Egbert, J. (2017). Corpus linguistics and language testing: Navigating uncharted waters. *Language Testing*, *34*(4), 555-564.

Gabriele, A. (2009). Transfer and transition in the SLA of aspect: A bidirectional study of learners of English and Japanese. *Studies in Second Language Acquisition, 31*(3), 371-402.

Grömping, U. (2007). Relative importance for linear regression in R: the package relaimpo. *Journal of statistical software*, *17*, 1-27.

Hawkins, R., & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, *118*(4), 595-612.

Hopp, H. (2013). The development of L2 morphology. *Second Language Research*, *29*(1), 3–6. https://doi.org/10.1177/0267658312465304

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*(2), 163-175.

Ishikawa, S. I. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner corpus studies in Asia and the world*, *1*, 91-118.

Kim, J. Y. (2014). Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A Corpus-Based Study. *English teaching*, *69*(4).

Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, 16(4), 776-791.

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*(4), 513-535.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, *102*(2), 333-349.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior research methods*, *50*, 1030-1046.

Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, *60*(1), 221-230.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, *44*, 325-343.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. *The handbook of second language acquisition*, 487-535.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development* (pp. 16-30). Palgrave Macmillan UK.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development* (pp. 16-30). Palgrave Macmillan UK.

Muroya, A. (2019). L1 Transfer in L2 Acquisition of English Verbal Morphology by Japanese Young Instructed Learners. *Languages*, *4*(1), 1. https://doi.org/10.3390/languages4010001

Navarrete, C. B., Soares, F. C., & Navarrete, M. C. B. (2020, January). *Package 'dominanceanalysis'*.

Oğuz, E., & Kırkıcı, B. (2023). The processing of morphologically complex words by developing readers of Turkish: a masked priming study. *Reading and Writing, 36*(8), 2053-2080.

Oğuz, E., & Özge, D. (2020). A Developmental Study of the Turkish Connectives. *Dilbilim Araştırmaları Dergisi, 31*(1), 77-100.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117-134.

Qian, L., Cheng, Y., & Zhao, Y. (2021). Use of Linguistic Complexity in Writing Among Chinese EFL Learners in High-Stakes Tests: Insights From a Corpus of TOEFL iBT. *Frontiers in Psychology*, *12*, 765983. https://doi.org/10.3389/fpsyg.2021.765983

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Schepens, J., van der Slik, F., & van Hout, R. (2013). Learning Complex Features: A Morphological Account of L2 Learnability. *Language Dynamics and Change*, *3*(2), 218–244. https://doi.org/10.1163/22105832-13030203

Selinker, L. (1972). INTERLANGUAGE. *International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209-232. https://doi.org/10.1515/iral.1972.10.1-4.209

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and cognition*, *11*(2), 245-260.

Slabakova, R. (2009). What is easy and what is hard to acquire in a second language. In *10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)*.

Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, *39*(3), 302-327.

Tywoniw, R., & Crossley, S. (2020). Morphological complexity of L2 discourse. In *The Routledge handbook of corpus approaches to discourse analysis* (pp. 269-297). Routledge.

Ullman, Michael T. "The declarative/procedural model and the shallow structure hypothesis." *Applied Psycholinguistics* 27, no. 1 (2006): 97-105.

Uzun, K. (2025). Morphological Complexity Development Across Proficiency Levels in English as a Foreign Language. In *Exploration of the Intersection of Corpus Linguistics and Language Science* (pp. 73-88). IGI Global Scientific Publishing.

Van der Slik, F., van Hout, R., & Schepens, J. (2019). The role of morphological complexity in predicting the learnability of an additional language: The case of La (additional language) Dutch. *Second Language Research*, *35*(1), 47–70. https://doi.org/10.1177/0267658317691322

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239-263.

Wang, G., Wang, H., & Wang, L. (2022). Kolmogorov complexity metrics in assessing L2 proficiency: An information-theoretic approach. *Frontiers in Psychology*, *13*, 1024147.

Wang, G., Wang, H., & Wang, L. (2022). Kolmogorov complexity metrics in assessing L2 proficiency: An information-theoretic approach. *Frontiers in Psychology*, *13*, 1024147. https://doi.org/10.3389/fpsyg.2022.1024147

West, D. E. (2014). Assessing L2 lexical versus inflectional accuracy across skill levels. *Journal of Psycholinguistic Research*, *43*(5), 535–554. https://doi.org/10.1007/s10936-013-9268-0

Wickham, H. (2011). ggplot2. *Wiley interdisciplinary reviews: computational statistics*, *3*(2), 180-185.

Wood, S., & Wood, M. S. (2015). Package 'mgcv'. *R package version*, *1*(29), 729.

Xanthos, A., & Gillis, S. (2010). Quantifying the development of inflectional diversity. *First language*, *30*(2), 175-198.

Xanthos, A., & Gillis, S. (2010). Quantifying the Development of Inflectional Diversity. *First Language*, *30*(2), 175–198. https://doi.org/10.1177/0142723709359236

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics*, *31*(2), 236-259.

Chapter 5:
Motor Activation in Action Language: Task-Dependent but Non-Specific Effects

Hazel Zeynep Kurada

Abstract

Recent research in embodied cognition suggests that language comprehension—particularly for action-related content—is grounded in sensorimotor systems. However, the extent to which concurrent motor activity modulates semantic access remains unclear, especially across different task contexts and languages with complex morphological structures. This study examined whether real-time motor activation facilitates the processing of hand-related action verbs and whether such facilitation is modulated by task demands. Two behavioral experiments were conducted with Turkish-speaking participants. In Experiment 1, participants completed a lexical decision task involving action verbs and pseudowords under two motor conditions: spring squeezing (motor activation) and resting. In Experiment 2, participants judged verbs as either motor or mental actions using the same motor manipulation. Reaction times (RTs) and accuracy rates were measured across both tasks. Results revealed that motor activation significantly reduced RTs across both tasks, indicating a general processing benefit. However, no significant interaction was found between motor condition and verb type, suggesting that motor activity did not selectively facilitate the processing of motor-action verbs. Accuracy remained stable across all conditions, confirming that motor effects were limited to processing speed rather than response accuracy. These findings provide behavioral evidence that embodiment effects are not uniformly elicited, but may depend on factors such as semantic depth, cognitive load, and task structure. The study contributes to simulation-based theories of conceptual representation and expands embodied language research to Turkish—an agglutinative and morphologically rich language—highlighting the importance of cross-linguistic perspectives in grounded cognition research.

Keywords: embodiment, embodied cognition, Turkish, action language, sensorimotor simulation.

1. INTRODUCTION

In traditional semantic models, meaning is often explained through abstract symbols and amodal representations. However, these approaches fail to fully account for how abstract meaning is connected to bodily experience and action. Over the past few decades, theories of embodied cognition have offered a compelling alternative by proposing that conceptual knowledge is deeply grounded in sensorimotor systems and shaped by physical interactions with the world. According to this view, language comprehension—particularly for action-related content—involves partial reactivation of the same neural and cognitive mechanisms that support actual perception and action (Barsalou, 1999, 2008; Glenberg, 1997; Gallese & Lakoff, 2005; Pulvermüller, 2013). Rather than being a secondary byproduct, motor activation is considered an integral component of linguistic processing. Supporting this view, neuroimaging studies have shown that reading action verbs elicits activation in somatotopically organized motor cortices (Hauk, Johnsrude, & Pulvermüller, 2004), and that transient disruption of motor areas via transcranial magnetic stimulation (TMS) can impair the processing of motor-related language (Pulvermüller et al., 2005).

The embodied perspective further emphasizes that cognition is not detached from bodily experience but rather formed through continuous interaction with the environment (Gibbs, 2005; Wilson, 2002). From this standpoint, the body serves both as a medium for experiencing the world and as a foundation for structuring conceptual knowledge. This notion aligns with the theory of *simulation*, which posits that conceptual processing involves reactivating perceptual and motor experiences stored in the brain (Barsalou, 1999; 2008; Gallese & Lakoff, 2005). For instance, reading a verb such as "*to grasp"* can automatically engage motor circuits associated with hand movements (Hauk et al., 2004; Glenberg & Kaschak, 2002). Experimental findings suggest that action-related language consistently activates sensorimotor regions implicated in actual movement execution (Aziz-Zadeh et al., 2006). These representations appear to be stored in, and accessed through, the same neural systems in which the original sensorimotor experiences occurred (Pulvermüller, 2005; Gallese, 2009). For example, reading verbs like *lick, pick*, or *kick* has been shown to activate motor areas corresponding to the tongue, hand, and foot, respectively (Hauk et al., 2004). Such findings indicate a close coupling between linguistic processing and sensorimotor representation.

Evidence for motor system involvement has also emerged across various input modalities and task types. Action-related activation has been observed not only during reading, but also when participants listen to spoken action phrases (Tettamanti et al., 2005), make similarity judgments about motor actions (Kemmerer et al., 2008), or perform semantic tasks under motor interference conditions. Notably, disrupting motor areas with TMS has been found to selectively impair access to action-related semantics (Kurada et al., 2024; Repetto et al., 2013; Pulvermüller et al., 2005). These findings collectively support the idea that motor cortices contribute to the semantic processing of action language.

Nevertheless, the literature remains divided. Despite considerable empirical support, recent reviews and meta-analyses have questioned both the replicability and interpretation of embodiment effects (e.g., Goldinger et al., 2016; Kemmerer et al., 2013; Mahon, 2015a; Mahon & Caramazza, 2008; Maieron et al., 2013; Meteyard et al., 2012; Wilson, 2002). Even among studies employing similar methodologies, the reported effects have often lacked consistency (e.g., Boulenger et al., 2009; Raposo et al., 2009). Some scholars have therefore proposed that sensorimotor activation is not automatically elicited by linguistic input and may function as an accessory rather than a necessary mechanism in language comprehension (Tomasino & Rumiati, 2013). Supporting this interpretation, large-scale meta-analyses of neuroimaging data have found little converging evidence for reliable activation in motor and premotor cortices during the processing of action verbs (Crepaldi et al., 2013; Watson et al., 2013). Together, these findings suggest that language, perception, and motor activity are not isolated cognitive systems, but rather operate as interconnected processes. Yet, the degree to which bodily states and motor engagement modulate semantic processing appears to depend on contextual factors such as task demands, stimulus modality, and linguistic structure. Understanding the boundary conditions of embodiment effects remains a central challenge for theories of grounded cognition.

While many studies support the view that motor activation facilitates action word processing, the absence of a verb-type-specific facilitation effect—particularly for abstract or mental action verbs—raises important theoretical questions. According to simulation theory, embodied representations are reactivated when processing semantically grounded content; however, this reactivation is not obligatory or uniform across all linguistic input. Contextual variables, such as task demands, verb concreteness, and the degree of sensorimotor experience associated with a concept, can modulate the extent and nature of simulation (Barsalou, 2008; Mahon & Caramazza, 2008; Tomasino & Rumiati, 2013). Therefore, a lack of differential effects between motor and non-motor verbs may not necessarily contradict embodiment theories but rather highlight the flexible and context-sensitive character of grounded cognition. In this study, simulation theory provides a framework for testing whether externally induced motor activation can enhance semantic access for action verbs, and whether this facilitation is contingent upon the type of task or verb semantics. By manipulating sensorimotor engagement at the time of word processing, we aim to probe the conditions under which simulation mechanisms are triggered and to clarify their functional role in comprehension.

1.1. Current Study

Despite the growing body of research on embodied language processing, the precise mechanisms through which motor activation influences word comprehension remain unclear. While many studies have shown facilitative effects of motor system engagement on the processing of semantically congruent action words, others report potential interference effects, possibly due to competition for shared cognitive or neural resources. Moreover, many existing studies have been conducted in Indo-European languages—particularly English and German—leaving a gap in our understanding of how embodiment operates in agglutinative, morphologically complex languages such as Turkish, which may follow distinct processing dynamics.

The present study investigates whether engaging in a concurrent hand motor action modulates the processing of action-related words, in terms of both speed and accuracy. Specifically, we test the hypothesis that motor activation congruent with the semantic content of a verb (e.g., squeezing a spring while processing the verb to grasp) facilitates faster and more accurate processing compared to a resting condition. Unlike prior research that often-employed sentence- or discourse-level stimuli, the current design focuses on isolated word-level processing to minimize cognitive load and to target motor-language interactions more precisely. The study is guided by the following research questions:

1. Does concurrent activation of the hand motor cortex enhance the processing speed or accuracy of hand-related action verbs?

2. Is this effect modulated by the type of linguistic task (e.g., lexical decision vs. semantic judgment)?
3. Can findings from Turkish contribute to the broader embodied semantics literature by providing cross-linguistic validation from a typologically distinct language?

By addressing these questions, this study aims to deepen our understanding of how real-time sensorimotor experiences shape language comprehension. Furthermore, it seeks to clarify the functional contribution of motor systems to semantic access and to extend the scope of embodied cognition theories through novel evidence from Turkish.

2. METHODOLOGY

2.1 Participants

Thirty undergraduate students (18 females, 12 males; Mean Age: 23,42) participated in the study. All participants were enrolled at Ankara Medipol University. Participants were recruited on a voluntary basis from the Departments of Psychology, Speech and Language Therapy and Audiology at the Faculty of Health Sciences. The sample size was determined via a priori power analysis using G*Power software. Participants were eligible to take part in the study if they were native and monolingual speakers of Turkish and right-handed. Volunteers were excluded if they had a history of neurological disorders, uncorrected hearing or vision impairments, or any diagnosed speech or language disorders. All participants meeting the inclusion criteria completed a standardized hand dominance inventory (Nalçacı et al., 2002), and only those confirmed to be right-handed were included in the experiment. The study was approved by the Ethics Committee of Ankara Medipol University, and all procedures were conducted in accordance with the Declaration of Helsinki.

2.2. Experimental Stimuli

The study consisted of two experimental sessions, each designed to assess the impact of concurrent motor activity on semantic processing at the word level. A 2 × 2 within-subjects factorial design was employed for both experiments.

2.2.1. Experiment 1: Lexical Decision Task

The first experiment aimed to examine whether hand motor activation modulates the speed and accuracy of lexical access, particularly for action-related verbs. In the first experiment, participants were presented with a total of 100 stimulus items, comprising 50 meaningful hand-related action verbs (e.g., "*tut-*" *to hold, "yaz-" to write*), and 50 pseudowords, created to be phonotactically plausible in Turkish but without any semantic representation (e.g., *gos-, föh-*). To ensure formal consistency with real Turkish verbs, all pseudowords were suffixed with the infinitive morpheme -*mAk*, despite having no semantic content. The experiment design followed a 2 × 2 mixed factorial structure: Motor Condition: Motor Action (spring squeezing) vs. No motor action (resting state), Linguistic Condition: Meaningful hand motor verbs vs. Pseudowords. The design and conditions of the experiment are detailed in Table 13.

Table 13. Sample Stimuli Used in the Lexical Decision Task

Condition	Motor Condition	Stimuli
Cond. 1	spring squeezing	*Tut-*
Cond. 2	spring squeezing	*Düşün-*
Cond. 3	Rest (control condition)	*Yaz-*
Cond. 4	Rest (control condition)	*Kork-*

The pseudowords used in Experiment 1 were generated to be phonotactically legal and morphologically plausible in Turkish. All items (1) respected Turkish syllable structure rules (predominantly CVC or CVCV), (2) contained no semantic representation in Turkish or any known dialect, (3) were constructed manually by altering one or two phonemes in real verbs while preserving overall prosodic and morphological structure, (4) included the infinitive suffix -mAk to parallel real verbs and maintain consistency in visual processing load. Pseudowords were piloted informally with 5 native Turkish speakers to confirm their lack of meaning and their phonological naturalness.

Each participant completed the first experiment in two sessions. In one session, participants performed a spring-squeezing task using the thumb and index finger of their left (non-dominant) hand throughout the task. In the other block, participants were in a rest condition, with no hand movement performed. In each trial, a word or pseudoword appeared in the center of the screen. Participants were instructed to press one of two buttons to indicate whether the stimulus was a real Turkish word or a non-word. Both reaction time and response accuracy were recorded with SuperLab.5 software. The presentation of the motor and rest blocks was counterbalanced across participants to control for order effects.

2.2.2. Experiment 2: Word Judgment Task for Action-Related Content

The second task aimed to determine whether motor activation selectively facilitates the processing of motor-related semantic content, as opposed to abstract concepts that lack a direct sensorimotor grounding. A total of 100 Turkish verbs were used, However, this time, all stimuli were meaningful words, equally divided into two categories: 50 concrete hand-related action verbs (e.g., *to grasp, to write, to squeeze*) and 50 abstract mental-action verbs, which do not involve any physical movement (e.g., *to think, to imagine, to fear*). The second experiment design followed a 2 × 2 mixed factorial structure: Motor Condition: Movement vs. Rest, Linguistic Condition: Motor-action verbs vs. Mental verbs. Thus, the goal was to compare semantic processing of action-related versus abstract concepts under motor activation versus rest conditions. The design and conditions of the experiment are detailed in Table 14.

Table 14. Sample Stimuli Used in the Word Judgment Task

Condition	Motor Condition	Stimuli
Cond. 1	spring squeezing	*Tut-*
Cond. 2	spring squeezing	*Gos-*
Cond. 3	Rest (control condition)	*Yaz-*
Cond. 4	Rest (control condition)	*Föh-*

To ensure semantic validity, all verbs were rated by five native Turkish speakers on a 5-point Likert scale (1 = not at all action-related, 5 = highly action-related). Inter-rater agreement was high (mean agreement = 87%). Items with inter-rater agreement below 80% were excluded or revised. Final categories included: Motor-action verbs involving explicit hand/arm movement (e.g., to grasp, to squeeze, to carry) and Mental-action verbs involving abstract cognitive or emotional processes (e.g., to imagine, to fear, to think). This validation ensured that the two categories represented distinct semantic domains, with minimal conceptual overlap.

In the second experiment, participants were again asked to complete two sessions under the same motor conditions: one involving spring squeezing, and one involving rest. However, in this session, all stimuli were meaningful Turkish verbs, categorized based on their semantic content. In each trial, a word was presented on the screen, and participants were asked to judge whether the verb described a physical hand-related action or a non-motor mental action, responding via button press. As in the first experiment, reaction times and accuracy scores were recorded. The order of motor condition blocks (squeezing vs. rest) was again counterbalanced across participants.

Overall, the stimuli used in both experiments consisted of Turkish action verbs and pseudowords, which were carefully selected and normed to ensure consistency across linguistic variables. A total of 200 stimulus items were used across the two experiments (100 in each), distributed as follows: Experiment 1: 50 meaningful hand-related action verbs and 50 pseudowords; Experiment 2: 50 motor-action verbs and 50 abstract mental-action verbs.

To minimize confounding variables and reduce cognitive load, both experiments were designed at the word level, excluding syntactic or contextual cues. Furthermore, by isolating the motor simulation effect via physical manipulation (spring squeezing), the design enabled a focused investigation into how motor system activation influences semantic processing.

Lexical Properties of Experimental Stimuli

In order to ensure experimental control and eliminate low-level linguistic confounds, all stimulus words were analyzed in terms of frequency, length, and syllabic structure. To ensure that all stimuli were psycholinguistically controlled, a comprehensive analysis of lexical features was performed. All meaningful verbs were selected based on the following criteria: Word frequency values were derived from the Turkish National Corpus (TUD; Aksan et al., 2012). Verbs with extreme high or low frequency were excluded to avoid familiarity bias. All words were infinitive verb forms, consistently ending with the suffix -mak / -mek, which is the canonical citation form in Turkish. To ensure comparability across experimental conditions, all stimuli were carefully matched in terms of key psycholinguistic properties. The three word groups—motor-action verbs, mental-action verbs, and pseudowords—were analyzed for word frequency, word length (number of letters), and syllable count. Statistical comparisons revealed no significant differences among the stimulus groups on any of these variables (ps > .05). This matching process ensured that any observed effects could be attributed to the semantic content and motor congruency of the words, rather than to low-level linguistic differences.

2.3. Procedure

Participants completed both experiments in a soundproof laboratory setting at Ankara Medipol University, Audiology Laboratory. The experiments were run using SuperLab 5.0 software, and responses were recorded with a Cedrus RB-740 response pad. At the beginning of each experiment session, participants received a detailed on-screen instruction. Each block was preceded by a cue indicating whether they should squeeze the spring or remain at rest (e.g., "*In this block, squeeze the spring" / "In this block, do not squeeze the spring*"). Stimuli were presented on a 60 Hz monitor. During experiments, each trial began with a fixation cross ("+") displayed for 500 ms, followed by the presentation of a target word. Participants were instructed to respond as quickly and accurately as possible. In Experiment 1, participants judged whether the presented item was a real Turkish word or not by pressing corresponding buttons. In Experiment 2, participants judged whether the presented word described a physical or a non-physical (mental) action.

Reaction times and accuracy were automatically recorded by the software. The study was designed with blocked motor conditions (i.e., separate blocks for squeezing and rest) to ensure consistent data collection and motor state control. Practice trials were included before each experiment to familiarize participants with the tasks. All reaction times (RTs) and accuracy rates were automatically recorded by the software.

2.3.1 Spring Squeezing Task

Previous studies have shown that passively holding an object may induce minimal activation in the hand motor region, whereas dynamic and intentional motor actions (e.g., writing, drawing, or squeezing) result in significantly higher activation levels (Gallese, 2009). Furthermore, the thumb and index fingers occupy a larger cortical representation in the primary motor cortex compared to other fingers. Therefore, in this study, participants were instructed to perform a spring-squeezing task using the thumb and index fingers of the non-dominant (left) hand, while responding with the dominant (right) hand during the experiment (see Figure 10).

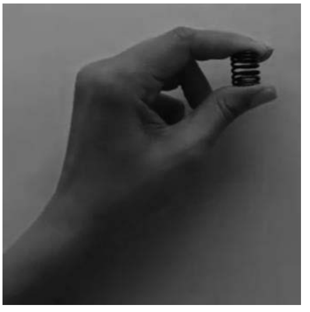


Figure 10. Spring Squeezing Task Performed by Participants During the Experiments

2.4 Data Analysis

All statistical analyses were conducted using IBM SPSS Statistics. The primary dependent variables were reaction times (RTs) and accuracy rates (% correct responses). Separate analyses were carried out for Experiment 1 (Lexical Decision Task) and Experiment 2 (Word Judgment Task for Action-Related Content). Prior to analysis, individual trials with RTs shorter than 200 ms or longer than 3000 ms were excluded, as such responses were considered anticipatory or inattentive outliers. Additionally, for each participant, responses exceeding ±2.5 standard deviations from their mean RT per condition were discarded.

For the statistical analysis, a series of 2 (Motor Condition: Squeezing vs. Rest) × 2 (Linguistic Condition). Repeated-measures ANOVA were conducted separately for each experiment. In Experiment 1, the linguistic conditions were: Meaningful hand-related action verbs and Pseudowords. In Experiment 2, the linguistic conditions were: Motor-action verbs and non-motor (mental) action verbs. A significance level of $p < .05$ was adopted for all statistical tests. The same factorial ANOVA structure was applied for both reaction time and accuracy as dependent variables. Where necessary, Greenhouse–Geisser corrections were applied to adjust for violations of sphericity. Main effects and interaction effects were reported with F-values, degrees of freedom, p-values, and partial eta squared (η^2_p) as a measure of effect size. For significant interaction effects, Bonferroni-corrected pairwise comparisons were performed to further explore the simple main effects between motor and linguistic conditions. These comparisons aimed to determine whether reaction times were significantly faster during spring-squeezing blocks for motor-related words compared to rest blocks and whether accuracy rates were significantly higher for congruent motor-linguistic pairings. *Hypothesis Testing Overview:* The following hypotheses were formally tested: H1: Reaction times would be faster in the motor activation condition compared to the rest condition, specifically for motor-related words. H2: Accuracy rates would be higher for motor-related words when processed under concurrent motor activity. H3: Motor activation would not significantly influence the processing of pseudowords or non-motor mental verbs.

Prior to running the ANOVAs, assumptions of normality and homogeneity of variances were checked. RT distributions were visually inspected and subjected to Shapiro-Wilk tests, which indicated no significant deviation from normality across conditions (all ps > .05). Levene's test for equality of variances was also non-significant for each factor. Where necessary, Greenhouse–Geisser corrections were applied in cases of sphericity violation, and the ε values are reported where applicable. All ANOVAs include partial eta squared (η^2_p) values as a measure of effect size, even for non-significant effects, in line with best practices in statistical reporting.

3. RESULTS

Experiment 1: Lexical Decision Task

A 2 (Motor Condition: squeezing vs. rest) × 2 (Verb Type: action verbs vs. pseudowords) repeated-measures ANOVA was conducted on reaction times (RTs). There was no significant main effect of Motor Condition, $F(1, 29) = 2.14$, $p = .15$, $\eta^2_p = .07$, indicating that overall RTs did not significantly differ between the squeezing (M = 713 ms, SD = 85) and rest conditions (M = 726 ms, SD = 88). In contrast, a significant main effect of Verb Type was observed, $F(1, 29) = 64.92$, $p < .001$, $\eta^2_p = .69$, with faster responses for action verbs (M = 702 ms, SD = 75) than for pseudowords (M = 784 ms, SD = 87). The interaction between Motor Condition and Verb Type was not significant, $F(1, 29) = 1.38$, $p = .25$, $\eta^2_p = .05$, indicating that motor activation did not differentially affect the processing of action verbs compared to pseudowords. For accuracy, no significant main effects or interaction were found: Motor Condition, $F(1, 29) = 1.26$, $p = .27$, $\eta^2_p = .04$; Verb type, $F(1, 29) = 2.04$, $p = .16$, $\eta^2_p = .07$; Interaction, $F(1, 29) = 0.94$, $p = .34$, $\eta^2_p = .03$.

Experiment 2: Word Judgment Task for Action-Related Content

A 2 × 2 repeated-measures ANOVA was conducted on response times (RTs), with Motor Condition (squeezing vs. rest) and Verb Type (motor-related vs. mental-related verbs) as within-subject factors. There was a significant main effect of Motor Condition, $F(1, 29) = 27.83$, $p < .001$, $\eta^2_p = .49$, indicating that participants responded faster in the active motor condition (M = 1019 ms, SD = 101) than in the no-motor condition (M = 1349 ms, SD = 188). A significant main effect of Verb Type was also observed, $F(1, 29) = 10.03$, $p = .004$, $\eta^2_p = .26$, with motor-related verbs (M = 1032 ms, SD = 175) processed faster than mental-related verbs (M = 1422 ms, SD = 195). However, the interaction between Motor Condition and Verb Type was not significant, $F(1, 29) = 1.19$, $p = .30$, $\eta^2_p = .04$, indicating that the effect of motor activation was not selectively greater for motor verbs compared to mental verbs. For accuracy data, a similar 2 × 2 repeated-measures ANOVA showed no significant main effects or interaction: Motor Condition, $F(1, 29) = 0.44$, $p = .51$, $\eta^2_p = .02$; Verb Type, $F(1, 29) = 0.17$, $p = .68$, $\eta^2_p = .01$; Interaction, $F(1, 29) = 0.22$, $p = .64$, $\eta^2_p = .01$. Overall accuracy remained stable across conditions (M = 96.5%, SD = 2.4%).

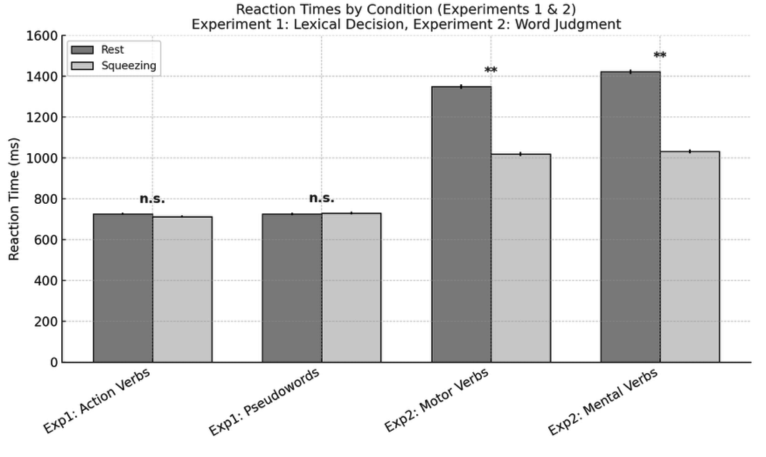


Figure 11. Mean reaction times (RTs) across motor and rest conditions in Experiments 1 and 2. Error bars indicate standard errors of the mean. In Experiment 1 (Lexical Decision Task), no significant effects of motor condition were observed, although action verbs elicited faster RTs than pseudowords. In Experiment 2 (Word Judgment Task), motor activity led to significantly faster RTs across both verb types. However, no interaction was found, suggesting that the motor effect was not verb-type-specific. n.s. = non-significant; ** = p < .01.

4. DISCUSSION

The present study aimed to examine whether concurrent motor activity modulates the processing of action-related language, as predicted by embodied cognition theories. Two behavioural experiments were conducted with Turkish-speaking participants, using distinct task demands—lexical decision (Experiment 1) and semantic categorization (Experiment 2)—to assess whether motor-language interactions are influenced by task structure and processing depth. Across both experiments, participants performed a spring-squeezing task in one block and completed the same linguistic task at rest in another. The dependent measures were reaction time (RT) and accuracy. The study also sought to broaden the empirical base of embodied cognition by testing Turkish, a morphologically rich and typologically distinct agglutinative language.

In both experiments, motor activation led to faster response times overall, regardless of verb type. However, contrary to our initial hypotheses, the findings revealed no evidence of a selective motor facilitation effect for action verbs under motor activation. While participants responded faster overall during the motor condition compared to rest, this effect was observed regardless of verb type, and the interaction between motor condition and verb type was not statistically significant in either experiment. These results suggest that motor activation, in and of itself, does not selectively enhance the processing of motor-related language, at least when stimuli are presented in isolation at the word level. Additionally, presenting the stimuli in isolation—without syntactic or contextual embedding—may have further reduced the likelihood of triggering full semantic simulation. Prior research suggests that embodied effects are often more robust in sentence or discourse-level contexts, where conceptual depth is richer and simulation demands are higher (Glenberg et al., 2008). Thus, no evidence was found to support the hypothesis that concurrent motor activity selectively facilitates the processing of motor-action verbs over mental-action verbs or pseudowords.

These findings challenge the notion that embodiment effects are automatic or obligatory features of language processing. Although earlier neuroimaging and TMS studies have provided support for the idea that understanding action words recruits motor areas in the brain (Hauk et al., 2004; Pulvermüller et al., 2005), more recent research—including large-scale meta-analyses—has raised concerns about the consistency and interpretability of such effects (Mahon & Caramazza, 2008; Crepaldi et al., 2013; Watson et al., 2013). The current findings align with this latter body of work by demonstrating that motor engagement alone is insufficient to produce a robust facilitation effect specific to motor-action semantics.

Several factors may account for this null result. First, the motor task—squeezing a spring—may not have been semantically specific enough to activate effector-specific motor representations at a fine-grained level. While the task likely engaged the primary motor cortex, it may not have generated simulation-like activation patterns that map onto the semantic representations of verbs like to grasp or to write. Second, as argued in recent theoretical work (Barsalou, 2008; Tomasino & Rumiati, 2013), the depth of semantic processing may determine whether embodiment effects emerge. The lexical decision task in Experiment 1, in particular, may have relied more on orthographic or lexical familiarity judgments than on access to conceptual content, thereby failing to invoke simulation processes. Even in Experiment 2, which required participants to explicitly judge verbs as motor- or mental-related, no significant interaction emerged between motor condition and verb type. This is especially telling, as prior studies using categorization tasks have reported embodiment effects under similar conditions (e.g., Glenberg & Kaschak, 2002; Aziz-Zadeh et al., 2006). One possible explanation is that sensorimotor simulation may be context-sensitive and easily disrupted by concurrent cognitive demands. It is plausible that the motor task itself introduced an additional attentional load or competition for cognitive resources, thus attenuating the potential benefits of congruent motor-language coupling. This possibility aligns with theories suggesting that motor simulation is a goal-directed, resource-dependent process, rather than a reflexive one (Gallese & Lakoff, 2005; Glenberg & Gallese, 2012).

The consistent main effect of motor condition on RTs in both experiments does, however, indicate that motor activity influenced overall processing speed, albeit not in a semantically selective manner. This finding is important in itself. It shows that motor activation during language tasks can alter general processing dynamics, even if it does not produce differential effects based on semantic congruency. Moreover, the absence of any significant effects on accuracy reinforces the interpretation that motor activity affected the efficiency, rather than the quality, of linguistic responses.

From a cross-linguistic perspective, these findings contribute a critical data point to the growing recognition that language typology and morphological structure may shape the expression of embodiment effects. While prior studies have largely focused on English or German, the current research situates Turkish—a morphologically complex and agglutinative language—within this theoretical framework. The absence of a motor-verb-specific facilitation effect in Turkish may indicate that semantic activation patterns are more distributed or delayed in morphologically dense contexts, and thus less amenable to enhancement through concurrent motor activity. This pattern appears consistent with findings from other agglutinative languages such as Korean or Japanese, where the morphological richness of verbs may alter access patterns to lexical-semantic content. Such cross-linguistic parallels emphasize the need to integrate typological variation into embodiment models more systematically.

Methodologically, the study employed a blocked design with clearly defined motor and rest conditions and utilized a dynamic motor task (spring squeezing) involving high-cortical-representation finger movements (Gallese, 2009). However, despite the ecological validity of the motor manipulation, its semantic alignment with specific verb meanings remains questionable. Future studies could benefit from incorporating effector-specific motor tasks that more directly mirror the conceptual representations of the target verbs. Nonetheless, several limitations should be acknowledged. First, only hand-related verbs were used; future work should examine whether effects differ for other effector domains (e.g., foot, mouth). Second, the absence of neuroimaging measures limits conclusions about neural mechanisms; combining behavioral paradigms with EEG, fMRI, or TMS would allow for a more precise understanding of the neural dynamics underlying motor-language coupling. Finally, the study did not include incongruent motor-linguistic pairings, which may offer further insight into interference effects predicted by some simulation models. Lastly, all linguistic stimuli in the present study were isolated words. It remains an open question whether richer semantic context—such as that provided by sentence-level or discourse-level stimuli—might amplify embodiment effects by increasing conceptual depth and simulation demands. Future studies should explore whether embedding action-related language within syntactic or pragmatic context elicits stronger motor-language coupling.

In summary, the present findings highlight the non-automatic and context-sensitive nature of motor-language interactions. Although embodied cognition theories propose that language comprehension partially relies on sensorimotor simulation, the current results suggest that such simulation is not obligatorily triggered by motor language nor easily enhanced by simple motor activation. Rather, embodiment effects appear to be conditional on task demands, cognitive load, and the semantic richness of the stimuli. These insights provide a more nuanced understanding of embodied language processing and underscore the need for further research in typologically diverse languages and experimental contexts.

5. CONCLUSION

The present study offers empirical insights into how bodily states interact with linguistic processing, drawing on the framework of embodied cognition. Across two behavioural experiments, we examined whether concurrent motor activation modulates the processing of action-related language in Turkish—a morphologically complex and typologically distinct language that remains underrepresented in this literature. Participants completed lexical decision and semantic categorization tasks under both motor (spring squeezing) and resting conditions, allowing us to isolate the behavioural impact of sensorimotor engagement during language comprehension. Contrary to early theoretical predictions, the findings revealed no evidence for a semantically selective facilitation effect. While motor activation significantly reduced reaction times across both experiments, this effect was not modulated by verb type—motor-action and mental-action verbs were processed equally faster under motor engagement. These results challenge the assumption that sensorimotor simulation is automatically or preferentially engaged by motor-related language and instead point to a non-specific, general processing boost associated with motor activity. These findings are consistent with a growing body of research questioning the automaticity and replicability of embodiment effects (Goldinger et al., 2016; Mahon & Caramazza, 2008; Tomasino & Rumiati, 2013). Although prior studies have demonstrated motor-specific cortical activation in response to action language (Hauk et al., 2004; Pulvermüller et al., 2005), such effects may not reliably translate to behavioural facilitation, particularly in tasks involving isolated words and low semantic demands. Indeed, lexical decision tasks may not engage deep semantic processing sufficiently to elicit sensorimotor simulation, and even explicit verb categorization—used in Experiment 2—did not result in a verb-type-specific motor effect. Importantly, the observed general RT facilitation suggests that motor activity influences the dynamics of language processing, though not in a semantically targeted way. This raises the possibility that bodily engagement may modulate attention or arousal, thereby accelerating cognitive responses broadly, rather than activating effector-specific semantic representations. It also suggests that simple motor tasks such as spring squeezing may not be sufficient to trigger meaningful semantic resonance at the word level. Furthermore, the concurrent motor demand may have introduced cognitive load, attenuating any subtle simulation-based effects. From a theoretical perspective, these findings invite a more nuanced view of embodied language processing as flexible and context-dependent, rather than fixed or automatic. The absence of semantic specificity does not disprove simulation-based models but rather highlights the need to identify boundary conditions—such as task complexity, depth of semantic engagement, and the level of linguistic context—that shape when and how sensorimotor systems contribute to meaning.

The present study also expands the embodied cognition literature by demonstrating these patterns in Turkish. That no motor-verb-specific advantage emerged in this typologically distinct language suggests that embodiment effects are not universally expressed and may be shaped by linguistic and morphological features. This highlights the value of cross-linguistic work in testing the generalizability of cognitive theories. Methodologically, the use of a physically active motor task and a blocked design enabled a clean manipulation of motor engagement. However, several limitations must be noted. First, the use of isolated words may have reduced semantic richness; sentence-level or discourse-level stimuli might evoke more robust simulation effects. Second, neurophysiological methods were not employed, limiting inferences about the underlying neural dynamics. Finally, only hand-related verbs were tested; future work could explore different effectors or incongruent motor-language pairings.

In conclusion, the study underscores that language comprehension is not an abstract or disembodied process, but one that unfolds within the embodied mind. However, the degree to which bodily systems participate in meaning construction is not uniform—it depends on task demands, cognitive resources, and linguistic structure. These findings highlight the importance of investigating when and how the body contributes to language, rather than assuming it always does. Future research should continue to refine the conditions under which embodiment arises, integrating behavioural and neural measures, diverse languages, and richer stimuli to advance a more comprehensive theory of grounded semantics.

REFERENCES

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., ... & Kurtoglu, Ö. (2012, May). Construction of the Turkish National Corpus (TNC). In *LREC* (pp. 3223-3227).

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current biology*, *16*(18), 1818-1823.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and brain sciences, 22(04), 577660.

Barsalou, L. W. (2008). Grounded cognition. Annu. Rev. Psychol., 59, 617-645.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral cortex*, *19*(8), 1905-1914.

Crepaldi, D., Berlingeri, M., Cattinelli, I., Borghese, N. A., Luzzatti, C., & Paulesu, E. (2013). Clustering the lexicon in the brain: a meta-analysis of the neurofunctional evidence on noun and verb processing. *Frontiers in human neuroscience*, *7*, 303.

Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic dialogues*, *19*(5), 519-536.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, *22*(3-4), 455-479.

Gibbs Jr, R. W. (2005). Embodiment and cognitive science. Cambridge University Press.

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 495–522). American Psychological Association.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *cortex*, *48*(7), 905-922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin & review*, *9*(3), 558-565.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic bulletin & review*, *23*, 959-978.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and language*, *107*(1), 16-43.

Kemmerer, D., Miller, L., MacPherson, M. K., Huber, J., & Tranel, D. (2013). An investigation of semantic similarity judgments about action and non-action verbs in Parkinson's disease: implications for the Embodied Cognition Framework. *Frontiers in human neuroscience*, *7*, 146.

Kurada, H. Z., Jiménez-Bravo, M., Giacobbe, C., & Obeso, I. (2024). Context, transparency and culture in motor resonance phenomena: Causal evidence of the motor cortex. *Cortex*, *179*, 25-34.

Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *69*(2), 172.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of physiology-Paris*, *102*(1-3), 59-70.

Maieron, M., Marin, D., Fabbro, F., & Skrap, M. (2013). Seeking a bridge between language and motor cortices: a PPI study. *Frontiers in Human Neuroscience*, *7*, 249.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788-804.

Nalçaci, E., Kalaycioğlu, C., Güneş, E., & Çıçek, M. (2002). El Tercihi Anketinin Geçerlik ve Güvenilirliği. *Türk Psikiyatri Dergisi*.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences*, *17*(9), 458-470.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388-396.

Repetto, C., Colombo, B., Cipresso, P., & Riva, G. (2013). The effects of rTMS over the primary motor cortex: The link between action and language. Neuropsychologia, 51(1), 8-13.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., ... & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of cognitive neuroscience*, *17*(2), 273-281.

Tomasino, B., & Rumiati, R. I. (2013). At the mercy of strategies: the role of motor representations in language understanding. *Frontiers in Psychology*, *4*, 27.

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of cognitive neuroscience*, *25*(8), 1191-1205.

Wilson, M. (2002). Six views of embodied cognition. Psychonomic bulletin & review, 9(4), 625636.

Chapter 6:
Subject Control in Temporal Converb Constructions in Turkish: Evidence from the Turkish National Corpus (TNC)

Doğan Baydal

ABSTRACT

The aim of this study is to investigate the subject control structure in temporal converb clause constructions in Turkish by referring to the Turkish National Corpus (TNC) (Aksan et al., 2012). The study is based on the typology of Nedjalkov (1995) who distinguishes three types of converbs in terms of control structure: (i) same-subject converbs, (ii) different-subject converbs and (iii) varying-subject converbs. The temporal converbal suffixes analysed in this study are -(y)IncA (when), -DIğI zaman (when), -DIğIndA (when), -(A/I) r…-mAz (as soon as), -DIğIndAn beri (since), -DIktAn sonra (after), -mAdAn önce (before), -ken (while) and -DıkçA (whenever) in Turkish. 1000 instances from each of the above clauses were randomly selected and in total 9000 instances were analysed in this study. The results of the study show that when the subject of the converb clause is implicit, it can either be the same as the subject in the main clause or a different subject from the subject of the main clause. In certain instances, subject agreement markers assist Turkish speakers in identifying the subject of the converb clause in constructions where the subject is implicit. Another result is that converbs with agreement morphology may have an explicit subject. Thus, it is safe to state that Turkish has the varying subject control in temporal converb constructions.

Keywords: temporal converb constructions, control structure, Turkish, subject, corpus

1. INTRODUCTION

A frequent topic in modern linguistics, which has been extensively studied, concerns the interpretation of sentences like those in example (1), where the subordinate and main clauses have their own subjects and example (2), where the implied subject of the subordinate clause is interpreted as being co-indexed with the subject of the main clause.

(1) Ayşe [el-in-de-ki tepsi-yi düşür-ünce]
Ayşe hand-POSS-LOC-POSS tray-ACC drop-CVB[1]
Akif gül-me-ye başla-dı.
Akif cry-INF-DAT start-PST-3SG
'Akif started laughing when Ayşe dropped the tray in her hand.'

(2) Ayşe [el-in-de-ki tepsi-yi düşür-ünce]
Ayşe hand-POSS-LOC-POSS tray-ACC drop-CVB
ağla-ma-ya başla-dı.
cry-INF-DAT start-PST-3SG
'Ayşe started crying when she dropped the tray in her hand.'
(Çetintaş Yıldırım, 2004, p. 98)

In example (2), the verb in the subordinate clause appears to have its valence fulfilled even though it is non-finite. These types of constructions are referred to as control constructions, where a main clause argument, known as the controller, determines or provides the interpretation of the subordinate clause's subject, known as the controlee (Yasavul, 2009). Control constructions have attracted significant attention since the early days of Transformational Grammar (Chomsky, 1965). One of the first analyses of control can be traced back to Postal (1974) within the framework of Transformational Grammar, where both finite and non-finite structures are generated from the same deep structure by applying different transformations. In Government and Binding (GB) Theory (Chomsky, 1980), the implicit subject of a subordinate clause is labelled as PRO. GB Theory employs control theory, along with other additional modules like government, binding, case, and theta theory, to determine the distribution and interpretation of PRO. With the introduction of Minimalist Program (MP), the phenomenon of control structure is analysed as instances of movement, generated through move and merge operations (Yücel, 2009). As well as transformational grammar framework, control structure has been analysed in non-transformational grammar frameworks, such as Lexical-Functional Grammar, Montague Grammar and Head-driven phrase structure

[1] Leipzig Glossing Rules were applied.

grammar. All these grammar frameworks have tried to allow us to gain a deeper understanding of the control structure phenomenon.

The aim of this study is to investigate the subject control structure in temporal converb clause constructions in Turkish by referring to Turkish National Corpus (TNC) and the research question guiding this study aligns with its aim:

1. How are the subject control strategies realized in temporal converb constructions in Turkish?

2. BACKGROUND OF THE STUDY

2.1 Temporal Converb Constructions

Haspelmath (1995) defines converb as "a non-finite verb form whose main function is to mark adverbial subordination" (p. 4). Johanson (1995) states that many characteristics commonly associated with converbs in modern languages are already present in the earliest known Turkic texts. In converb constructions, the converb segment is considered as a non-finite unit that is structurally subordinate to the matrix segment. Converb segments, which at minimum include a verb form but can expand into full clauses, are marked with suffixes that function similarly to subordinating conjunctions in European languages. Example (3) below is an example of a converb construction from Turkish.

(3)	Ayşe	sınav-ı	geç-ince	Ahmet	sevin-di.
	Ayşe	exam-ACC	pass-CVB	Ahmet	be.happy-PST-1SG
	'When Ayşe passed the exam, Ahmet was happy.'				

In example (3), the converb is marked by an affix, namely "-(y)IncA", which is attached to the verb stem, "-geç (to pass)" and the subordinate clause gives temporal meaning to the whole construction. Banguoğlu (1995) explains that temporal converbial suffixes are added to verbs in subordinate clauses to express a time relationship, linking these subordinate clauses to the main clause. According to the classifications of Göksel and Kerslake (2005) and Kornfilt (1997), the temporal converb suffixes in Turkish are: *-(y)IncA* (when), *DIğIndA* and -*DIğI zaman* (when), -(*y*)*ken* (while, as), (*A/I)r*…-*mAz*, -*DIğI gibi* (as soon as), -(y)AlI (beri), -*DIğIndAn beri* (since), *-mAdAn (önce)* (before), -*DIktAn sonra (after)*, *-(y)IncAyA kadar / dek* (until), *-DIkçA* (whenever) and -*DIğI sürece/ müddetçe* (throughout the time).

Nedjalkov (1990) categorizes temporal converb clauses into three groups, with the first group indicating a relationship of simultaneity between the main clause and the subordinate clause. The suffixes of this category are *-ken* (while), *-DIğI zaman* (when), *-DIğIndA* (when), *-DıkçA* (whenever) and *-DIğI sürece/ müddetçe* (throughout the time). The second group conveys an anteriority relationship between the subordinate clause and the main clause. The suffixes of this category are *-(y)IncA* (when), *-DIğIndA* (when), *-DIğI zaman* (when), *-(A/I) r…-mAz* (as soon as), *-DIğIndAn beri* (since) and *-DIktAn sonra* (after). The third group represents a posteriority relationship between the subordinate clause and the main clause. The suffixes of this category are *-DığIndA* (when), *-DIğI zaman* (when), *-mAdAn önce* (before) and *-(y)IncAyA kadar / dek* (until). Table 15 shows the temporal converb suffixes in Turkish according to the meaning relationships they have.

Table 15 Temporal Converb Suffixes in Turkish According to the Meaning Relationships (Adopted From Kornfilt (1997) and Göksel and Kerslake (2005))

Meaning relationships	Temporal converb suffixes
Simultaneity	-ken (while), -DIğI zaman (when), -DIğIndA (when), -DıkçA (whenever) and -DIğI sürece/ müddetçe (throughout the time).
Anteriority	-(y)IncA (when), -DIğIndA (when), -DIğI zaman (when), -(A/I) r…-mAz (as soon as), -DIğIndAn beri (since) and -DIktAn sonra (after).
Posteriority	-DığIndA (when), -DIğI zaman (when), -mAdAn önce (before) and -(y)IncAyA kadar / dek (until)

It is seen from Table 15 that some of the converb suffixes are in more than one category because the aspectual characteristics of the main clause influence the interpretation of the converb clause.

2.2 Subject Control in Temporal Converb Constructions

Erguvanlı Taylan (1996) suggests that subject control occurs in four distinct forms of non-finite subordinate clause constructions. These are (i) object complements, (ii) subject complements, (iii) purpose clauses and (iv) gerunds. Examples below show subject control in non-finite subordinate constructions:

(4)	Ben$_i$	[Ø$_i$ ayakta durma-ya]	alışkın-ım.
	I	stand-COMP	be.used.to-1SG

'I am used to standing.'

(5) [Ø Sigara içmek] yasak-tır.
smoking forbidden-PRS
'It is forbidden to smoke here.'

(6) Ben*i* [Ø*i* maç-a git-mek için] para
I match-DAT go-PURP money
biriktir-iyor-um.
save-PROG-1SG
'I am saving money in order to go to match.'

(7) [Ø*i* Okul-dan gel-ince] ben*i* ders çalış-tı-m.
school-ABL come-CVB I lesson study-PST-1SG
'When I came from school, I studied my lessons.'

In example (4), object complement has empty category PRO; in example (5), subject complement has empty category PRO; in example (6), purpose clause has empty category PRO and in example (7), the converb has empty category PRO and the subject is co-indexed with the subject of the matrix clause.

In terms of subject control in converb constructions, Haspelmath (1995) distinguishes three types of converbs. These are (i) implicit-subject converbs, (ii) explicit-subject converbs and (iii) free-subject converbs. In *implicit- subject* converbs, subjects may not be expressed explicitly. A related example is given from Russian below.

(8) Ona prigotoviv zavtrak Zamira razbudila
she prepare-PRF-CVB breakfast Zamira wake.up-PST
detej.
children
'Having prepared breakfast, Zamira woke up the children.'
(Haspelmath, 1995, p. 10)

In example (8), the subject of the subordinate clause is not overtly expressed, it is only expressed with the main clause verb agreement. In *explicit-subject* converbs, subjects are expressed explicitly. A related example is given from Lithuanian below.

(9) Rut-ai išej-us iš mišk-o, patekejo saule
Ruta-DAT go-out-CVB from forest-GEN rise-PST sun
'When Ruta went out of the forest, the sun rose.'

(Haspelmath, 1995, p. 10)

In example (9), the subjects of the subordinate clause and main clause are overtly expressed. In *free-subject* converbs, the subject of the subordinate clause can be expressed explicitly, but it's not necessary. A related example is given from Lezgian below.

(10) (Ĉeb) mašhur Samarkanddi-z agaq'-daldi
selves well-known Samarkand-DAT reach-CVB
aburu-z req'-e gzaf zat'-ar aku-na.
they-DAT way many thing-PL see-PRS
'Before they reached well-known Samarkand, they saw a lot of things on the way.'

(Haspelmath, 1995, p. 10)

In example (10), the subject of the subordinate clause is identical to the subject of the main clause, but in other cases, it may be different from the subject of the main clause. Nedjalkov (1995) also divides converbs into three categories in terms of referential types of the subject in converb constructions. These categories are (i) same-subject converbs, in which the subjects of the subordinate clause and main clause are coreferential; (ii) different-subject converbs, in which the subjects of the subordinate clause and main clause are non-coreferential and (iii) varying-subject converbs, in which the subjects of the subordinate clause and main clause may be coreferential or non-coreferential. The classification of Nedjalkov (1995) is not independent of the typology of Haspelmath (1995) and the connections can be seen in Table 16.

Table 16 Subject Reference in Converbs (Haspelmath, 1995, P. 10)

	Same-subject	Different-subject	Varying-subject
Implicit-subject	typical	unusual	unusual
Explicit-subject	unusual	typical	unusual
Free-subject	unusual	unusual	typical

Table 16 shows that when the subject is implied, its reference can only be identified through a same-subject reference. Conversely, when the subject differs from the main clause constituents, its reference can only be identified through explicit expression. This study is based on the classifications of Haspelmath (1995) and Nedjalkov (1995) in analysing the subject control in temporal converb constructions in Turkish.

2.3. Previous Research

Kortman (1995) analyses control structure in adverbial participles in English from a typological point of view and finds that the typical English converb is the "*-ing*" form, where the subject is implied and determined by the subject of the main clause, without any additional markers. However, English converbs can also use other non-finite forms or verbless constructions, may have their own explicit subject, and may be introduced by a prepositional phrase. Weiss (1995) examines control properties in Russian and explores which nouns can control the implicit subject of a converb. He notes that the issue has often been approached from a prescriptive standpoint and observes that, despite grammarians' attempts to restrict overt controllers to subjects, non-subject controllers have been regularly used over the past 150 years. These controllers appear not only in newspapers but also in the works of renowned authors. Weiss further points out that Russian displays notable flexibility in permitting non-subject controllers, as long as they hold semantic relevance. Groot (1995) examines Hungarian converbs ending in "*-va/-ve*" and states that in the adverbial use of the converb, the subject and the object of the converb may be coreferential with the subject of the main clause. He further states that the converb in Hungarian may also be used as a predicative adjunct. In that case, the subject of the converb may be coreferential with either the subject or the object of the main clause. Haspelmath (1995) examines control in Lezgian converbal constructions, finding that control relations are not highly grammaticalized in the language, even though contextual converbs frequently omit and control subjects. He also notes that the lack of grammaticalization extends to the omission and control of non-subject arguments, allowing them to be omitted and controlled as well. Nedjalkov (1995) examines the control structure of converbs in Evenki, revealing that the language uses both same-subject and varying-subject converbs but lacks different-subject control. The study also indicates that same-subject converbs only use the plural suffix "*-l*" and do not take any other subject agreement suffixes. Varying-subject converbs, however, may take the markers "*-vi/-var*" for reflexive possession in same-subject constructions or personal possession markers in varying-subject constructions. Alpatov and Podlesskaya (1995) investigate the control structure of primary and secondary converbs in Japanese. Their findings indicate that nearly all primary and secondary converbs can function in two ways: (i) as same-subject constructions, where the implicit subject is controlled by the main clause subject, and (ii) as different-subject constructions, where the subject is explicitly stated, similar to an independent clause. The studies above are from different languages on control structure on converb constructions. There are limited studies on control structure on temporal converb constructions in Turkish, one of which is by Çetintaş Yıldırım (2004). She finds that Turkish temporal converb clauses can have subjects that are either co-referential or non-coreferential with the matrix clause. If the subjects are non-coreferential, leaving out a lexical subject in the converb clause results in ungrammatical sentences. Therefore, including a lexical subject in the converb clause is obligatory in such cases. Moreover, she states that when there is no overt lexical subject in the temporal converb clause, the converb clause subject is regarded as being co-indexed with the subject of the main clause.

3. METHODOLOGY

This section provides a detailed explanation of the study's methodology. It begins by presenting the converb clauses examined in the research. Following this, the data collection process is outlined including the sample size. After that, data analysis is provided including the coding process.

3.1 Temporal Converb Clauses Analysed in The Study

The temporal converb clauses examined in this study fall into three categories: (i) converbs that indicate a relationship of simultaneity, (ii) converbs that indicate a relationship of anteriority and (iii) converbs that indicate a relationship of posteriority. The temporal converbal suffixes of the first category that are analysed in this study are *-DIğIndA* (when), *-DIğI zaman* (when), *-ken* (while) and *-DıkçA* (whenever). The temporal converbal suffixes of the second category that are analysed in this study are *-(y)IncA* (when), *-DIğIndA* (when), *-DIğI zaman* (when), *-(A/I) r...-mAz* (as soon as), *- DIğIndAn beri* (since) and *-DIktAn sonra* (after). The temporal converbal suffixes of the third category that are analysed in this study are *-DIğIndA* (when), *-DIğI zaman* (when) and *-mAdAn önce* (before).

3.2 Data Collection

The study's data were gathered from the Turkish National Corpus (TNC) (Aksan et al., 2012), which consists of 50 million words. This corpus includes a wide range of samples, covering a 24-year period from 1990 to 2013. For this study, 1000 instances (both from written and spoken corpus) from each of the following clauses were randomly selected: *-(y)IncA* (when), *-DIğIndA* (when), -DIğI zaman (when), -ken (while), *-(A/I) r…-mAz* (as soon as), *-DIğIndAn beri* (since), *-mAdAn önce* (before), *-DIktAn sonra* (after), and *-DıkçA* (whenever). Simple random sampling, a basic sampling method, was employed to select the example sentences. In this technique, each sentence in the sampling frame is assigned a unique number, and the sample is then chosen using a random number table (McEnery et al., 2006). In total, 9000 instances were analysed in this study.

The reason for analysing these nine specific temporal converbal suffixes is that the standardized frequency values of these nine converbal endings are higher when compared to other temporal converbal suffixes in Turkish. Biber et al. (1998) assert that standardized values are essential for adjusting raw frequency counts from texts of differing lengths, enabling accurate comparisons. Table 17 below shows the standardized and observed frequency values of the temporal converbal suffixes for this study.

Table 17 The Standardized and Observed Frequency Values of the Temporal Converbal Suffixes

Converbal Suffixes	Observed Frequency	Normalised Frequency
-(y)IncA (when)	41976	828,9
-ken (while)	40117	811
-DIğIndA (when)	34609	699,6
-mAdAn önce (before)	32221	651,4
-DIğI zaman (when)	31344	633,6
-DIktAn sonra (after)	30993	626,5
-DIkçA (whenever)	27543	556,8
-DIğIndAn beri (since)	23670	478,5
-(A/I) r…-mAz (as soon as)	20019	404,7
-DIğI gibi (as soon as)	1336	27
-(y)IncAyA kadar (until)	865	17,4
-DIğI sürece(throughout the time)	533	10,7
-(y)IncAyA dek (until)	391	7,9
-DIğI müddetçe (throughout the time)	223	4,5

Table 17 shows that the standardized frequency values of *-(y)IncA* (when), *-DIğIndA* (when), *-DIğI zaman* (when), *-ken* (while), -(A/I) r…-mAz (as soon as), *-DIğIndAn beri* (since), *-mAdAn önce* (before), *-DIktAn sonra* (after), and *-DıkçA* (whenever) converbal suffixes are higher than 400 and observed frequency values are more than 20.000. The standardized frequency values of *-DIğI gibi* (as soon as), *-(y)IncAyA kadar/dek* (until) and *-DIğI sürece/müddetçe* (throughout the time) are less than 27 and observed frequency values are less than 1400. Thus, these nine specific temporal converbal suffixes are selected to be analysed in this study.

3.3 Data Analysis

Each temporal converbal instance was analysed to evaluate two aspects: (i) implicit subject converb where the subject of the converb construction is not expressed and (ii) explicit subject converb where the subject of the converb construction is expressed overtly. In implicit subject converb constructions, if the converb construction allows agreement morphology, each temporal converbal instance was analysed to evaluate two more aspects: (i) implicit subject converb construction with agreement marker and (ii) implicit subject converb construction without agreement marker. Similarly, in explicit subject converb constructions, if the converb construction allows agreement morphology, each temporal converbal instance was analysed to evaluate two more aspects: (i) explicit subject converb construction with agreement marker and (ii) explicit subject converb construction without agreement marker. The coding process is shown in Figure 12 below.

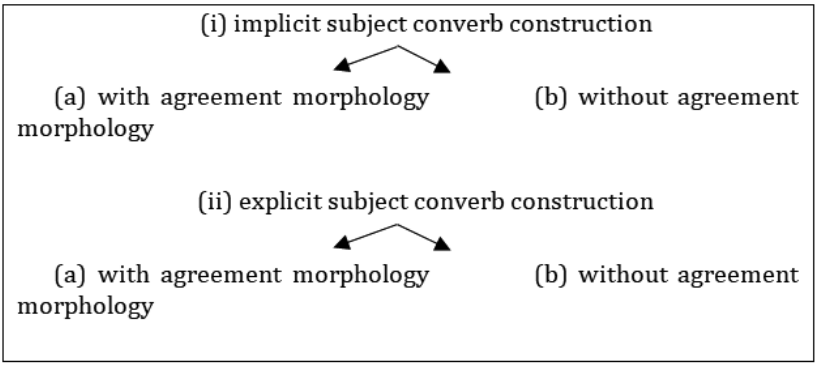


Figure 12. The coding process of the temporal converbal constructions.

Figure 12 shows that there are six different parameters in the coding process. Two researchers independently conducted the coding process. After the two test sets were independently coded by the raters, the consistency of their coding was assessed using Cohen's Kappa. For the coding between implicit subject converb and explicit subject converb, Cohen's κ = .999; for the coding between implicit subject converb with agreement morphology and implicit subject converb without agreement morphology, Cohen's κ = .993; and for the coding between explicit subject converb with agreement morphology and explicit subject converb without agreement morphology, Cohen's κ = .990. McHugh (2012) states that if the results of the Cohen's Kappa are between 0.81–1.00, there is an almost perfect agreement between the raters. Thus, it is safe to state that the results of the coding process are reliable.

4. FINDINGS AND DISCUSSION

The study's findings are analysed individually for each converbal suffix. Table 18 shows the subject control structure in -ken (while) converb construction.

Table 18 Control Structure in -ken (While) Converb Construction

	Implicit Subject Converb 59.1% (n=591)		**Explicit Subject Converb** 40.9% (n=409)	
	With agreement marker	Without agreement marker	With agreement marker	Without agreement marker
	11. 8 % (n=70)	88.2 % (n=521)	9.2 % (n=38)	90.8 % (n=371)
Total	100 % (n=1000)			

It is seen from Table 18 that out of the -ken (while) temporal converb constructions; 59.1% have implicit subject converb and 40.9% have explicit subject converb. It is seen that implicit subject converb occurs significantly more frequently than explicit subject converb in -ken (while) constructions ($\chi2$ (1, N = 1000) = 33.12, p < .001).

In implicit subject converb constructions, 11.8 % have agreement morphology on themselves and 88.2 % do not have agreement morphology. In explicit subject converb constructions, 9.2 % have agreement morphology on themselves and 90.8 % do not have agreement morphology. Related examples are given from (11) to (14) below. All the examples were drawn from the TNC.

(11)	Tabii	[bu kadar	şık	ve	modern	bir	bina-ya
	of course	[such	stylish	and	modern	a	building-DAT

gel-ir-ler-ken] kıyafet meselesi de
come-PRS-agr-CVB] clothing problem also
önem taşıma-ya başla-mış-tı.
be.crucial-DAT start-PRF-PST (W-QE37C4A-0402-434)[3]
'Of course, [when coming to such a stylish and modern building], the issue of clothing also became important.'

(12) [Bura-ya gel-ir-ken] Levent-in şimdi konu ettiği
[here-DAT come-PRS-CVB] Levent-GEN now state-REL
metn-in-den bir bölüm-ü bura-ya getir-di-m.
text-GEN-ABL one part-ACC here-DAT bring-PST-1SG
(W-UG03A3A-2712-652)
'[While I was coming here] I brought here a part from Levent's text that he is talking about now.'

(13) [Onlar İngiltere-den gel-ir-ler-ken],
[they England-ABL come-PRS-AGR-CVB]
eşya-lar-ı da Singapur-dan gel-iyor.
belonging-PL-ACC also Singapore-ABL come-PROG
(W-YI45F1C-5071-297)
'[While they are coming from England], their belongings are also coming from Singapore.'

(14) [Ben yürü-r-ken] göz-ler-i-ni ban-a
[I walk-PRS-CVB] eye-PL-GEN-ACC me-DAT
dik-miş-ti.
stare-PRF-PST
'[While I was coming], he was staring at me.'

In example (11), the subject of the subordinate clause is implicit and different from the main clause and there is an agreement marker "*-ler*" which marks that the subject of the subordinate clause is third person plural. In example (12), the subject of the subordinate clause is implicit and same with the main clause. There is no agreement morphology on the subordinate clause. In example (13), the subject of the subordinate clause is explicit and different from the main clause and there is an agreement marker "*-ler*" which marks that the subject of the subordinate clause is third person plural. In example

[3] The tags within the parentheses were provided by the Turkish National Corpus (TNC) (Aksan et al., 2012). "S" stands for spoken data and "W" stands for written data.

(14), the subject of the subordinate clause is explicit and different from the main clause and there is no agreement morphology on the subordinate clause. Moreover, Göksel and Kerslake (2005) states that "where an adverbial clause contains no overt subject there is potential for ambiguity" (p. 470). In example (12), the subordinate clause could also mean 'While I / you / he / she / we was / were coming here ...'. This potential ambiguity is prevented in example (11) by the agreement marker "*-ler*".

Table 19 below shows the subject control structure in -DIğIndA (when) converb construction.

Table 19 Control Structure in -DIğIndA (When) Converb Construction

	Implicit Subject Converb 71% (n=710)		**Explicit Subject Converb** 29% (n=290)	
	With agreement marker	Without agreement marker	With agreement marker	Without agreement marker
	100 % (n=710)	0 % (n=0)	100 % (n=290)	0 % (n=0)
Total	100 % (n=1000)			

It is seen from Table 19 that out of the -DIğIndA (when) temporal converb constructions; 71% have implicit subject converb and 29% have explicit subject converb. It is seen that implicit subject converb occurs significantly more frequently than explicit subject converb in -DIğIndA (when) constructions ($\chi 2$ (1, N = 1000) = 176.40, p < .001.)

Both in implicit subject converb and explicit subject converb constructions, all data have agreement morphology on themselves. Related examples are given from (15) to (17) below.

(15) [Görev-e geldiği-m-de] sıcağı sıcağına,
[position-DAT take.up-AGR-CVB immediately
olağanüstü özlü biçim-de dile getir-diği düşünce-ler-i
very.concise way-ABL express-REL idea-PL-ACC
cümle âlem-i derin-den etkile-miş-ti.
whole.world-ACC profound-ABL effect-PRF-PST.
(W-MA16B3A-0457-1103)
'[When I took up the position] the ideas he/she expressed immediately in a very concise way had a profound effect on the whole world.'

(16) [Telefon-un başına geldiği-n-de] el-ler-i

[phone-GEN get.on-AGR-CVB] hand-PL-GEN
titri-yor-du.
shake-PROG-PST (W-TA16B0A-0093-1640)
'[When he/she got to the phone] his/her hands were shaking'

(17) [Öksüz biz-e ilk geldiği-n-de] Şükran Hemşire
[Öksüz us-DAT first come-AGR-CVB] Şükran sister
ilgilen-miş-ti.
take.care-PERF-PST (W-NA16B4A-0040-2339)
'[When Öksüz first came to us] Sister Şükran took care of him.'

In example (15), the subject of the subordinate clause is implicit and different from the main clause and there is an agreement marker "*-m*" which marks that the subject of the subordinate clause is first person singular. In example (16), the subject of the subordinate clause is implicit and same with the main clause. Because the subject of the subordinate clause is implicit, it is not clear whether the subject of the subordinate clause is second person singular or third person singular. This ambiguity is prevented in example (15) by the agreement marker "*-m*". In example (17), the subject of the subordinate clause is explicit and different from the main clause and there is an agreement morphology "*Ø*" which enables the speakers of Turkish determine the subject of the subordinate clause.

Table 20 below shows the subject control structure in -DIğI zaman (when) converb construction.

Table 20 Control Structure in -DIğI Zaman (When) Converb Construction

	Implicit Subject Converb 63.3% (n=633)		**Explicit Subject Converb** 36.7% (n=367)	
	With agreement marker	Without agreement marker	With agreement marker	Without agreement marker
	100 % (n=633)	0 % (n=0)	100 % (n=367)	0 % (n=0)
Total	100 % (n=1000)			

It is seen from Table 20 that out of -DIğI zaman (when) temporal converb constructions; 63.3% have implicit subject converb and 36.7% have explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -DIğI zaman (when) constructions ($\chi2$ (1, N = 1000) = 70.76, p < .001.) Both in implicit subject converb and explicit subject converb constructions, all data have agreement morphology on themselves. Related examples are given from (18) to (20) below.

(18) [Balık-lar-a bak-tığı-mız zaman] asla doy-maz-lar
[fish-PL-DAT look-AGR-CVB] never get.full-NEG-AGR
ya da doy-mak ne-dir bil-mez-ler.
or get.full-INF mean-Q know-NEG-AGR
(S-ADABAo-0368-83)
'[When we look at fish] they never get full, or they don't know what it means to get full.'

(19) Zaten [gel-diğ-i zaman] konuş-maz-dı.
anyway [come-AGR-CVB] talk-NEG-COP-AGR
'She wouldn't talk anyway [when she came]'
(W-LH09C2A-0276-542)

(20) [Yayılma durduğ-u zaman] ateş düş-er.
[spread stop-AGR-CVB] tension subside-PRS
'[When the spread stops] the tension subsides.'
(W-RC01A2A-1353-786)

In example (18), the subject of the subordinate clause is implicit and different from the main clause and there is an agreement marker "*-mız*" which marks that the subject of the subordinate clause is first person plural. In example (19), the subject of the subordinate clause is implicit and same with the main clause. Although the subject of the subordinate clause is implicit, it is clear that the subject is third person singular because of the agreement morphology "*Ø*" and there is no potential ambiguity. In example (20), the subject of the subordinate clause is explicit and different from the main clause and there is an agreement morphology "*Ø*".

Table 21 shows the subject control structure in -DIğIndAn beri (since) converb construction.

Table 21 Control Structure in -DIğIndAn beri (Since) Converb Construction

	Implicit Subject Converb 69.6% (n=696)		**Explicit Subject Converb** 30.4% (n=304)	
	With agreement marker	Without agreement marker	With agreement marker	Without agreement marker
	100 % (n=696)	0 % (n=0)	100 % (n=304)	0 % (n=0)
Total	100 % (n=1000)			

It is seen from Table 21 that out of -DIğIndAn beri (since) temporal converb constructions; 69.6% have implicit subject converb and 30.4% have explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -DIğIndAn beri (since) constructions ($\chi 2$ (1, N = 1000) = 153.66, p < .001. Both in implicit subject converb and explicit subject converb constructions, all data have agreement morphology on themselves. Related examples are given in (21) and (22) below.

(21) [Bura-ya geldiği-n-den beri] Kurfüstendamm aşağı
[here-DAT come-AGR-CVB] Kurfürstendamm down
Kurfürstendamm yukarı, tek başına dolaşmak-tan
Kurfürstendamm up by.herself wandering-ABL
sıkıl-dı.
get.bored (W-HA16B4A-1363-290)
'[Since he/she came here] he/she has got bored of wandering around Kurfürstendamm down and Kurfürstendamm up all by herself.'

(22) [Sen geldiği-n-den beri] bit-e-mi-yor-um.
[you come-AGR-CVB] finish-COP-NEG-PROG-1SG
'[Since you came] I can't finish.' (S-BEABXW-0069-4)

In example (21), the subject of the subordinate clause is implicit and same with the main clause. The agreement morphology on the subordinate clause "-*n*" could also mean "Since you / he / she...", thus there is potential for ambiguity. In example (22), the subject of the subordinate clause is explicit and different from the main clause and there is an agreement morphology "-*n*". The presence of overt subject in example (22) prevents the potential for ambiguity.

Table 22 shows the subject control structure in -(y)IncA (when) converb construction.

Table 22 Control Structure in -(y)Inca (When) Converb Construction

	Implicit Subject Converb	**Explicit Subject Converb**
	59.7% (n=597)	40.3% (n=403)
Total	100 % (n=1000)	

Table 22 shows that among the -(y)IncA (when) temporal converb constructions, 59.7% have an implicit subject converb, while 40.3% have an explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -(y)IncA (when) constructions ($\chi2$ (1, N = 1000) = 37.64, p < .001). Both in implicit subject converb and explicit subject converb constructions, no data have agreement morphology on themselves. Related examples are given in (23) and (24) below.

(23) [Ev-e gelince] yüzüğ-ü parmağı-m-dan
[home-dat come-CVB] ring-ACC finger-GEN-DAT
çıkar-dı-m.
take-PST-1SG (W-GA16B1A-0643-383)
'[When I got home] I took the ring off my finger.'

(24) [Un hafif-çe sarar-ınca] süt-ü ve
[flour slight-ADV turn.yellow-CVB] milk-ACC and
suy-u karıştır-arak ilave ed-in.
water-ACC mix-ADV add-IMP (W-SI41C1A-1454-1287)
'[When the flour turns slightly yellow] add the milk and water by mixing.

In example (23), the subject of the subordinate clause is implicit and same with the main clause and there is no agreement marker in the subordinate clause. The subject of the main clause could also mean "When I / you / he / she / we / they got home…", thus there is potential for ambiguity. In example (24), the subject of the subordinate clause is explicit and different from the main clause and there is no agreement morphology in the subordinate clause.

Table 23 shows the subject control structure in -(A/I) r…-mAz (as soon as) converb construction.

Table 23 Control Structure in -(A/I) R…-maz (as soon as) Converb Construction

	Implicit Subject Converb 74% (n=740)	**Explicit Subject Converb** 26% (n=260)
Total	100 % (n=1000)	

It is seen from Table 23 that out of -(A/I) r…-mAz (as soon as) temporal converb constructions; 74% have implicit subject converb and 26% have explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -(A/I) r…-mAz (as soon as) constructions ($\chi2$ (1, N = 1000) = 230.40, p < .001). Both in implicit subject converb and explicit subject converb constructions, no data have agreement morphology on themselves. Related examples are given in (25) and (25) below.

(25) [Gider gitmez] yat-acağ-ım.
[leave-CVB] go.to.bed.-FUT-1SG
(W-JA16B4A-0854-57)
'[As soon as I leave] I will go to bed.'

(26) [O gid-er gitmez] Enise Hanım Aylin-in
[he/she go-CVB] Mis. Enise Aylin-GEN
merakı-nı gider-di.
cruosity-ACC satisfy-PST (W-PA16B2A-0748-80)
'(As soon as he/she left) Ms. Enise satisfied Aylin's curiosity.'

In example (25), the subject of the subordinate clause is implicit and same with the main clause and there is no agreement marker in the subordinate clause. The subject of the main clause could also mean "As soon as I / you / he / she / we / they leave…", thus there is potential for ambiguity. In example (26), the subject of the subordinate clause is explicit and different from the main clause and there is no agreement morphology in the subordinate clause.

Table 24 shows the subject control structure in -mAdAn önce (before) converb construction.

Table 24 Control Structure in -Madan Önce (Before) Converb Construction

	Implicit Subject Converb 58.9% (n=589)	**Explicit Subject Converb** 41.1% (n=411)
Total	100 % (n=1000)	

Table 24 shows that among the -mAdAn önce (before) temporal converb constructions, 58.9% have an implicit subject converb, while 41.1% have an explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -mAdAn önce (before) constructions ($\chi 2$ (1, N = 1000) = 31.68, $p < .001$). In both implicit and explicit subject converb constructions, no data displays agreement morphology. Relevant examples are provided in (27) and (28) below.

(27) [Bura-ya gel-meden önce] komutan

[here-ACC come-CVB] commander

karn-ın-ı doyur-muş-tu.

stomach-GEN-ACC fill-prf-pst (W-EA16B2A-1563-828)

‘[Before coming here] the commander had had her fill.’

(28) [Biz bura-ya gel-meden önce] onlar durum-u

[we here-DAT come-CVB] they situation-ACC

bil-mi-yor-du.

know-NEG-PROG-PST (W-ME39C3A-0581-182)

‘They didn’t know the situation [before we came here].’

In example (27), the subject of the subordinate clause is implicit and same with the main clause and there is no agreement marker in the subordinate clause. The subject of the main clause could also mean “Before I / you / he / she / we / they come here…”, thus there is potential for ambiguity. In example (28), the subject of the subordinate clause is explicitly stated and differs from the subject of the main clause, with no agreement morphology present in the subordinate clause.

Table 25 below shows the subject control structure in -DIktAn sonra (after) converb construction.

Table 25 Control Structure in -Dıktan Sonra (After) Converb Construction

	Implicit Subject Converb 66% (n=660)	**Explicit Subject Converb** 34% (n=340)
Total	100 % (n=1000)	

It is seen from Table 25 that out of -DIktAn sonra (after) temporal converb constructions; 66% have implicit subject converb and 34% have explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -DIktAn sonra (after) constructions ($\chi 2$ (1, N = 1000) = 102.40, p < .001). Both in implicit subject converb and explicit subject converb constructions, no data have agreement morphology on themselves. Related examples are given in (29) and (30) below.

(29) Dur-mu-yor-um [uyan-dıktan sonra].

stop-NEG-PROG-1SG [wake.up-CVB]

‘I don't stop [after I wake up].’ (S-AEABTZ-0337-1530)

(30) [Baba-m ev-den ayrıl-dıktan sonra da] öyle
[father-GEN home-ABL leave-CVB even] that
yap-tı-m.
do-PST-1SG (W-RA16B1A-1214-1423)
'[Even after my father left home] I did that.'

In example (29), the subject of the subordinate clause is implicit and is the same as in the main clause, with no agreement marker present in the subordinate clause. The subject of the main clause could also mean "…after I / you / he / she / we / they wake up" and there is potential for ambiguity. In example (30), the subject of the subordinate clause is explicitly stated and differs from the subject of the main clause, with no agreement morphology present in the subordinate clause.

Table 26 shows the subject control structure in -DıkçA (whenever) converb construction.

Table 26 Control Structure in -Dıkça (Whenever) Converb Construction

	Implicit Subject Converb 57.5% (n=575)	**Explicit Subject Converb** 42.5% (n=425)
Total	100 % (n=1000)	

Table 26 shows that among the -DıkçA (whenever) temporal converb constructions, 57.5% have an implicit subject converb, while 42.5% have an explicit subject converb. Implicit subject converb occurs significantly more frequently than explicit subject converb in -DıkçA (whenever) constructions ($\chi 2$ (1, N = 1000) = 22.50, p < .00). In both implicit and explicit subject converb constructions, no data displays agreement morphology. Relevant examples are provided in (31) and (32) below.

(31) [Kafayı bul-dukça] ben-i mutlu ed-en
[get.high-CVB] me-ACC happy make-REL
hayal-ler-imin peşine takıl-ıyor-du-m.
dream-PL-GEN chase-PROG-PST-1SG
(W-LH09C2A-0276-1504)
'[Whenever I got high] I was chasing my dreams that made me happy.'

(32) [Yol uza-dıkça] azık kıymetlen-ir.

[journey get.longer-CVB] provision become-valuable-AOR
'[Whenever the journey gets longer] provisions become more valuable.'
(W-UI22C4A-0820-1568)

In example (31), the subject of the subordinate clause is implicit and same with the main clause and there is no agreement marker in the subordinate clause. The subject of the main clause could also mean "Whenever I /you / he / she / we / they got high…", thus there is potential for ambiguity. In example 32, the subject of the subordinate clause is explicit and different from the main clause and there is no agreement morphology in the subordinate clause.

5. CONCLUSION

As mentioned above, the aim of this study is to analyse the subject control structure in temporal converb clause constructions in Turkish, using the Turkish National Corpus (TNC) as a reference. Based on the findings of the study, the research question is answered as follows.

RQ: How are the subject control strategies realized in temporal converb constructions in Turkish?

First, when the random sample extracted from the corpus is analysed, it is seen that implicit subject converbs are used more than explicit subject converbs in Turkish (In -ken (while) converb construction, 59.1 % have implicit subject converb and 40.9 % have explicit subject converb; in -DIğIndA (when) temporal converb constructions; 71% have implicit subject converb and 29% have explicit subject converb; in -DIğI zaman (when) temporal converb constructions; 63.3 % have implicit subject converb and 36.7% have explicit subject converb; in -DIğIndAn beri (since) temporal converb constructions; 69.6% have implicit subject converb and 30.4% have explicit subject converb; in -(y)IncA (when) temporal converb constructions, 59.7% have an implicit subject converb, while 40.3% have an explicit subject converb; in -(A/I) r…-mAz (as soon as) temporal converb constructions; 74% have implicit subject converb and 26% have explicit subject converb; in -mAdAn önce (before) temporal converb constructions, 58.9% have an implicit subject converb, while 41.1% have an explicit subject converb; in -DIktAn sonra (after) temporal converb constructions; 66% have implicit subject converb and 34% have explicit subject converb and in -DıkçA (whenever) temporal converb constructions, 57.5% have an implicit subject converb, while 42.5% have an explicit subject converb).

Second, the corpus data show that when the subject of the subordinate clause is implicit, it is not clear whether it is co-referential with the main clause or not as it was suggested by Haspelmath (1995) and Çetintaş Yıldırım (2004). Haspelmath (1995) states that in implicit - subject converbs, subjects may not be expressed explicitly. Çetintaş Yıldırım (2004) states that when the converb clause does not have an overt subject, its subject reference is interpreted as being co-indexed with the subject of the main clause. However, the corpus data show that when a converb clause contains no overt subject, there is potential for ambiguity. In order to prevent the ambiguity, either an explicit subject should appear in the subordinate clause or in some cases as in -ken (while) temporal converb constructions, the ambiguity is prevented by the agreement marker. In the light of this findings, it can be stated that when the converb clause does not have an overt subject, its subject reference is not always interpreted as being co-indexed with the subject of the main clause in Turkish.

Third, the subject agreement markers may enable the speakers of Turkish determine the subject of the converb clause in implicit converbs, but it is not possible in every temporal converb construction. The temporal converb constructions, namely -(y)IncA (when), -(A/I) r…-mAz (as soon as), -mAdAn önce (before), -DIktAn sonra (after) and -DıkçA (whenever) do not have explicit subject agreement marking inherently. The temporal converb construction, -ken (while) may have subject agreement marking before itself, but it can only have the third person plural, and this marking helps the speakers of Turkish understand the referentiality between the subject of the main clause and subordinate clause. The temporal converb constructions, namely -DIğIndA (when), -DIğI zaman (when) and -DIğIndAn beri (since) have subject agreement markers on themselves. In -DIğIndA (when) temporal converb constructions; the agreement marker of first person singular, first person plural and second person plural help speaker of Turkish determine the subject of the subordinate clause and prevent potential ambiguity. In -DIğI zaman (when) temporal converb constructions, the agreement marker of first person singular, second person singular, first person plural and second person plural help speaker of Turkish determine the subject of the subordinate clause and prevent potential ambiguity. In -DIğIndAn beri (since) temporal converb constructions, the agreement marker of first person singular, first person plural and second person plural help speaker of Turkish determine the subject of the subordinate clause and prevent potential ambiguity.

In conclusion, the functional motivation of the subject control in temporal converb clauses in Turkish is that when the subject of the converb clause is implicit, it may be the same subject with the main clause, or it may be different subject from the main clause. In some cases, subject agreement markers help the speakers of Turkish understand the referent of the converb clause subject in implicit converb constructions. Another finding is that the converbs which have agreement morphology on themselves may have explicit subject. In the light of this information, it is safe to state that Turkish has the varying subject control, in which the subjects of the subordinate clause and main clause may be coreferential or non-coreferential in the classification of Nedjalkov (1995).

5.1 Limitations

In Turkish, the dative or accusative experiencers, as well as genitive constructions in converb clauses may indicate the referent of the subject in the converb clause. These constructions were not analysed in this study. Moreover, the subject of the converb clause may be determined from the context. Thus, future research is suggested to incorporate these parameters to be analysed in terms of control structure in temporal converb constructions in Turkish.

REFERENCES

Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yılmazer, 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 (ELRA).

Alpatov, V., & Podlesskaya, V. (1995). Converbs in Japanese. In M. Haspelmath & E. König (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms - adverbial participles, gerunds* (pp. 465–486). De Gruyter Mouton.

Banguoğlu, T. (1995). *Türkçenin grameri*. Türk Dil Kurumu.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1980). On binding. *Linguistic Inquiry, 11*(1), 1–47.

Çetintaş Yıldırım, F. (2004). *Türkçe zaman ulaçlarının sözdizim ve anlambilim çözümlemesi* (Master's thesis). Mersin University.

Erguvanlı Taylan, E. (1996). Aspects of control in Turkish. In B. Rona (Ed.), *Current issues in Turkish linguistics: Proceedings of the 5th ICTL, London, 15–17 August 1990* (pp. 46–60). Ankara.

Groot, C. (1995). The Hungarian converb or verbal adverbial in -va/-ve. In M. Haspelmath & E. König (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – adverbial participles, gerunds* (pp. 283–312). De Gruyter Mouton.

Haspelmath, M. (1995a). The converb as a cross-linguistically valid category. In *Converbs in cross-linguistic perspective* (pp. 1–56). De Gruyter Mouton. https://doi.org/10.1515/9783110884463-003

Haspelmath, M. (1995b). Contextual and specialized converbs in Lezgian. In M. Haspelmath & E. König (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms - adverbial participles, gerunds* (pp. 415–440). De Gruyter Mouton.

Johanson, L. (1995). On Turkic converb clauses. In *Converbs in cross-linguistic perspective* (pp. 313–348). De Gruyter Mouton.

McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Nedjalkov, V. (1990). Some typological parameters of converbs. In *Converbs in cross-linguistic perspective* (pp. 97–136). De Gruyter Mouton. https://doi.org/10.1515/9783110884463-005

Nedjalkov, V. (1995). Converbs in Evenki. In M. Haspelmath & E. König (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms - adverbial participles, gerunds* (pp. 441–464). De Gruyter Mouton.

Postal, P. (1974). *On raising: One rule of English and its theoretical implications*. MIT Press.

Yasavul, Ş. M. (2009). *Aspects of control and complementation in Turkish* (Master's thesis). Middle East Technical University.

Yücel, Ö. (2009). Control relations in Turkish infinitival complements. In *Essays on Turkish linguistics: Proceedings of the 14th International Conference on Turkish Linguistics, August 6–8, 2008*.

Weiss, D. (1995). Russian converbs: A typological outline. In M. Haspelmath & E. König (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms - adverbial participles, gerunds* (pp. 239–282). De Gruyter Mouton.

Chapter 7:
Good-Enough Parsing in Turkish: Task Effects in Online Processing of Role-Reversals

Onur Keleş & Nazik Dinçtopal Deniz

Abstract

The good-enough processing framework proposes that language comprehension can occasionally involve rapid, plausibility-driven interpretations instead of detailed syntactic analyses. We investigated this phenomenon in Turkish, a morphologically rich, case-marking language, via two self-paced reading experiments which employed either a Semantic Plausibility Judgment (PJ) task or an Agent-Patient Detection (APD) task. Our experimental sentences involved active or passive constructions where the verb had either two animate arguments (e.g., the dog bit the man) or one animate one inanimate argument (e.g., the chef wore the apron). The results of the PJ task showed that the participants' accuracy decreased and their decision times increased when real-world plausibility information conflicted with morphosyntactic information, particularly when the arguments of the verb were both animate. In the APD task, accuracy was near-ceiling across all conditions but processing times increased especially in conditions where argument roles were reversed. The results of the two experiments show that the cognitive demands that a task requires can affect if the parse is detailed or "good enough". The results further show that employment of good-enough heuristics is also affected by how lexical and morphological information is weighted as well as similarity-based cue interference.

Keywords: Good Enough processing, Turkish, task effects, sentence processing

1. INTRODUCTION

Understanding language in real time involves cognitive shortcuts as much as it involves detailed analyses. Unlike the traditional models that assume fully specified syntactic and semantic parses, the *good-enough processing* framework proposes that comprehenders occasionally rely on shallow interpretations informed by real-world knowledge (Ferreira et al., 2002; Ferreira & Patson, 2007). This approach, like several others, highlights efficiency in language processing, where interpretations are constructed rapidly but may not always align with the sentence's full structural detail.

Good-enough processing is mostly observed in contexts that involve *role-reversals*, where listeners or readers rely on real-world knowledge in assigning thematic roles, ignoring the conflicting syntactic cues. This observation is more common for constructions with more complex syntax such as passives as in *The dog was bitten by the man* which can be misinterpreted as *The dog bit the man* (Ferreira, 2003), as opposed to actives such as *The man bit the dog* for which the misfit for thematic roles against the real-world plausibility is easily detected. Electrophysiological studies further suggest that when they lead to *good-enough* parsing, such thematic misfits do not consistently trigger early signals showing detection of semantic anomaly (e.g., N400), but elicit later components (P600) indicating syntactic reanalysis (Kim & Osterhout, 2005; Kuperberg et al., 2003; Chow et al., 2016).

Good-enough parsing, however, has been reported to be modulated by task demands. It was observed that when tasks require only general comprehension or plausibility judgments, comprehenders tend to employ heuristics -based interpretations more often. But if the task requires attention to syntactic detail, such as identifying thematic roles, language users usually engage in detailed syntactic processing (Swets et al., 2008; Ehrenhofer et al., 2018). Nakamura et al. (2024) report that task and timing effects in argument role sensitivity emerges through a monitoring mechanism that serially inhibits role-inappropriate candidates. That is, argument role sensitivity appears when sufficient processing time is available and when production tasks demand assigning proper thematic roles.

We examine structural complexity and task demand effects in *good-enough* parsing in Turkish, a language, to our knowledge, underexplored in this context. Turkish employs morphology very productively; it also has flexible word order and overt case system, which provide clear and early morphosyntactic cues to thematic structure. We test, specifically, if these cues prevent good-enough parsing in varying task demands.

The present study builds on our prior work which used a Semantic Plausibility Judgment (PJ) task ((Keleş & Deniz, 2025). In our previous study Turkish native speakers read sentences and judged if the sentence described a plausible or implausible event. More specifically, the participants (*N* = 26) read sentences in which voice (active versus passive) and argument order (canonical versus reversed) were systematically manipulated (see below, Section 2.1.2 for example experimental sentences). The results showed that when both arguments were animate (e.g., *the men* and *the dogs*) but one was more strongly associated with the agent role (e.g., *the men*), reversing the arguments led to decrease in accuracy and an increase in decision times. In contrast, in the irreversible set, where one argument was inanimate (e.g., *aprons* vs. *chefs*), accuracy remained high and reversed sentences elicited slightly faster plausibility judgments, suggesting that clear animacy contrasts facilitated thematic role assignment and reduced the probability of good-enough parsing. We took those results to indicate that Turkish speakers may rely on plausibility-based strategies but only in cases when there was a match in animacy cue for the two NPs that can potentially serve as the agent of the event.

In the present study, we increase the participant size for the PJ task to ensure statistical generalizability and further examine sentence processing routines in role-reversal contexts with a cognitively more engaging Agent-Patient Detection (APD) task. In the APD task, participants were asked to explicitly identify the agent or the patient of each sentence. This would allow us to examine if a task that requires identification of the thematic roles would lead to syntactically detailed parses and prevent good-enough processing. Our results showed a clear task effect: in the PJ task, accuracy in plausibility judgments were reduced and decision times were longer when plausibility conflicted with morphosyntactic cues, especially in sentences whose arguments were both animate and could exchange agent/patient roles, indicating some reliance on a “good-enough” heuristic. In the APD task, accuracy in agent/patient detection was high showing that the participants could identify role-reversal errors. Unsurprisingly, this was associated with longer reading and decision times, revealing that comprehenders engaged in slower, morphology-driven parsing when the task required to do so. We take our results to confirm that good-enough processing is possible in Turkish, despite the early morpho-syntactic cues to the agent/patient of the event but it is modulated by task demands (Swets et al., 2008; Ehrenhofer et al., 2018), similarity-based retrieval interference (Lewis & Vasishth, 2005) and cue-informativeness.

2. GOOD-ENOUGH SENTENCE COMPREHENSION

Early models of sentence processing, such as Kimball's seven (1973) parsing principles, maintain that syntactic parsing operates through encapsulated, domain-specific mechanisms that proceed in a largely bottom-up and linear fashion. The Garden Path Model and the "sausage machine" model of Frazier and Fodor (1978) further refined this view by introducing a two-stage parser that first constructs an initial phrase structure, mainly influenced by syntactic constraints, and only revises it in light of incoming input that conflicts with the structure built thus far. In the Garden Path model, non-syntactic information is used at the second stage, where a thematic processor is employed. Subsequent theories challenged these modular assumptions. Constraint-based lexicalist models (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994) argued that syntactic parsing is not an isolated process but one in which multiple sources of information, syntactic, semantic, lexical, and pragmatic, interact from the earliest stages of processing. Later, the cue-based memory retrieval model (e.g., Lewis & Vasishth, 2005) emphasized the role of working memory, proposing that sentence comprehension relies on retrieval operations that access linguistic representations through noisy cue-matching mechanisms. The good-enough processing approach (Ferreira et al., 2002; Ferreira & Patson, 2007) differs from these perspectives in that it asserts that the language comprehension system can sometimes prioritize efficiency over accuracy. Rather than building fully specified, hierarchically rich representations in all cases, the parser can occasionally settle for "good enough" interpretations that are shallow in their syntactic detail but sufficient for the communicative demands at hand.

Classic examples of *good-enough* processing include semantic illusions such as the *Moses illusion* in which when addressed the question, *How many animals of each kind did Moses take on the Ark?,* participants often answer "two" despite the inconsistency of the sentence-level information against that in the Bible (*Moses* instead of *Noah*), indicating prioritization of coherence and familiarity over strict accuracy. In sentence processing contexts, in addition to the example given above for the processing of thematic information in passive/active constructions, there are good-enough parsing effects observed for garden-path sentences like *While Anna bathed the baby played in the crib* (Christianson et al., 2001). For such sentences although comprehenders show reanalysis (i.e., garden path) effects at the disambiguating region, i.e., at the verb *played,* they frequently retain the initial, incorrect analysis where the NP, *the baby,* is interpreted as the object of the embedded verb *bathed*. These findings suggest that initial interpretations based on surface heuristics, such as linear order or real-world plausibility, can persist despite the grammatical cues to the contrary.

The central heuristics underlying many of these misinterpretations is the NVN strategy (Bever, 1970), where a Noun-Verb-Noun sequence is interpreted such that the first noun is the agent and the second is the patient. This strategy reflects the canonical mapping of thematic roles in English and many other languages. When a sentence aligns with this default order, comprehension is fast and accurate, but when it deviates, as in passive constructions, readers are more prone to errors.

Ferreira (2003) examined good-enough processing in role reversal contexts across three experiments in which she manipulated the voice (active vs. passive) and real-world plausibility of thematic role assignment (schema-consistent vs. schema-inconsistent) and probed the participants' ability to accurately identify agents and patients in English. In her Experiment 1, the participants listened to plausible/implausible active/passive sentences and were asked to name out loud either the agent or the patient. Their decision accuracy and RTs were recorded. Overall, for all sentence types, passives were found to be more difficult to understand than the actives. Furthermore, in implausible and passive sentences in the biased set, where animacy was not a strong cue, like *The dog was bitten by the man*, there was a 74% accuracy in agent and 85% accuracy in patient detection indicating that the participants reversed the roles, treating *the dog* as the agent, which is a more plausible but a syntactically unsupported interpretation. The accuracy was higher in the passive and implausible condition in the irreversible set, e.g., *The mouse was eaten by the cheese*, where animacy could be used as a cue. The decision time data showed a parallel pattern. Experiments 2 and 3 further explored whether processing difficulty stems from syntactic complexity or from violations to the canonical agent-first thematic role assignment pattern. Experiment 2 compared the processing of subject-clefts (which maintain agent-first order despite syntactic complexity) with passives and Experiment 3 compared the processing of object-clefts (which violate agent-first order) with subject-clefts. The results showed that thematic role order, rather than surface frequency or syntactic complexity, affected processing ease. Ferreira and colleagues concluded that any complete theory explaining how people understand language needs to recognize that shortcuts are employed alongside, and sometimes in place of, deep syntactic algorithms during comprehension.

Good-enough processing effects have been reported for several languages for different syntactic structures and with different behavioral methodologies since then (see Swets et al., 2008; Von Der Malsburg & Vasishth, 2013; Nakamura & Arai, 2016; Chromy, 2022; Paape et al., 2024; Kharkwal & Stromswold, 2014; Lassotta et al., 2016). However, there has also been work has shown that good-enough parsing effects are not always robust or can be due to factors other than shallow or underspecified parsing. For example, it has been argued that errors attributed to good-enough processing may be related to task demands (Swets et al., 2008) or memory-related retrieval difficulties (Qian et al., 2018), rather than genuine incomplete sentence representations (Logačev & Vasishth, 2016; Bader & Meng, 2018). Studies have shown that when comprehension tasks are shallow, e.g., requiring plausibility ratings, readers are more likely to underspecify sentence structure and prefer fast, heuristic interpretations. Swets et al. (2008) and Tan and Foltz (2020) found that participants exhibited faster reading times for ambiguous versus unambiguous sentences when only superficial comprehension was expected, a phenomenon known as the ambiguity advantage. However, this advantage disappeared when tasks required deeper processing, such as identifying the agent or patient. Similarly, Gilbert et al. (2021) showed that ambiguity resolution is postponed unless the task explicitly demanded disambiguation, reinforcing the view that sentence processing is adaptive and goal-sensitive, rather than exhaustive by default (cf., Logačev, Aydın, & Tuncer (2022), for the generalizability of underspecification to Turkish).

Similarly, Bornkessel-Schlesewsky et al. (2011) report in a series of ERP experiments that while English and Dutch exhibit the classic P600-only pattern for thematic role reversal anomalies, languages with rich case marking or flexible word order, such as Turkish, German, and Chinese, show an N400 effect, indicating earlier sensitivity to argument role misassignment. Icelandic, however, revealed both patterns depending on verb type: N400s were observed when case marking determined argument roles, and P600s were elicited when roles were positionally assigned. These findings support the argument that the neurocognitive mechanisms underlying good-enough processing and the detection of thematic role anomalies are affected by language-specific properties.

Turkish is an agglutinative SOV language with overt case marking, free word order, and rich verbal morphology. It is predicted that these cues can inform the parser on the thematic structure of a sentence. For instance, in a sentence with a transitive verb in active voice, such as *Adam kadın-ı öp-tü* [(man woman-(ACC kiss-PST 'The man kissed the woman')], the definite object, *kadın-ı*, is marked by accusative case marking, *-I*. (The nominative case on the subject is usually morphologically null; but cf. embedded clause constructions in the form of genitive-possessive NPs, Göksel & Kerslake, 2005). Such case marking, if used predicatively for sentence structure (see Özge, Marinis, & Zeyrek, 2016), can help prevent good-enough assignment of thematic roles as it provides cues for the patient of the event before the upcoming verb. Passive constructions in Turkish, unlike English, also involve affixation (with the passive morpheme *-Il*) on the verb without modifications to the word order unless the agent is specifically mentioned through the postposition *tarafından* ("by"), a lexical, rather than a functional marker, as in *Kadın (adam tarafından) öp-ül-dü* (woman (man by) kiss-PASS-PST 'The woman was kissed (by the man)'). This lexical postposition also provides information about thematic relations before the verb.

Since Turkish is a head-final language, comprehending sentences in Turkish presents certain integration challenges. According to Levy's (2008) surprisal framework, processing head-final languages or constructions requires arguments to be held in memory until the clause-final verb, which inherently involves predictive processing for that verb. Expectation-based strategies are considered, by some, to be language-specific and more operative in head-final languages (Vasishth, 2010). Whether or not this is so, under Levy's framework, the ease of integration is inversely related to the verb's surprisal: highly predictable verbs entail lower integration costs, potentially facilitating accurate thematic role assignment. Conversely, higher surprisal, and thus increased processing difficulty, arises with less predictable structures. We predict that the presence or absence of morphological cues in Turkish can inform the parser on the word-order of the sentence and contribute to the predictive processes for the clause-final verb's voice. This may result in preventing good-enough processing during online sentence comprehension.

Cross-linguistic neurocognitive research suggests that Turkish comprehenders rely less on heuristic strategies like the agent-first NVN bias (Bever, 1970) and instead show early sensitivity to argument role mismatches (Bornkessel-Schlesewsky et al., 2011). Passive constructions may also not be as prone to heuristic misinterpretation in Turkish as they are in English (Ferreira, 2003) because morphological (e.g.,morphological case marking on NPs) and lexical information (the postpositional *tarafından*) may lead to accurate thematic role detection even when plausibility and syntax conflict.

2. THE PRESENT STUDY

2.1 Experiment 1: Plausibility Judgment (PJ)

2.1.1 Participants

Fifty-three native speakers of Turkish (41 female) participated in the PJ experiment. All were first-year undergraduate students (M_{age} = 19) at Boğaziçi University, enrolled in an introductory linguistics course, and received course credit in exchange for their participation. Participants were tested online using the PCIbex Farm platform (Zehr & Schwarz, 2018), which allows for browser-based self-paced reading experiments. All participants provided informed consent before the experiment commenced.

2.1.2 Materials

(1) A total of 42 experimental and 21 control sentences were taken from Ferreira (2003) and translated into Turkish with small adaptations to ensure naturality and to prevent introduction of unintended ambiguity. Care was given to preserve the relevant structural and semantic properties. Each sentence was systematically manipulated for syntactic structure (active vs. passive) and word order (canonical vs. reversed). Half of the experimental sentences contained two animate arguments, yielding reversible thematic roles but they were biased towards one argument receiving the agent role and the other receiving the patient role as in (1) and the other half included an animate and an inanimate argument, creating irreversible thematic-role assignments as in (2). This manipulation enabled a controlled investigation of how morphosyntactic marking interacts with plausibility-based interpretation in Turkish. Each

sentence ended with a content-neutral word (e.g., *sanırım* 'I think'[1] as in (1,2) to prevent wrap-up effects.

(2) Biased-but-reversible Set

a. Active, canonical

Köpek-ler adam-ı ısır-dı san-ır-ım.
dog-PL man-ACC bite-PST think-AOR-1SG
'I think the dogs bit the man.'

b. Passive, canonical

Adam-lar köpek tarafından ısır-ıl-dı.
man-PL dog by bite-PASS-PST
san-ır-ım

think-AOR-1SG
'I think the men were bitten by the dog.'

c. Active, reversed

Adam-lar köpeğ-i ısır-dı san-ır-ım..
man-PL dog-ACC bite-PST think-1SG
'I think the men bit the dog.'

d. Passive, reversed

Köpek-ler adam tarafından ısır-ıl-dı san-ır-ım.
dog-PL man by bite-PASS-PST think-1SG
'I think the dogs were bitten by the man.'

(2) Irreversible Set

a. Active, canonical

Şef-ler önlüğ-ü giy-di san-ır-ım.
chef-PL apron-ACC wear-PST think-1SG
'I think the chefs wore the apron.'

b. Passive, canonical

Önlük-ler şef tarafından giy-il-di san-ır-ım.
apron-PL chef by wear-PASS-PST think-1SG
'I think the aprons were worn by the chef.'

c. Active, reversed

[1] Other content-neutral words were words expressing epistemic modality such as *perhaps*, *maybe* or factive adverbs such as *certainly* and *really*.

Önlük-ler şef-i giy-di san-ır-ım.
apron-PL chef-ACC wear-PST think-1SG
'I think the aprons wore the chef.'

d. Passive, reversed

Şef-ler önlük tarafından giy-il-di san-ır-ım.
chef-PL apron by wear-PASS-PST think-1SG
'I think the chefs were worn by the apron

The forty-two experimental sentence sets, 21 from the biased-but-reversible and 21 from the irreversible set, were distributed across four reading lists, counterbalancing for syntactic structure and word order. In each list, there were an additional 21 symmetrical controls, in which argument reversals yielded equally plausible alternatives such as (*The boy kissed the girl*). In each list, there were also six practice sentences at the beginning of the experiment.

We predict that while the rich morphosyntactic and lexical cues in Turkish can facilitate accurate thematic role assignment, good-enough processing effects, reflected as reliance on plausibility over conflicting syntax, may still occur. The Biased-but-Reversible set includes two animate arguments and can, therefore, be more vulnerable to plausibility-driven misinterpretations due to similarity-based cue interference from the two animate NPs (Lewis & Vasishth, 2005) when syntax indicates an implausible event (e.g., 1c,d), resulting in lower accuracy and longer decision times. The Irreversible set (e.g., sentences 2c,d), however, includes one animate and one inanimate argument whose reversals result in implausible events as in (2c,d). The mismatching animacy information for the two NPs can prevent good-enough parsing, yielding high accuracy and faster judgments for implausible role assignments in (2c,d) compared to the reversible and relatively more plausible role assignments in (1c,d). These effects may be more prominent in passive constructions (1d) as their structure is more complex than their active counterparts which is predicted to increase the processing cost (Ferreira, 2003). As such, interference from matching animacy cue may be more prominent (Bakay & Deniz, 2021) and result in employment of good-enough processing strategies more often.

This prediction, however, has the reservation that passive structures in Turkish involve a lexical cue *tarafından* indicating the agent of the event (cf., the function word *by* in English). This lexical agent marker may function as a particularly salient cue to thematic roles and prevent good-enough parsing in

Turkish passive constructions. It can also potentially prevent plausibility-driven misinterpretations compared to active sentences (e.g., 1c) where the only cue to thematic roles in conflicting scenarios before the sentence-final verb is morpho-syntactic case marking, which may not be as prominent as the lexical cue *tarafından*. If this is the case, we would predict higher accuracy in discerning the syntactically correct, albeit implausible, thematic roles for the *tarafından*-marked passives when plausibility and syntax conflict, even as their processing may still reflect underlying structural complexities.

Morphosyntactic cues have been reported to prevent good-enough processing. The morphosyntactic case marking like accusative case (e.g., on *köpeğ-i* in 1c) in the above Turkish sentences can similarly inform on the sentence's word order and the voice for the upcoming verb (Hale, 2006; Levy, 2008) and aid detection of thematic roles properly. However, it is not clear if they will prevent an overall plausibility-based interpretation sentence-finally, specifically for the Biased-but-Reversible implausible conditions (1c, 1d), due to other, perhaps stronger, cues such as animacy which may potentially result in similarity-based interference (Lewis & Vasishth, 2005) and can lead to good-enough heuristics.

Finally, reversed word order is generally expected to increase processing difficulty (longer reading and decision times) across conditions. Crucially, we predict a dissociation based on task demands: in the PJ task the participants are predicted to show more good-enough processing (lower accuracy in conflicting conditions like 1c, 1d). (See below how this prediction changes for the Agent-Patient Detection (APD) task).

2.1.3 Procedure

Experiment 1 (and Experiment 2) was designed in PCIbex Farm (Zehr & Schwarz, 2018) as a word-by-word self-paced reading (SPR) task. Each word appeared sequentially at the center of the screen in 22-point Verdana font. The participants were asked to press the space bar to reveal each new word, while previously displayed words disappeared from view. Each sentence was followed by a comprehension question prompting if the sentence was plausible or implausible in Turkish. They responded to the comprehension question by clicking on one of two noun phrases on the screen, whose position was also counterbalanced. The experiment lasted approximately 25 minutes.

2.1.4 Data Analysis

Raw PCIbex logs were parsed in R with a *read.pcibex* wrapper (Zehr & Schwarz, 2018). Word reading times (RTs) below 150 ms or above 3000 ms, and decision times (DTs) below 150 ms or above 8000 ms were discarded (corresponding approximately to < 1% of data) following standard practices for outlier removal in self-paced reading experiments (Jegerski, 2014). Participants (N = 1) whose mean accuracy fell below 80% were also excluded from the analyses.

The data were examined for response accuracy, word-by-word reading times (RTs), and end-of-sentence decision times (DTs). Generalized linear mixed-effects models (binomial link) were used to analyze the accuracy data, and linear mixed-effects models were used to analyze the RT/DT data. The RT/DT data were log-transformed before the main analyses as their distribution was heavily skewed. Shapiro-Wilk normality tests confirmed that both measures significantly deviated from normal distribution in their raw form (DTs: W = 0.830, p < 0.001; RTs: W = 0.653, p < 0.001), while log-transformation substantially improved normality (DTs: W = 0.997, p < 0.001; RTs: W = 0.960, p < 0.001), justifying the use of log-transformed data in the analyses.

Structure (active, passive), Word Order (canonical, reversed), and Set (biased, irreversible) were entered into the models as fixed effects; subjects and items were random effects. The analyses on RTs also included word length (measured by the number of letters) for the current and preceding word as an additional predictor. The regions of interest for the RT analyses were the verb and the second NP (the accusative-marked object in active sentences and the unmarked agent adjunct in passive sentences).

The data were analyzed in R (R Core Team, 2023) with the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages.

2.1.5 Results

The results are reported for accuracy, sentence-final DTs, and word-level RTs, separately, below.

Accuracy results. Table 27 presents the percent accuracy for the sentence-final plausibility decisions.

Table 27. Percent accuracy of the plausibility judgments in Experiment 1.

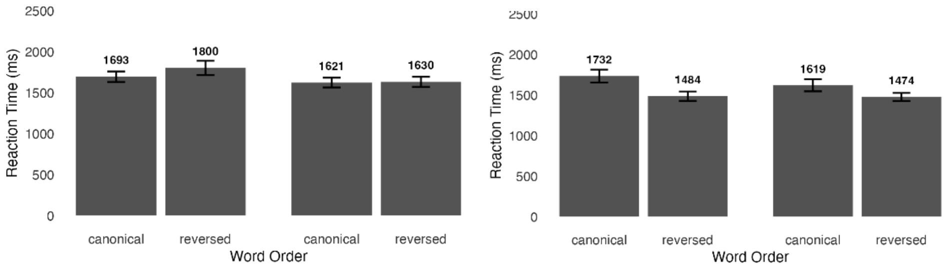
Set	Word Order	Structure	Percent Accuracy
Biased	Canonical	Active	97
		Passive	98
	Reversed	Active	74
		Passive	83
Irreversible	Canonical	Active	96
		Passive	93
	Reversed	Active	97
		Passive	99

The analyses on the accuracy data showed that models that involved an interaction of word order and set (*OR* = 28.80, 95% *CI* [11.18, 69.81], $t = 7.44$, $p < .001$) explained the data better than a simple model with set or only word order as predictors, $\chi^2 = 63.8$, $p < .001$. As such, we analyzed the data in each set separately.

For the biased set, accuracy for canonical sentences were high, almost at ceiling level ($M = .98$, $SE = .01$), but was reduced in the reversed condition ($M = .79$, $SE = .023$), $OR = 0.05$, 95% *CI* [0.03, 0.10], $t = -8.59$, $p < .001$. There was also a main effect of Structure where passives were more likely to be judged correctly than actives; $OR = 2.07$, 95% *CI* [1.28, 3.33], $t = 2.99$, $p = .003$.

For the irreversible set, the direction of the word-order effect was reversed. Accuracy for canonical sentences was high ($M = .95$, $SE = .019$), but was further increased in the reversed condition ($M = .98$, $SE = .014$), $OR = 1.27$, 95% *CI* [0.48, 3.33], $t = 0.48$, $p = .628$. There was an effect of Structure where actives were more likely to be judged correctly than passives, but this difference did not reach the conventional statistical significance levels. $OR = 0.48$, 95% *CI* [0.21, 1.08], $t = -1.77$, $p = .076$. There was a Word-Order × Structure interaction $OR = 4.31$, 95% *CI* [0.99, 18.78], $t = 1.95$, $p = .052$.

Decision time results. Figure 13 presents the DT results for the PJ task.



(a) Biased Set (b) Irreversible Set

Figure 13. Sentence-final DTs in the PJ Task by Set, Structure, and Word Order.

The analyses on DTs for sentence plausibility also showed that the models which included an interaction of Set and Word Order fit to the data better than the simpler models with Set and Word Order as main effects, χ^2 = 25.84, *df* = 1, *p* < .001. Hence, as in accuracy data, the DT data were examined separately for each set.

In the biased set, there was a main effect of Word Order where reversed sentences were processed more slowly than canonical sentences (*β* = 0.12, 95% *CI* [0.06, 0.18], *t* = 3.76, *p* < .001), corresponding to an increase from approximately 1,480 ms to 1,660 ms. There was also a main effect of Structure where passives were processed faster than actives (*β* = -0.07, 95% *CI* [-0.13, -0.01], *t* = -2.16, *p* = .031).

In the irreversible set, the opposite was observed. There was a main effect of Word Order where reversed sentences were processed faster than canonical sentences (*β* = -0.09, 95% *CI* [-0.15, -0.04], *t* = -3.51, *p* = .001), corresponding to a decrease from approximately 1,440 ms to 1,310 ms. There was no effect of Structure; *β* = -0.01, 95% *CI* [-0.07, 0.04], *t* = -0.50, *p* = .614.

Word-level RT results. Figure 14 presents the word-level RT results for the PJ task.

The analyses on RTs were conducted on two regions: the verb region and the region that corresponded to the second NP (the accusative object in active clauses and the adjunct agent in passives).

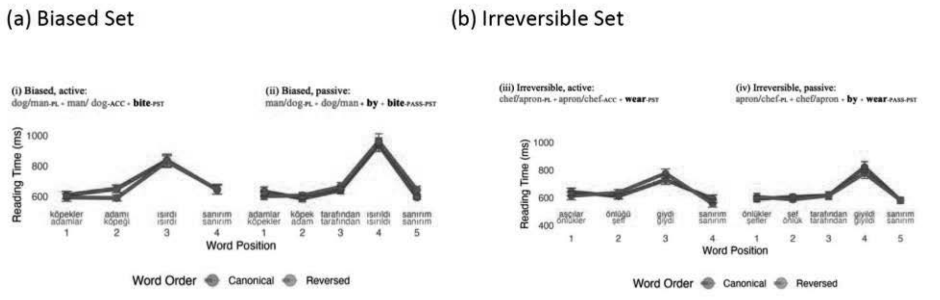


Figure 14. Word-by-word RT in the PJ Task by Set, Structure, and Word Order.

Verb region. The RTs increased with word length (β = 0.06, 95% *CI* [0.02, 0.10], t = 3.02, p = .003) but showed no effect of Word Order; β = -0.01, 95% *CI* [-0.07, 0.05], t = -0.21, p = .830. There was a marginal effect of Structure where RTs were longer in passives than actives; β = 0.06, 95% *CI* [-0.00, 0.13], t = 1.90, p = .058.

For irreversible sentences, there was a significant Word Order × Structure interaction; β = -0.12, 95% *CI* [-0.23, -0.00], t = -2.03, p = .042. RTs increased with word length (β = 0.04, 95% *CI* [0.00, 0.08], t = 2.03, p = .043), and there was a marginal effect of reversed order; β = 0.08, 95% *CI* [-0.00, 0.15], t = 1.88, p = .060. The interaction indicated that the Word Order slowdown was restricted to active sentences. Structure alone showed no effect; β = 0.03, 95% *CI* [-0.06, 0.11], t = 0.64, p = .520.

NP2 region. For biased-reversible sentences, there was a significant Word Order × Structure interaction; β = 0.10, 95% *CI* [0.02, 0.18], t = 2.48, p = .013. Word length increased RTs; β = 0.04, 95% *CI* [0.01, 0.06], t = 2.98, p = .003. A Word Order main effect indicated that reversed sequences were faster than canonical ones; β = -0.07, 95% *CI* [-0.12, -0.01], t = -2.36, p = .018. Structure alone was not significant; β = -0.05, 95% *CI* [-0.10, 0.01], t = -1.54, p = .124. The interaction showed that the facilitation for the reversed order disappeared in passives.

For the irreversible sentences, none of the factors affected the RTs: Word length (β = 0.03, 95% *CI* [-0.00, 0.06], t = 1.74, p = .082), Word Order (β = -0.01, 95% *CI* [-0.05, 0.03], t = -0.30, p = .767), and Structure (β = -0.02, 95% *CI* [-0.06, 0.03], t = -0.493, p = .463) were all null.

2.1.6. Discussion

The PJ data overall revealed good-enough parsing effects in Turkish, as reflected in the participants' accuracy patterns. Accuracy in plausibility judgments was lower when the syntactic structure conflicted with real-world plausibility, a pattern similar to findings in English (e.g., Ferreira, 2003). Notably, however, we observed this effect in both active and passive constructions, a result that contrasts with our morphological predictions based on prior work (Bornkessel-Schlesewsky et al., 2011). We argue, in the interim, that real-world plausibility overrides morphosyntactic information for sentence-final decisions. We will return to this conclusion in more detail in General Discussion.

The PJ data also showed that the lexical cue by the postposition *tarafından* (by) and the presence of an animacy contrast in the irreversible set (i.e., animate subject and inanimate object) prevented good-enough processing and guided the readers towards the correct syntactic interpretation. The morpho-syntactic cues were used, online, at the NP2 region but they did not influence if a good-enough parsing strategy was employed to the same extent as animacy cues.

2.2 Experiment 2: Agent-Patient Detection (APD)

2.2.1 Participants

Seventy-four native speakers of Turkish (55 female) participated in Experiment 2, which included an APD task. All the participants were first-year undergraduate students at Boğaziçi University, enrolled in an introductory linguistics course, and received course credit in exchange for their participation. As in Experiment 1, the participants were tested online using PCIbex Farm platform (Zehr & Schwarz, 2018). All participants provided informed consent before the experiment commenced.

2.2.2 Materials

The materials were the same as in Experiment 1.

We predict, for Experiment 2, that if task effects influence the depth of processing (see the procedures below) as shown in previous work (e.g., Swets et al., 2008), the explicit requirement to identify thematic roles will reduce good-enough processing effects and induce a syntax-driven parsing strategy.

This should yield higher accuracy in agent and patient identification across all conditions compared to the Plausibility Judgment (PJ) task, particularly for sentences where syntax and plausibility normally conflict (e.g., 1c, 1d). The animacy cues in the reversed conditions of the Irreversible set (e.g., sentences 2c, 2d) are predicted to contribute to high accuracy with relatively efficient processing for thematic role assignment.

Since the APD task is predicted to encourage more immediate and detailed online syntactic analysis, word-level reading times at verbs and sentence-final decision times can be longer than those in the PJ task because participants are predicted to engage proactively to resolve argument roles. Reversed word order may also continue to impose a general processing cost, even if final accuracy in role assignment remains high due to the task's focus on syntactic detail.

However, if good-enough parsing strategies occur irrespective of task requirements, the predictions are like those proposed for Experiment 1.

2.2.3 Procedure

The procedure was the same as in Experiment 1, except that each sentence was followed by a comprehension question prompting the explicit identification of the agent or patient by mouse-clicking one of two noun phrases displayed on the screen. Participants were asked to identify the agent 50% of the time and the patient 50% of the time, with question type (e.g., … *ısır-dı* 'bit' versus … *ısır-ıl-dı* 'was bitten') counterbalanced across trials.

2.2.4 Data Analysis

The same data analysis procedures as in Experiment 1 were employed. Two participants in the APD task were removed from the analyses as their accuracy was below 80% (corresponding to ~1.5% of the data).

2.2.5 Results

As in Experiment 1, the data were analyzed using R (R Core Team, 2023) with the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages.

The results are reported for accuracy, sentence-final DTs, and word-level RTs, separately, below.

Accuracy results. Table 28 presents the percent accuracy for the sentence-final agent-patient detection.

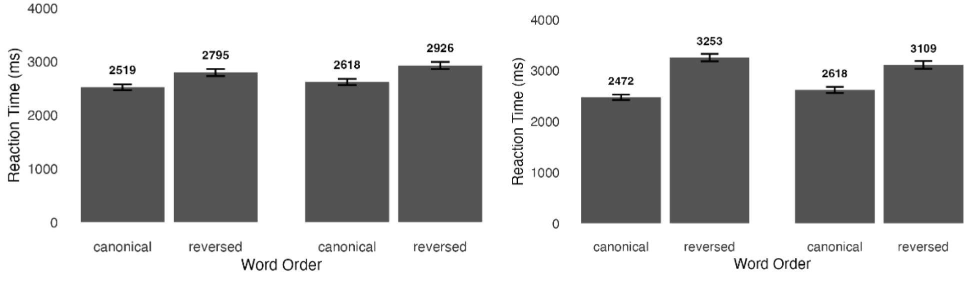
Table 28. Percent accuracy on the agent-patient detection in Experiment 2.

Set	Word Order	Structure	Percent Accuracy
Biased	Canonical	Active	92
		Passive	96
	Reversed	Active	91
		Passive	92
Irreversible	Canonical	Active	97
		Passive	96
	Reversed	Active	94
		Passive	94

The analyses on the accuracy data showed that models that involved an interaction of Word Order and Set did not explain the data better than a simple model with Word Order, Set, and Structure as predictors, $\chi^2 = 0.71$, $p = .398$. As such, we analyzed the data without splitting by Set.

Overall, there was a main effect of Word Order where reversed sentences were less likely to be judged correctly than canonical sentences, *OR* = 0.62, 95% *CI* [0.44, 0.88], $t = -2.71$, $p = .007$. There were no effects of Set (*OR* = 1.01, 95% *CI* [0.49, 2.09], $t = 0.03$, $p = .979$) or Structure (*OR* = 1.20, 95% *CI* [0.85, 1.68], $t = 1.03$, $p = .303$).

Decision time results. Figure 3 presents the DT results for the APD task.



(a) Biased Set (b) Irreversible Set

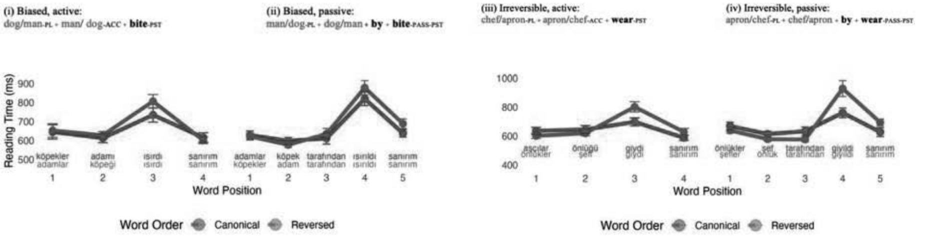
Figure 15. Sentence-final DTs in the APD Task by Set, Structure, and Word Order

The analyses on DTs for sentence plausibility showed that the models that included an interaction of Word Order and Set fit the data better than the simpler models with Word Order, Set, and Structure as main effects, $\chi^2 = 20.53$, $df = 1$, $p < .001$. Hence, the DT data were examined separately for each set.

In the biased set, there was a main effect of Word Order where reversed sentences were processed more slowly than canonical sentences ($\beta = 0.10$, 95% *CI* [0.06, 0.13], $t = 5.23$, $p < .001$), corresponding to an increase from approximately 2,380 ms to 2,618 ms. There was also a main effect of Structure where passives were processed more slowly than actives ($\beta = 0.05$, 95% *CI* [0.01, 0.08], $t = 2.63$, $p = .009$), increasing decision times by approximately 121 ms.

In the irreversible set, there was a significant Word Order × Structure interaction; $\beta = -0.10$, 95% *CI* [-0.18, -0.03], $t = -2.80$, $p = .005$. There was a main effect of Word Order where reversed sentences were processed more slowly than canonical sentences ($\beta = 0.27$, 95% *CI* [0.22, 0.32], $t = 10.10$, $p < .001$), corresponding to an increase from approximately 2,300 ms to 3,017 ms. Structure showed a marginal effect; $\beta = 0.05$, 95% *CI* [-0.00, 0.10], $t = 1.81$, $p = .070$. The interaction indicated that the Word Order slowdown was reduced in passive sentences.

Word-level RT results. Figure 16 presents the word-level RT results for the APD task.



(a) Biased Set **(b) Irreversible Set**

Figure 16. Word-by-word RT in the APD Task by Set, Structure, and Word Order.

Verb Region. For the biased sentences, RTs showed no effect of word length (β = 0.02, 95% *CI* [-0.01, 0.05], t = 1.38, p = .166) but they increased with reversed order (β = 0.05, 95% *CI* [0.01, 0.09], t = 2.51, p = .012), corresponding to an increase from approximately 600 ms to 630 ms. There was also a main effect of Structure where passives increased RTs compared to actives (β = 0.06, 95% *CI* [0.01, 0.11], t = 2.55, p = .011).

For irreversible sentences, RTs increased with word length (β = 0.04, 95% *CI* [0.01, 0.07], t = 2.72, p = .007) and with reversed order (β = 0.06, 95% *CI* [0.02, 0.10], t = 2.94, p = .003), corresponding to an increase from approximately 607 ms to 644 ms. No effect of Structure was observed; β = -0.01, 95% *CI* [-0.05, 0.04], t = -0.24, p = .810.

NP2 Region. For biased sentences, none of the fixed effects approached significance: Word length (β = 0.02, 95% *CI* [-0.00, 0.04], t = 1.76, p = .078), Word Order (β = 0.01, 95% *CI* [-0.03, 0.04], t = 0.43, p = .665), and Structure (β = 0.00, 95% *CI* [-0.04, 0.04], t = 0.06, p = .952) all had null effects.

For irreversible sentences, similarly, none of the fixed effects approached significance: Word length (β = 0.01, 95% *CI* [-0.02, 0.04], t = 0.67, p = .504), Word Order (β = -0.01, 95% *CI* [-0.04, 0.03], t = -0.50, p = .615), and Structure (β = -0.01, 95% *CI* [-0.05, 0.03], t = -0.69, p = .489) all had null effects.

2.2.6. Discussion

The results of the APD task show that accuracy was high in all conditions. The sentence-final DT data showed that when word order was reversed or was passive, the participants took longer to identify agents and patients of described events. Similarly, the RT data for the verb region showed that the participants spent longer time in reversed order sentences. These results, alongside those in the PJ task, indicate that the cognitive requirements of a task influence the depth of syntactic analysis in sentence processing. The explicit demands to identify the agent or patient in the APD task led the participants to engage in detailed, morphology/syntax-driven parsing, resulting in correctly assigning thematic roles despite conflicting real-world plausibility information. These processes took the participants longer to read the sentences or make decisions sentence-finally compared to the PJ task. This outcome supports the conclusion that explicit syntactic tasks suppress plausibility-based heuristics, in line with findings from other languages (Swets et al., 2008; Ehrenhofer et al., 2018).

Taken together with the results of the PJ task, the results of the APD task show that the depth of syntactic analysis in sentence comprehension, at least in Turkish, is adaptively scaled to task demands. The cost of deeper syntactic analyses reflects a natural tradeoff between speed and accuracy in language processing (cf., McElree, 2006), where more detailed analyses require additional processing time.

3. GENERAL DISCUSSION

The results for the PJ task showed that the participants' accuracy for real-world plausibility judgments was reduced in the biased set when the word order was reversed as in *The men bit the dog/The dog were bitten by the men*. The participants were highly accurate in the irreversible set, where the reversing of the arguments (one animate, one inanimate) resulted in implausible events (e.g., where *aprons wear chefs* as opposed to where *chefs wear aprons*). The results confirm our predictions regarding similarity-based cue interference in sentence comprehension and its consequences for good-enough parsing. We had predicted that the overlapping animacy cue for the two argument NPs in the biased set would cause similarity-based interference (Lewis & Vasishth, 2005) and result in difficulty with detecting thematic role assignments based on syntactic structure especially when they conflicted with real world plausibility. The participants judgments indicated that they relied more on real-world plausibility information in these conditions. When the animacy cue from the two NPs did not overlap, as in the irreversible set, the participants were more likely to detect the thematic roles based on syntactic structure, which aligns with findings that strong animacy information supports accurate thematic assignment and mitigates reliance on plausibility-based heuristics (Schlesewsky & Bornkessel, 2006; Zhou et al., 2018).

These results were predicted by the good-enough parsing model, but the fact that it was observed for active as well as passive constructions is unlike the English data (Ferreira, 2003). It is not entirely clear why the participants engaged in good-enough heuristics in the syntactically less complex active constructions. It is possible that the participants judged the real-world plausibility of the event, i.e., their ratings reflected how likely the agent (e.g., *the men*) could perform the action (e.g., *biting*) to the patient (*the dogs*). However, the participants were not more likely to employ real-world plausibility heuristics in the passive constructions than their active counterparts. The participants' plausibility ratings for the passive counterparts of these conditions were both numerically and statistically lower (18% plausible) than their active counterparts (26% plausible). It is possible that the postposition *tarafından* (by), a lexical item in Turkish (unlike its functional counterpart in English), which marks the agent phrase in the passive construction was more informative about thematic roles than the accusative case marking on the object phrase in the active construction. This would suggest that lexical (as well as semantic) cues are more informative for thematic role assignments than morphological cues and would explain why the passive sentences in Turkish (involving a lexical cue to thematic roles) did not involve more good-enough parsing than active sentences (involving a morphological cue to thematic roles).

These results suggest that a good-enough parsing strategy may indeed be related to cue-based memory retrievals (Lewis & Vasishth, 2005) and to the informativeness of lexical, semantic and morphological cues to thematic role assignment. When there is similarity-based interference (as in overlapping animacy cues) and/or syntactic difficulty (as in passive as opposed to active constructions in English), language users can be more vulnerable to good-enough heuristics due to the distributed cognitive resources. The lexical and morphological cues to the overall syntactic structure and thematic roles appear to be weighted where lexical cues are taken to be more informative than morphological cues. In cases when there are additional and informative cues to the thematic structure, the parser can detect thematic roles more easily and can ignore interference from real-world plausibility information.

The sentence-final DTs for the PJ task were mainly in line with the participants' accuracy data. The DTs were shorter for canonical word orders in the biased set but were so for the non-canonical word orders in the irreversible set. This is probably due, again, to the overlapping/non-overlapping animacy cue from the NPs in each set (Lewis & Vasishth, 2005). When the animacy cues did not overlap, as in the irreversible set, and when word orders were reversed, the participants detected the implausible events much faster. When the animacy cues overlapped, as in the biased set, the participants had difficulty detecting the plausibility of the event.

The APD data showed no reliable differences between conditions for accuracy: performance was near-ceiling across all sets and word orders. This replicates findings that when task demands require detailed parses (such as explicit agent/patient identification), comprehenders engage in syntactically detailed parsing, and not in plausibility-based heuristics (Ehrenhofer et al., 2018;

Swets et al., 2008). However, these detailed analyses were costly in decision and reading times, especially for reversed orders and passives in the biased set, where both NPs were animate. This pattern is consistent with evidence from speed-accuracy tradeoff studies, which show that retrieving more complex or integrated information, such as that required for deeper syntactic or conceptual analysis, naturally incurs greater processing time, reflecting a tradeoff between speed and accuracy in comprehension (McElree et al., 2006).

Taken together, our results suggest that good-enough processing could be modulated by linguistic cue retrievals, cue-informativeness and task demands. This is in line with cross-linguistic evidence indicating that semantic features such as animacy and lexical cues significantly influence heuristic processing. For instance, animacy hierarchies have been shown to guide role assignment across languages, often overriding syntactic cues in certain contexts (Bornkessel-Schlesewsky et al., 2008; Kyriaki et al., 2021). Similarly, lexical semantic cues can mitigate interference effects during sentence processing (Frances, 2024). Our results are also consistent with theoretical accounts arguing that plausibility judgment tasks, by their nature, encourage surface-level, heuristic processing over detailed syntactic analysis (Swets et al., 2008; Tan & Foltz, 2020; Gilbert et al., 2021).

It must be noted that were not able to replicate the previous effects of morphology on good-enough processing. Recall that Bornkessel-Schlesewsky et al., (2011) had reported for Turkish (as well as German and Chinese) that good-enough parsing was less pronounced due to earlier or more reliable detection of reversal anomalies than languages such as English, where word order tends to dominate thematic assignment (Bornkessel-Schlesewsky et al., 2011). We believe that the lexical cue in our corresponding passive constructions were stronger than the morphological cue in the active constructions to thematic role assignments. We aim to investigate, in future work, in a more systematic manner, if morpho-syntactic information such as case marking can affect employment of good-enough heuristics in Turkish.

Future work can also explore timing effects more systematically by manipulating intervals between critical regions or using graded comprehension tasks to better understand the temporal dynamics of good-enough processing. Neurophysiological methods such as ERP can also provide more direct evidence of whether the task effects we observed reflect early differences in parsing strategies or later monitoring processes. Finally, Turkish's flexible word order offers opportunities to examine how different argument orders (SOV, OSV, etc.) interact with case marking and animacy cues, potentially revealing whether certain orders are more susceptible to plausibility-driven heuristics and shedding light on the interaction between structural complexity and semantic processing strategies.

4. CONCLUSION

The results confirm previous work that task demands can influence the extent of good-enough heuristics. When the task requires explicit analysis of sentence structure, such as identifying agents and patients, Turkish speakers engage in syntactically detailed parsing. But when the task does not require a syntactically detailed parse, they tend to employ heuristic, plausibility-driven strategies. The results also show that certain cues such as overlapping semantic cues (e.g., the animacy of the NPs) or lexical cues to the thematic roles can influence the extent of good-enough processing, suggesting that good-enough parsing may indeed be related to similarity-based cue interference and can be examined in the context of cue-interference, cue-weighting and cue-informativeness (Lewis & Vasishth, 2005; Frances, 2024).

References

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

Bakay, Ö., & Deniz, N. D. (2021). *Case interference and phrase length effects in processing Turkish center-embeddings.* 34th Annual CUNY Conference on Human Sentence Processing, University of Pennsylvania.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*(3), 133–152. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Chow, W.-Y., Lau ,Ellen, Wang, Suiping, & and Phillips, C. (2018). Wait a second! Delayed impact of argument roles on online verb prediction. *Language, Cognition and Neuroscience*, *33*(7), 803–828. https://doi.org/10.1080/23273798.2018.1427878

Christianson, K., Dempsey, Jack, M. Deshaies, Sarah-Elizabeth, Tsiola, Anna, & Valderrama, L. P. (2023). Do readers misassign thematic roles? Evidence from a trailing boundary-change paradigm. *Language, Cognition and Neuroscience*, *38*(6), 872–892. https://doi.org/10.1080/23273798.2023.2171071

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Chromý, J. (2022). When readers fail to form a coherent representation of garden-path sentences. *Quarterly Journal of Experimental Psychology*, *75*(1), 169–190. https://doi.org/10.1177/17470218211037152

Ehrenhofer, L. M. I. (n.d.). *Argument Roles in Adult and Child Comprehension* [Ph.D.]. Retrieved May 30, 2025, from https://www.proquest.com/docview/2116153898/abstract/30F8732EDA8F4C84PQ/1

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Frances, C. (2024). Good enough processing: What have we learned in the 20 years since Ferreira et al. (2002)? *Frontiers in Psychology*, *15*, 1323700. https://doi.org/10.3389/fpsyg.2024.1323700

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, Language, Brain*, *2000*, 95–126.

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2021). The relationship between sentence comprehension and lexical-semantic retuning. *Journal of Memory and Language*, *116*, 104188. https://doi.org/10.1016/j.jml.2020.104188

Göksel, A., & Kerslake, C. (2004). *Turkish: A Comprehensive Grammar* (1st ed.). Routledge. https://doi.org/10.4324/9780203340769

Hale, J. (2006). Uncertainty About the Rest of the Sentence. *Cognitive Science*, *30*(4), 643–672. https://doi.org/10.1207/s15516709cog0000_64

Jegerski, J. (2013). Self-Paced Reading. In *Research Methods in Second Language Psycholinguistics*. Routledge.

Keleş, O., & Deniz, N. D. (2025). When men bite dogs: Testing good-enough parsing in Turkish with humans and large language models. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (pp. 219–231). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.cmcl-1.26

Kharkwal, G., & Stromswold, K. (2014). Good-Enough Language Processing: Evidence from Sentence-Video Matching. *Journal of Psycholinguistic Research*, *43*(1), 27–43. https://doi.org/10.1007/s10936-013-9239-5

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. https://doi.org/10.1016/0010-0277(72)90028-5

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129. https://doi.org/10.1016/S0926-6410(03)00086-7

Kyriaki, L., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2021). Thematic role assignment in non-default verb classes: A cross-linguistic comparison of English and German. *Glossa: A Journal of General Linguistics*, *6*(1), Article 1. https://doi.org/10.16995/glossa.5728

Lassotta, R., Omaki, A., & Franck, J. (2016). Developmental changes in misinterpretation of garden-path wh-questions in French. *Quarterly Journal of Experimental Psychology*, *69*(5), 829–854. https://doi.org/10.1080/17470218.2015.1054845

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Logačev, P., Aydın ,Özgur, & and Tuncer, A. M. (2022). Absence of evidence for underspecification in prenominal relative clause attachment. *Journal of Cognitive Psychology*, *34*(5), 575–591. https://doi.org/10.1080/20445911.2022.2029868

Logačev, P., & Vasishth, S. (2016). A Multiple-Channel Model of Task-Dependent Ambiguity Resolution in Sentence Comprehension. *Cognitive Science*, *40*(2), 266–298. https://doi.org/10.1111/cogs.12228

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

McElree, B., Murphy, G. L., & Ochoa, T. (2006). Time-Course of Retrieving Conceptual Information: A Speed-Accuracy Tradeoff Study. *Psychonomic Bulletin & Review*, *13*(5), 848–853. https://doi.org/10.3758/bf03194008

Nakamura, M., Momma, S., Sakai, H. and Phillips, C. (2024), Task and Timing Effects in Argument Role Sensitivity: Evidence From Production, EEG, and Computational Modeling. Cognitive Science, 48: e70023. https://doi.org/10.1111/cogs.70023

Nakamura, C., & Arai, M. (2016). Persistence of Initial Misanalysis With No Referential Ambiguity. *Cognitive Science*, *40*(4), 909–940. https://doi.org/10.1111/cogs.12266

Paape, D. (2024). How do linguistic illusions arise? Rational inference and good-enough processing as competing latent processes within individuals. *Language, Cognition and Neuroscience*, *39*(10), 1334–1365. https://doi.org/10.1080/23273798.2024.2387226

Qian, Z., Garnsey ,Susan, & and Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254. https://doi.org/10.1080/23273798.2017.1379606

Schlesewsky, M., & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: Electrophysiological evidence from subject–object ambiguities in language comprehension. *Brain Research*, *1098*(1), 139–152. https://doi.org/10.1016/j.brainres.2006.04.080

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Tan, M., & Foltz, A. (2020). Task Sensitivity in L2 English Speakers' Syntactic Processing: Evidence for Good-Enough Processing in Self-Paced Reading. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.575847

Vasishth, S., Suckow ,Katja, Lewis ,Richard L., & and Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567. https://doi.org/10.1080/01690960903310587

von der Malsburg, T., & and Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578. https://doi.org/10.1080/01690965.2012.728232

Zehr, J., & Schwarz, F. (2018). *PennController for internet based experiments (IBEX)*.

Zhou, P., Yao, Y., & Christianson, K. (2018). When Structure Competes with Semantics: Reading Chinese Relative Clauses. *Collabra: Psychology*, *4*(1), 22. https://doi.org/10.1525/collabra.131

Chapter 8:
What Eye Movements Reveal about Pre-Verbal Focus Processing in Turkish: An Eye-Tracking Study on Sentence Comprehension during Reading

İpek Pınar Uzun, Seçkin Arslan, & Özgür Aydın

Abstract

The present study reports on an eye-movement monitoring experiment investigating the focus (immediate vs. non-immediate preverbal) and argument (subject vs. object noun phrases (NPs)) position during silent reading. Turkish exhibits flexible word order where the inherent focus position is immediately pre-verbal. We examined whether processing differences occur when the focused constituent is positioned either at canonical neutral focus (immediate preverbal) or elsewhere in the preverbal area, and if so, whether these processing differences interact with the order of subject and object arguments (i.e. syntactic position). Our participants were asked to read sets of question-answer pairs, while their eye-movements were recorded, and to respond to an end-of-trial felicity judgement task. Our results showed that focused-elements were largely preferred in the immediate preverbal position. However, longer fixation durations were observed for those in the non-immediate preverbal position, and subject focus was consistently judged to be more acceptable than object focus in both immediate and non-immediate preverbal regions. Further, longer fixations and more regressions were observed for non-focused subjects than focused subjects. We suggest that there is a processing advantage for subject-scrambling to the immediate pre-verbal position in Turkish, which makes focus marking rather more salient during reading. This processing pattern is not applicable to object phrases. Our findings are compatible with theories which hold that focused-elements require greater processing demands, and that the given information is expected before the new during focus processing.

Keywords: Focus, sentence comprehension, eye-tracking during reading, Turkish

1. INTRODUCTION

In sentence processing research, much attention has recently been placed on processing what is beyond the meanings of words in a sentence (see e.g., Hagoort, 2019). Anyone with the intact ability to comprehend language may interpret the sentence ‘*My dog chewed my homework*’, without providing much effort. However, interpreting features hidden behind the sheer words in a sentence, including pragmatic and information structure level features, is only possible when reader identifies the correct focus position. Compare the difference between ‘$_{F}$ MY DOG *chewed my homework*’ and ‘*My dog chewed* $_{F}$ MY HOMEWORK’. The speaker in the first sentence places focus on the dog –who chewed the homework– not a cat or giraffe, and this was probably not expected from this dog. One can easily read between the lines and identify these ‘hidden meanings’ during listening because overt marking of pitch accent on certain words makes what is focused clear to the listener. This notion is known as ‘focus marking’ which signals the newly asserted information by assigning contrastive information between current information and its alternatives (Halliday, 1967; Jackendoff, 1972; Rochemont & Culicover, 1990; Rooth, 1992; Selkirk, 1995). What about during silent reading? Is it possible to identify focus position during silent reading? This partly depends on the language, as focus marking strategies vary across languages. Some languages with relatively less flexible word order rely more on prosodic features, such as accentuation and/or sentential prosody (e.g., German, English), while some others, like Turkish, recruit both prosody and word order to mark focused elements (see Bader, 1998; Birch & Rayner, 1997; Stolterfoht et al., 2007). Thus, for a Turkish reader, the sentence ‘$_{F}$ MY DOG *chewed my homework*’ would correspond to ‘*my homework* $_{F}$ MY DOG *chewed*’. Accordingly, the critical pre-verbal position of the Noun Phrase (NP) ‘my dog’ signals that it is the focused constituent, although the reader does not actually hear the prosody of the sentence. Focus position has been shown to be an important determiner in successful sentence interpretation in auditory sentence listening tasks (Cutler et al., 1997; Hruska et al., 2000; Schafer et al., 2000). Although written text does not often explicitly convey prosodic information, numerous studies have shown that readers assign prosodic and phonological representations during silent reading (Ashby & Clifton, 2005; Bader, 1998; Chafe, 1988; Fodor, 1998, 2002; Rayner et al., 2012; Savill et al., 2011).[1] This is based on the Implicit Prosody Hypothesis (IPH; Fodor, 2002), which holds that readers activate representations of prosodic focus during silent reading. This explanation stands on support from a number of studies reporting that focus structure effects are visible when skilled readers silently read sentence material (Birch & Rayner, 1997; Kentner & Vasishth, 2016; Kitagawa et al., 2013; Paterson et al., 2007; Stolterfoht & Bader, 2004; Stolterfoht et al., 2007). Nonetheless, our current understanding remains limited regarding how focus structure is processed in flexible word-order languages, particularly those that also mark focus syntactically. The present study reports on an eye-movement monitoring experiment during reading to investigate the moment-by-moment incremental processing of preverbal focus in Turkish.

1.1. Processing Focus Marking during Sentence Comprehension

Focus assignment is relevant to pragmatic and semantic interpretation of an utterance and is often through prosodic and/or syntactic means (e.g., Krifka, 1992; Selkirk, 1995; Zubizarreta, 1998). A widely renowned approach to focus marking is the dichotomy between contrastive and presentational focus (Selkirk, 2002; Vallduví & Engdahl, 1996; Zimmermann, 2008, but see also Kiss, 1998 for alternative focus dichotomies). According to the contrastive account, focus marking functions to mark new as opposed to contrastive information, see question-answer pairs in (1) for an illustration.

(1) a. Q: What did John buy in the store?   A: John bought $[\text{a book}]_{F}$ in the store.

b. Q: Did John buy a journal in the store?   A: (No), John bought $[\text{a book}]_{CF}$ in the store.

The answer in (1a) illustrates a sentence that marks new information representing a proposition from a set of alternatives (a *book* but not *shoes* or *bags* or *anything else*), whereas (1b) exhibits a contrastive focus where the newly inserted information refers to an only true exhaustively contrasted proposition. An important issue here is that *wh*-questions provide partial information regarding something John bought, which leads the reader/hearer to evaluate and map grammatical focus realization in answers (Büring, 2010; Zimmermann & Onea, 2011).

[1] Following the phonological coding account, we assume that phonological and prosodic representations are built up during silent sentence reading. However, how these phonological processes come about during reading is out of the scope of this paper.

It is well-established that sentence interpretation is sensitive to contrastive focus during sentence listening (Braun & Biezma, 2019; Breen et al., 2010; Carlson et al., 2009; Watson et al., 2008), and sentence reading (Bader, 1998; Birch & Rayner, 1997; Gotzner & Spalek, 2019; Gross et al., 2014; Kentner & Vasishth, 2016; Kitagawa et al., 2013; Paterson et al., 2007; Sauermann et al., 2013; Stolterfoht et al., 2007). Birch and Rayner (1997) using an eye-movement monitoring during reading design in sentences where focused elements are manipulated syntactically or by preceding context, and they reported that the readers have longer reading times when a region is focused than when not. Bader (1998) examined processing local syntactic ambiguities with focus particles and sentential adverbials in word-by-word reading experiments and argued that readers need to reanalyze prosodic structure which in turn impacts on syntactic processing, making it hard to recover from syntactically anomalies due to prosodic constraints. Paterson et al. (2007) conducted eye-movement monitoring during reading experiments with the English focus participle '*only*' manipulating the presence and placement position of the participle in sentence comprehension. The authors reported that reading times increased when *only* replaced in an incongruent position with its constituent; as this effect dispersed when the focus particle was absent, the authors affiliated increased reading times with focus processing. Using a similar design, Sauermann et al. (2013) examined English sentences with remnant elliptical constructions (e.g., *John wondered who Sally would pass the apples/the children. Sally passed [only] the children the apples but not the grownups*…) where the answer is either congruent or incongruent with the *wh*-question. The authors found that focus processing is disrupted when focus was specified on the indirect object (e.g., the children) suggesting that the readers expect given information before the new (see also Arnold et al., 2000; Clifton & Frazier, 2004). Kitagawa et al. (2013) conducted an acceptability experiments using focus-sensitive particle -*dake* 'only' and polarity item *daremo* 'anyone' in Japanese sentences with either spoken and visual or with visual-only presentation. The authors reported that acceptability ratings increased when readers heard the sentence prosody at the same time and that intervention effects occur due to prosodic constraints.

Bornkessel et al. (2003), in their event-related potentials study with question-answer pairs in German, showed that both subject and object focused constituents evoked an early positivity (i.e. between 280–480ms), which the authors discuss as the indicator of focus integration. Stolterfoht et al. (2007), using an event-related potentials design during silent reading, examined

German sentences with contrastive ellipses (e.g., *Am Dienstag hat der Direktor [nur] den Schüler getadelt, und nicht den/der Lehrer.* "On Tuesday, the principal criticized [only] the pupil, and not the teacher.") Depending on which article form the elliptical constituent receives, interpretation of the sentence shifts as to whether or not the teacher was criticized. Their results unveiled that the sentences in which the focus particle *nur* 'only' was absent elicited lower acceptability scores than those containing the particle, and a sustained positivity with an onset at about 350ms, while the sentences requiring implicit prosodic revision evoked a negativity that occurred between 400-600ms time-range. Kentner and Vasishth (2016), using a behavioral oral reading and an eye-tracking during reading experiments, investigated German sentences with ambiguous focus particle *auch* ('too') where the prosodic structure and contextual information are manipulated. In their oral reading results, the authors found that German readers preferred to realize focus on object NPs leaving *auch* unaccented; the eye-tracking data indicated that reading times on the focus particle increased when the context required focus to realize on subject NP, suggesting that ambiguity resolution in focus particles requires both context and prosodic sources to be involved in processing.

1.2 Focus Marking in Turkish and The Current Study

In Turkish, the focus structure can be grammatically marked through syntactic and prosodic means. Although Turkish has a base word order of Subject-Object-Verb, its word order is flexible allowing for constituent scrambling, and hence, subject and object NPs can move from their base-generated positions (İşsever, 2003; Kural, 1992; Taylan, 1984). The immediate preverbal position has traditionally been analyzed as the canonical focus position, as illustrated in (2a), (Erkü, 1983; Kennelly, 2003). With the availability of flexible word order, focus can also be realized in other constituent positions within the preverbal, see (2b), not in the post-verbal area, however, see (2c).

(2)	a.	Müşteri	[TEZGAHTAR-I]$_{\text{FOC}}$	suçla-dı
		customer$_{\text{-NOM}}$	salesclerk$_{\text{-ACC}}$	blame$_{\text{-PAST.3SG}}$
	b.	[MÜŞTERI]$_{\text{FOC}}$	tezgahtar-ı	suçla-dı
		customer$_{\text{-NOM}}$	salesclerk$_{\text{-ACC}}$	blame$_{\text{-PAST.3SG}}$
	c.	*Tezgahtar-ı	suçla-dı	[MÜŞTERI]$_{\text{FOC}}$
		salesclerk$_{\text{-ACC}}$	blame$_{\text{-PAST.3SG}}$	customer$_{\text{-NOM}}$

All: '*The customer blamed the salesclerk*'

In both (2a) and (2b), focus marking is prosodically prominent, which is indicated by high pitch (H*). In (2c), focus cannot be realized in the post-verbal area for structural constraints, and therefore, prosodical prominence is not applicable. The validity of a definite focus position has been put under scrutiny with more recent theoretical definitions of focus realization in Turkish (Göksel, 2013; Göksel & Özsoy, 2000; Güneş, 2013; Kamali & Krifka, 2020; Özge & Bozşahin, 2010). According to the Focus Field Hypothesis (FFH; Göksel & Özsoy, 2000), focus cannot be assigned freely to any phrase in Turkish, as per instance, post-verbal area is banned from receiving focus. The FFH considers the immediate pre-verbal position to be reserved for sentential stress but not as a designated focus position, suggesting that a focused element may in fact occur anywhere within the entire pre-verbal area. According to this point of view, focus field contains the area denoting information which is non-recoverable from the previous discourse (i.e., non-presupposed or new information) between the constituent that receives focal stress (H*) and region including the verb (V), see (3) for an illustration. The FFH rules out potential occurrences of elements denoting non-recoverable information or hosting focus in any other domain outside of the focus field.

(3) [$_{\text{pre-H*}}$ ….. [H*….. V] …..   ]

pre-H   focus field   postverbal*

Consider (4) and (5). Both sentences present a syntactically canonical SOV sentence with preverbal focus; in (4) focus is realized in the immediate preverbal position whereas in (5) focus occurs in the sentence initial position but not canonically in the immediate preverbal position.

(4)	[Müşteri	[TEZGAHTAR-I]$_{\text{FOC}}$]	suçla-dı SOV/canonical preverbal focus
	customer$_{\text{-NOM}}$	salesclerk$_{\text{-ACC}}$	blame-PAST.3SG

'The customer blamed the salesclerk'

(4') a. Who did the customer blame?

b. What did the customer do?

c. What happened? Any news?

(5)	[MÜŞTERI]$_{\text{FOC}}$	tezgahtar-ı	suçla-dı SOV/ non-canonical preverbal focus
	customer$_{\text{-NOM}}$	salesclerk$_{\text{-ACC}}$	blame-PAST.3SG

'The customer blamed the salesclerk'

(5') a. Who blamed the salesclerk?

b. #What happened to the salesclerk?

c. #What happened? Any news?

One can question whether focus position corresponds to the allowed contextual presentation following Büring (2010), who claimed that *wh*-phrases in question-answer pairs are supposed to correspond to the focused constituent in the answer. The focused elements in (4) correspond to the *wh*-phrases in (4'), that is, both narrow and wide focus readings are allowed. However, the focused element in (5) allows for a narrow focus reading compatible only with the *wh*-phrase in (5'a). According to the FFH (Göksel & Özsoy, 2000) which claims that contextually non-recoverable elements should always be located within the focus field, the subject NP *müşteri* 'customer' in (4) receives recoverable interpretation while focused element (i.e., *tezgahtarı* ''salesclerk') receives non-recoverable/ non-presupposed interpretation. Nonetheless, when the focused subject *müşteri* 'customer' is a leftmost focus phrase in the focus field as seen in (5), both focused element and object *tezgahtarı* 'salesclerk' receives non-presupposed interpretation. Consider syntactically non-canonical OSV sentences in (6)-(7). Particularly in (6), the leftmost part of the focus field hosts the object NP *tezgahtarı* 'salesclerk' which contains the presupposed information compatible with both wide and narrow focus readings, see (6'). This is based on the idea that constituents positioned within the focus field may host recoverable or presupposed information (Göksel & Özsoy, 2000; İşsever, 2019). By contrast, in (7), focused object and subject (*müşteri* 'customer') receives obligatory non-presupposed readings, compatible with a narrow focus reading only, see (7').

(6) Tezgahtar-ı [MÜŞTERI]$_{FOC}$ suçla-dı] OSV/immediate preverbal focus
salesclerk$_{-ACC}$ customer$_{-NOM}$ blame$_{-PAST.3SG}$
'*The customer blamed the salesclerk*'

(6') a. Who blamed the salesclerk?
b. What happened to the salesclerk?
c. What happened? Any news?

(7) [TEZGAHTAR-I]$_{FOC}$ *müşteri* suçla-dı OVS/ non-immediate preverbal focus
salesclerk$_{-ACC}$ customer$_{-NOM}$ blame$_{-PAST.3SG}$
'*The customer blamed the salesclerk*'

(7') a. Who did the customer blame?
b. #What did the customer do?
c. #What happened? Any news?

Following the articulated Complementizer Phrase (CP) structure proposed by Rizzi (2004), we assume that the [+focus] (focused) and [-focus] (non-focused) features occur within the CP domain (Şener, 2019). Structural distance of [focus] head from the original position of focused object is longer than that of the original position of focused subject independently of the accounts for focus movement to CP or Agree mechanism. That is, focus agrees with features of the focused NPs. This is also compatible with the Structural Distance Hypothesis (SDH) (Hawkins, 1999; O'Grady, 1999).

Experimental investigations on focus processing in Turkish are very scarce, and most of the available studies investigated impact of prosodic cues on comprehending ambiguous syntactic structures without contemplating much on focus marking (Uzun et al., 2021; Atasoy, 2022; Aydın & Uzun, 2023; Deniz, 2025; Dinçtopal-Deniz & Fodor, 2019; Zora et al., 2025). Of particular relevance, Kurt and Deniz (2023) reported results from an eye-tracking during reading and a behavioral sentence processing experiment with sentences involving the so-called replacive phrases. The broad vs. narrow focus was manipulated with syntactic position of constituents (*Defne Hanım [çiçeği] [yönetmene] verdi* 'Miss Defne gave the flowers to the director') which were followed by replacive phrases either compatible or incompatible with the given sentence (i.e. *aktör değil/sunucu değil* 'not the actor/not the presenter'). Their data indicated that for narrow focus conditions, reading times increased in both immediate pre-verbal and the replacive phrase regions. Associating narrow focus with non-canonical syntactic focus position, the authors suggest that immediate preverbal position is the preferred broad focus position in Turkish.

To sum up, focus processing during sentence reading has been characterized by longer reading/fixation times on focused than non-focused constituents (Bader, 1998; Birch & Rayner, 1997; Paterson et al., 2007); readers often expect given information before the new in English (Clifton & Frazier, 2004; Sauermann et al., 2013). While this strategy has been widely cited in languages like English, the findings from the current study suggest that Turkish readers may also exhibit sensitivity to the given-before-new principle, particularly when subjects are in focus.

It is also important to note that readers' increased sensitivity to non-neutral prosody during processing focused constituents rather strongly occurs in early measures. There is evidence for focus position to influence sentence comprehension during listening (see Cutler et al., 1997 for an overview);

however, less has been understood about syntactic position, and particularly, in reading when explicit prosody cannot aid comprehension. Research on flexible word-order languages other than English and German are sporadic, leaving gaps in our understanding how syntactic position of constituents interact with focus position. Turkish is one such language, allowing for rather more flexible word order than English or German, offering a suitable testing ground for readers' sensitivity to focus while being able to manipulate focus position and syntactic position of constituents in the preverbal area. This is the topic of our study; we are seeking insights into the per-millisecond processing of preverbal focus in Turkish using an eye-movements monitoring experiment during naturalistic reading.

The overarching aims of the current study are two-fold. Our first aim is to unveil the extent to which Turkish readers are sensitive to the effects of focus position within the pre-verbal area. The rationale behind this is that if we follow the traditional view for Turkish which posits that the canonical focus position is taken to be the immediate preverbal position (Kural, 1992; Taylan, 1984), we may observe reading disruptions when focus is hosted on an another constituent within the preverbal area. Alternatively, according to the FFH (Göksel & Özsoy, 2000), focus can occur on any constituent within the entire pre-verbal area, and if this view holds true, there should be fewer reading disruptions when focus is realized periphrastically to the immediate preverbal position. Our second aim is to uncover whether the syntactic position of constituents influences focus processing. Recall that canonical word order in Turkish is SOV, and that scrambling within the preverbal area (i.e. OS vs. SO) is fully perspicuous to Turkish readers, making it possible to position a subject constituent to the immediate pre-verbal position to mark focus. If this study finds greater reading times for focused object constituents than their subject counterparts without any clear effects of focus position, the processing cost may be affiliated with syntactic sources. However, an interaction between focus position and syntactic position would indicate simultaneous processing resources at play.

2. METHODS

2.1 Participants

A group of 50 native Turkish speakers aged 20–35 (mean = 22.94; SD = 3.81, 35 females), who were bachelors or master's students at Ankara University at the time of testing, participated in the study. All of the participants reported to have normal or corrected to normal vision; none of them reported to have extensive stays abroad or neurological/psychiatric disorder that may impact language comprehension ability in Turkish. The participants were informed that their participation was voluntary and were asked to give their consent allowing us to anonymously process their experimental data for scientific purposes. The procedures in this study conformed to ethical principles and were approved by the Ethical Board of Ankara University (*No:13-430-12*).

2.2 Materials and Experimental Design

A total of 24 question-answer pairs were constructed with four experimental conditions where the focus position within the preverbal area and the syntactic position (i.e. the order of subject vs. object arguments) were manipulated. Each trial began with an interrogative sentence followed by a target answer sentence. Position and/or argument type of the *wh*-phrase in the question clauses was manipulated to evoke different foci in the target answer clause by using nominative *who* pronoun for subject positions and accusative *whom* pronoun for object positions, see (8)-(10) (critical areas of interest (AOIs) are bolded). The position of *wh*-phrases, (i.e. either preverbal object or preverbal subject) elicited an inherent focus position in the answer response.

(8) *Non-i*Pre-Verbal: Non-immediate Preverbal Position

a. *Focused-subject Condition*

A: Toplantıda kim yazarı eleştirdi?
meeting-LOC who writer-ACC criticize-PAST.3SG

'At the meeting, *who* criticized the writer?'

B: Toplantıda **ressam** **yazarı** eleştirdi.
meeting-LOC [artist]FOC writer-ACC criticize-PAST.3SG

'At the meeting, *the artist* criticized the writer.'

b. *Focused-object Condition*

A: Toplantıda kimi ressam eleştirdi?

meeting-LOC who-ACC artist criticize-PAST.3SG

'At the meeting, *who* did the artist criticize?'

B: Toplantıda **yazarı** **ressam** eleştirdi.

meeting-LOC [writer-ACC]FOC artist-ACC criticize-PAST.3SG

'At the meeting, the artist criticized the *writer.*'

(9) *i*Pre-Verbal: Immediate Preverbal Position

a. *Focused-subject Condition*

A: Toplantıda ressamı kim eleştirdi?

meeting-LOC artist-ACC who criticize-PAST.3SG

'At the meeting, *who* criticized the artist?'

B: Toplantıda **ressamı** **yazar** eleştirdi.

meeting-LOC artist-ACC [writer]FOC criticize-PAST.3SG

'At the meeting, the *writer* criticized the artist.'

b. *Focused-object Condition*

A: Toplantıda yazar kimi eleştirdi?

meeting-LOC writer who-ACC criticize-PAST.3SG

'At the meeting, *who* did the writer criticize?

B: Toplantıda **yazar** **ressamı** eleştirdi.

meeting-LOC writer [artist-ACC]FOC criticize-PAST.3SG

'At the meeting, the writer criticized the *artist.*'

The length of experimental sentences was similar to four words, none of the words were shorter than two syllables. In order to rule out potential effects of thematic role assignment and verb semantics influence on sentence comprehension; we used reversible verbs only. Subject NPs were marked with nominative and object NPs were marked with accusative case consistently.

In all experimental sentences, we used transitive verbs allowing for subject and object NPs. These subject and object NPs contained generic profession names denoting human entities (i.e., writer, artist, salesclerk, etc.). In our experimental conditions, the focus was elicited on either of the NPs which corresponds to the *who*-pronoun in the question. The sentence materials started with a prepositional phrase to avoid critical NPs (i.e. *ormanda* 'in the forest') appearing in the sentence-initial position to reduce 'corner effects'. 96 experimental sentences were divided into two lists, counterbalancing focus position across participants and participants saw 48 experimental trials with equal number of items from each condition. In addition, there were 72 filler items in each list. Fillers contained 12 congruent and 60 incongruent question-answer pairs so as to balance out the number of (in)congruent trials. For the incongruent fillers, either the position of focused constituent did not allow for a natural reading (i.e. in the post-verbal area) or syntactic argument position did not match the *wh*-phrase, and thus, eliciting an inappropriate focus reading, see (10). In total, one participant saw 120 trials in a randomized order.

(10) *Fillers*

Q:	Toplantıda	oyuncuyu	kim	alkışladı?
	meeting.$_{\text{LOC}}$	performer.$_{\text{ACC}}$	who	applaud.$_{\text{PAST}}$
	'*Who applauded the performer in the meeting*?'			
A1:	Toplantıda	oyuncuyu	seyirci	alkışladı. (congruent)
	meeting.$_{\text{LOC}}$	performer.$_{\text{ACC}}$	[audience]$_{\text{FOC}}$	applaud.$_{\text{PAST}}$
A2:	#Toplantıda	oyuncuyu	alkışladı	seyirci. incongruent)
	meeting.$_{\text{LOC}}$	performer.$_{\text{ACC}}$	applaud.$_{\text{PAST}}$	[audience]$_{\text{FOC}}$
	Both: '*The audience applauded the performer in the meeting.*'			

2.3 Procedure and Analysis

All the experiments took place at Ankara University Linguistics Laboratory (diLab, http://dilab.ankara.edu.tr/). The participants were admitted to a dimly lit room and seated within a comfortable reading distance from a 1680 x 1050 pixels 22-inch high-quality LCD monitor equipped with an SMI RED-500 IView-X (SensoMotoric Instruments GmbH) eye-tracking unit positioned underneath the monitor. The eye-tracker sampled data at the rate of 500 Hz. Stimuli were programmed and presented using the SMI Experiment Building software tool in a way that the participants saw one question-answer pair in a single trial. The participants were asked to position their head within a chin rest placed about 60 cm afar from the monitor to avoid excessive head movements and off-screen looks. For an optimal reading experience, each experimental sentence material (i.e. question–answer clauses) was positioned on the top left to the middle of a trail screen and were presented in 36-point size with Courier New fonts so as to ascertain that letter characters have equal width. Sentence materials were presented black fonts in grey background. Each trial started with a fixation cross for 500 ms followed by the *wh*-question stimulus, which stayed on the screen until the participants pressed the spacebar. For the subsequent target answer stimulus to appear on the screen, participants needed to fixate on a top-left-aligned fixation cross for 500 ms so as to be able to neutralize eye fixations to a natural reading direction in Turkish (i.e. left to right). The target answer stimulus stayed on the screen until the participants responded with a spacebar press, following the response, an end-of-trial judgment task appeared. The judgement task required the participants to judge whether the answer stimulus was appropriate in reference to the question stimulus. The participants logged their responses by clicking on *evet* ('yes') or *hayır* ('no') response options presented on the screen.

The participants were instructed that they were about to see series of question-answer pairs which they should read silently for comprehension and respond to an end-of-trial judgment task for each trial by clicking on the most appropriate response option. The participants were reminded to keep their gaze on the screen as much as possible. The experiment started with a five-point calibration; the eye-tracker was then re-calibrated if necessary. At the beginning of each experiment, there were 10 practice trials to make sure that the participants understood the task. A total number of four breaks were programmed in the experiments. Each experiment took approximately 30 minutes to complete. There was a one-month interval between the two experimental blocks. We recorded end-of-sentence acceptance rates, and eye-movement measures reflecting both early processing stages (first fixation and first-pass duration in milliseconds) and integration stages (second-pass, dwell-time durations in millisecond and the total number of regressions per word). Object and subject NP positions in the answer stimulus were defined as critical AOIs, and the averaged fixation/pass times in milliseconds and number of regressions were extracted and analyzed for those critical AOIs. Eye-movement data that exceeds the mean by more than 2.5 SD for each condition were removed from analyses. The removed data were minimal and corresponded to 2.04% for the first fixation, 2.33% for the first-pass duration, 0.72% for the second-pass duration and 2.47% for the dwell-time durations.

The behavioral end-of-sentence acceptance rates were analyzed using the generalized linear mixed-effects regression models and the eye-movements data were analyzed using the linear mixed-effects regression models using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2025). Statistical significance was set to $p<0.05$, and the p-values for linear mixed models were obtained with the Satterthwaite approximation (Satterthwaite, 1946). An initial global model was computed with fixed-effects of FocusPosition×SyntacticPosition×AOI; individual participants and items were included as random intercepts and/or slopes where appropriate (Baayen, 2008). We included region length (i.e. number of characters per AOI across items) as a fixed-effect in mixed-effects regression models with continuous eye-movement data to be able to control for varying word lengths. Sum-coding was used for categorical variables (i.e. -0.5 vs. 0.5 instead of binary 1-0 coding for Focus/SyntacticPosition) to avoid potential biases. All fixation durations data were log-transformed to avoid problems with heteroscedasticity. In addition to critical AOIs analyzed in the current study, spillover regions were also examined for potential cascading effects, but no systematic patterns observed.

3. RESULTS

End-of-trial acceptance rates showed that the Turkish readers accepted the Immediate Pre-verbal (iPre-verbal) Subject condition 99% of the time (SD = 0.07), the iPre-Verbal Object condition 96% of the time (SD = 0.20), while in conditions where the focus position was not immediately pre-verbal, their acceptance rates were 93% for the Non-immediate Pre-verbal (non-iPre-verbal) Subject (SD = 0.26) and 85% for the non-iPre-verbal Object (SD = 0.35) conditions. A set of statistical outputs from a generalized mixed-effects regression model indicated significant fixed-effects of FocusType (ß = -2.22, SE = 0.22, z = -10.24, p <.001) and FocusPosition (ß = 1.54, SE = 0.22, z = 7.11, p <.001), and an interaction between the two (ß = -1.09, SE = 0.43, z = -2.52, p <.05). These outputs clearly suggest that the Turkish readers favored sentences where focus occurred immediately preverbally (98% [SD=0.15] both the iPre-verbal Subject and iPre-verbal Object conditions) than when focus occurred prior to its expected location (89% [SD=0.32] both the non-iPre-verbal Subject and non-iPre-verbal Object conditions). Furthermore, focused subject NPs were favored more often, as high as 96% (SD=0.19), as compared focused object NPs (90% [SD=0.29]). As all condition comparisons were statistically significant, we suggest the following scale of acceptability:

immediate preverbal focus on [Subject NP > Object NP] non-immediate preverbal focus on [Subject NP > Object NP].

Table 29. Descriptive means per condition in milliseconds and standard errors in parentheses (in milliseconds: FFD=First fixation duration, FPD=First Pass Duration, SPD=Second Pass Duration, DT=Dwell Time; Reg.=Number of regressions into critical AOI)

	SOV				OSV			
	Non-immediate pre-verbal position (Subject)		*Immediate Pre-verbal position (Object)*		*Non-immediate pre-verbal position (Object)*		*Immediate Pre-verbal position (Subject)*	
	[**S**]OV	[S]**O**V	S[**O**]V	**S**[O]V	[**O**]SV	[O]**S**V	O[**S**]V	**O**[S]V
FFD	171 (2.32)	149 (1.71)	195 (2.16)	197 (2.12)	172 (2.20)	154 (1.66)	202 (2.35)	220 (3.33)
FPD	199 (3.09)	166 (2.25)	284 (4.07)	283 (4.28)	239 (4.00)	189 (2.66)	260 (3.62)	288 (5.02)
SPD	353 (8.01)	276 (5.19)	374 (7.56)	427 (9.34)	487 (11.73)	219 (3.50)	302 (7.40)	453 (10.42)
DT	374 (7.36)	284 (4.96)	401 (6.93)	424 (8.53)	517 (11.00)	262 (3.96)	328 (5.59)	453 (10.00)
Reg	1.24 (0.01)	1.17 (0.01)	1.00 (0.00)	1.12 (0.01)	1.32 (0.02)	1.00 (0.00)	1.09 (0.01)	1.07 (0.01)

Note: The focused element is underlined and bold. The AOI for which the value in the table is given is shown in square brackets. For example, [**S**]OV indicates that the AOI is the subject and the subject is focused, while [S]**O**V indicates that the AOI is the subject, but the focused element is the object.

Table 30. Statistical outputs from global mixed-effects regression models computed with eye movements data (Foc=FocusPosition, SP=SyntacticPosition, AOI=Area of Interest, FFD=First Fixation Duration, FPD=First Pass Duration, SPD=Second Pass Duration, DT=Dwell Time, Reg=Regressions into AOI)

	FFD		FPD		SPD		DT		Reg	
	β(SE)	t	β(SE)	t	β(SE)	t	β(SE)	t	β(SE)	z
Intercept	4.92 (0.02)	229.94 ***	5.00 (0.03)	187.85 ***	5.36 (0.05)	114.73 ***	5.48 (0.04)	143.76 ***	0.15 (0.04)	3.76 ***
Foc	0.13 (0.02)	6.75 ***	0.30 (0.02)	13.40 ***	0.48 (0.04)	11.98 ***	0.54 (0.03)	21.18 ***	0.13 (0.05)	2.41 *
SP	0.04 (0.02)	2.08 *	0.12 (0.02)	5.31 ***	-0.13 (0.05)	-2.70 **	-0.04 (0.03)	-1.58	-0.15 (0.06)	-2.39 *
AOI	0.27 (0.02)	14.63 ***	0.52 (0.02)	23.01 ***	0.31 (0.05)	6.42 ***	0.34 (0.03)	13.50 ***	-0.15 (0.08)	-1.85
Foc × SP	-0.06 (0.03)	-2.30 *	-0.28 (0.03)	-8.79 ***	-0.15 (0.06)	-2.39 *	-0.26 (0.04)	-7.19 ***	0.09 (0.08)	1.11
Foc × AOI	-0.05 (0.03)	-1.76	-0.33 (0.03)	-10.43 ***	-0.38 (0.06)	-6.01 ***	-0.51 (0.04)	-14.32 ***	-0.06 (0.11)	-0.52
SP × AOI	-0.01 (0.03)	-0.19	-0.20 (0.03)	-6.21 ***	-0.15 (0.07)	-1.96 *	-0.15 (0.04)	-4.06 ***	0.24 (0.12)	1.99 *
Foc × SP × AOI	-0.03 (0.04)	-0.93	0.38 (0.05)	8.41 ***	0.37 (0.10)	3.86 ***	0.43 (0.05)	8.43 ***	-0.13 (0.16)	-0.82

** p<0.05, ** p<0.01, *** p<0.001*

For the eye-movement data, Table 29 presents mean statistics for each variable selected for analysis, and Table 30 exhibits outputs from a series of global (generalized) linear mixed-effects regression models with FocusPosition, SyntacticPosition and Area of Interest (both critical AOIs). For both the early eye-movement measurements (first fixation and first pass durations), the initial set of mixed-effects regression models showed significant fixed-effects of FocusPosition, SyntacticPosition, AOI and an interaction effect of FocusPosition×SyntacticPosition. Two-way interactions effects of FocusPosition×AOI, SyntacticPosition×AOI and a three-way interaction effect of FocusPosition×SyntacticPosition×AOI were proved significant. For second-pass duration, we found significant fixed-effects of FocusPosition, SyntacticPosition, AOI and all interaction effects between those factors. For dwell-time, fixed-effects of FocusPosition and SyntacticPosition were significant, and we observed a three-way interaction of FocusPosition × SyntacticPosition × AOI. These suggest that Turkish readers are sensitive to focus manipulations during both early and late reading processes. For the regressions into critical AOIs, fixed effects of FOCUSPOSITION and SYNTACTICPOSITION proved significant, and we observed significant interaction effects of SYNTACTICPOSITION×AOI, suggesting that the participants' eye-movements regressed back into the critical regions depending on the word region and SYNTACTICPOSITION (i.e. object vs subject). Longer fixations were recorded at the focused elements in the immediate pre-verbal region than focused elements in the non-immediate preverbal region.

Since there was a significant three-way interaction effect for first/second pass and dwell time, and a two-way interaction for regression data indicating different reading patterns emerge in each AOI, we further analyzed the data per AOI. Table 31 presents two subsequent mixed-effects regression models computed on eye-movement data from each AOI, and Figure 1 demonstrates per-region eye-movements data across conditions.

Table 31. Statistical outputs from subsequent mixed-effects regression models computed with eye-movements data from immediate and non-immediate preverbal AOIs (Foc=FocusPosition, SP=SyntacticPosition, AOI=Area of Interest, FFD=First Fixation Duration, FPD=First Pass Duration, SPD=Second Pass Duration, DT=Dwell Time)

	FFD		FPD		SPD		DT		Reg.	
	β(SE)	t	β(SE)	t	β(SE)	t	β(SE)	t	β(SE)	z
AOI-1 (non-immediate preverbal)										
Intercept	4.99 (0.04)	130.09 ***	4.85 (0.04)	108.04 ***	5.18 (0.08)	65.72 ***	5.14 (0.06)	93.14 ***	0.10 (0.09)	1.14
Word Length	-0.01 (0.00)	-2.18 *	-0.02 (0.01)	4.21 ***	0.03 (0.01)	2.93 **	0.05 (0.01)	8.64 ***	0.01 (0.01)	0.54
Foc	0.14 (0.02)	7.06 ***	0.27 (0.02)	11.28 ***	0.45 (0.04)	11.20 ***	0.47 (0.03)	18.19 ***	0.12 (0.06)	2.12 *
SP	0.05 (0.02)	2.46 *	0.09 (0.02)	3.61 ***	-0.17 (0.05)	-3.54 ***	-0.11 (0.03)	-4.15 ***	-0.16 (0.07)	-2.46 *
Foc × SP	-0.09 (0.03)	-2.99 **	-0.21 (0.03)	-6.07 ***	-0.08 (0.06)	-1.23	-0.12 (0.04)	-3.17 **	0.11 (0.09)	1.26
AOI-2 (immediate preverbal)										
Intercept	5.43 (0.04)	132.51 ***	5.22 (0.05)	100.09 ***	5.45 (0.11)	48.85 ***	5.33 (0.06)	86.56 ***	0.02 (0.19)	0.09
Word Length	-0.03 (0.00)	-7.25 ***	0.04 (0.00)	7.51 ***	0.03 (0.01)	2.40 *	0.06 (0.01)	10.59 ***	0.00 (0.02)	0.92
Foc	0.04 (0.02)	1.95 *	0.02 (0.02)	0.83	0.15 (0.06)	2.69 **	0.10 (0.03)	3.96 ***	0.07 (0.10)	0.63
SP	0.00 (0.02)	-0.15	-0.03 (0.02)	-1.36	-0.24 (0.06)	-3.83 ***	-0.11 (0.03)	-4.18 ***	0.08 (0.10)	0.79
Foc × SP	-0.02 (0.03)	-0.55	0.00 (0.03)	0.04	0.16 (0.09)	1.83	0.01 (0.04)	0.22	-0.04 (0.15)	-0.25

* p<0.05, ** p<0.01, *** p<0.001

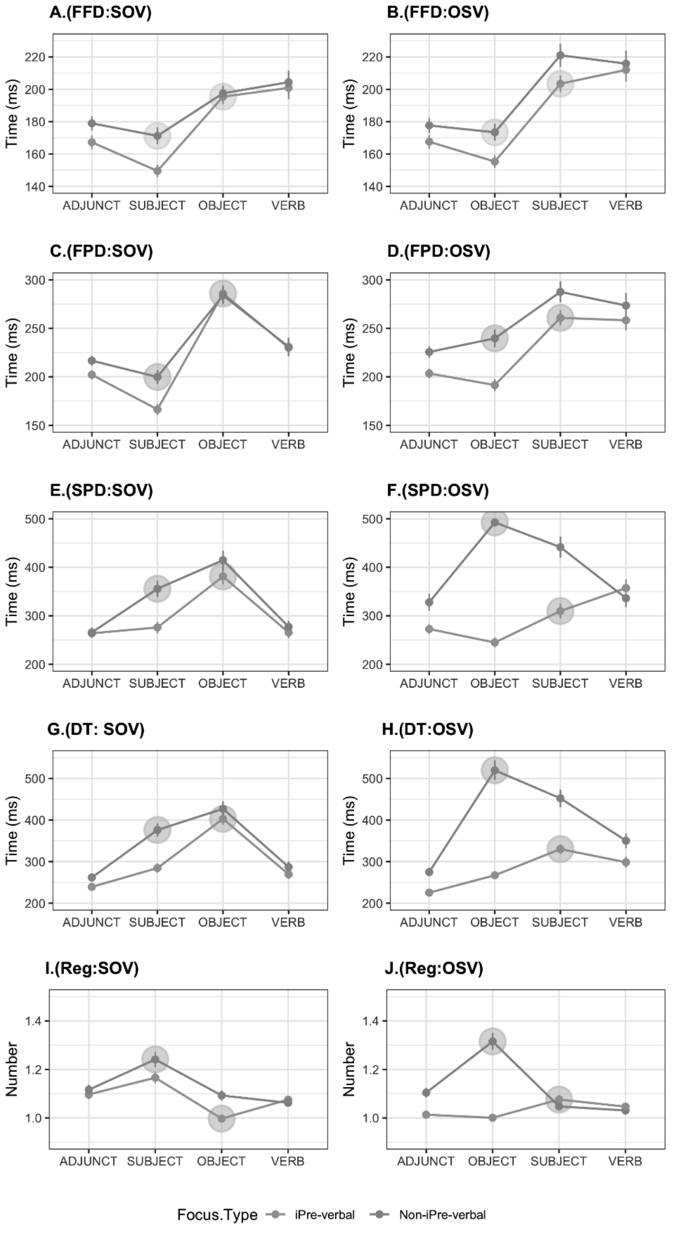


Figure 17. Per-region eye-movements data (fixation durations and regressions back into region) FFD=First Fixation Duration, FPD=First Pass Duration, SPD=Second Pass Duration, DT=Dwell Time, Reg=Regressions back into region. Object and Subject positions mark critical AOI. Dark shaded areas in figures demonstrate the position of focus.

3.1. AOI Region 1: Non-Immediate Pre-Verbal

A subsequent model computed with fixed effects of FOCUSPOSITION and SYNTACTICPOSITION for the non-immediate preverbal AOI data revealed significant effects of both factors on first fixation and first pass durations, as well as on second pass duration and dwell time (see Table 3). This evidence shows that focused elements in this region attracted higher fixation durations both early and late stages of processing compared to non-focused elements, and that object NPs received greater fixations than subject NPs (see Figure 1). We found significant interactions between these two factors, particularly in the early measures of processing (first fixation and pass durations) and dwell-time, but not in second-pass durations. A set of postdoc-tests confirmed that, compared to non-focused subjects, focused subjects attracted increased first fixation duration (β = -0.05, SE = 0.02, z = -2.67, p<.05, see Panel A in Figure 1), second-pass duration (β = -0.37, SE = 0.05, z = -7.99, p<.001, see Panel E in Figure 1), dwell-time (β = -0.35, SE = 0.03, z = -13.59, p<.001, see Panel G in Figure 1) and greater number of regressions back into this region (β = -0.23, SE = 0.06, z = -3.55, p<.01, see Panel I in Figure 1). There were no significant differences in first-pass durations between focused and non-focused subject in non-immediate preverbal region (β = -0.06, SE = 0.02, z = -2.34, p=.088, see Panel C in Figure 1). Similar patterns emerged for object NPs, focused objects received higher amount of first fixation duration (β = -0.14, SE = 0.02, z = -7.06, p<.001, see Panel B in Figure 1), first-pass duration (β = -0.27, SE = 0.02, z = -11.28, p<.001, see Panel D in Figure 1), second-pass duration (β = -0.45, SE = 0.04, z = -11.18, p<001, see Panel F in Figure 1), dwell-time (β = -0.47, SE = 0.03, z = -18.19, p<.001, see Panel H in Figure 1). There were no significant differences in number of regressions between focused and non-focused object in non-immediate preverbal region (β = -0.12, SE = 0.06, z = -2.12, p=.146, see Panel J in Figure 1). Greater reading disruptions in object NPs over subject NPs were visible in focused elements (β = -0.09, SE = 0.02, z = -3.61, p<.01 for FPD; β = 0.17, SE = 0.05, z = 3.53, p<.01 for SPD; β = 0.11, SE = 0.03, z = 4.15, p<.01 for DT) and non-focused elements (β = 0.13, SE = 0.02, z = 5.36, p<.001 for FPD; β = 0.25, SE = 0.04, z = 6.23, p<.001 for SPD; β = 0.23, SE = 0.03, z = 8.90, p<.001 for DT). However, for first fixation duration and regressions, there was no difference between object NPs and subject NPs in both focused (β = -0.05, SE = 0.02, z = -2.46, p=.067 for FFD; β = 0.16, SE = 0.07, z = 2.46, p=.066 for Reg) and unfocussed elements (β = 0.04, SE = 0.02, z = 1.96, p=.203 for FFD; β = 0.05, SE = 0.05, z = 0.98, p=.761 for Reg).

3.2. AOI Region 2: Immediate Pre-verbal

A set of mixed-effects regression models computed with the eye-movements data from the immediate preverbal region have shown significant effects of FOCUSPOSITION in first fixation duration (see Table 3). In second-pass duration and dwell-time, we found significant fixed-effects of FOCUSPOSITION and SYNTACTICPOSITION without an interaction between the two. An impact of FOCUSPOSITION indicates here non-focused elements caused reading disruptions (see Figure 1). This was true for subject NPs, in which fixation durations to non-focused subjects were greater than focused ones in second-pass duration (β = -0.31, SE = 0.06, z = -4.96, p<.001, see Panel F in Figure 1), and dwell-time (β = -0.11, SE = 0.03, z = -4.30, p<.001, see Panel H in Figure 1). This pattern was observed for object NPs only in second-pass durations (β = -0.15, SE = 0.06, z = -2.69, p<.05, see Panel E in Figure 1) and dwell-time (β = -0.10, SE = 0.03, z = -3.96, p<.001, see Panel H in Figure 1). However, there are no significant differences between focused and non-focused elements in the early processing measures for both subject (e.g., first-fixation duration: β = -0.02, SE = 0.02, z = -1.15, p=.659) or first-pass duration (β = -0.02, SE = 0.02, z = -0.89, p=.809) and object elements (e.g., first-fixation duration: β = -0.04, SE = 0.02, z = -1.95, p=.206) or first-pass duration (β = -0.02, SE = 0.02, z = -0.83, p=.841). In the immediate pre-verbal region, we observed a strong impact of SYNTACTICPOSITION on second-pass durations and total dwell-time. This was because object NPs in this region elicited longer fixations than subject NPs. Longer fixations to objects in this region were observed in focused elements (second-pass duration: β = 0.24, SE = 0.06, z = 3.82, p<.001; dwell-time: β = 0.11, SE = 0.03, z = 4.18, p<.001) but not in non-focused elements in second-pass duration (β = 0.08, SE = 0.05, z = 1.54, p=.415). In non-focused elements, we only observe an object-subject hierarchy in dwell-time: β = 0.10, SE = 0.03, z = 3.83, p<.001).

4. DISCUSSION

This study investigated the processing of pre-verbal focus in Turkish by using an eye-movement monitoring experiment during naturalistic reading with an analogous end-of-trial acceptance task. The experiment reported here was designed to examine potential effects of focus position and syntactic position. Our first aim was to unveil whether Turkish readers process sentence contexts with an immediate pre-verbal focus differentially in comparison to sentence contexts with non-immediate pre-verbal focus. Our second aim was to determine whether the syntactic position (i.e. subject and object) influence focus processing during naturalistic reading. Summarizing, our end-of-trial acceptance data have shown that focused-elements in the immediate pre-verbal position were judged to be more acceptable than non-immediate pre-verbal focus. Fixation duration data reflected a similar pattern with longer fixations in the non-immediate pre-verbal conditions than in the immediate pre-verbal conditions. Furthermore, non-focused subjects in the non-immediate pre-verbal condition led to greater reading disruptions as measured with increased fixation durations and greater number of regressions than focused subjects in the immediate pre-verbal condition. We observed a processing advantage when subjects scrambled to their syntactically non-canonical immediate pre-verbal position within the focus field to receive focus; this advantage was, however, not apparent for object NPs.

Regarding the first aim, the end-of-trial acceptance task data indicated that Turkish readers favored immediate preverbal focus more acceptable (both the iPre-Verb Subject and iPre-Verb Object conditions) as compared to non-immediate pre-verbal focus (i.e. the Pre-Verb Subject and Pre-verb Object conditions). Further, the eye-movement data indicated that in the non-immediate pre-verbal regions, focused elements attracted higher fixation durations in both early and late processing stages compared to non-focused elements. These findings seem compatible with previous studies that showed focused elements elicit longer response times and/or fixation durations than non-focused elements (e.g., Bader, 1998; Birch, & Rayner, 1997; Fodor, 1998; Patterson et al., 2007; Stolterfoht et al., 2007). An interpretation of our data reported here, suggesting that the immediate preverbal position is the preferred broad focus position in Turkish, seems fully reconcilable with Kurt and Deniz (2023). In the immediate pre-verbal region, however, interesting patterns emerged in that fixation durations increased in non-focused subjects than focused subjects in OSV sentences, while there was no impact of focus when the immediate pre-verbal argument was an object NP in SOV sentences (see Figure 1). In other words, the Turkish readers reported here took longer to read focused than non-focused elements in the non-immediate preverbal region, whereas focus processing was interrupted when a non-focused subject NP is positioned in the immediate pre-verbal region, compared to its focused counterpart. Following the FFH (Göksel & Özsoy, 2000), which holds that focus can be realized on any constituent within the entire pre-verbal area, we expected minimum or no reading disruptions when focus is placed in non-immediate preverbal regions. This is not what we found, however. Our findings indicate a clear disadvantage placing focus in non-immediate pre-verbal regions, as measured by reduced acceptance rates and elevated fixation durations, as compared to when focus is realized in the immediate preverbal position. However, although it is very likely that readers were not expecting non-focused constituents at the immediate preverbal position, it should be noted that processing disruptions within the preverbal area cannot be attributed to syntactic position of focus alone. Göksel and Özsoy (2000) posit that the entire preverbal area in Turkish is reserved for non-recoverable information, building upon that, we clearly observe that when contextually non-recoverable constituents, especially subject NPs, cause reading disruptions, in line with Göksel and Özsoy's focus field apporach, we will turn to this issue below.

Regarding our second aim, we found an impact of syntactic position on focus processing. In their end-of-sentence acceptance task responses, the Turkish skilled-readers favored sentences with focused-subjects as compared to focused-object NPs overall. However, focus position modulates Turkish readers' preferences for object/subject focused-constituents: for the immediate pre-verbal region a subject NP is preferred over an object NP (O[***S***]V>S[***O***]V) and this hierarchy is also present in the non-immediate pre-verbal regions ([***S***]OV>[***O***]SV). Fixation durations data showed that non-focused subjects received increased fixations compared to focused subjects during both first and second pass reading and greater number of regressions, suggesting that non-focus subjects lead to reading disruptions. Furthermore, fixation durations were similar for both non-focused and focused objects when objects are positioned canonically in the immediate pre-verbal region. It is therefore conceivable that Turkish readers are expecting a focused-element to occur in the immediate pre-verbal region, when this element is a subject NP, and hence moved to the immediate preverbal region to receive focus, Turkish readers judge this as felicitous with ceiling acceptance rates. In other words, non-focused subject readings were less preferred and associated with longer fixation durations compared to focused subjects. One would imagine that focused subject elements, denoting contextually non-recoverable new information, are default in Turkish. In fact, our finding that scrambled subjects within the focus field were rated as largely acceptable is also compatible with the account predicting the given information before new (e.g., Clifton & Frazier, 2004; Sauermann et al., 2013). However, we should mention that the given-before-new strategy holds in Turkish when subjects scramble to immediate pre-verbal position but not for focused objects, as we found no difference in almost all fixation duration measures, except for the second pass fixation durations, to focused and non-focused objects in the immediate preverbal region. The reason for the lack of differences in eye-movement measures to focused and non-focused objects in this neutral focus position is obviously stemming from syntactic canonicity effects of word order. That is, following the canonical word order in Turkish, the immediate pre-verbal position would naturally receive an object argument, and hence, placing the focus on the canonical object NP makes little or no difference, at least in naturalistic silent reading. Even if not directly tested in the current study, it is possible that when subjects move to the immediate preverbal position to receive focus, they are inherently marked through implicit prosody. The marked status of subject focus is on fact not unheard of, for instance, Kentner and Vasishth (2016) found that skilled-readers of German, another relatively flexible word order language, show increased fixation duration on objects NPs in subject focus conditions. This is fully reconcilable with our results here.

One particular finding worth further contemplating is that, regardless of word order, whether immediate or the non-immediate pre-verbal positions, focused objects elicited increased processing demands, especially in later stages of reading. This was observed by dwell time and longer second-pass durations for focused objects across both regions. Object-subject asymmetries are commonly observed in sentence processing research in general. However, a purely syntactic explanation based on word order effects may not fully account for the processing difficulty with focused objects. This is because, when the focus was not on the critical region, we did not observe significant differences for fixation durations between object and subject NPs in the immediate pre-verbal position. Following the cartographic approach to CP structure proposed in Rizzi (2004), one can claim that the [focus] features occur at the CP domain (Şener, 2019). Regardless of the discussion on whether or not the (syntactic) focus movement to CP exist in Turkish, a structural distance between Foc^0 and the original position of focused objects is longer than the structural distance between Foc^0 and the original position of focused subjects. Therefore, it is possible to propose that a subject focus is structurally closer to the head than object focus would be. This may explain why object focus imposes greater processing demands than subject focus, in line with the SDH. Although this explanation goes beyond the empirical scope of the current study, it suggests a theoretical framework for future research on the observed object–subject asymmetry.

A final point worth contemplation is that we report slightly differential time courses for focus and syntactic processing in our eye-movement data recorded at the immediate pre-verbal position. It was clear that the impact of focus position was significant in first fixation durations (i.e. early processing measures) while the effects of syntactic position only emerged in second fixation durations and total dwell times (i.e. late processing measures). This pattern suggests that in the immediate preverbal position, focus may be processed relatively early, possibly preceding syntactic re-analysis, while syntactic re-analysis and focus information seem to be rather synchronous in late processing stages. An early effect of focus may be affiliated with the fact that readers access implicit prosody rather quickly to parse focus information, however, syntactic position of focus requires a re-analysis stage during sentence processing.

5. CONCLUSION

Our findings from our eye-movement-monitoring during reading experiment suggest that Turkish readers are sensitive to focus marking and show a preference for focus to occur at the immediate pre-verbal region rather than elsewhere in the pre-verbal area, compatible with the given-before-new processing strategy. Overall, focused constituents required greater processing demands than non-focused constituents; however, non-focused subjects at the immediate preverbal region caused noticeable reading disruptions, particularly in later processing measures. We interpreted these findings as suggesting that within the preverbal area in Turkish, subject-focus tend to function as the default strategy, especially when aligned with the immediate pre-verbal position. In conclusion, our findings should be interpreted in the light of our certain methodological limitations, particularly on the reliance of written input. Future studies may address these constraints by adopting complementary approaches to prosodic focus marking in Turkish.

REFERENCES

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*(1), 28-55. https://doi.org/10.2307/417392

Ashby, J. & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89-B100. https://doi.org/10.1016/j.cognition.2004.12.006

Aydın, Ö., & Uzun, İ.P. (2023). Pupil dilation response to prosody and syntax during auditory sentence processing. *Journal of Psycholinguistic Research* 52, 153–177. https://doi.org/10.1007/s10936-021-09830-y

Atasoy, A. (2022). *Production, perception, and processing of focus in Turkish* (Doctoral dissertation). University of Potsdam. Retrieved from https://doi.org/0.25932/publishup-54815

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Springer.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition, 25*(5), 653-660. https://doi.org/10.3758/BF03211306

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter-versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 871. https://doi.org/10.1037/0278-7393.29.5.871

Braun, B., & Biezma, M. (2019). Prenuclear L*+ H activates alternatives for the accented word. *Frontiers in Psychology, 10*, 1993. https://doi.org/10.3389/fpsyg.2019.01993

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7-9), 1044-1098. https://doi.org/10.1080/01690965.2010.504378

Büring, D. (2010). Towards a typology of focus realization. In M. Zimmermann & M. Féry (Eds.), *Information structure* (pp. 177–205). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199570959.003.0008

Carlson, K., Dickey, M. W., Frazier, L., & Clifton, C. (2009). Information structure expectations in sentence comprehension. *Quarterly Journal of Experimental Psychology, 62*(1), 114-139. https://doi.org/10.1080/1747021070188017

Chafe, W. (1988). Punctuation and the prosody of written language. *Written communication, 5*(4), 395-426. https://doi.org/10.1177/0741088388005004001

Clifton, C., & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory & Cognition, 32*(6), 886-895. https://doi.org/10.3758/BF03196867

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech, 40*(2), 141-201. https://doi.org/10.1177/002383099704000203

Deniz, N. (2025). Sentence processing in Turkish: A review and future directions. *Lingua*, 313, 103835, https://doi.org/10.1016/j.lingua.2024.103835

Dinçtopal-Deniz, N., & Fodor, J. D. (2019). Timing of Syntactic and Rhythmic Effects on Ambiguity Resolution in Turkish: A Phoneme Restoration Study. *Language and speech*, 0023830919894614. https://doi.org/10.1177/0023830919894614

Erkü, F. (1983). *Discourse pragmatics and word order in Turkish* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Fodor, J. D. (1998). Learning to parse? *Journal of psycholinguistic research, 27*(2), 285-319. https://doi.org/10.1023/A:1023258301588

Fodor, J. D. (2002, April). *Psycholinguistics cannot escape prosody*. Paper presented at the Speech Prosody 2002 International Conference, Aix-en-Provence, France.

Göksel, A. (2013). Flexible word order and anchors of the clause. *SOAS Studies in Linguistics, 16*, 3-25.

Göksel, A., & Özsoy, A. S. (2000). Is there a focus position in Turkish? In A. Göksel & C. Kerslake (Eds.), *Studies on Turkish and Turkic languages* (pp. 219–228). Wiesbaden: Harrassowitz.

Gotzner, N., & Spalek, K. (2019). The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass, 13*(2), e12310. https://doi.org/10.1111/lnc3.12310

Gross, J., Millett, A. L., Bartek, B., Bredell, K. H., & Winegard, B. (2014). Evidence for prosody in silent reading. *Reading Research Quarterly, 49*(2), 189-208. https://doi.org/10.1002/rrq.67

Güneş, G. (2013). *On the role of prosodic constituency in Turkish*. In U. Özge (Ed.), Proceedings of the WAFL8 MIT Working Papers in Linguistics. Cambridge, MA.

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science, 366*(6461), 55-58. https://doi.org/10.1126/science.aax0289

Halliday, M. A. (1967). Notes on transitivity and theme in English: Part 2. *Journal of linguistics, 3*(2), 199-244. https://doi.org/10.1017/S0022226700016613

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 244-285. https://doi.org/10.2307/417261

Hruska, C., Alter, K., Steinhauer, K., & Steube, A. (2000). Can wrong prosodic information be mistaken by the brain? *Journal of cognitive neuroscience*(Suppl. S), 122.

İşsever, S. (2003). Information structure in Turkish: the word order–prosody interface. *Lingua, 113*(11), 1025-1053. https://doi.org/10.1016/S0024-3841(03)00012-3

İşsever, S. (2019). On the ban on postverbal wh phrases in Turkish: A syntactic account. In S. Özsoy (Ed.), *Word order in Turkish* (pp. 67–90). Springer.

Jackendoff, R. S. (1972). Semantic interpretation in generative grammar. MIT Press.

Kamali, B., & Krifka, M. (2020). Focus and Contrastive Topic in Questions and Answers, with particular reference to Turkish. *Theoretical Linguistics, 46*(1-2), 1-71. https://doi.org/10.1515/tl-2020-0001

Kennelly, S. D. (2003). The implications of quantification for the role of focus in discourse structure. *Lingua, 113*(11), 1055-1088.

Kentner, G., & Vasishth, S. (2016). Prosodic focus marking in silent reading: effects of discourse context and rhythm. *Frontiers in Psychology, 7*, 319. https://doi.org/10.3389/fpsyg.2016.00319

Kiss, K. É. (1998). Identificational focus versus information focus. *Language, 74*(2), 245-273. https://doi.org/10.2307/417867

Kitagawa, Y., Tamaoka, K., & Tomioka, S. (2013). Prosodic matters in intervention effects in Japanese: An experimental study. *Lingua, 124*, 41-63. https://doi.org/10.1016/j.lingua.2012.01.006

Krifka, M. (1992). A compositional semantics for multiple focus constructions *Informationsstruktur und grammatik* (pp. 17-53): Springer.

Kural, M. (1992). *Properties of scrambling in Turkish* (Unpublished manuscript). University of California, Los Angeles.

Kurt, D., & Deniz, N. D. (2023). Processing focus in Turkish. *Languages*, *8*, 38. https://doi.org/10.3390/languages8010038

O'Grady, W. (1999). Toward a new nativism. *Studies in Second Language Acquisition*, 621-633. https://www.jstor.org/stable/44486832

Özge, U., & Bozşahin, C. (2010). Intonation in the grammar of Turkish. *Lingua, 120*(1), 132-175. https://doi.org/10.1016/j.lingua.2009.05.001

Paterson, K. B., Liversedge, S. P., Filik, R., Juhasz, B. J., White, S. J., & Rayner, K. (2007). Focus identification during sentence comprehension: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 60*(10), 1423-1445. https://doi.org/10.1080/17470210601100563

R Core Team. (2025). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading* (2nd ed.). Psychology Press.

Rizzi, L. (2004). Locality and left periphery. *Structures and beyond: The cartography of syntactic structures, 3*, 223-251.

Rochemont, M. S., & Culicover, P. W. (1990). *English focus constructions and the theory of grammar* (Vol. 52). Cambridge University Press.

Rooth, M. (1992). A theory of focus interpretation. *Natural language semantics, 1*(1), 75-116. https://www.jstor.org/stable/23748778

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110-114.

Sauermann, A., Filik, R., & Paterson, K. B. (2013). Processing contextual and lexical cues to focus: Evidence from eye movements in reading. *Language and Cognitive Processes, 28*(6), 875-903. https://doi.org/10.1080/01690965.2012.668197

Savill, N., Lindell, A., Booth, A., West, G., & Thierry, G. (2011). Literate humans sound out words during silent reading. *NeuroReport, 22*(3), 116-120. https://doi.org/10.1097/WNR.0b013e328342ed27

Schafer, A., Carlson, K., Clifton, C., & Frazier, L. (2000). Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and speech, 43*(1), 75-105. https://doi.org/10.1177/00238309000430010301

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Blackwell.

Selkirk, E. (2002, April). Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English. Paper presented at the Speech Prosody 2002, International Conference, Aix-en-Provence, France.

Stolterfoht, B., & Bader, M. (2004). Focus structure and the processing of word order variations in German. *Information structure: Theoretical and empirical aspects*, 259-275.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition, 104*(3), 565-590. https://doi.org/10.1016/j.cognition.2006.08.001

Şener, S. (2019). Eliminating Scrambling: The variation of word order in Turkish. In S. Özsoy (Ed.), *Word Order in Turkish* (pp. 91-138): Springer.

Taylan, E. E. (1984). *The function of word order in Turkish grammar*. Berkeley: University of California Press.

Uzun, İP., Aydın, Ö., Kalaycıoğlu, C., & Ergenç, İ. (2021). Prosodic focus marking in Turkish: An electrophysiological study. *Studies in Psychology*, 41(1), 331–364. https://doi.org/10.26650/SP2020-0065

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics, 34*(3), 459-520. https://doi.org/10.1515/ling.1996.34.3.459

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H. *Cognitive Science, 32*(7), 1232-1244. https://doi.org/10.1080/03640210802138755

Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica (Since 2017 Acta Linguistica Academica), 55*(3-4), 347-360. https://doi.org/10.1556/aling.55.2008.3-4.9

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua, 121*(11), 1651-1670. https://doi.org/10.1016/j.lingua.2011.06.002

Zora, H., Kabak, B., & Hagoort, P. (2025). Relevance of prosodic focus and lexical stress for discourse comprehension in Turkish: Evidence from psychometric and electrophysiological data. *Journal of Cognitive Neuroscience*, 37(3), 693-736. 10.1162/jocn_a_02262.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order*: MIT Press.

Chapter 9:
The Role of Syntactic and Semantic Information in Cross-Linguistic Processing: A Translation Priming Study

Orhan Demir

ABSTRACT

In this chapter, I investigated how syntactic and semantic information jointly influence translation choices in bilingual processing by employing a structural priming paradigm with Turkish-English bilinguals, varying verb type and syntactic structure under short and long stimulus onset asynchrony conditions. My results revealed clear structural priming effects, demonstrating that participants were influenced by both syntactic repetition and syntax–semantics interactions; high-proficiency individuals showed greater sensitivity to grammatical constraints, while less experienced participants exhibited strong facilitation based on surface similarity. Reaction time patterns supported a form-before-meaning account, and translation experience emerged as a key modulator, with experienced translators being less influenced by prime's structural properties, suggesting reliance on entrenched production routines. These findings reinforce the view that translation priming offers a naturalistic window into bilingual structural alignment and cognitive control, underscoring its potential to investigate the interaction of linguistic knowledge, processing dynamics, and bilingual experience, and suggesting future studies diversify structural phenomena and language pairs, treat proficiency and experience as continuous variables, and integrate neurocognitive methods to capture processing time course.

Keywords: structural priming, translation, congruency, reaction time

1. INTRODUCTION

Priming refers to the cognitive phenomenon whereby prior exposure to a stimulus facilitates the processing of a subsequent, related stimulus, typically without conscious awareness (Trofimovich & McDonough, 2011). In psycholinguistics, priming has proven to be a powerful experimental tool, offering insights into the implicit and automatic mechanisms that support language comprehension and production. It is particularly valued for its methodological versatility, as it enables researchers to infer the structure and dynamics of linguistic knowledge in real-time, often in more ecologically valid contexts than traditional (offline) measures.

Over the past several decades, we see that priming has been productively employed across a range of linguistic domains. In phonetics, for example, exposure to specific speech sounds has been shown to influence subsequent articulation, revealing processes of phonetic convergence (e.g., Goldinger, 1998). In the morphological domain, priming studies were observed to have provided evidence on how complex words—whether affixed or compound—are organized in the mental lexicon (Marslen-Wilson et al., 1994; Rastle et al., 2000, 2004). Similarly, semantic priming has been instrumental in uncovering how related concepts become co-activated during lexical access, shedding light on the structure of conceptual memory (Neely, 1991).

For a couple of decades, beyond these domains, priming has become increasingly prominent in second language (L2) research. Because L2 users often exhibit distinct processing profiles from native speakers, priming paradigms are particularly well-suited to revealing subtle effects related to proficiency, developmental stage, or cross-linguistic transfer. For instance, a number of studies in semantic priming have shown how bilinguals activate meaning across languages (e.g., Jiang, 1999; Schoonbaert et al., 2007), while morphological priming has helped clarify how L2ers store and process morphologically complex words (e.g., Diependaele et al., 2005; Silva & Clahsen, 2008). Besides, phonological priming studies have also revealed that L2 learners' awareness/sensitivity to sound patterns plays a crucial role in lexical access and word recognition, particularly when the L1 and L2 phonologies diverge (e.g., Gor, 2018). Collectively, such findings have contributed to a more nuanced picture of how linguistic knowledge is structured and accessed in bilingual minds.

Among the various forms of priming, *structural priming*—the tendency to reuse a syntactic structure after recent exposure—has emerged as a particularly robust and theoretically productive phenomenon. Since the seminal work of Bock (1986), structural priming has been used to argue for the psychological reality of abstract syntactic representations. Through literature in L2, we see the compelling evidence that non-native speakers can form and rely on abstract syntactic knowledge (e.g., Hartsuiker et al., 2008), especially when the L2 shares structural properties with the native language. Structural priming thus offers a window into core questions in second language acquisition (SLA): How abstract are syntactic representations in L2? To what extent do they resemble those of native speakers? How do learners balance syntactic form with meaning during real-time processing?

Despite its strengths, structural priming research—especially in the context of L2—faces important theoretical and methodological challenges. Chief among these is the tendency to treat syntax as an isolated, autonomous module. However, we argue that sentence processing is a richly interactive process, and accumulating evidence indicates that structural priming is influenced by other levels of linguistic representation, particularly semantics. Factors such as verb meaning, thematic roles, and contextual plausibility can shape whether and how priming occurs. In other words, syntactic repetition does not happen in a vacuum; it is tightly coupled with meaning.

Several studies highlight this interdependence. For instance, Hare et al. (2004) showed that event knowledge activated during verb comprehension affects syntactic expectations. Similarly, Gracanin-Yüksek et al. (2017) demonstrated that verb semantics and contextual fit modulate syntactic processing in both native and non-native speakers. These findings challenge models that posit a purely abstract syntactic level of representation and call for a more integrated account—one that incorporates both syntactic and semantic constraints.

In L2 contexts, this integration becomes even more critical. L2 users are assumed to often rely more heavily on semantic cues due to reduced syntactic automaticity, and their structural choices may reflect a complex negotiation between source and target language norms. Several studies have shown that priming in L2 is not merely a function of frequency or exposure, but also of semantic congruence and morphosyntactic compatibility across languages (e.g., Hopp, 2015). Accordingly, we can make an inference that understanding structural priming in L2 learners demands a methodological shift—one that captures the interplay between form and meaning.

One promising avenue for such research is *translation priming*. Translation tasks inherently involve the mapping of meaning from one language to another, requiring speakers to make syntactic decisions under semantic constraints. Because the intended message remains constant across languages, translation provides a controlled yet naturalistic method for observing how L2 speakers negotiate cross-linguistic syntactic options. We should especially point out that it allows researchers to ask how syntactic preferences are shaped when form must align with meaning—both within and across languages.

The remainder of this chapter proceeds as follows. First, I provide a brief overview of the interaction between syntactic and semantic information in sentence processing, with a particular emphasis on how this interaction has been conceptualized in psycholinguistic models. This theoretical background sets the stage for a discussion of structural priming, particularly in the context of L2 acquisition, and highlights how semantic features—such as verb class and event structure—can modulate priming effects. Then I introduce translation as a methodological tool to investigate these phenomena, arguing that it offers a naturalistic yet controlled means of tapping into L2 structural preferences under semantic constraints. Following this, the methodological framework of the present study is described, including participant selection, materials, and design. The chapter concludes with a detailed presentation of the results and a discussion of their implications for current models of L2 sentence processing and syntactic alignment.

1.1. Beyond Modularity: Syntax–Semantics Integration in Psycholinguistic Models

The interaction between syntax and semantics is a foundational topic in psycholinguistics as it is central to understanding how humans comprehend language. Historically, generative linguistic theories defended a modular architecture, where syntactic processing operated independently before feeding into semantic interpretation (e.g., Chomsky, 1965). However, contemporary psycholinguistic frameworks which we frequently encounter in the literature increasingly advocate for an interactive architecture that integrates syntactic and semantic information in real time during sentence processing (MacDonald et al., 1994; Trueswell et al., 1994). Constraint-based models, for instance, propose that language users simultaneously evaluate multiple syntactic possibilities, with semantic plausibility, discourse context, and lexical biases probabilistically contributing to disambiguation (Pickering et al., 2001). Within this interactive view, argument structure knowledge and thematic role expectations play a special role. Verbs inherently carry rich semantic and syntactic constraints that guide the parser in creating plausible interpretations even when it has to deal with structural ambiguity. While specific theoretical and computational models differ in their implementation, we see the broad consensus that syntactic and semantic information work together in continuous mutual constraint.

In first language (L1) sentence processing, a number of empirical studies support the interdependence of syntactic and semantic processing mechanisms. Classic garden-path sentences, such as *The horse raced past the barn fell*, demonstrate how the parser starts to work with a syntactic interpretation that may later conflict with semantic expectations, requiring costly reanalysis (Frazier & Rayner, 1982; Ferreira & Clifton, 1986). Similarly, thematic fit effects provide substantial evidence for real-time semantic involvement: verbs like *arrest* or *examine* evoke strong role expectations that shape syntactic assignments (McRae et al., 1998). Findings from ERP studies further support this interaction. Semantic anomalies typically elicit an N400 component, while syntactic violations trigger P600 effects; in some cases, both emerge, indicating a layered response to conflict or ambiguity (Kuperberg et al., 2003). There are additional factors such as concretness, animacy, frequency, and world knowledge that are claimed to systematically influence syntactic decisions in areas like attachment ambiguity and pronoun resolution (Altmann & Kamide, 1999). Based on these findings, we can conclude that L1 sentence processing involves a highly adaptive system that integrates multiple information streams probabilistically and dynamically. Computational models based on constraint satisfaction (e.g., Spivey & Tanenhaus, 1998) capture this integration well and show how candidate parses are continuously evaluated based on syntactic fit and semantic plausibility.

In second language (L2) sentence processing, several learner-specific variables such as proficiency, L1 transfer, and cognitive processing capacity are considered to be the determining factors of the dynamics of syntax–semantics interaction. A consistent finding is that L2 learners—especially those at lower proficiency levels—tend to rely more heavily on semantic cues when syntactic structures are complex or unfamiliar. This asymmetry has been formalized in *the Shallow Structure Hypothesis*, which argue that L2 learners often build less detailed syntactic representations and instead benefit from semantic or lexical information (Clahsen & Felser, 2006). For example, learners may misinterpret agent-patient relations in sentences with embedded relative clauses, such as *The doctor that the patient thanked was empathetic*, due to difficulty in tracking long-distance dependencies (Papadopoulou, 2005). Moreover, L2 learners have been observed to show enhanced sensitivity to semantic cues like animacy and favor semantically plausible interpretations over grammatically accurate ones (Dussias & Cramer Scaltz, 2008). Despite these differences, proficiency plays a crucial role: as L2 users gain experience, they begin to demonstrate native-like syntactic sensitivity, especially when lexical-semantic mappings are well established (Hopp, 2010). Moreover, research on how bilinguals understand sentences shows that both their first and second languages can be active at the same time and affect each other during reading or listening (Weber & Broersma, 2012). This leads us to the argument that a shared but adaptable architecture is needed for syntactic-semantic integration. Understanding these processes in L2 speakers provides critical insight into the flexibility of the language system and the constraints imposed by experience and learning.

1.2. Structural Priming as a Tool for Understanding Syntax

Structural priming—the tendency for speakers to repeat syntactic structures across utterances—has become a cornerstone methodology in psycholinguistics for uncovering the architecture of grammatical knowledge. Initially demonstrated in production by Bock (1986), who showed that exposure to a passive sentence increases the likelihood of producing another passive, the effect has since been replicated in comprehension using self-paced reading, eye-tracking, and neuroimaging (e.g., Tooley & Traxler, 2010; Segaert et al., 2012). These findings support the claim that priming reflects the activation of abstract syntactic representations, not merely superficial repetition. As Tooley (2023, p. 882) notes, structural priming allows researchers to probe the cognitive links that enable fluent language use—links between syntactic templates, conceptual structure, and lexical information. In addition to informing linguistic theory, structural priming research has implications for language acquisition, processing efficiency, and even long-term grammatical change (Pickering & Ferreira, 2008).

In L2 contexts, structural priming offers a powerful lens through which to examine how grammatical knowledge develops and interacts across languages. A central concern has been the extent to which L2 learners share syntactic representations with their L1. Results are mixed: While some studies report fully shared representations across L1 and L2 (e.g., Desmet & Declerq, 2006; Schoonbaert et al., 2007; Kantola & Van Gompel, 2011), others emphasize constraints introduced by lexical overlap (Hartsuiker et al., 2008), proficiency (Bernolet et al., 2013), and semantic equivalence (Cai et al., 2011). The presence or absence of a lexical boost—stronger priming when prime and target share verbs—has emerged as a key diagnostic for representational overlap. Structural priming in L2 thus serves as a crucial tool not only for evaluating syntactic convergence between languages, but also for testing broader models of bilingual memory architecture (see van Gompel & Arai, 2018 for review). Importantly, structural priming does not operate in isolation; it often interacts with morphosyntactic complexity, processing limitations, and semantic interpretation, all of which are magnified in bilingual settings.

Given the unique ability of structural priming paradigms to reveal implicit grammatical knowledge, it is natural to explore whether they can be extended beyond traditional production and comprehension tasks. In particular, translation—the dynamic act of rendering meaning across languages in real time—presents an ideal context for testing structural alignment and cross-linguistic influence. The next section turns to this novel application of priming, introducing *translation priming* as a methodological bridge between psycholinguistics and bilingual language use.

1.3. Translation Priming: A Methodological Bridge to Bilingual Processing

In recent years, the structural priming paradigm has been applied to translation, offering insights into how bilinguals manage grammatical mappings between languages. By using the source sentence as a prime and the translation as a target, translation priming adapts structural priming to a bilingual context, highlighting how underlying syntactic structures influence the translation process in real time. While early research focused on lexical repetition and cognate effects (e.g., Finkbeiner et al., 2004; van Hell & de Groot, 2008), more recent studies have investigated the syntactic aspect. For instance, Jensen et al. (2009) observed that incongruent word order between source and target languages affected eye movements, suggesting increased processing costs when syntactic restructuring is necessary. These findings challenge the notion of translation as solely a lexical or semantic activity.

Similarly, Schaeffer and Carl (2013) demonstrated that syntactic congruency impacts both form and meaning, reinforcing that translation is influenced by structural alignment.

Maier et al. (2017) conducted an influential study using a speeded translation paradigm, showing that bilinguals, particularly highly proficient ones, tend to maintain the syntactic structure of the source sentence, regardless of translation direction (L1 to L2 or vice versa). This persistence, especially when both languages allow the same structure, supports the horizontal view of translation, where source-language syntax remains active during target-language production (cf. Ruiz et al., 2008). However, Jacob et al. (2024) recently re-examined these assumptions using online methods, finding limited support for traditional structural priming and instead proposing a model based on serial lexical co-activation. This presents a critical question: are observed syntactic echoes truly structural, or are they a result of lexically driven activation? Regardless, the consistent occurrence of such effects in translation tasks warrants further investigation.

Building on this, Demir and Uygar (2023) extended the structural priming framework to English and Turkish, a typologically diverse language pair, adapting Maier et al.'s (2017) speeded translation task. Their design addressed the ongoing debate between horizontal and vertical translation models. By manipulating stimulus onset asynchrony (SOA: 1.4s vs. 3s) and considering individual differences like L2 proficiency and translation experience, they found robust priming effects for both double-object (DO) and prepositional-object (PO) dative structures. These effects were stronger at shorter SOAs, supporting the horizontal account. However, experienced translators showed a preference for DO structures and processing delays when translating PO primes, suggesting an internalized, strategy-based approach more aligned with the vertical model. Reaction time analyses confirmed that repeating syntactic structures facilitated faster processing. These results indicate that translation priming is not merely an abstract laboratory phenomenon but is linked to real-world experience and cognitive timing.

Despite these advancements, the translation priming paradigm remains underutilized. Many existing studies have small sample sizes, limited variable control, or rely on offline methods. Nevertheless, its potential is significant. Translation priming can reveal how comprehension and production interact in bilingual minds and reflects real-world translation decisions. Furthermore, it can be combined with fine-grained methods like eye-tracking or EEG to explore the timing and localization of syntactic effects, opening new avenues for investigating translation expertise, syntactic transfer, and broader bilingual grammatical processing. While Demir and Uygar (2023) sought how typological distance, timing, and experience influence priming effects, the current focus shifts to the interplay of syntax and semantics. Specifically, the questions posed are: How do meaning and form interact during real-time translation? Can semantic roles modulate syntactic repetition in a way that reflects deeper structural mappings across languages? These inquiries guide the exploration for a more complete understanding of how bilinguals manage the complex task of translating between languages.

2. THE PRESENT STUDY

In this study, my aim is to investigate how the meaning of verbs influences the grammatical structure chosen during translation from Turkish to English, specifically focusing on the interplay between structural priming and target language rules. While Turkish commonly uses prepositional-like structures for ditransitive events and has flexible verb usage, English is stricter. For instance, I note that English *pure-transfer verbs* like *give* and *send* allow both double object (DO) and prepositional object (PO) constructions (e.g., *She gave him the book* vs. *She gave the book to him*). However, *manner-of-transfer verbs* such as *carry* and *push* are generally restricted to PO constructions (e.g., *She carried the bag to him*), with the DO alternative (*She carried him the bag*) being considered ungrammatical.

In this experiment, Turkish sentences serve as primes, and participants are instructed to translate them into English. Because Turkish does not mark the DO–PO distinction morphosyntactically, a sentence like *Kadın çantayı ona verdi* (*The woman gave the bag to him*) can be translated into either DO or PO constructions in English, depending on lexical and contextual factors. Likewise, *Kadın çantayı ona taşıdı* (*The woman carried the bag to him*) has only one grammatical equivalent in English—the PO form—since the DO alternative (*She carried him the bag*) is not licensed by the semantic properties of the verb *carry*. This design creates a context in which the structure of the Turkish prime may either align with or conflict with the syntactic constraints imposed by the English target verb.

The experiment is designed to test whether structural priming effects originating from Turkish primes influence syntactic choices in English translation, even when the primed structure is incompatible with the lexical properties of the English verb. In congruent conditions, the syntactic structure implied by the Turkish prime and the English verb's argument structure are aligned (e.g., a PO-like Turkish structure followed by a PO-only verb such as *carry*). In incongruent conditions, the Turkish prime encourages a DO-like rendering, but the target verb in English does not permit DO constructions (e.g., *taşıdı* translated with *carry*). These conditions allow us to assess whether participants are guided by structural priming from the source language, or whether they adapt their syntactic choices based on the grammatical requirements of the target language.

The study also considers the role of individual differences, specifically L2 English proficiency and translation experience, as potential modulators of the priming effect. Prior research suggests that bilinguals with higher proficiency and more translation experience are more sensitive to the grammatical and semantic constraints of the target language and may therefore be more likely to override structurally primed forms when they are not appropriate. In contrast, less proficient or less experienced participants may be more susceptible to syntactic persistence, even at the expense of grammatical well-formedness. By investigating the interaction between source-language priming and target-language constraints, this study aims to clarify how bilinguals manage competing demands during translation. It contributes to a more nuanced understanding of structural priming in bilingual contexts and offers empirical evidence on how semantic features of verbs influence syntactic realization across languages with differing grammatical systems. To investigate these issues, the study addresses the following research questions:

- Does structural priming from Turkish influence syntactic choices in English translation, depending on the verb's semantic constraints?
- To what extent do L2 proficiency and translation experience modulate the interaction between structural priming and verb semantics in translation?

2.1. Participants

A total of 86 Turkish-speaking learners of English participated in the study. Six participants took part only during the piloting phase of the translation task. The majority (n = 80; 46 female, 34 male) completed all components of the experiment, including a multiple-choice cloze test and a background questionnaire used to assess English proficiency.

The participants ranged in age from 18 to 36 years, with a mean age of 24.1 (SD = 4.1), and all volunteered to take part in the study. All reported acquiring Turkish as their first language from birth and being introduced to English at school, indicating a shared L1–L2 acquisition profile. Across the entire group, self-reported use of English in daily life averaged 37.8%, suggesting lower usage of L2 compared to their dominant language, Turkish. Of the 80 participants in the main study, 42 were professional translators with specialized academic training in translation, while the remaining 38 were bilinguals without professional translation experience. These untrained participants, similar to those described in Maier et al. (2017), were either students or graduates from a variety of academic programs, representing a broad range of disciplinary backgrounds. When grouped by translation experience, professional translators reported slightly higher English usage (40.9%) than their untrained counterparts (38.7%), though both groups showed similar overall profiles in terms of L2 exposure.

Participants' proficiency levels were assessed through both subjective and objective measures. Self-reported data on overall English proficiency and daily language use were collected via a linguistic background questionnaire, and a standardized cloze test developed by Ionin and Montrul (2010) was administered. Equal weight was assigned to both sources of data. The total proficiency score was calculated on a 100-point scale: 50 points were allocated to the cloze test, 30 points to self-rated overall proficiency, and 10 points each to self-reported frequency of English use and comfort with comprehension and production. Composite scores were then computed for each participant. Using a median-split procedure, participants with scores above 82 were classified as highly proficient (n = 41), while those scoring 82 or below were categorized as low proficient (n = 39).

2.2. Materials and Design

The experimental materials were developed to examine how the semantic properties of verbs constrain syntactic realization during translation from Turkish into English, particularly in the context of structural priming. The item set included 24 critical Turkish sentences, all featuring ditransitive events and structured to resemble either double object (DO) or prepositional object (PO) constructions in English. These items were evenly divided based on verb class: 12 sentences contained pure-transfer verbs (e.g., *vermek* "to give", *göndermek* "to send"), which are compatible with both DO and PO constructions in

English (e.g., *She gave him the book* / *She gave the book to him*), while the other 12 included manner-of-transfer verbs (e.g., *taşımak* "to carry", *itmek* "to push"), which are typically only compatible with PO constructions in English (e.g., *She carried the bag to him*, but not *She carried him the bag*).

Each Turkish sentence was designed to be semantically and pragmatically neutral with regard to animacy, definiteness, and givenness, to ensure that structural choices in the English translations would be driven primarily by priming effects and verb semantics. The sentences ranged from 4 to 6 words in length (M = 4.8, SD = .9) and included both masculine and feminine referents to avoid gender bias. The Turkish source sentences were created in a prepositional-like form (e.g., *Kadın çantayı ona verdi*) to simulate the PO construction without making the DO–PO distinction morpho-syntactically explicit, as Turkish lacks this grammatical opposition.

To obscure the focus on ditransitive structures and reduce response strategies, 36 filler sentences were added. Half of these featured transitive verbs (e.g., *Adam kapıyı kapattı*) and the other half included intransitive verbs (e.g., *Çocuk ağladı*). The full set of 60 sentences, including both critical items and fillers, was pseudo-randomized to prevent semantic or structural overlap between successive items and to avoid consecutive presentation of more than two critical sentences from the same verb condition. This ensured that priming effects would not be driven by superficial associations or predictable patterns in the input stream.

To examine the temporal dynamics of priming effects, the 20 critical items were presented under two different stimulus onset asynchrony (SOA) conditions: a short SOA of 1.4 seconds and a long SOA of 3 seconds. This manipulation was intended to probe whether priming effects emerge more strongly in early or delayed stages of sentence planning during translation. Finally, to control for order effects and participant fatigue, a second version of the experiment was created with the item order reversed.

2.3. Procedure

Participants completed individual experimental sessions in a quiet, controlled environment. Stimuli, consisting of English sentences, were displayed in white 18-point Courier New font on a black background for optimal visibility. OpenSesame 4.0. (Mathôt et al., 2012) was used on a 15.6-inch laptop to present stimuli, precisely control timing, and record preparation times during the translation task.

Before starting, participants were instructed to translate English sentences into Turkish quickly and accurately. They were told sentences would appear briefly, and they should press the SPACEBAR when ready to translate and again to proceed. Each trial began with a fixation cross, then the sentence appeared for either 1.4 or 3 seconds, based on the SOA condition. Three practice trials ensured familiarity. A short break was given mid-session to prevent fatigue. The 25-minute experiment concluded with participants signing a consent form.

3. RESULTS

3.1. Translation Outcomes

Before the statistical analyses, a data cleaning procedure was applied to prepare the dataset for analysis. Following standard conventions in psycholinguistic research, trials were excluded if they contained translation errors, did not allow for a dative alternative, or had response times falling more than three standard deviations from a participant's mean. This process led to the removal of 10.2% of the data. To assess general task performance, participants' responses to unambiguous filler items were examined. Over 95% of these translations were accurate in terms of meaning, and approximately 92% retained the syntactic structure of the original English sentences. These results indicate that participants engaged with both the semantic and structural aspects of the stimuli during the speeded translation task.

In the initial phase of the analysis, participants' performance was examined irrespective of their proficiency levels or translation experience. Overall accuracy across all experimental items was relatively high, averaging 84.2%. However, a more fine-grained analysis based on verb type revealed significant differences tied to the semantic congruency between source and target constructions. For congruent items involving pure transfer verbs, participants demonstrated near-ceiling accuracy (97%) and showed strong alignment with the source text's syntactic structure, reproducing the ditransitive construction in 86.3% of the acceptable translations. In contrast, incongruent items—featuring manner-of-transfer verbs used with double object (DO) constructions in the source language—resulted in a substantial drop in performance. The overall accuracy in these cases declined to 64.5%, and participants deviated from the source ditransitive structure in 34.9% of the translations deemed acceptable, opting instead for alternative syntactic forms.

These differences were statistically confirmed by repeated-measures ANOVAs, which revealed a significant effect of verb type on both accuracy, ($F(1, 79) = 9.875$, $p < .001$), and structural adherence, ($F(1, 79) = 7.421$, $p < .001$).

To investigate the influence of temporal dynamics on translation performance, the data were analyzed with respect to stimulus onset asynchrony (SOA), comparing short (1.4 seconds) and long (3 seconds) intervals. Overall, a longer SOA led to significantly higher translation accuracy. Specifically, when the SOA was 1.4 seconds, the mean accuracy across all items dropped to 77.8%, whereas it increased to 90.5% at the 3-second interval. This difference was statistically significant, $F(1, 79) = 10.318$, $p < .001$. The effect of SOA was particularly pronounced in incongruent items, where semantic and syntactic mappings were misaligned. Under the short SOA condition (1.4 s), the mean accuracy for these items was 52.1%, while it rose to 74.5% when the SOA was extended to 3 seconds—a statistically robust difference, $F(1, 79) = 14.197$, $p < .001$. Interestingly, the translation priming effect, measured in terms of structural adherence to the source ditransitive form, appeared to be stronger at shorter SOAs. For congruent items, participants used the ditransitive construction in 79.4% of acceptable translations when the SOA was 1.4 seconds, compared to 92.0% at 3 seconds ($F(1, 79) = 7.471$, $p = .020$). A similar pattern emerged in incongruent items, where ditransitive usage decreased from 60.4% (SOA = 3 s) to 48.6% (SOA = 1.4 s), again reaching statistical significance, $F(1, 79) = 15.830$, $p < .001$. These results suggest that while longer preparation time enhances overall accuracy, shorter SOAs appear to strengthen the syntactic priming effect, particularly in structurally challenging or semantically incongruent contexts.

3.2. Translation Outcomes by Proficiency and Experience

To examine the potential impact of individual differences on translation accuracy, participants' performance was analyzed based on their language proficiency and translation experience. A significant effect of proficiency was observed, with high-proficiency participants outperforming their low-proficiency peers. The mean accuracy rate for the high-proficiency group was 93.0%, compared to 75.9% for the low-proficiency group, a difference that reached statistical significance, $F(1, 79) = 13.206$, $p < .001$. In contrast, translation experience did not yield a significant effect. Participants with prior translation experience demonstrated an average accuracy of 84.6%, while those without such experience averaged 82.8%. This difference was not statistically significant, $F(1, 79) = 0.751$, $p = .610$.

In terms of the priming effect, the pattern of results differed depending on both proficiency and the stimulus onset asynchrony (SOA). When acceptable translations were examined across proficiency levels, both high- and low-proficiency groups exhibited a priming effect, with no statistically significant difference between them. The proportion of primed responses was 68.6% in the high-proficiency group and 71.4% in the low-proficiency group, $F(1, 79) = 1.003$, $p = .841$. However, a more nuanced pattern emerged when SOA conditions were considered. Under the short SOA condition (1.4 seconds), the priming effect significantly differed by proficiency. The low-proficiency group produced more primed responses (74.5%) compared to the high-proficiency group (62.5%), a difference that reached statistical significance, $F(1, 79) = 4.163$, $p < .04$. In contrast, no significant difference was observed under the long SOA condition (3 seconds), with similar rates of primed responses in the low- (69.7%) and high-proficiency groups (71.6%), $F(1, 79) = 0.108$, $p = .830$.

When translation experience was taken into account, the data revealed a clear difference in priming effects between experienced and non-experienced participants. In both SOA conditions, participants without prior translation experience exhibited robust priming effects. Specifically, the proportion of primed responses was 79.0% in the short SOA condition (1.4 seconds) and 82.6% in the long SOA condition (3 seconds). In contrast, the experienced group showed considerably weaker priming effects, with only 63.5% primed responses under the short SOA and 57.4% under the long SOA. These differences between groups were statistically significant in both conditions: $F(1, 79) = 12.004$, $p < .001$ for the 1.4-second SOA, and $F(1, 79) = 18.935$, $p < .001$ for the 3-second SOA.

To examine potential interactions among the main variables, a repeated-measures ANOVA was conducted with priming (primed vs. unprimed), group (based on either proficiency or translation experience), and SOA (1.4 seconds vs. 3 seconds) as factors. The overall analysis revealed no significant three-way interaction among these variables. To further explore the data, separate repeated-measures ANOVAs were conducted including either proficiency or experience as between-subject factors. These analyses revealed a significant interaction between priming and experience, $F(1, 79) = 19.193$, $p < .001$, as well as between priming and SOA, $F(1, 79) = 13.280$, $p < .001$. These findings support the interpretation that the strength of the priming effect varied depending on participants' proficiency level and the duration of exposure to the source items.

3.3. Reaction Time Patterns

In the second analytic phase, attention was directed toward reaction time (RT) measures in order to assess whether the priming effects observed in participants' translation preferences also extended to temporal aspects of processing. Reaction time was operationalized as the duration between the initial display of the source sentence and the activation of the SPACEBAR, which signaled that the participant was ready to initiate the translation. Prior to analysis, the dataset was subjected to a series of standard preprocessing procedures. Instances were excluded if the translated output lacked a dative construction (either DO or PO) or contained content-related inaccuracies. Furthermore, RT values falling beyond ±2 standard deviations from each participant's mean were discarded to reduce the impact of statistical outliers. As a result of these steps, approximately 18% of the trials were removed from the dataset.

Descriptive statistics were then calculated, collapsing across individual differences, to capture general tendencies in response behavior. It was observed that RTs were shorter when the syntactic form of the prime sentence was retained in the translation, compared to when an alternative structure was produced. This facilitative effect appeared under both pure-transfer verbs and manner-of-transfer verbs. Additionally, shorter RTs were recorded in trials with longer stimulus onset asynchrony (SOA), indicating that extended exposure to the prime may have supported faster structural processing. These trends are summarized in the table below.

To examine whether the magnitude of the priming effect varied as a function of prime type, separate repeated-measures ANOVAs were conducted for each SOA condition. The analyses revealed no significant difference in the size of the accelerative effect between the two prime types in either SOA condition (SOA: 1.4 s, $F(1, 79) = 1.573$, $p = .35$; SOA: 3 s, $F(1, 79) = 0.648$, $p = .858$). In contrast, when the accelerative effect was compared across SOA conditions, a significant difference emerged. The priming effect was found to be significantly stronger at the shorter SOA, indicating that reduced exposure duration led to more pronounced facilitation ($F(1, 79) = 26.009$, $p < .001$).

Table 32: Mean Reaction Times by Priming Condition (Priming vs. No Priming)

SOA: 1.4 seconds			
	Priming RT	No Priming RT	Priming Effect
Prime: pure-transfer	1804 ms.	2151 ms.	347 ms.
Prime:manner-of-transfer	1860 ms.	2177 ms.	317ms.
SOA: 3 seconds			
	Priming RT	No Priming RT	Priming Effect
Prime: pure-transfer	1493 ms.	1688 ms.	195 ms.
Prime:manner-of-transfer	1465 ms.	1673 ms.	208 ms.

3.4. Reaction Time Analysis by Proficiency and Experience

Participants' reaction times were compared across groups defined by proficiency and translation experience to examine the general pattern of the priming effect. When participants were grouped by proficiency, both high and low proficiency individuals showed evidence of an accelerative priming effect, responding faster when the target translation was structurally similar to the prime. The magnitude of this facilitation did not differ significantly between the groups at either SOA. At SOA 1.4 s, the mean priming effect was 319 ms for high proficiency participants and 330 ms for low proficiency participants, $F(1, 79) = -2.217$, $p = .216$. At SOA 3 s, the corresponding values were 189 ms and 302 ms, respectively, $F(1, 79) = -1.740$, $p = .305$.

A different pattern was observed when participants were grouped based on translation experience. Inexperienced participants demonstrated a strong accelerative effect across both SOAs (SOA 1.4 s: 692 ms; SOA 3 s: 427 ms). In contrast, experienced participants exhibited an inhibitory effect (SOA 1.4 s: -55 ms; SOA 3 s: -39 ms), with significant group differences in both conditions, $F(1, 79) = 45.003$, $p < .001$; $F(1, 79) = 36.979$, $p < .001$ respectively. These results indicate that while inexperienced participants were highly sensitive to structural overlap between prime and target, experienced participants were not facilitated by structural similarity, suggesting a reliance on internalized translation strategies over form-based processing.

To examine potential interactions among variables, a series of repeated-measures ANOVAs were conducted with a 2 (prime type) × 2 (group: proficiency vs. experience) × 2 (SOA: 1.4 s vs. 3 s) design. No significant interaction effects were identified across these factors. However, when follow-up repeated-measures ANOVAs were conducted separately for each grouping variable, a main effect of translation experience emerged, $F(1, 79) = 28.361$, $p < .001$. This finding aligns with the earlier group-based comparisons, further supporting the conclusion that the magnitude and direction of the priming effect are modulated by participants' level of experience. Specifically, the degree of facilitation or inhibition appears to be influenced by the extent of participants' prior exposure to translation tasks and the duration of their engagement with the source language input.

4. DISCUSSION

This study investigated structural priming in cross-linguistic translation, revealing that participants were sensitive to both purely syntactic overlap and syntax-semantics interactions between source and target constructions. This indicates that a person's knowledge of sentence structure (syntactic knowledge) and how syntax interacts with meaning (interface-level knowledge) both influence translation choices. A key finding was that language proficiency affected sensitivity to grammatical rules. Highly proficient individuals were more likely to avoid grammatically incorrect double object (DO) constructions, suggesting a more refined syntactic awareness. Reaction time data supported this: participants responded faster when their translations structurally matched the source in cases of purely syntactic overlap, but responses were slower when semantic considerations were involved. Interestingly, when priming occurred despite semantic complexity, neither translation preferences nor reaction times differed, suggesting successful integration of form and meaning. Additionally, both stimulus onset asynchrony (SOA)—the time difference between prime and target presentation—and translation experience influenced the priming effect. Inexperienced participants showed a significant acceleration in their responses due to priming, while experienced participants exhibited an inhibitory pattern. This highlights the importance of temporal factors and long-term experience in understanding priming mechanisms during translation.

These findings align with existing psycholinguistic research on the timing of semantic effects in language processing. Several studies propose that structural processing precedes semantic integration, meaning that form-based priming effects appear earlier than meaning-based effects (e.g., Rastle et al., 2000, 2004). My findings support this by showing that purely syntactic overlap led to more immediate facilitation compared to cases requiring deeper semantic processing. The delayed influence of semantic information in translation mirrors the "form-then-meaning" trajectory often observed in studies of a single language, now extended to bilingual translation.

Finally, the results are consistent with Demir and Uygar ‘s (2023) proposal regarding the role of internalized translation routines over external linguistic input. Experienced translators in this study were less influenced by the structural properties of the priming stimulus and more guided by fixed mental templates developed through extensive translation practice. This strengthens the argument that translators constitute a unique subgroup within bilingual populations—neither solely guided by the source language nor operating purely as second-language users (Jacob et al., 2024). Their behavior underscores the hybrid nature of expert translation as their performance displayed structural sensitivity as well as strategic autonomy. I argue that the reduced priming effect among experienced participants suggests the consolidation of individualized production patterns that can override momentary structural cues from the source language. In this regard, translation priming not only offers insights into structural alignment but also reflects bilingual expertise shaped by long-term experience.

5. CONCLUSION

This study demonstrated that translation priming occurs when there is structural overlap between source and target sentences, with both purely syntactic constructions and those involving syntax–semantics interaction eliciting priming effects. Proficiency and translation experience modulated the priming magnitude and direction, particularly in the presence of ungrammatical constructions and under varying temporal conditions. While inexperienced participants were more responsive to structural cues, experienced translators appeared to rely more on internalized routines shaped by prior exposure.

Despite these insights, the study has several limitations. First, it focused exclusively on two languages and a single syntactic phenomenon (dative alternation), limiting the generalizability of the findings. Future research should incorporate a wider range of languages and structures to determine the extent to which these effects apply across typologically diverse systems. Second, although proficiency and translation experience were treated categorically, these dimensions are best conceptualized as continuous. Large-scale studies employing gradient measures of experience and proficiency could offer more nuanced insights into how individual differences shape cross-linguistic priming. Third, the current methodology relied on reaction time and translation outcomes, which provide indirect evidence about cognitive processing. To more precisely examine the underlying mechanisms, neuropsychological methods such as eye-tracking, ERP (event-related potentials), or fMRI should be integrated to capture both temporal and spatial dynamics of the translation process.

In addition, expanding the participant pool beyond university students and including professional translators at various stages of their careers would enhance ecological validity. Combining experimental approaches with qualitative data, such as think-aloud protocols or retrospective interviews, could also offer complementary perspectives on decision-making during translation. Overall, this study contributes to our understanding of structural priming in bilingual contexts and highlights the need for methodological and linguistic diversity in future research on translation cognition.

REFERENCES

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition, 127*, 287–306.

Bock, J.K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Cai, Z. G., Pickering, M. J., Yan, H., & Branigan, H. P. (2011). Lexical and syntactic representations between closely related languages: Evidence from Cantonese-Mandarin bilinguals. *Journal of Memory and Language, 65*, 431–445.

Chomsky, N.A. (1965). *Aspects of the theory of syntax*. New York: MIT Press. 447- 475.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Demir, O. & Uygar, M. (2023). Cross-Linguistic Structural Priming Effects Between TypologicallyDifferent Language Pairs: An Experimental Study on Turkish-English Translation Priming. 21st International Conference on Turkic Languages.

Desmet, T., & Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language, 54*(4), 610–632.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming : Unravelling morpho-orthographic and morpho- semantic influences in early word recognition. *Language and Cognitive Processes, 20* (1–2), 75–114.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta psychologica, 128*(3), 501–513.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51*(1), 1–22.

Frazier, L., & Rayner, K. (1982). Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology, 14*, 178-210.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Gor, K. (2018). Phonological priming and the role of phonology in nonnative word recognition. *Bilingualism: Language and Cognition, 21*(3), 437–442.

Gračanin-Yuksek, M., Lago, S., Şafak, D.F., Demir, O., & Kırkıcı, B. (2017). The interaction of contextual and syntactic information in the processing of Turkish anaphors. *Journal of Psycholinguistic Research, 46*, 1397–1425.

Hare, M., McRae, K., & Elman, J. L. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes, 19*, 181–224.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language, 58*, 214–238.

Hopp, H. (2010). Ultimate attainment in L2 inflectional morphology: Performance similarities between non-native and native speakers. *Lingua, 120*, 901–931.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3), 277-306.

Ionin, T., & Montrul, S. (2010). The role of L1-transfer in the interpretation of articles with definite plurals in L2-English. *Language Learning, 60*, 877–925.

Jacob, G., Schaeffer, M. J., Oster, K., & Hansen-Schirra, S. (2024). The psycholinguistics of shining-through effects in translation: cross-linguistic structural priming or serial lexical co-activation? *Applied Psycholinguistics, 45*(3), 542–566.

Jensen, K.T.H., Sjørup, A.C., &Winther Balling, L. (2009). Effects of L1 Syntax on L2 Translation. In F. Alves, S. Göpferich, & I.M. Mees (Eds.), *Methodology, technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (pp. 319–336). Copenhagen: Samfundslitteratur.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition, 2*(1), 59-75.

Kantola, L., & van Gompel, R. P. G. (2011). Between- and within language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition, 39*, 276–290.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272–293.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Maier, R. M., Pickering, M. J., & Hartsuiker, R. J. (2017). Does translation involve structural priming? *Quarterly Journal of Experimental Psychology, 70*(8), 1575–1589.

Marslen-Wilson, W., Tyler, K. L., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101*(1), 3 – 33.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*(2), 314–324.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension*. Journal of Memory and Language, 38*(3), 283–312.

Neely, J. H. (1991b). Semantic priming in visual word recognition: A selective review of current theories and findings. In D. Besner & G. W. Humphrey (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Lawrence Erlbaum Associates, Inc.

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research, 21*(2), 98-120.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin, 134*(3), 427–459.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2001). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language, 43*,

Rastle, K., Davis, M.H. & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review, 11*(6), 1090–1098.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D., & Tyler, L.K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. Language and Cognitive Processes, 15(4/5), 507-537.

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica, 128*, 490–500.

Schaeffer, M.J. & Carl, M. (2013). Shared representations and the translation process: A recursive model. *Translation and Interpreting Studies, 8*, 169–190.

Schoonbaert, S., Duyck, W., Brysbaert, M. & Hartsuiker, R. J. (2007). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37*(5), 569-586.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension--an FMRI study. *Cerebral cortex, 22*(7), 1662–1670.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition, 11*, 245–260.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1521–1543.

Tooley K. M. (2023). Structural priming during comprehension: A pattern from many pieces. *Psychonomic bulletin & review, 30*(3), 882–896.

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass, 4*(10), 925–937.

Trofimovich, P., & McDonough, K. (Eds.). (2011). *Insights from psycholinguistics: Applying priming research to L2 learning and teaching*. Amsterdam: John Benjamins.

Trueswell, J., Tanenhaus, M. K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285–318.

van Gompel, R. P. G., & Arai, M. (2018). Structural priming in bilinguals. *Bilingualism: Language and Cognition, 21*(3), 448–455.

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*, 431–451.

Weber, A., & Broersma, M. (2012). Spoken word recognition in second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Bognor Regis: WileyBlackwell.

Chapter 10:
The Effect of Cognateness on Tip-of-The-Tongue (TOT) States and Their Resolution in Turkish EFL Learners

İpek Çelik Gençer & S. Çiğdem Sağın-Şimşek

ABSTRACT

Tip-of-the-tongue (TOT) states are defined as having a temporary challenge in remembering and bringing a word back to use during speech (Brown & McNeill, 1966). TOT state experiences are also frequently reported with names of people, places or objects in daily life. TOT states are common and experienced by people from all language backgrounds, proficiency levels and ages. TOT states occur even in sign language. Many studies (Gollan & Silverberg, 2001; Gollan & Acenas, 2004; Ivanova & Costa, 2008; Pyers et al., 2009) have revealed that bilinguals experience more TOT states compared to monolinguals due to the complexity caused by having two languages' lexicon stored in the brain as suggested by Bialystok et al. (2010). Borodkin and Faust (2013) proposed cross-lexical influence to be a modulating factor on the occurrences of TOT states. The aim of the present study was to reveal the effect of proficiency and cognateness on TOT state experiences of Turkish EFL learners and the effect of first syllables on TOT state resolution. A picture naming task with 26 cognate and 26 noncognate words along with 52 filler cognate and noncognate words, totaling 104 items, was conducted with 102 Turkish EFL learners. Reaction times to assess latency and key responses were analyzed. The interaction between cognateness and proficiency checked through. Findings showed a significant proficiency effect such that the higher proficient group was faster and experienced fewer TOT states. The cognateness effect was not foud to be significant although noncognates were revealed to induce fewer TOT states. The resolution of TOT states was successful after seeing the first syllables. Regarding the effect of cognateness on resolution, cognate words elicited more resolutions. Lastly, less proficient participants resolved more from their TOT states with non-cognate words.

Keywords: Tip-of-the-tongue states, cross-linguistic similarity, cognateness, lexical retrieval, picture-naming task, resolution, proficiency

1. INTRODUCTION

How humans access and utter words along with how these words are organized in the mental lexicon has long been a central topic in psycholinguistics. When the lexical retrieval process breaks down temporarily, speakers experience retrieval failures called Tip-of-the-tongue (TOT) states. Rather than a competence failure, Gollan and Acenas (2004) suggested that TOT states are performance failures where speakers have information stored in the mental lexicon, whereas they only have partial access to words' phonological forms. While words' phonological information is not accessible and activated, semantic information of the word is present and accessible during lexical retrieval. Ecke (2008) put forward that the possible reason of retrieval failures stems from the unsuccessful encoding of words' representations at all levels.

TOT states are highly prevalent and experienced across speakers of all language backgrounds, proficiency levels, ages and genders experience TOT states frequently (Warriner & Humphreys, 2008). Abrams and Davis (2016) even report occurrences of TOT state experiences in sign languages, which are called "tip-of-the-fingers" (Pyers et al., 2009). Indeed, D'Angelo and Humphreys (2015) claimed that if a TOT state occurred with a word or name, it is highly likely that the same word may cause a TOT state later even if it was resolved before. Burke et al. (1991) revealed in a diary study that most of the words that their participants noted were people's names. Speakers find TOT states annoying as Gollan and Brown (2006) suggest because finding words they search for often takes time or results in failure. Thus, suspension in speech and effort spent during retrieval may be frustrating because of being sure target words are stored somewhere in the brain not being able to bring it back to use. In that regard, calling these retrieval failures "tip-of-the-mind" states which were originally termed for the cognitive and conceptualizing role of gestures during speech by Hostetter and Alibali (2004) would also be proper for TOT states.

Numerous studies have consistently demonstrated that bilinguals have been found to be less advantaged in terms of experiencing TOT states compared to monolinguals (Gollan & Silverberg, 2001; Gollan & Acenas, 2004; Ivanova & Costa, 2008; Pyers, et al., 2009). In comparison to monolingual lexical processing where only one lexicon is represented in the brain, Bialystok et al., (2010) proposed bilingual processing is naturally more complex since there are two lexicons for each language and bilinguals supposedly know more vocabulary items. Moreover, bilinguals use two languages in daily life which causes both languages to be used less often than a monolingual speaker who uses words from a language (Altman et al., 2018; Ivanova & Costa, 2008). In addition, Pureza et.al. (2015) suggested that bilinguals experience more TOT states in their L2 because of infrequent speaking practices.

The Frequency-lag Hypothesis (Gollan et al., 2008, 2011) and Transmission Deficit Hypothesis (Burke, 1991) have been regarded as possible explanations to uncover the mechanisms and reasons causing TOT states. The Frequency-lag Hypothesis suggests that weaker connections between less frequently used word forms and their meanings make the retrieval process more susceptible to failures. Similarly, the Transmission Deficit Hypothesis posits that TOT states occur when semantic and phonological nodes are not activated simultaneously because of the deficit in stronger links and inadequate priming between each due to aging, recency and frequency of use. Considering the effect of aging on TOT experiences, Salthouse and Mandell (2013) discuss that older people experience more TOT states because they know more words compared to younger people, which causes breakdowns in retrieval processes trying to find target words.

The proficiency level of bilinguals has been revealed to modulate the pace of lexical processing in their second languages such that higher proficient speakers can naturally remember more quickly and accurately (Mathison, 2017) thanks to the increase in language experience The higher amount of exposure to target language and frequency of language provides stronger links during encoding of words representations at phonological, orthographical and semantic levels, which minimizes the risk of TOT states among higher proficient speakers. Low proficiency related yerine caused by/ stemming from limited language skills, on the contrary, account for weaker links between words phonological and semantic representation in lexicon due to insufficient priming which thereby cause latency in remembering target words and occurrences of TOT states according to Transmission Deficit Hypothesis (van Hell & Tanner, 2012). Taking these into consideration, low proficiency speakers are assumed to experience more TOT states and slower processing in their second languages.

The cross-linguistic similarity between each language affects occurrences of TOT states. The degree of difference or similarity between two languages is a determining factor in TOT states (Lijewska, 2020). Semantically and phonologically similar words across two languages, called cognates, may either help or cause latency in remembering. So as to reveal the cross-language similarity effect, cognate words have been prominently used in studies (Costa et al., 2000; Kroll et al., 2000; Gollan & Acenas, 2004; Haigh & Jared, 2007; Hoshino & Kroll, 2008; Colome & Miozzo, 2010; Poarch &Van Hell, 2012; Pureza et.al., 2015;). Cognates are the shared words between two languages. Thus, Haigh and Jared (2007) called cognates interlingual homophones. Non-cognate words are translation equivalents between two languages. To exemplify, cognates between English and Turkish can be *bazaar* and *pazar*. *Soup* and *sopa* are also cognates between English and Spanish.

The current literature presents predominantly positive evidence towards the facilitative role of cognates in recalling more accurately and quickly as well as experiencing TOT states among bilinguals (Gollan & Acenas, 2004; Hoshino & Kroll, 2008). On the other hand, cognates may also have an inhibitory effect on lexical retrieval because bilingual speakers need to be able to "turn one lexicon off" as Gollan and Silverberg (2001) claimed. Not being able to inhibit the unwanted words which are frequently semantically related words causes uttering unintended words or having a TOT state. The concurrent activation and retrieval of cognates have been explained within the Bilingual Interactive Activation+ model (van Heuven et al., 1998), which proposes an integrated mental lexicon for both languages of bilinguals and parallel activation of both lexicons simultaneously. The inhibition abilities at the lexical competition level are responsible for uttering only wanted words. Failure in inhibition may cause utterances not intended. Jones (1989) named unwanted words interlopers. Remembering an interloper can be because it is possibly and partially related to words being searched in terms of either meaning or phonological aspects.

Gollan and Acenas's (2004) study provides evidence for the cognate facilitation effect on TOT states. Two experiments with two different groups of bilinguals against monolinguals were conducted using a picture naming task. Participants in Experiment 1 were English- Spanish bilinguals and English monolinguals while Experiment 2 was conducted with English Tagalog bilinguals and English monolinguals who did not take part in the first study. In both experiments, 139 cognate words out of 442 experimental items were tested. Findings revealed that bilinguals experienced more TOT states compared to monolinguals. The reported number TOT states with cognate words among bilinguals was significantly less than non-cognate words. Thus, the cross-linguistic lexical similarity was shown to ease retrieval. In addition, fewer TOT states occurred with higher frequency words, confirming the frequency-lag hypothesis. This study is crucial both for the bilingualism effect effect on TOT states and the facilitative role of cognate words.

Aiming to verify the prevalent consensus on the cognate facilitation effect in literature and take a step further, Hoshino and Kroll (2008) conducted a picture naming task with cognate words and Spanish-English and Japanese-English bilinguals. The cognate words and picture-naming task paradigm were used to reveal the cross-linguistic effect because they are suitable for activating phonological level word representations and preventing the effect of orthography on lexical processing. Findings revealed that phonological representations are activated even when scripts differ. Thereby, orthographic differences do not impede phonological co-activation among bilinguals from different language backgrounds. This study holds the significance of revealing the cross-language facilitation effect among bilinguals even when the orthographical systems of two languages are completely different.

The types of tasks and procedures commonly employed to uncover why TOT states are experienced and how one can resolve TOT states can be listed as keeping a cognitive diary or structured diaries (Ecke 1998, 2004, 2008; Burke et al., 1991; Gollan et al., 2005), general information questions and answering task (Schwartz, 2001; Schwartz & Cleary, 2016), naming imaginary or created beings task (TOTimals) (Smith et al., 1991), picture-naming tasks (Gollan & Acenas, 2004; Hoshino & Kroll, 2008; Borodkin & Faust, 2014), definition-translation elicitation tasks (Warriner & Humphreys 2008; Hofferberth, 2014; Oliver & Humhreys, 2019), visual-lexical decision task (Pureza et al., 2013), naming famous people (Hanley & Cowell, 1988; Gollan & Bonanni, 2005) and oral production task under a lab condition (Schmank & James, 2020). Considering the similarity of cognates at phonological and orthographical levels, picture naming tasks to induce TOT states can lead to the activation of cognate words in both languages of speakers, which may have a facilitative effect on recalling words.

Many studies (James & Burke, 2000; Ecke, 2000; Abrams et al., 2003; Farrel & Abrams, 2011; Oberle & James, 2013; Hofferberth, 2014; 2015; Sauer, 2016) have investigated the salient attributes that help resolution of TOT states in addition to uncovering mechanisms and factors causing them. A common feature of TOT states is the awareness of knowing words because words are partially accessible but not in complete forms as put forward by Brown (1991). Aiming to reveal which parts of words are generally accessible or which salient attributes help speakers bring the full forms of words into use is crucial in terms of revealing how the mental lexicon is organized and guiding us to test the validity of the partial access theories. Regarding the resolution of temporary lexical access failures, it has been reported that speakers can successfully recall target items when cues are shown or presented to them (Schwartz & Cleary, 2016). Burke and James (2000) claimed that the key to successful cueing and making speakers retrieve wanted words rather than any interloper or unintended yet either semantically or phonologically related words is the similarity of the provided cue to target words.

Different kinds of cues have been used in different experimental paradigms such as syllables (Farrel & Abrams, 2011; Hofferberth, 2014; Sauer, 2016), phonemes (Oberle & James, 2013), pseudo homophones (Pureza et al., 2013), translation equivalents or associates (Lee et al., 2018), phonologically related words (Burke & James, 2000) and sound segments from target words (Abrams et al., 2003). Among all these different types of cues, syllables have been the most helpful ones. Specifically, using the first syllables of target words has been regarded to be more triggering (White & Abrams, 2002) compared to middle and final syllables which have also been utilized in many studies. Hofferberth (2012) referred to the first syllables as keys to the rest of words. While first syllables may facilitate activating the remaining syllables within words, interloper words sharing phonologically similar syllables or sound segments may also come along causing a more effortful retrieval and latency (Oberle & James, 2013). Cognate words between two languages make first syllables a perfect fit for resolution of TOT states.

The study by Abrams et al. (2003) was conducted to find out which salient attributes help resolution of TOT states. The salient attributes under investigation were phonological primes. Primes were the first letters of words, three types of syllables as first, middle and final along with first phonemes. Three separate experiments with a single type of prime were conducted. Findings from each of the three experiments revealed that entire first syllables led to more resolution compared to letters, phonemes and middle or final syllables. The Transmission Deficit Hypothesis was discussed by the authors to explain the results such that the first syllables activate the rest of the words by activating the phonological nodes and minimizing lexical competition. Moreover, participants who scored lower in vocabulary tasks resolved less from TOT states, which was attributed to weaker links in the lexicon.

Pureza et al. (2015) investigated the role of cognateness, syllabic position as initial versus last and word length as two or three syllables long on TOT states and their resolution in a visual lexical decision task, containing a total of 80 pictures of cognate and non-cognate target words. Each target word was matched with 8 related and 8 unrelated pseudowords with similar initial and final syllables as primes. Three groups as control, first and last syllables were formed. The control group saw only 8 unrelated primes. The first syllable group saw 4 primes out of 8 shared the first syllable and the last-syllable group saw 4 of 8 primes sharing the final syllable with the target. During the visual lexical decision task, participants were presented with 4 primes and then target words. If they could not remember target words, these were assumed to be TOT states. A following recognition task was implemented to verify the occurrences of TOT states. The main results confirmed the facilitative effect of cross-linguistic phonological overlaps. Moreover, word length also modulated TOT states such that more TOT states were elicited with words composed of more syllables. In contrast to Pureza et al. (2013) who found the last syllable's effect on resolution to be stronger, the first syllables were found to be more effective than the last syllables in this study. This discrepancy was discussed to stem from the word length differences. While the study in 2013 used four syllables, the 2015 study used two and three-syllable words. The authors discussed that the last syllable can be more helpful with longer words for speakers. However, the first syllables in shorter words can actually trigger the rest of the words which consist of fewer syllables.

1.1. The Present Study

The current study set out to investigate the TOT state experiences of Turkish EFL learners. In order to uncover the effect of cross-linguistic similarity on the occurrences and resolution of TOT states, cognate and non-cognate words were used in a lexical decision task. Another aim of this study was to find out if proficiency was an effective factor in TOT states. Lastly, the first syllables of words were used as cues to help participants retrieve target words.

The fact that TOT states are commonly experienced often and by speakers of any language background and at any age makes this phenomenon highly crucial. Uncovering the mechanisms that lead to TOT states and what strategies can be utilized to resolve partial retrieval failures can also help us reveal how the mental lexicon of bilinguals is represented and how words are accessed and processed. The current literature has ample studies on different language combinations such as English German, English Spanish and English-

Hebrew. More studies in distinct languages and bilingual or even multilingual contexts are needed. In the Turkish context, no other study has been done on TOT states on any participant sample either on monolinguals or bilinguals, with different types of words as cognates and non-cognates and in either L1 or L2. This study aimed to contribute to the field by being the first study investigating TOT states in Turkish English bilingual context. In addition, regarding English being a foreign language rather than a second language, the amount of interaction and communication in English in daily life is not adequate for consolidated learning because learners do not have real-life encounters where they can use words they have learned in classes. Not recalling words in class can be regarded as normal because of infrequent use and exposure. The vocabulary learning and teaching practices in English classrooms in the Turkish context make this study highly significant because the findings of the current study hold the potential for classroom implications.

This study seeks to address the following research questions (i) to what extent cognateness affects TOT state experiences of Turkish EFL learners, (ii) whether cognate words affect the retrieval pace, (iii) to what extent proficiency affects TOT states that Turkish EFL learners experience and (iv) whether first-syllables helps Turkish EFL learners resolve from their TOT states. Taking the theoretical background and findings of earlier studies into consideration, it was hypothesized that cognate words would cause fewer TOT states while more TOT states would be induced with non-cognate words. Similarly, cognate words were predicted to be retrieved faster due to the orthographic similarity in both languages. The lower proficiency participants were expected to experience more TOT states. On the other hand, the higher proficiency participants would also experience more TOT states due to latency caused by lexical competition within a larger lexicon as suggested by Gollan and Kroll (2014). The last hypothesis was that participants would resolve from TOT state experiences upon seeing syllables.

2. METHOD

2.1. Participants

102 native speakers of Turkish and English learners (Mean age=20.209, SD=2.331, 55 females) participated in the main experiment. All participants had normal or corrected-to-normal vision, and no language or learning disorders. All participants were right-handed and used their dominant hands

to press buttons on the keyboard. An imageability rating task was conducted with 75 participants who did not take part in the main experiment.

The participants took a standardized placement test prepared by Ionian and Montrul (2010). The median-split technique was utilized to create two proficiency groups. The median was 82.5 (n=102). The participants who scored 82,5 and above were grouped as high (n=52) while 80 and below were grouped as low proficiency (n=50).

This study was approved by Human Subjects Ethics Committee of METU with the protocol code 0048-ODTUİAEK-2023.

2.2. Materials

An imageability rating task containing a total of 114 cognate and non-cognate words was run to choose the highest imageable words. 75 participants who did not take part in the main experiment rated words' imageability from 1 to 7. The highest imageable 32 cognate and 32 non-cognate words were selected based on the criteria for how well and easily these words create a mental image (Bird, Franklin, and Howard (2001). After the 46 pilot study sessions, the final version of the experiment items were 26 cognate and 26 non-cognate target words, alongside 52 filler items. The imageability ratings ($t(50)=-1{,}140$, $p=.260$) and frequency ($t(50)=-9, -.09$, $p=.929$) of cognate and non-cognate words were matched. The filler words were one-syllable 26 cognate and 26 noncognate words. The order of the words was randomized for each participant automatically. The first syllables of words were identified. The first sound segments of filler words were used since they already have one syllable. The best-fitting pictures of all words were selected and confirmed for naming consistency by participants whether pictures aroused only one word upon seeing during the piloting sessions. Non-cognate words were selected from the list named The Oxford 3000 by CEFR level (Oxford University Press, n.d.) Cognate words were chosen from the list prepared by Uzun and Salihoğlu (2021). An example of the word list can be seen in Table 33.

To assess participants' English proficiency, a cloze test (Ionian & Montrul, 2010) which consisted of 40 multiple-choice questions was administered. Additionally, a linguistic background questionnaire was used to gather information about participants' language use, exposure, and education.

Table 33. Sample Targets

Condition	Word	The first syllable prime
Cognate	tunnel	tu-
Noncognate	recipe	re-
Filler-cognate	cream	cre-
Filler-noncognate	pipe	pi-

2.3. Procedure

A picture-naming task was designed to elicit TOT (Tip-of-the-Tongue) states. The experiment was created using OpenSesame (v3.3.1.4) (Mathôt & March 2022) and deployed via the open-source platform JATOS (Lange et al., 2015). JATOS' robust backend provides a stable server environment for hosting and managing experiments, facilitating data collection and storage. It is a reliable platform for online experiments collecting reaction time (RT) data because it measures RT on participants' browsers thanks to its use of client-side timing. The client-side timing feature allows locally measured RT (Lange et al., 2015). Variability or delay due to internet connection speed does not affect timing data but only the uploading process. Once the download of the experiment into a browser is complete, data is not affected by internet conditions until the upload of results. More cognitive and behavioral researchers have started using web-based online software for collecting reaction time data because it provides consistency with laboratory data and millisecond-level precision with only minimal and deviations (Bridges et al., 2020; de Leeuw & Motz, 2016). For each participant, an one-time only link of the experiment was created. Participants used links to download the experiment to their servers. The researcher and participants met at ZOOM sessions where they concurrently joined the experiment. All participants were tested individually in live sessions. A consent form was given before beginning to start the experiment. After the consent was given, the definition of TOT state in Turkish was explained. In order to familiarize the participants, a trial loop was added with 10 words.

Trials started with a fixation dot remaining in the middle for 500ms. Pictures of words followed and stayed for 5000ms or until participants responded. If no answer was given within 5000ms, the next trial started. The picture-naming task took approximately 15 to 20 minutes. For the experiment, participants used their laptop or desktop keyboards to respond. The participants were asked to respond as quickly and accurately as they could by using their dominant hand. After seeing pictures, they were asked to press H if they knew and at once remembered the target. If they were sure they did not know words, they pressed K. When they were sure they knew but could not remember words, they pressed U. After participants utter words, the researcher pressed F for correct and J for incorrect responses from her own keyboard. All responses were recorded, and reaction times (RTs) were measured from stimulus onset to participant response. Figure 1 illustrates the full procedure and codes.

The key responses of participants were analyzed as follows. Upon pressing H, got was coded for correct answers. When participants pressed know key but did not utter target words, they were shown the first syllables of words. Then, if they could remember correctly, cuegot was coded but if they could not find words, they were shown target words and were asked if they knew or not. If they stated they actually knew, it was coded notgot+. If they stated they did not know, notgot- was coded. After pressing K, target words were immediately shown and they were asked if they knew the words. Again, if they stated they actually knew, it was coded notgot+. If they stated they did not know, notgot- was coded. When they could not remember, the first syllables were shown. If they could remember upon seeing the first syllables, it was coded resolvedtot. If they could not remember, target words were presented to them. If they stated that they actually knew words upon seeing them, they were asked a subsequent question whether they were searching for those words or not. If they were searching for it but could not retrieve it at that moment, it was coded as tot+. However, if they stated that they were actually trying to remember another word but they also knew the target words, it was coded as tot-. Lastly, when they stated that they did not know target words, it was coded as notgot-.

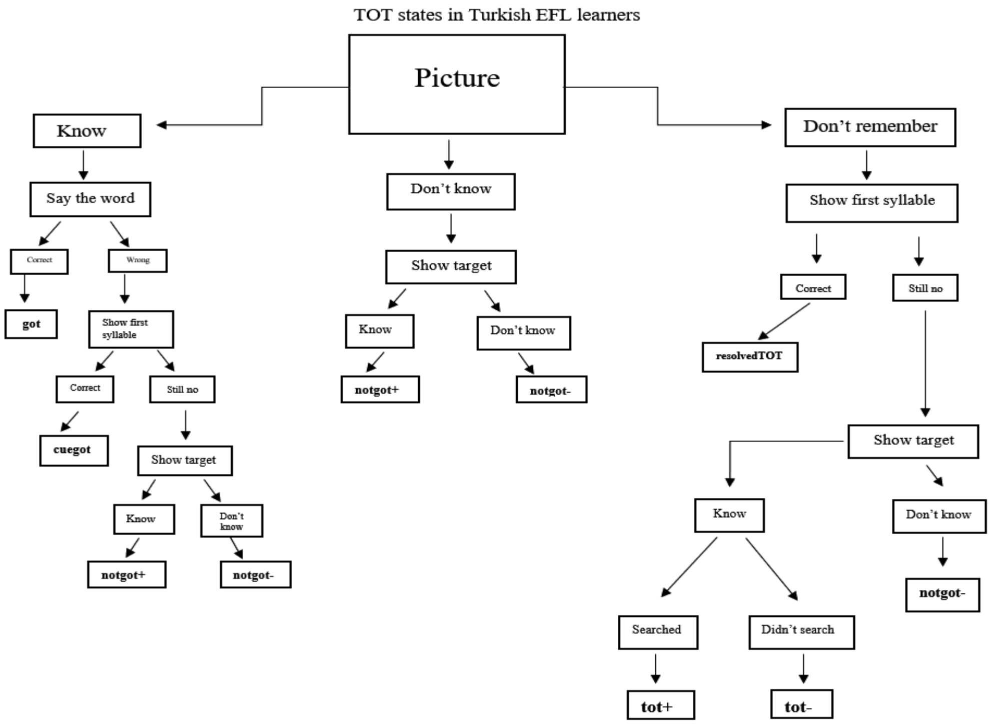


Figure 18: TOT Experiment Procedure

3. RESULTS

The final data was analyzed with JASP Version 0.17.1. Firstly, As RT from data was positively skewed, RT data was log-transformed to be normalized because the linear models require the data to have a normal distribution. Also, reaction times below 200ms were excluded since it is impossible for a human to respond in such an extremely short time (Baayen, 2010). Lastly, two separate variables for cognateness and proficiency were computed for dummy coding in the models for better estimation. For proficiency, high proficiency was coded as 0 while low proficiency was coded as 1. Similarly, cognate was coded as 0 while noncognate was coded as 1. As the participants pressed only three keys, so reaction times were taken only from these first three responses.

Log-transformed reaction times with standardized placement (Zplacement) were analyzed using a linear Mixed Effect Model (LLM). To find the final model that converges, random slopes were reduced until it did not give any errors. The model convergence was attained by random intercepts for words, and cognateness as random slopes and intercepts for subjects were specified to account for variability across items and participants. A significant effect of proficiency was observed on reaction times such that the higher proficient group had faster reaction times (β=-0.066, SE=0.021, t=-3.089, p.<0.005). Although cognate words were found to cause more latency compared to noncognates, it was not found to be significant (β=-0.072, SE=0.054, t=-1.329, p.=0.190). The effect of the interaction between proficiency and cognateness was found to be marginally significant (β=-0.023, SE=0.012, t=-1.1976, p.=0.051) such that the higher proficiency group responded faster to noncognate words.

A separate LLM model was run for reaction times of each key response which awas coded as coded_new, proficiency and cognateness condition along with all interactions among them. The model included random intercepts for both participants (subject) and items (word) to account for repeated measures. The full model formula is Model= `log(RT)` ~ coded_new * prof * condition + (1 | subject) + (1 | word). Both notgot (- and +) responses were faster compared to the other two responses (β=-0.199, SE=0.018, t=-11.339, p<0.00) such that participants responded faster when they were sure they did not know target words in pictures shown to them. notgot responses for non-cognates were faster (β=-0.036, SE=0.026, t=-1.400, p=0.161)but not significant. Even though the effect was not significant, TOT key responses caused latency (β=0.049, SE=0.027, t=1.825, p>.005), which may be due to the need to think and decide on whether they knew or not. Lastly, no significant effect of proficiency was found on three key responses (p>.005 in all cases) even though the low proficient group responded slower in notgot responses (β=0.015, SE=0.031, t=0.488, p=0.626). Additionally, the low proficient group was faster for TOT key upon seeing a picture of a noncognate word (β=-0.068, SE=0.048, t=-1.422, p=0.155). This finding demonstrates that it took time for the low proficient participants to search for an unknown word but they could quickly decide if they knew a word even though these words were not accessible at that moment. No interaction among given responses, proficiency and cognateness showed a significant effect on reaction times.

The responses of the participants were analyzed with two different models namely Generalized Linear Mixed Model (GLMM) and Paired Samples T-test. Figure 2 and 3 present the descriptives for responses given by participants without cross-classifying data meaning raw responses were just divided by condition and proficiency group. In order to conduct GLLM analysis with cross-classified data, both cognateness condition and proficiency were added to models for reliable estimates. The raw descriptive statistics show that the total number of responses was 5166. The number of each response type along with cognateness condition and proficiency can be seen in Figure 19 and Figure 20.

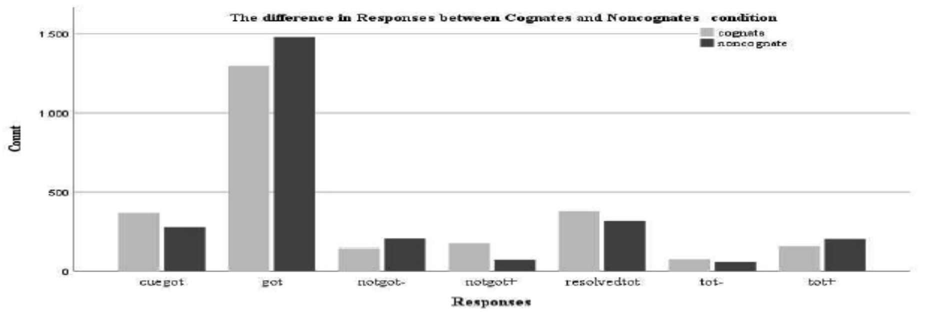


Figure 19: Responses for Cognate and Noncognate Words

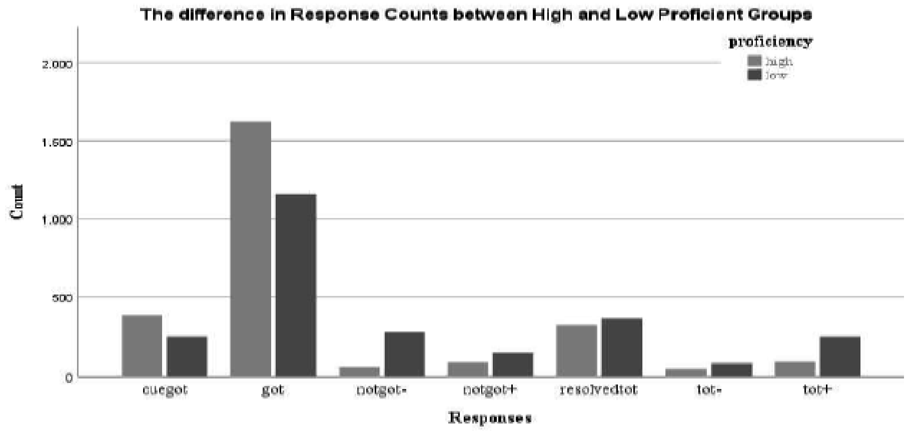


Figure 20: Responses by High and Low Proficiency Groups

The first GLMM model was run for the effect of cognateness and proficiency on the occurrences of TOT states. The full model formula is as follows: Model= total_tots~ prof * condition + (1 | subject) + (1 + prof | word). The analysis of total TOT states (tot+ n=347, tot- n=133, resolvedtot n=690, total= 1170) has revealed a significant effect of proficiency (β=0.677, SE=0.214, t=3.163, p<0.005) such that the low proficient group (n=700) experienced more TOT states than the high proficient group (n=470). Cognate words caused more TOTs but there was not a significant effect (β=-0.313, SE=0.338, t=-0.926, p=0.354). Although the cognate words were found to cause more TOT states and the noncognate words cause fewer TOT states, it was not estimated to be significant. No interaction was significant (p>0.455). Even though the lower proficient group experienced more TOT states with noncognate words (<c.190, SE=0.254, t=0.748, p=0.455), no significance was attained in the model.

Paired Sample T-test was used to find out if there is a significant difference between resolved and unresolved TOT states. To compute a new variable for unresolved TOT states, tot+ and tot- responses were combined into a single response as is.unresolvedtot. There was a significant difference (MD=0.179, t=6.238, p<0.001) between resolved (M=0.590) and unresolved (M=0.410) TOT. This finding indicates that cueing has a positive effect on resolution meaning that participants are resolved from TOT states. The effect of cognateness and proficiency on the resolution of TOT states was analyzed by using another GLLM model. The full formula for the model is Model= \`is.resolvedtot\` ~ condition * prof + (1 | subject) + (1 + prof | word). The model revealed that proficiency did not have a significant effect on resolution but the low proficient participants resolved more (β=0.299, SE=0.226, t=1.327, p=0.185). Noncognates were resolved less compared to cognates but there was not a significant effect (β=-0.304, SE=0.279, t=-1.090, p=0.276). Lastly, the effect of the interaction between cognateness and proficiency was revealed to be insignificant (p=0.904).

4. DISCUSSION

Bilingualism has been discussed to affect the number and frequency of occurrences of TOT states. Bilinguals are expected to experience more TOT states compared to monolinguals (Gollan & Silverberg, 2001). The presence of dual lexicons in the brain complicates lexical processing due to the possibility of activation of words from both lexicons (Brown, 1991). In that regard, lexical processing may encounter retrieval failures, especially TOT states. Reaction times and keyboard responses were used to analyze latencies in recalling the pictures shown to participants, thereby revealing mechanisms causing TOT states. The impact of the proficiency level of participants and cross-lexical similarities on the occurrences of TOT states along with a resolution of TOT states with the help of first syllables.

In line with the literature on the effect of proficiency on TOT state experiences and lexical processing pace, the findings of this study provided further evidence on the matter such that higher proficiency was revealed to accelerate the processing. It was expected for higher proficiency speakers to respond faster to pictures shown to them and they were faster. On the contrary, the lower proficient group showed significant latency. Confirming the hypotheses further based on the third research question, the number of TOT state experiences was larger for the lower proficient speakers and they also showed latency similar to the previous study by Borodkin and Faust (2014). As a consequence of low proficiency, speakers do not use words as frequently as they do in their native languages, thereby insufficient priming is provided for words in mental representation.

In line with the frequency-lag hypothesis (Gollan et al., 2011), latency and more retrieval failures can arise from weaker links between words' phonological, semantic and orthographic representations. Borodkin et al. (2017) discuss that regarding TOT states as retrieval failures at phonological level representations along with lower proficiency as modulating phonological encoding abilities, weaker phonological skills can induce more TOT states among low proficient speakers. Moreover, less proficient speakers also know fewer words and have a smaller lexicon compared to higher proficiency speakers. The number of notgot- responses from the low proficiency group (n=279) was significantly higher. This finding approves the prevalent consensus in the literature. In addition to knowing fewer words, they also showed latency in deciding if they were in a TOT state and if they actually knew words. On the other hand, they pressed don't know key key faster, which may be because of the less confidence in their language abilities and awareness of the possibility of knowing a small number of words yet.

A unique aspect of this study is to investigate the role of interlopers in lexical retrieval. Common experimental paradigms resemble each other for inducing TOT states where participants reply to a stimulus after making a decision on knowing, not knowing or not remembering a word. However, studies focused mainly on words not remembered, the response for TOT state. Disregarding the co-activation of other possible competitors with the help of first syllables presented to participants was because these responses were not expected to elicit TOT states. However, the uttered non-target words upon pressing 'know' key can shed light on how words sharing similar semantic or phonological properties are stored in the mental lexicon. Jones (1989) highlighted the inhibitory role of recalling a similar word to the target in experiencing TOT states because target word search is blocked. Aiming to see the role of competitors and interlopers, participants were presented with syllable cues if the words they said were not target words but either semantically or phonologically similar words. In a picture naming task, uttering an unintended word yet successfully recalling it after seeing a cue can only elicit a translation equivalent or a synonym. Findings on the 'cuegot' responses revealed that the high proficiency group experienced more instances of retrieving an interloper rather than targets. The possible explanation for this finding is the relationship between the size of lexicon and bilingualism, and directly to proficiency level. Supporting the hypothesis proposed by Gollan and Kroll (2014) that higher proficient speakers tend to experience more retrieval problems since their search process is more complex as a direct consequence of a larger lexicon containing a higher number of words. Their responses after getting a cue were not coded as TOT states because they were not actually searching for targets and having difficulty retrieving these words.

Cognate words between Turkish and English words in a picture naming task were expected to uncover an effect on the retrieval process. The findings of the present study did not reveal either facilitation or interference impact of cross-linguistic similarities due to the number of responses given for both cognate and non-cognate words were not found to be significantly different. Regarding the proficiency level of participants, no significant difference between both groups was found. It was hypothesized to find a facilitative role in cases where first syllables between cognates in both languages could trigger the rest of words' syllables. On the contrary, non-cognate words were remembered faster, which may be because participants were EFL learners and they may have encountered non-cognate words more recently and frequently in classroom settings.

Unlike Gollan and Acenas (2004) whose study provided evidence for the cognate facilitation effect with fewer TOT states, more TOT states were reported with cognate words in the current study even though it was not significant. Gollan and Acenas (2004) discuss this unexpected finding through the need to know cognates in both languages. Bilinguals need to know cognate words in both of their languages to remember faster and more easily. In line with Pureza et al. (2015), the low proficient participants reported having more TOT states with non-cognates because cognates are easier to learn while noncognate words require more effort for learning due to their unfamiliarity and weaker language skills. On the other hand, the number of tot- and tot+ responses indicates that participants actually knew words but they were not remembering that word or they were trying to remember another word at the moment. Also, the number of induced TOT states was not expected to be this low, which may have prevented the emergence of the cognate effect. Another point that can be made is the need for cognates to be used equally frequently in both languages so as to find an effect. However, the lack of an up-to-date and working Turkish corpus hindered checking the frequency of cognates in Turkish in addition to their frequency in English.

Regarding the resolution of elicited TOT states, first syllables facilitated remembering target words. Aligning with the current literature, the first syllable as a key or triggering force to the activation of words remaining syllables (White & Abrams, 2002; Hofferberth, 2012; Oberle & James, 2013). Lastly, participants retrieved more cognate words they were searching after seeing the first syllables though it was not found to be significant. The shared phonological and semantic nature of cognates makes it easier to recall them when a shared syllable is provided.

Limitations

The most important limitation was the selection of noncognate words from Oxford 3000 words list (Oxford University Press, n.d.). This may have caused a recency effect because these words are often used in coursebooks. Since participants were EFL learners, they may have encountered words in learning contexts recently, which may have induced fewer TOT states. Future studies may prepare an item list with more variety or from a dictionary in order to prevent the risk of recency. Also, both cognate and noncognate words can be selected from the lists of the previous studies. Secondly, frequencies of cognates in their Turkish correspondences should also be checked from an up-to-date Turkish corpus to match the frequencies of both conditions. The adjustments of key buttons should be carefully done because K button for 'don't know' response was placed on the right side, which may have caused faster reaction times for notgot responses as the participants were right-hand dominant. Lastly, definition task design can also be suitable for examining the cognateness effect. In a definition task design, false friends can be added to items. The picture-naming task in this study did not allow the use of false friends because they have similar or nearly the same orthographical forms but two distinct semantic representations. The integration of false friends would have caused imageability problems.

Acknowledgment

We would like to express my gratitude to Asst. Prof. Dr. Orhan Demir, Assoc. Prof. Dr. Duygu Sarısoy, and Semih Can Aktepe for their support and valuable insights throughout the process. We are truly thankful for their contributions.

REFERENCES

Abrams, L., & Davis, S. K. (2016). Tip-of-the-tongue states. In G. Hickok & S. L. Small (Eds.), *The neurobiology of language* (pp. 307–322). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00025-4

Abrams, L., White, K., & Eitel, S. L. (2003). Isolating phonological components that increase tip-of-the-tongue resolution. *Memory & Cognition*, 31(8), 1153–1162.

Altman, C., Goldstein, T., & Armon-Lotem, S. (2018). Language dominance in bilingual children: Challenges and directions for future research. *International Journal of Bilingual Education and Bilingualism*, 21(4), 506–518. https://doi.org/10.1080/13670050.2018.1428221

Baayen, R. H. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33(1), 73–79. https://doi.org/10.3758/BF03195349

Borodkin, K., & Faust, M. (2013). Tip-of-the-tongue (TOT) states and cross-linguistic transfer. *Bilingualism: Language and Cognition*, 16(4), 914–923.

Borodkin, K., & Faust, M. (2014). Naming abilities in low-proficiency second language learners. *Journal of Learning Disabilities*, 47(3), 237–253.

Borodkin, K., Maliniak, N., & Faust, M. (2017). Phonological processing in second language users: Evidence from the tip-of-the-tongue paradigm. *Second Language Research*, 33(4), 455–472. https://doi.org/10.1177/0267658317704266

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ,* 8, e9414. https://doi.org/10.7717/peerj.9414

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, 109(2), 204–223.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30(5), 542–579. https://doi.org/10.1016/0749-596X(91)90026-G

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99(2), 205–235. https://doi.org/10.1016/j.cognition.2005.01.009

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48, 1–12. https://doi.org/10.3758/s13428-015-0567-2

Ecke, P. (2000). Tip-of-the-tongue states with foreign language words: Resolution types and word search strategies. *Arizona Working Papers in SLAT*, 7, 55–69.

Ecke, P. (2004). Words on the tip of the tongue: A study of lexical retrieval failures in Spanish-English bilinguals. *Southwest Journal of Linguistics*, 23(2), 33–63.

Ecke, P. (2008). Cross-linguistic influence on word search in tip-of-the-tongue states. *TESOL Quarterly*, 42(3), 515–527.

Farrell, M. T., & Abrams, L. (2011). Tip-of-the-tongue states reveal age differences in the syllable frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 277–285.

Gollan, T. H., & Acenas, L. A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 246–269. https://doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General*, 135(3), 462–483.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, 4(1), 63–83. https://doi.org/10.1017/S136672890100013X

Gollan, T. H., Bonanni, M. P., & Montoya, R. I. (2005). Proper names get stuck on bilingual and monolingual speakers' tip-of-the-tongue equally often. *Neuropsychology*, 19(3), 278–287.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, 16(4), 562–576.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787–814.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186–209. https://doi.org/10.1037/a0022256

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530.

Hanley, J. R., & Cowell, E. S. (1988). The effects of different types of retrieval cues on the recall of names of famous faces. *Memory & Cognition*, 16, 545–555.

Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213–229. https://doi.org/10.1017/S1366728998000364

Hofferberth, N. J. (2012). On the role of the syllable in tip-of-the-tongue states. *In Proceedings of the International Conference of Experimental Linguistics,* 27–29, Athens, Greece.

Hofferberth, N. J. (2014). Resolution of lexical retrieval failures: Reaction time data in the tip-of-the-tongue paradigm. *In 10th International Seminar on Speech Production (ISSP)*, 194–197, Cologne, Germany.

Hofferberth-Sauer, N. J., & Abrams, L. (2014). Resolving tip-of-the-tongue states with syllable cues. In V. Torrens & L. Escobar (Eds.), *The processing of lexicon and morphosyntax* (pp. 43–68). Cambridge Scholars Publishing.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script?. *Cognition*, 106(1), 501–511. https://doi.org/10.1016/j.cognition.2007.02.001

Hostetter, A. B., & Alibali, M. W. (2004). On the tip of the mind: Gesture as a key to conceptualization. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 26(26).

Ionin, T., & Montrul, S. (2010). The role of L1-transfer in the interpretation of articles with definite plurals in L2-English. *Language Learning*, 877–925.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127, 277–288.

James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1378–1391.

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, 17, 69–76.

Kroll, J. F., Dijkstra, T., Janssen, T., & Schriefers, H. (2000, November). Selecting the language in which to speak: Experiments on lexical access in bilingual production. Paper presented at *the 41st Annual Meeting of the Psychonomic Society*, New Orleans, LA.

Lange, K., Kühn, S., & Filevich, E. (2015). Just Another Tool for Online Studies (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLoS ONE*, 10(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lijewska, A. (2020). Bilingual lexical processing: The role of cross-linguistic similarity and language experience. *International Journal of Bilingualism*, 24(1), 88–103. https://doi.org/10.1177/1367006918808027

Mathôt, S., & March, J. (2022). Conducting linguistic experiments online with OpenSesame and OSWeb. *Language Learning*. https://doi.org/10.1111/lang.12509

Oberle, S., & James, L. E. (2013). Semantically- and phonologically-related primes improve name retrieval in young and older adults. *Language and Cognitive Processes*, 28(9), 1378–1392.

Oliver, L. K., & Humphreys, K. R. (2019). Phonological interlopers tend to repeat when tip-of-the-tongue states repeat. *Frontiers in Psychology*, 10, 341.

Pureza, R., Soares, A. P., & Comesana, M. (2013). Syllabic pseudohomophone priming in tip-of-the-tongue states resolution: The role of syllabic position and number of syllables. *Quarterly Journal of Experimental Psychology*, 66(5), 910–926.

Pureza, R., Soares, A. P., & Comesana, M. (2015). Cognate status, syllable position and word length on bilingual tip-of-the-tongue states induction and resolution. Bilingualism: *Language and Cognition*, 19(3), 533–549.

Pyers, J. E., Gollan, T. H., & Emmorey, K. (2009). Bimodal bilinguals reveal the source of tip-of-the-tongue states. *Cognition*, 112(2), 323–329. https://doi.org/10.1016/j.cognition.2009.05.010

Salthouse, T. A., Mandell, A. R. (2013). Do age-related increases in tip-of-the-tongue experiences signify episodic memory impairments? *Psychological Sciences*, 24(12), 2489- 2497.

Schmank, C. J., & James, L. E. (2020). Adults of all ages experience increased tip-of-the-tongue states under ostensible evaluative observation. *Aging, Neuropsychology, and Cognition*, 27(4), 517–531.

Schwartz, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition*, 29(1), 117–126.

Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metacognitive experiences. In *The Oxford Handbook of Metamemory*. Oxford University Press.

Smith, S. M., Brown, J. M., & Balfour, S. P. (1991). TOTimals: A controlled experimental method for studying tip-of-the-tongue states. *Bulletin of the Psychonomic Society*, 29(5), 445–447.

Oxford University Press. (n.d.). *The Oxford 3000 by CEFR level*. Retrieved from https://www.oxfordlearnersdictionaries.com/wordlist/english/oxford3000

Uzun, L., & Salihoglu, U. (2021). A list of English-Turkish cognates and false cognates. *Poznan Studies in Contemporary Linguistics*, 57(2), 325–327.

Warriner, A. B., & Humphreys, K. R. (2008). Learning to fail: Reoccurring tip-of-the-tongue states. *The Quarterly Journal of Experimental Psychology*, 61(4), 535–542.

White, K. A., & Abrams, L. (2002). Does priming specific syllables during TOT states facilitate word retrieval in young and older adults? *Aging, Neuropsychology, and Cognition*, 9(4), 255–271. https://doi.org/10.1076/anec.9.4.255.13588