Bilinguals’ Logical Inference-Making and Language Tagging

by

Aslı Yurtsever

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

Psychology

KOÇ ÜNİVERSİTESİ

June 10, 2021

Bilinguals’ Logical Inference-Making and Language Tagging

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a master’s thesis by

Aslı Yurtsever

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

____

Prof. Dr. Sami Gülgöz (Advisor)

____

Assoc. Prof. Tilbe Göksun

____

Asst. Prof. Ercenur Ünal

Date 10.06.2021

To my family. For all of their languages, love, support, and cognitive abilities.

ABSTRACT

Bilinguals' Logical Inference-Making and Language Tagging

Aslı Yurtsever

Master of Arts in Psychology

June 10, 2021

Inference-making is a critical process in understanding and processing information daily. People synthesize inputs into a whole and retain the whole (gist), instead of specific parts (verbatim). False recognition of inferred information offers evidence for it. We conducted two studies to examine whether memory errors occur similarly when bilinguals are tested concurrently on two languages, and whether bilinguals remember the language in which the information is received (language tag). In two experiments, we tested Turkish-English bilingual students with Turkish, English, and dual-language sentence groups, which induced spatial inferences about objects. In Experiment 1, we found inferences were falsely recognized in Turkish and English conditions, but not in dual-language. we found no effect of individual cognitive differences. In Experiment 2, we found that false recognition of inferred sentences were predicted more by lower EF and higher L2 In both experiments, tag identification was overall accurate and partially predicted by higher cognitive functions. We conclude that inferring in L2 induces memory errors and inferences are tagged with the language of encoding.

ÖZETÇE

Çiftdillilikte Çıkarım Yapma Süreçleri

Aslı Yurtsever

Psikoloji, Yüksek Lisans

10 Haziran 2021

Çıkarım yapma bilgi işleme sürecimizin önemli bir parçasıdır. Deneyimlerimiz ile yeni bilginin sentezlenmesiyle oluştuğu için yanlış hatırlamalara sebebiyet verebilir. Yanlış hatırlamaların çiftdilli insanlarda nasıl gerçekleştiğini ve anıların dil etiketi ile saklanıp saklanmadığını araştırmak için iki deney yürüttük. Bu çalışmalarda Türkçe-İngilizce çiftdilli katılımcılar, Türkçe, İngilizce ve iki dilin aynı anda kullanıldığı cümeleri çalıştılar ve uzamsal çıkarımlard bulundular. Birinci deneyde katılımcıların Türkçe ve İngilizce durumlarda yaptıkları çıkarımları yanlış hatırladıklarını, ancak karışık dilli durumda yanlış hatırlamadıklarını bulduk. İkinci deneyde yanlış hatırlamanın düşük bilişsel beceriler ve yüksek İngilizce yeterliliği ile ilintli olduğunu bulduk. Her iki deneyde de katılımcıların dil etiketleme performansı yüksekti ve yüksek bilişsel becerilerle ilintili bulundu. Çıkarım yapmanın, çiftdillilerin iki dilinde de yanlış hatırlamaya sebep olduğu ve çıkarım yapma sürecindeki anıların dil etiketinin raporlanabildiği sonucuna varıldı.

AKNOWLEDGEMENTS

I would like to thank my parents, Drs. Mine and Ersin Yurtsever for all of their love, support, and everything I could never finish listing nor repaying. Dr. Zeynep Yurtsever, to whom I owe my existence and many more, for proofreading everything I send and appreciating comic reliefs. Let it be known that I retain the right to hold this dedication over her head forever.

My beloved advisors, Drs. Sami Gülgöz and Tilbe Göksun, not only for the invaluable experiences and skills I gained working with them but also for the working environments they set up. Thank you for creating the best labs, the best friendships, and hosting the best dinners. You set the bar incredibly high! I am not the first, nor will be the last, to understand why we can check-out any time we like, but can never leave.

Many helped me execute the projects that are presented in this thesis. I am deeply appreciative of Buğra Sipahioğlu for helping me code the experiment, to whom I significantly undersold how much time it would take. I thank him for his hard work, kindness, and patience in dealing with my unceasing questions throughout. I have had the pleasure of mentoring great research assistants: Hilal Öztürk, Nilay Rodoplu, and Aylin Kaçmaz. They aided me in data collecting and organization. Without their quick and meticulous work, I would have never finished this thesis during the pandemic.

Last but not least, I would like to express my gratitude for my dear friends and colleagues, members of KURAM and LC labs, for repeatedly listening to my presentations and venting. You have shown me what a great support system should look like, be it virtually or with karaoke. We will always have SOS Mutfak!

TABLE OF CONTENTS

List of Tables

List of Figures

Abbreviations

WM	Working Memory
EF	Executive Functions
RHM	Revised Hierarchical Model
L1	First language
L2	Second language
DL	Dual-language

Chapter 1:
INTRODUCTION

People do not merely encode separate units of information but synthesize them to create a meaningful whole. The constructive and inferential nature of human memory and information processing has been investigated extensively, mostly with memory errors, and in people's first language (L1). The question of the current study is whether the inference mechanisms differ when the information is received in the second language or in two languages concurrently. A secondary question is whether the language of information is retained when information is presented in two languages. We consider that the literature on bilingual lexical organization would benefit from understanding how the inferential processes are sustained in bilinguals' second language (L2) and in the dual-language context.

1.1 Bilingualism Models

Research into the bilingual organization of information investigated if bilinguals have different networks for knowledge acquired in their first language (L1) versus the language later acquired (L2). The Revised Hierarchical model (RHM), a verbal model devised as a concession to Kolers (1963) and Potter et al. (1984)'s word association model and concept mediation model, posited that the lexical memory systems are separate, but the concepts are stored in a common, shared system (Kroll & de Groot, 1997; Kroll & Stewart, 1994). The model conceptualized that the L2 lexicon is smaller than L1 and has weaker links to concepts, thus requiring additional access time. L1-L2 paths are conceptually mediated, where L1 cues automatically activate concepts. L2-L1 paths, on the other hand, are lexically mediated. As L2 concepts are learned, they get translated and associated with L1.

In 2010, Kroll and colleagues revised their model to suggest that words from either language can access concepts non-selectively, but the link from concept to L2 lexicon remains weak, especially for production tasks. They argued the possibility that L2 to L1 mediation occurs *after* already accessing the concept. They emphasized high cognateness [1] and concreteness of the words translated could render both directions conceptually mediated. Another review was on the L2-L1 path is developmental changes with proficiency increase. L2-concept links get stronger with increased proficiency and rely less on L1 translation. The asymmetry in translation was also explained by the Inhibitory Control Model (Bialystok et al., 2009; Green, 1998). It was argued that L1-L2 path is slower because L1 activation is more difficult to inhibit. The fact that bilinguals cannot switch a language off at will further supports that languages are not processed selectively, but are rather integrated (Dijkstra et al., 2019). Code-switching suggests that both languages are active at all times, regardless of L2 proficiency or L1 automaticity (Kroll & Bialystok, 2013).

Bilingual memory characteristics depend on two opposing accounts of language processing: automaticity and desirable difficulty. The automaticity account suggested accessing concepts through lexical representation is more automatic in L1, whereas the desirable difficulty account suggested difficulty of L2 lexical access requires more processing effort and translates into memory advantages (Arndt & Beato, 2017; Bjork, 1994). It is also argued that as L2 proficiency increases, the processing approaches that

[1]Cognate describes words similar in form and meaning across languages. E.g., tomato in English, tomaat in Dutch.

of L1 and becomes more automatic as well. Arndt and Beato (2017) showed that false memories are produced more in L1 compared to L2, and that high L2 proficiency participants have higher L2 false memories than low proficiency, which they attributed to automaticity account. Lastly, they found that for those less proficient, true memory performance is higher in L2, which they found aligned with the desirable difficulty account. Mortensen et al. (2015) have found faster retrieval with an L1 cue (vs. L2) and attributed it to the time it takes to translate L2 input into L1. If L1 access is automatic and L2 may be translated into L1 to be processed, we can argue that when learning in L2, the encoding can still occur in L1, or mixed with two languages, especially if L2 proficiency is low (Esposito & Baker-Ward, 2016).

The bilingual mind is organized differently from the monolingual counterpart on the account of having lexicon and concepts in different languages. The activation and proficiency predict how automatically the languages are (differentially) processed. This, in return, affect the subsequent memory production for information coming in different languages, such as language tagging.

1.2 *Bilingual Memories*

Researchers investigating the lexical organization of languages put forth the notion that bilinguals tag memories with the language of encoding, and awareness of linguistic networks in dual-language situations might be an indication for it. The language tag is viewed as part of non-semantic information about items, metalinguistic knowledge, and an awareness of language membership (Dornic, 1979; Paradis, 2004). Tags help people account for stored concepts, and they can be used as retrieval cues due to having links to the memory (Altarriba & Soltano, 1996). They have also been researched to gain insight into the bilingual organization. Better recall of unilingual lists compared to dual-language ones would support the common-storage hypothesis. Superior dual-language list recall, and perfect tagging ability would suggest that language-specific organization strategies are employed (Liepmann & Saegert, 1974).Their results. Saegert et al. (1975), attributed the results supporting the common-storage hypothesis to a contextual effect. They reasoned it is easier to attend to the language of words in isolation, as opposed to in sentences where semantic or contextual information need be encoded as well.

Some researchers opposed using tag processes to infer the lexical organization. Dornic (1979) argued that both common-storage and separate-storage mechanisms can be activated depending on the task demands. Lastly, Grosjean (2001) argued that bilinguals vary on the continuum of "language modes" and the organization of semantic representations would depend on how much each language is activated.

When discussing the findings in studies with bilinguals, it is important to consider learning histories, contexts of use, structural differences between languages, and relative dominance of each language (Riehl, 2010), and to refrain from treating bilingualism as a categorical or unitary construct (Kroll & Bialystok, 2013). Differences in L2 acquisition, level of immersion, and proficiency can produce results that contradict traditional models. Previous research and models have operated at the word-level to conclude the bilingual memory organization and language tagging. However, information processing at a more complex level may change how we use such models to investigate L1-L2 differences.

1.3 *Inference-Making and Memory Errors*

Inference-making is a process that allows us to acquire more information than we receive externally. As an internal source of information which uses our previous knowledge about the world and combines it with the provided information to create a coherent meaning, it facilitates comprehension (McKoon & Ratcliff, 1992). It is a complex ability, appropriate to study bilinguals' information processing both within languages and between, when two languages need to be synthesized to extract information. Its constructive characteristic allows basing on memory theories, while sentence-level paradigms allows for externally valid dual-language information.

There are several types of inferences. Pragmatic inferences are deductions that may be misaligned with the speaker's intentions. In the example, "The baby stayed awake all night," the receiver may use previous knowledge to infer that the baby cried, but it may or may not be an inaccurate representation of the event. On the other hand, logical inferences have preserved meaning, where one sentence necessarily implies the other (Brewer, 1977). In the example, "A is bigger than B," the logical inference that follows is, that B must be smaller than A. This type of inferences has been investigated using comparison and spatial paradigms, and it uses prior knowledge and assumptions about logic and negation, and is the focus of this study.

There are several theories attempting to explain how and when inferring occurs during reading comprehension. Bransford and Franks (1971), in their assimilation theory, argued comprehension is constructive, and semantic representations are derived from sentence information, context knowledge, and assimilative integrative operations. The caveat of constructing meaning is that we are unlikely to distinguish our interpretations from the input and that much more likely to construct erroneous memories about what was presented (Barclay, 1973; Bransford et al., 1972; Harris, 1974; Harris & Monaco, 1978; Johnson et al., 1973). It was found that logical inferences derived from consecutive sentences lead to false recognition, supporting their theory (Bransford et al., 1972).

Opposition to earlier constructionist studies' came from Reyna and Kiernan (1994) who, based on the Fuzzy Trace Theory, found that subjects sometimes falsely recognize meaning, despite having accurate verbatim memories, and that correct recognition of surface information and misrecognition of meaning are independent of each other. They argued that memory is not constructive since it is represented in two parallel forms, in gist form and in verbatim form (Reyna & Brainerd, 1995; Reyna & Lloyd, 1997). They suggested old and inferred sentences would have to be dependent to consider memory to be constructive, which Bransford et al. (1972) did not find. The researchers defined "verbatim" as the actual experience at the surface structure, whereas "gist" refers to our understanding of the experience and the story's general essence. They argued that memory performance is based more on gist than verbatim because individuals use flexible reasoning and inferring to process information (Reyna & Brainerd, 1995). Also because gist traces are more accessible and durable, and less effortful. However, they are also more malleable, as the representations can be phrased differently to yield the same meaning. Immediate testing revealed that while encoding is unbiased and both representations are stored, verbatim traces decay more rapidly and evolve to activate gist traces only (Brainerd & Reyna, 2002, 2004; Koriat et al., 2000; Sachs, 1967).

Due to internal suggestion, inference-making is a process that leads to gist processing, and subsequently, to memory errors when verbatim information is required (Reyna & Lloyd, 1997). If, on the other hand, the task does not demand verbatim or surface details (e.g. specific details, language), then the inference may be regarded as part of true memories.

It was found that false memories have similar phenomenological qualities as true ones, and can yield similar “remember” judgments (Roediger & McDermott, 1995), as opposed to “know” judgments (Tulving, 1985). If the gist traces are strengthened instead of verbatim, high confidence and remember judgments are observed because asking about inferences is tantamount to implanting false memories (Reyna et al., 2016). Indeed, it was argued that the memory's strength is different from the subjective measure, as participants often have metamemory illusions (Bjork, 1994), reporting high confidence for false memories (Payne et al., 1996). Bransford and Franks (1971) found that people are similarly confident in (falsely) recognizing new or inferred sentences as old sentences.

In this study, we are scrutinizing the inferring process for people who can process more than one language. We are bridging inference-making and gist representation in their relation to memory errors at the sentential level. This diverges from word-level paradigms investigating reaction time to translate, which reduces information complexity. The sentence level is appropriate to sustain experimental control and test the synthesis of information coming in in two languages. Proficiency in L2 creates automaticity in language access that we deem similar to gist extraction and reliance, evident by memory errors and relevant memory characteristics. The gist, conceptually-driven, should be retained in a common language store where language tags are not reportable. The verbatim, on the other hand, is tied to the surface details that include the language; therefore, a language-separate storage and high tagging performance are expected. The previous research have theorized for all bilinguals;

however, individual differences in how people process and retain linguistic information should also be considered.

1.4 Individual differences in Cognitive Functions

Executive functions (EF) refer to individual differences in mental processes that allow flexibility in thoughts, selective processing, and inhibition of information, updating, goal-directed behaviors, conflict monitoring, and conflict resolution (Hussey & Novick, 2012; Ye & Zhou, 2009). Three shared but distinct functions of EF have been identified: working memory, inhibition, and shifting. In linguistic goals, we need working memory (WM) to follow and comprehend the narrative as the message/goal unfolds, inhibition to resist competing meanings, and shifting to parse sentences (Key-DeLyria & Altmann, 2016; Ye & Zhou, 2009). WM ability is especially shown to be important for the storage of both gist and verbatim information (Brainerd & Reyna, 1990). The "unity and diversity" pattern explains why sometimes EF is referred to as a general ability and sometimes as dissociated functions with task impurity (Friedman & Miyake, 2017). In this study, we acknowledge this impurity and use several tasks to discuss a general association of cognitive functions with bilingual inference-making. However, we do not posit differential hypotheses for different functions.

EF components play an even more significant role in the context of bilingual processing. Bialystok (2009) argued that the attention and conflict resolution boosts of bilingual children are carried into adulthood, and bilinguals perform better at task switching with fewer mix costs and experience delayed cognitive aging; they, however, perform worse at lexical retrieval and verbal fluency tasks (Bialystok, 2007, 2009; Bialystok et al., 2012). She posited greater linguistic and cognitive outcomes are associated with bilingualism's higher (proficiency) degrees and even greater with multilingualism.

Bilinguals use working memory to sustain attention on the appropriate language while inhibiting the inappropriate language and words in its network ((Bialystok et al., 2009; Rodriguez-Fornells et al., 2002). This inhibition is necessary because both language representation systems are activated, even when only one is in use (Bialystok, 2007). Therefore, two active and competing systems need controlled attention to manage the linguistic representations and suppress the inappropriate one (Long et al., 2019). The Dual-mechanism framework refers to this process as two-fold: the “proactive” mode, where the bilingual tries to keep the appropriate language in mind, and the “reactive” mode, where the inappropriate language is inhibited to avoid interference (Braver, 2012), and that the bilingual advantage comes from parallel use of cognitive control (Costa et al., 2006). Liu et al. (2016) found that cognitive flexibility (shifting) plays a role in language switching, similar to task switching, by exercising inhibitory control.

The two languages may not always compete for activation. In a dual-language context, switching between languages occurs spontaneously (Green & Abutalebi, 2013). This context is deemed more demanding on cognitive mechanisms, such as goal maintenance, and requires better balancing of proactive and reactive modes to select, produce, or inhibit the appropriate language at all times. L1 generally has a privileged status since its activation is more automatic, and it needs to be inhibited to accommodate L2 production (Kroll & Bialystok, 2013).

EF is important in how they are linked to reading comprehension and bilingual cognition, and to be able to understand how bilingualism differs from monolingualism in its ties to cognitive systems. However, comprehension can surpass individual differences in linguistic processing, and include non-linguistic representation of information, such as visualization. Mental imagery is the second construct varying across individuals that plays a role in information processing and is shown to differ across bilinguals' languages, referring to internal representations of an experience. Good imagers are defined as those who can experience and report information as vivid images, as opposed to those who report a more abstract and phonology-based form (Keogh & Pearson, 2011).

Memory research shows that higher vividness of mental imagery predicts richer and clearer details in memory, more frequent retrieval of memories as visual images, faster retrieval, and an advantage for true recall of highly imageable, concrete words (D'Argembeau & Van der Linden, 2006; Reyna et al., 2016; Vannucci et al., 2016). However, relying more on visual mental imagery might also facilitate the stimulation of false memories as visualization creates memory traces that compete with the actual event (Eisen et al., 2013; Hyman & Pentland, 1996).

Imagery is also related to the organization of bilingual information. Conceptual Feature Model suggested that concrete words have an advantage over abstract and emotion words because they share more features (representations) across languages (De Groot, 1992). The Bilingual dual-coding theory posited related but independent verbal and perceptual modalities of information, and that separate L1 and L2 verbal systems for concepts and a common image system for translation for concrete words exist. The studies found bilingually coded items (translations) are recalled better than monolingual synonyms, demonstrating the additive effect of two storages (Paivio et al., 1988).

1.5 *The Present Study*

This study aims to understand whether inference-making leads to false recognition of encoded information differentially in the context of bilinguals, whether bilinguals undergo a similar process in both L1 and L2, and whether the memories are tagged with the language of encoding. As it involves retrieving prior knowledge and active meaning-making, inference-making should be regarded as a complex mental process, appropriate for exploring the memory networks in different languages.

When asked whether a sentence was previously presented, people are prone to falsely recognizing sentences representing the inferences, which we argue is a similar process to gist encoding. However, the process of logical inference-making has been studied almost exclusively in L1. To deduce whether this process differs for L1 and L2, we presented Turkish-English bilingual participants in Turkish and English unilingual, and dual-language (DL) sentences that allowed them to configure objects mentally and draw spatial inferences.

Among the various inference types previously introduced, we chose to study logical inferences, where accuracy is more easily quantifiable compared to other types, as opposed to pragmatics where inferences are limitless and highly subjective. We deemed a spatial task appropriate because the object positions would be unequivocal, the meaning would be preserved across sentences, and the concreteness and imageability in word processing could be kept a constant. The literature has focused on how bilinguals infer word meaning during reading in L1 and L2, and found that inferring in L2 is overall more difficult, while facilitated with higher proficiency (Karlsson, 2014; Prior et al., 2014). We aim to test a similar premise at the sentential level.

We included a dual-language condition in our task to simulate bilinguals' everyday experience, where they integrate and store different information and switch between languages both during comprehension and production. This naturally occurring context provides insight into bilingual memory organization, as well as the cognitive abilities associated with bilingualism.

Several individual differences on cognitive resources were explored with respect to the literature where we established their link with inference-making and bilingualism: L2 proficiency, executive functioning, and mental imagery. English proficiency measures would unveil the extent to which proficiency differences are related to performance between language conditions. To understand how EF plays a role overall, we incorporate the three main functions. Since inference-making involves synthesizing incoming information with priors, working memory ability would be an important predictor of task performance in all conditions. In the dual-language condition, this process would be additionally taxing; requiring participants to switch back and forth between languages and inhibit the competing language, exerting inhibition and shifting ability in both encoding and retrieval tasks. Inhibition and shifting abilities may predict performance in the L2-only condition, where participants need to inhibit processing or translating in L1 to retain L2 information. Lastly, in all conditions, the visualization strategy, which comes automatically to good imagers, may need to be inhibited to encode the verbatim information. Mental imagery was of interest to us, as the task involved configuring concrete objects as the basis for drawing inferences and understanding spatial relations.

We conducted two experiments to investigate if the sentences that were not presented in the study phase were recognized during testing due to inference-making. The second experiments differed in the stimulus exposure duration; designed to weaken verbatim traces, and increase task difficulty upon analyzing the data from the first experiment. Following the recognition task, we examined object configuration performance in different language conditions. We obtained accuracy, response time, and subjective confidence ratings from this task. Lastly, we investigated whether the participants recalled the language in which the sentence was presented (i.e., the language tag) in the dual-language condition.

Chapter 2:
EXPERIMENT 1

In Experiment 1, we examined if inferences were falsely recognized as inputs in one's L2 and a dual-language context, and if spatial configurations are made with similar accuracies in language conditions. Lastly, we investigated the existence of a language tag in a spatial inferential task.

Are inferred sentences falsely recognized as input?

We hypothesized that old and new sentences would be correctly recognized and rejected, respectively, whereas inferred sentences will be falsely recognized. In the Turkish condition, this will be aligned with gist representation in L1 due to the automaticity account (Bjork, 1994). Whether this account is valid for making inferences in L2 will depend on L2 proficiency. Therefore, in the English (L2) and dual-language (DL) conditions, we expect more highly proficient bilinguals to recognize inferred sentences more.

We also tested individual differences in cognitive resources that might affect false recognition of inferred sentences. All hypotheses concerning executive functioning components have the same direction. Working memory will help with the overall integration of input; and the ease of integration may facilitate gist representation over verbatim in L1 and L2. Especially in the DL condition, those who have higher working memory, inhibition, and shifting abilities will create the gist more easily and recognize inferred sentences. Lastly, we hypothesize that high imagery participants will recognize inferred sentences more, as they visualization is similar to relying on gist representations.

We will explore how response times in their relation to accuracy and performance. We will explore how the confidence ratings differ based on language condition. In the dual-language condition, we expect all participants' confidence scores to be overall lower than in Turkish and in English. Furthermore, we expect inferred sentences to be falsely recognized with similar confidence to old sentences.

How accurately are objects configured?

We expect higher configuration accuracy in unilingual conditions (Turkish and English) compared to dual-language. In the DL condition, higher proficiency in L2 will be associated with higher configuration accuracy.

We expect higher WM will predict higher accuracy in all conditions. In the dual-language condition, shifting may predict higher accuracy by allowing the two languages to be processed efficiently. We expect mental imagery to be a significant predictor in configuration accuracy, such that high imagery will predict higher configuration accuracy. This association will be stronger in the Turkish condition because of both encoding automaticity and ease of concept access in L1. We suppose this ease could be related to that of visualization also. Lastly, we will explore response times and confidence ratings.

Do we tag memories with the language of encoding during spatial inference-making?

We expect less proficient bilinguals to be more accurate in indicating the language tag due to the desirable difficulty account. Overall, we expect that L1 will be encoded in gist form, whereas L2 will be encoded as verbatim. As L2 proficiency increases, the nature of mental manipulations in L2 approaches L1. L2 proficiency, working memory, inhibition, shifting, and mental imagery facilitate the ability to configure the objects and integrate the gist. Difficulty in integrating the gist will result in awareness of the language tag.

2.1 Method

2.1.1 Participants

Thirty-four undergraduate students were recruited from Koç University subject pool (28 female). A prior power analysis using G-Power software indicated a sample size of 30 would be sufficient for analyses of 3 groups repeated-ANOVA with effect size of 0.2 and with power of .80 (Faul et al., 2007). Participants were rewarded with course credit. All participants were native Turkish speakers who spoke English as L2. Mean age of acquisition of English was 7 (*SD*=3.24), and acquisition took place in school settings. Exclusion criteria included scoring 2 SD higher and lower response time on 3 out of 9 trials of the Inference-Making task. After exclusions, data analysis was conducted with 30 participants whose ages ranged between 20-27 (M_{age}=23.33, *SD*=1.64).

2.1.2 Measures

Inference-Making Task. We created this task using PsychoPy3® software and administered it on the computer (Peirce et al., 2019). On the screen, we presented the participants with a group of four sentences in each trial. These sentences gave spatial information about four objects (e.g., The leaf is above the bookmark / The flower is above the leaf). We chose the objects to be stackable (to preclude any inference based on object property), concrete, and commonly known words in Turkish and English. The sentences remained on the screen for 90 seconds. We instructed the participants to study the sentences and to form the object configuration on a vertical plane (e.g., Flower/Leaf/Bookmark top to bottom) and to complete this task mentally. After the study duration, they were asked to indicate whether the following sentences were seen during the study phase. For each group of sentences, the participants were asked 8 Old/New questions. The program presented these sentences in random order from a pool of statements, such that three sentences would be old (e.g., The leaf is above the bookmark), two sentences would be new but inferable (e.g., The flower is above the bookmark), and three sentences would be new and not inferable (e.g., The bookmark is above the leaf). Participants responded by pressing keys on the keyboard. They rated their confidence on a 5-point scale after each response. We measured the accuracy of answers and response times.

We then asked the participants to indicate whether the following sentence was correct with respect to the configuration, even though they have not studied the exact sentence. In each trial, we presented the participants with 8 statements, four correct and four incorrect. The program presented these sentences in random order. Participants responded by pressing keys on the keyboard. They rated their confidence on a 5-point scale after each response.

This task consisted of three within-subjects language conditions. In English and Turkish conditions, all the sentences and the instructions were presented in that language. In the dual-language condition, half of the study sentences were in Turkish, and the other were in English (see Appendix A). The built-in instructions were in Turkish. Additionally, in the DL condition, the participants indicated in which language they had studied the sentence. They indicated this for all four sentences they had seen in the study phase.

Each language condition consisted of a practice trial and six test trials. The six within language trials were administered in random order. The language conditions were blocked and counterbalanced. Out of the six trials, two trials consisted of sentences only using the indication "above/üzerinde," two trials using "below/altında," and two trials where they were mixed (two above, two below). In the DL condition for the mixed indication trials, sentences with above and below were distributed equally. Overall, the participants completed 18 trials and three practice trials that took approximately 70 minutes. The participants were not instructed to respond as quickly as possible as we did not want speed-accuracy tradeoff; rather, we wanted to approach the everyday process of learning and responding. Therefore, the time measure we obtained is hereafter referred to as response time (RT).

2-back. Twenty-five numbers (ranging from 1 to 9) were presented back to back on the computer screen. The participant decided whether the stimulus matches what was shown two steps ago by pressing designated keys at each screen. We calculated d-prime accuracy to measure working memory ability.

Letter number sequencing (LNS). The participants were shown a sample of scrambled digits and letters on the computer screen. In the first half of the task, they were asked to recall the items in order by typing on the computer, exercising short term memory. In the second half, they were asked to recall the numbers in ascending order and the letters alphabetically. We summed accurate responses. This task was used to measure shifting ability.

Trail Making Test- Parts A and B. In Part A, participants were asked to connect the circled numbers by drawing lines as quickly as possible (i.e., 1-2-3). In Part B, they were asked to connect numbers and letters alternatingly (i.e., 1-A-2-B, see Appendix B). This task was found to measure visual scanning, graphomotor speed, mental flexibility, fluid cognition, and alternation (Misdraji & Gass, 2010; Salthouse, 2011). We administered this task with paper and pen, and we measured the time it takes to complete the task. Lower reaction times denote better performance.

Spontaneous Use of Imagery Scale (SUIS). The scale by Reisberg et al. (2003) consisted of 12 questions that measure the tendency to use visual mental imagery in daily life. The scale was administered on the computer, and the participants indicated the accuracy of the descriptions on a 5-point scale (e.g., "When I first hear a friend's voice, a visual image of him or her almost always springs to mind.") Total scores ranged between 12 to 60, higher scores indicating more use of mental imagery. The questionnaire had adequate reliability scores with Cronbach's alpha α=.83. We administered the scale's Turkish translation (Doenyas et al., in preparation; see Appendix C).

The Language Experience and Proficiency Questionnaire (LEAP-Q). The questionnaire by Marian et al. (2007) consisted of open-ended and point scale questions that assessed the self-report English proficiency of participants. It included questions about the age of acquisition, contexts of use, and proficiency in different linguistic domains. We administered the questionnaire on the computer.

2.1.3 Procedure

The experimental sessions took place in the computer lab. We assigned the participants to a predetermined counterbalanced order of tasks. We told them that they were going to have a test of memory, consisting of three parts. Upon signing the consent form, we told them that they would study a group of sentences and be tested on two types of questions (old/new and correct/incorrect). The inference-making task was administered in blocks. All language conditions started with a practice trial, followed by 6 test trials. The individual difference tasks were conducted between language blocks. Lastly, the participants completed the language questionnaire online. The experimental session lasted 90 minutes.

2.2 Results

2.2.1 Descriptive Statistics

Participants completed the EF tasks and self-reported their use of mental imagery and second language. In the language questionnaire, the participants indicated their age of acquisition for English as ranging from 0 to 11 years (*M*=7, *SD*=3.23) and years of language instruction as 4 to 19 (*M*=12.03, *SD*=4.33). We identified three dimensions of L2 use, using correlational and factor analyses. Scholastic use consisted of self-reported proficiency in understanding, speaking, reading, and writing in L2 and the frequency of mental translations. Use to communicate with people consisted of the frequency of L2 use with self, significant others, and friends. Lastly, daily life use consisted of the frequency in which participants read publications, consumed media, and the degree of emotiveness of L2, which they indicated on a 7-point Likert scale. We present the descriptive information and correlations in Table 2.1.

Table 2.1: Descriptive statistics and correlations of study Variables in Experiment 1.

Task	M	SD	2-back	Trail-A	Trail-B	SUIS	LNS-A	LNS-B	Scholastic	Comm	Daily life
2-back	21.77	2.66		-0.17	-0.20	-0.23	-0.10	0.28	0.25	0.01	0.31
Trail-A	26.58	8.30			**0.45***	-0.16	-0.27	-0.33	-0.12	**0.45***	-0.28
Trail-B	45.89	18.33				0.19	-0.22	-0.40	-0.09	0.16	-0.09
SUIS	3.34	0.63					0.09	0.18	**-0.41***	-0.05	0.07
LNS-A	61.14%	11.98						**0.43***	0.10	-0.24	-0.07
LNS-B	45.83%	17.70							-0.19	-0.23	0.04
Scholastic	5.54	0.88								0.08	**0.42***
Communicative	2.80	0.96									0.27
Daily life	5.36	1.06									

Notes. SUIS= Spontaneous Use of Imagery scale, LNS=Letter Number Sequencing, Comm=Communicative use.
*p<0.05.

2.2.2 Order effects

We did not find the effect of the language condition order on recognition or the configuration tasks. Therefore, it is not included in the subsequent analyses.

2.2.3 Sentence Recognition

We conducted a 3x3 within-subject ANOVA with language condition (Turkish, English, DL) and sentence type (old, inferred, new) to examine the likelihood of recognizing inferred sentences as seen. We calculated recognition percentages by dividing count of recognition the number of sentences presented in each type. We found that there was no significant main effect of language $F(2,56) = 2.706$, $p = .076$. However, there was a significant main effect of sentence type, $F(2,56)=185.62$, $p<.001$ $\eta^2=.869$, and a significant interaction of language and sentence type, $F(4,112)=5.405$, $p=.001$, $\eta^2=.162$, with observed power of .97. Pairwise comparisons conducted using Bonferroni correction ($\alpha=.05$) showed that there were no language differences within old and new sentences. However, inferred sentences were recognized significantly less in the DL condition than the Turkish and English conditions, and they were recognized similar to new sentences. The interaction is represented in Figure 2.1.

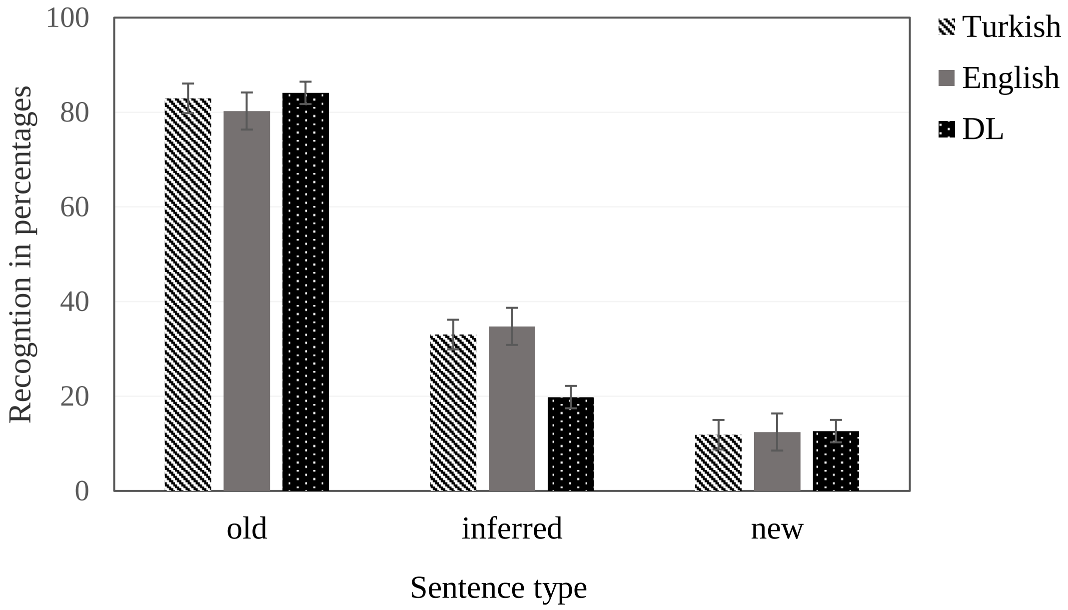


Figure 2.1: Percentage of recognized sentences for old, inferred, and new sentences. *Note.* DL=dual-language.

Before conducting separate within-subjects ANCOVAs for sentence type, we conducted correlation analyses to determine which covariates were appropriate (see Table 2.2). We included LNS-B as covariate for old sentences; and found it to be significant predictor, $F(2,46)=4.418$, $p=.018$, $\eta^2=.167$. We obtained the parameter estimates to examine the effect of the interval variable. Higher LNS-B scores predicted higher recognition in English, $\beta=.575$, $p=.006$, $\eta^2=.298$, but not in Turkish or dual-language. For inferred sentences we added 2-back, LNS-B, scholastic, and daily life use of English as covariates. We found significantly lower recognition of DL compared to the unilingual ones but there were no significant covariates. For new sentences we added 2-back, LNS-A, LNS-B, scholastic, and daily life as covariates. Only daily life use was significant, $F(2,36)=4.321$, $p=.021$, $\eta^2=.193$. Parameter estimates revealed higher scores in daily use of English predicted lower recognition in the English condition, β= -.570, *p*=.005, η^2=.366, but not Turkish or dual-language.

Table 2.2: Correlations of recognition performance with individual differences.

	2-back	Trail-A	Trail-B	SUIS	LNS-A	LNS-B	Scholastic	Comm	Daily life
TR old	0.22	0.12	-0.01	0.00	0.02	0.27	-0.11	0.06	0.01
ENG old	0.19	-0.15	-0.13	0.37	0.19	**0.55***	-0.12	-0.02	0.21
DL old	0.15	0.23	0.08	-0.25	-0.14	0.23	-0.13	-0.02	-0.28
TR inferred	**-0.42***	0.32	0.14	0.18	-0.23	**-0.41***	**-0.45***	0.17	**-0.46***
ENG inferred	**-0.40***	0.24	-0.09	-0.03	-0.25	-0.32	-0.19	0.31	-0.18
DL inferred	-0.17	0.10	0.10	0.01	-0.04	-0.34	0.04	0.23	0.13
TR new	**-0.43***	0.18	0.22	0.13	**-0.52***	-0.40	**-0.45***	0.05	**-0.45***
ENG new	-0.17	0.35	0.24	-0.21	**-0.47***	**-0.41***	-0.24	0.13	**-0.49***
DL new	-0.12	0.17	0.30	0.15	**-0.45***	**-0.68***	-0.22	0.18	-0.16

Table 2.3 shows correlations between sentence types for each language condition. We found that recognition of old and new were correlated in English, but not in Turkish and dual-language conditions. Two types of false recognition (inferred and new sentences) were correlated in Turkish and dual-language, but not in in English.

Note. *denotes p<.05.

Table 2.3: Correlations of true and false recognitions.

	TR inferred	TR new	ENG inferred	ENG new	DL inferred	DL new
TR old	0.003	-0.231				
TR inferred		0.523*				
ENG old			-0.119	-0.699*		
ENG inferred				0.324		
DL old					0.092	-0.141

DL inferred	0.371*

Note. *denotes $p<.05$.

2.2.4 Sentence Recognition Response Times

We present the descriptive statistics of response times in Table 2.4. Before conducting response time analyses, we conducted correlational analyses between recognition accuracy and their respective response times, and found no significant correlations.

We conducted between-subjects ANOVA[2] with language, sentence type (old, inferred, new), answer given (yes/no to recognition), and subjects as random factor to test how response times differed for sentence recognition. The analysis allowed us to factor out individual differences, and we found significant interaction of language and answer, $F(2, 81.44)=5.188$, $p=.008$, $\eta^2=.113$. “Yes” responses to recognition took significantly more time in DL (vs. English); “no” responses took the least time in DL (significantly less than English). While Turkish RTs did not differ across answers, in English, “yes” responses took less time to indicate compared to “no”, and the opposite pattern in DL (see Figure 2.2).

Table 2.4: Means and standard deviations of recognition times.

		Mean (sec)	Std. Deviation
Turkish	Old	3.68	1.18
	Inferred	4.01	1.26

[2] We deemed this analysis appropriate due to the censored dependent variables, where we did not observe data for certain type-answer interactions for fully accurate participants. Same method was preferred for remaining RT and confidence analyses.

	New	3.70	1.02
English	Old	3.98	1.35
	Inferred	4.36	1.65
	New	4.04	1.30
Dual-language	Old	4.13	1.36
	Inferred	4.01	1.58
	New	3.60	1.39

We explored response times to see if they were related to inferring and configuring performances. We found significant interaction of language and subject $F(61, 26.83)= 2.039$, $p=.022$, $\eta^2=.823$. We also found significant three-way interaction of language, answer, and subject, $F(57, 191.54)=1.560$, $p=.026$, $\eta^2=.467$, and sentence type, answer given, and subjects, $F(57, 104.39)=2.458$, $p<.001$, $\eta^2=.573$. Pairwise comparisons using Bonferroni correction revealed that "no" responses took the most time in old sentences (M=4.65, SD=.17), significantly less in inferred (M=4.14, SD=.09), and the least in new sentences (M=3.79, SD=.06).

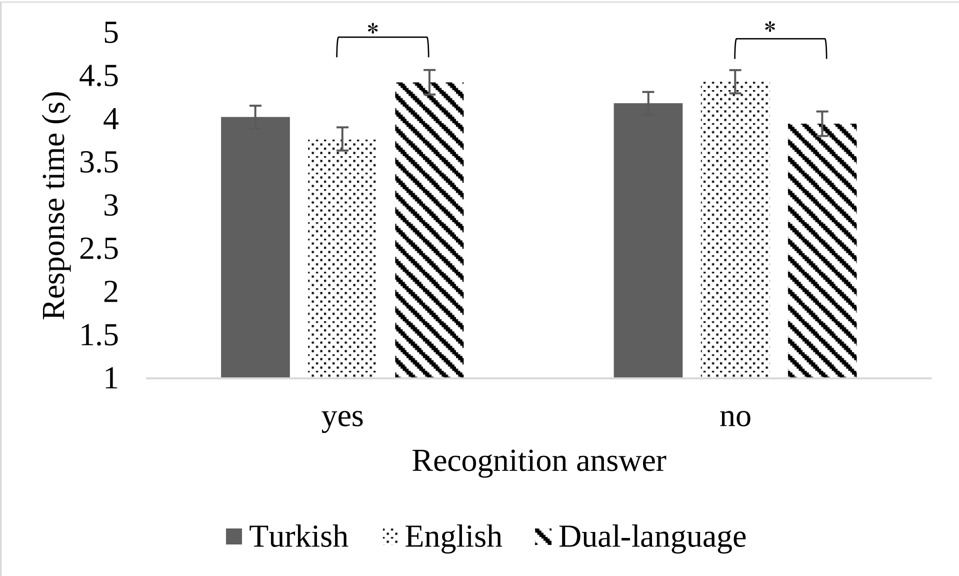


Figure 2.2: Recognition response times language by the answer given.

2.2.5 Correct Configuration

To examine how configuration accuracy differed across language conditions, we conducted within-subjects ANOVA. We obtained high accuracy in all conditions, and no difference between conditions, $F(2,56)=1.347$, $p=.268$, $\eta^2=.046$. To explore the effect of language proficiency and individual differences on accuracy, we conducted correlational analyses (See Table 2.5). In light of correlational analyses, we conducted within-subjects ANCOVA with language as predictor, and LNS-A, scholastic use, and daily use as covariates. None of the variables predicted difference between conditions; however, LNS-A emerged as a significant covariate $F(1,24)=7.736$, $p=.010$, $\eta^2=.244$, and observed power .761. Parameter estimates revealed significant values for $\beta_{Turkish}=.42$, $p=.016$, and $\beta_{English}=.51$, $p=.004$, but not for DL.

Table 2.5: Means of configuration accuracy by language and correlations with individual difference variables.

	M** (SD)	2-back	Trail-A	Trail-B	LNS-A	LNS-B	SUIS	Scholastic	Comm	Daily
TR	41.10 (1.36)	0.19	0.12	-0.26	0.36	0.16	-0.03	**0.42***	0.22	0.27
ENG	40.31 (1.70)	0.22	0.08	-0.19	**0.46***	0.20	0.03	**0.38***	0.23	0.27
DL	39.21 (1.55)	0.30	-0.13	-0.18	0.25	0.26	-0.02	0.33	0.04	**0.45***

Note. DL=dual-language, SUIS=Spontaneous Use of Imagery Scale, LNS=Letter-Number Sequencing, Comm=Communicative use, Daily = Daily use of English

* denotes p<.05.

**Means provided out of 48 sentences.

2.2.6 Correct Configuration Response Times

To test if configuration RTs were predicted by language condition, sentence type (correct-incorrect), answer given (yes/no), and subject (as random factor), we conducted between-subjects ANOVA. We found a main effect of answer, $F(1,54.26)=7.985$, $p=.007$, $\eta^2=.128$, saying “no” took longer time. The three-way interaction of language, type, and answer was not significant; however, the four-way interaction became significant with the addition of subject, $F(31,4233)=1.708$, $p=.009$, $\eta^2=.012$, see Figure 2.3. For correct English sentences, “no” responses took longer time than “yes”; and “no” responses took longer for correct sentences compared to incorrect.

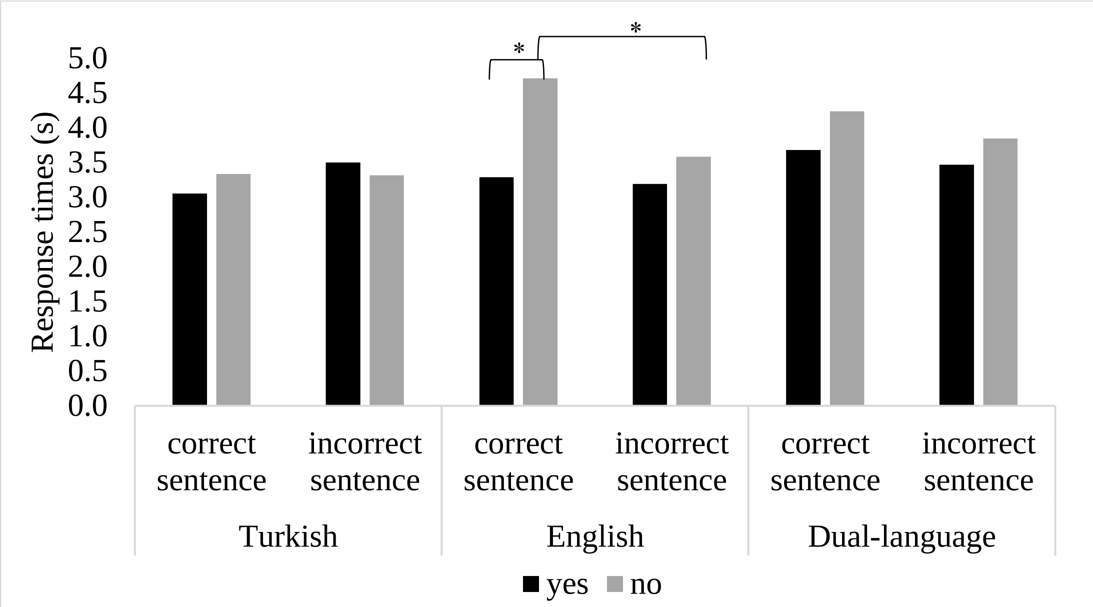


Figure 2.3: Configuration response time.

2.2.7 *Language Tag*

Most participants correctly identified the language tag. For Turkish sentences, 83.58% of language tags (*M*=10.03 out of 12) and English sentences, 81.33% (*M*=9.76 out of 12) tags were correctly identified. Paired t-tests revealed that the tag was not differently identified in Turkish and English, *t*(28)=.57, *p*=.573. While Turkish tag correlated with Trail-B and LNS-B tasks, regression analyses with the two covariates did not reveal they were significantly predicting Turkish tag performance, or difference between Turkish and English tags.

Paired T-tests on rection time revealed Turkish tag was identified faster (*M*=3.78, *SD*=1.24) than English tags (*M*=4.23, *SD*=1.26), *t*(28)=-2.231, *p*=0.34.

2.2.8 *Confidence Ratings*

Recognition Confidence

We explored confidence ratings for different sentence types, because false memories were also found to elicit high confidence ratings, or remember judgments

(Tulving, 1985). We analyzed the confidence ratings (on a 1-5 point scale) given to each recognition response. The ANOVA revealed significant two-way interactions of language and answer, $F(2,87.58)=4.203$, $p=.018$, $\eta^2=.088$, and sentence type and answer, $F(2, 76.65)=27.019$, $p<.001$, $\eta^2=.413$; and significant four-way interaction of language, sentence type, answer, and subject, $F(81,4039)= 1.844$, $p<.001$, $\eta^2=.036$. Pairwise comparisons using Bonferroni correction on the language by answer interaction revealed "yes" responses had lower confidence in DL ($M=3.99$, $SD=.05$), compared to English ($M=4.16$, $SD=.05$). "No" responses had significantly higher confidence in DL ($M=4.37$, $SD=.03$) compared to unilingual conditions ($M=4.20$, $SD=.03$). Type by answer interaction revealed "yes" responses had higher confidence in old ($M=4.35$, $SD=.02$), compared to inferred ($M=3.96$, $SD=.05$) and new ($M=3.85$, $SD=.07$), while "no" responses for new was rated with the highest confidence ($M=4.57$, $SD=.02$), inferred second ($M=4.29$, $SD=.03$), and old lowest ($M=3.89$, $SD=.05$).

Configuration confidence

We conducted 3x2x2 ANOVA to examine the confidence ratings for the configuration task, using subject as random factor. The analysis revealed two-way interactions of language and sentence type, $F(2, 67.48)=3.738$, $p= 0.029$, $\eta^2=.100$; language and answer type, $F(2, 65.47)= 4.101$, $p=0.021$, $\eta^2=.111$; and sentence type and answer, $F(1,30.75)=23.837$, $p<.001$, $\eta^2=.437$. The four-way interaction was also significant with the addition of subject, $F(31, 4233)=4.084$, $p<.001$, $\eta^2=0.029$.

We conducted pairwise comparisons using Bonferroni correction ($\alpha=.05$) to examine the two-way interactions. Language by sentence type revealed correct sentences had significant differences between all language pairs: Turkish had the highest confidence, while DL had the lowest (See Figure 2.4); incorrect sentences did not differ in their confidence ratings. Language by answer revealed no difference in "yes" responses, whereas for "no" responses, Turkish had higher confidence than English and DL. Sentence type by answer given revealed "yes" responses to correct sentences (vs. no) and "no" responses to incorrect sentences (vs. yes) had higher confidence ratings.

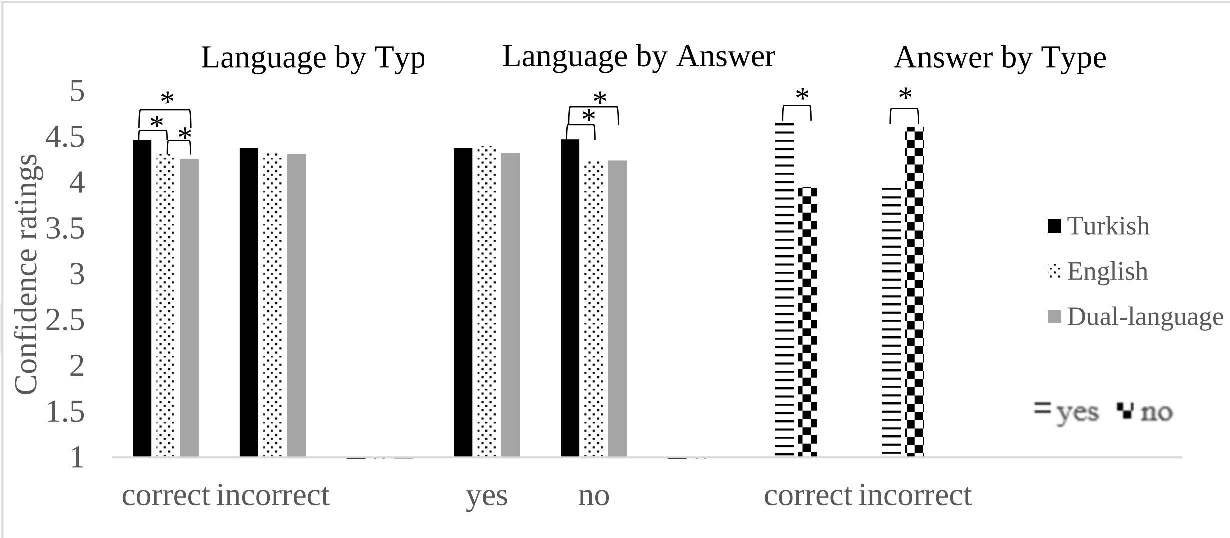


Figure 2.4: Confidence rating interactions.

2.3 *Discussion*

We investigated whether inferring spatial information would lead to sentences being falsely recognized when inferences are presented in L1, L2 and dual-language conditions. We confirmed our hypotheses that old and new sentences would be accurately recognized and rejected. Inferred sentences were indeed recognized more than new sentences, suggesting false recognition. Our recognition results are aligned with the Fuzzy Trace Theory's assumptions. We found inferred sentences were false recognized, but also correct recognition of verbatim information (old) and misrecognition of gist (inferred) were independent of each other in all conditions (Reyna & Lloyd, 1997). The percentages did not approach recognition of old sentences, suggesting while there is a pattern of gist encoding, the verbatim traces are stored alongside it.

We could not support our hypothesis that we are more prone to gist encoding and false recognition in L1. The automaticity and desirable difficulty accounts differentiating L1 and L2 are currently unsupported. It was found that participants can false recall more lure words in L2 if they were extensively exposed to L2 (Anastasi et al., 2005). The Revised Hierarchical Model claimed high L2 proficiency is what would explain L1-like automatic processing (Kroll & Stewart, 1994). Not only did we not find differences between languages, but also no moderating effect of proficiency. Since we selected everyday words participants would all know, the task may not have required high English proficiency to scrutinize the effect. We could only establish that false recognition of inferred sentences was lowest in the dual-language condition. The interaction suggests that verbatim traces of DL sentences were more salient and retained than the unilingual conditions. Furthermore, false recognition rate of inferred sentences was not associated with individual differences such any of the executive functioning components, or the use of mental imagery. Again, the task ease can be the reason and participants did not find it taxing, or used visualization as a facilitating strategy.

Response times in the recognition task revealed that answers indicating recognition, independent of whether the sentence was seen or not, took the longest in the DL, whereas rejections took the shortest. Confidence ratings followed a comparable pattern: recognition of sentences received lower ratings in DL compared to the English condition, whereas rejections were made with higher confidence than both unilingual conditions. Results can be indicative of the time necessary to search information in two languages, retrieving the configuration with both L1 and L2 translations of object words, resulting in lower confidences. Independent of the language condition, recognition of old sentences, and rejection of new sentences received higher confidence ratings, indicating good metacognitive accuracy.

Participants were highly accurate in the configuration task. The hypothesis that the DL condition would be configured less accurately was not confirmed. Neither L2 proficiency nor mental imagery moderated this association. A subjective proficiency rating might not have been valid and as sensitive to differences for our purposes. Surprisingly, working memory ability was associated with higher accuracy in the unilingual conditions, but not in the DL. Response times suggested that it takes longer time to indicate a sentence is inaccurate. The confidence ratings revealed that Turkish had the highest ratings for correct sentences, and DL the lowest, demonstrating participants are conscious of ease of L1 and cognitive demands of dual-language settings. Indicating accurate sentences are accurate, and incorrect sentences are inaccurate received higher confidence ratings, signaling good metacognitive accuracy.

Due to overall highly accurate tag performance, which we also attribute to long exposure duration, we did not find differences between Turkish and English identification, nor an association with L2 proficiency. This level of accuracy is aligned with separate-storage hypothesis. However, like we address this link with caution since Saegert et al. (1975) argued this conclusion may not be valid for sentence level tasks. We did find Turkish tags were identified faster, suggesting an L1 advantage in lexical access sentence activation.

The fact that inferences can be distinguished from true memories suggest that false recognition occurs. Carneiro and Fernandez (2010) argued that longer exposure to stimuli increased verbatim traces and discriminability from the gist, and facilitated rejection of false memories. We set the study duration as 90 seconds per sentence group liberally to ensure all participants had sufficient time to learn, especially in the DL condition, where the pilot participants took the most time. However, this duration might have been excessive and could explain why the inferred sentences were not falsely recognized as much as expected. If the time to process had been limited, it would have prompted the more efficient strategy of gist extraction. With more time to process, the strengthened verbatim traces were less likely to decay before test time. The long study duration and task requirements combined might have generated relatively low difficulty, that impeded any association with cognitive functions. We also pondered if our task did not require very advanced English proficiency, and if the advantage of L2 proficiency would emerge when time is restricted or when the task is more challenging. Although Luk and Bialystok (2013) argued that self-report measures of English use and proficiency were valid, we did not have objective measures to validate the reports in this sample. We addressed these issues in Experiment 2.

Chapter 3:
EXPERIMENT 2

Experiment 1 has demonstrated that inferences were falsely recognized in the unilingual conditions. However, they were still discriminable from true recognitions due to long study duration which led to overlearning and good encoding of verbatim traces, as seen in language tag performance as well. We failed to support our hypotheses concerning individual differences. The ample amount of time provided in the study phase may have facilitated the task and reduced the need for employing working memory, inhibition, and shifting abilities or use of mental imagery as a strategy. Furthermore, the 2-back test had a ceiling effect; therefore, the potential predictive value was undermined by lack of variance.

To address these issues, we reduced the stimuli exposure duration to 60 seconds per trial in Experiment 2. We expect the variability in dependent variables and the effects of cognitive functions to become more salient with shorter exposure to stimuli, which decreases discriminability and renders the task more difficult. To increase variance in working memory measures, we added the 3-back task. We added Peabody Picture Vocabulary Test for an objective measure of language proficiency against which we would test the LEAP-Q measures. We tested the same hypotheses as Experiment 1 because we attributed some of the rejected hypotheses to the task nature and validity of measures used.

3.1 Method

3.1.1 Participants

We sampled 48 Turkish-English bilinguals from Koç University subject pool (33 female, M_{age}=21.76, SD_{age}=4.41). Power analyses for repeated measures with G-power software yielded the necessary sample of 30 to detect an effect size of 0.25 with 95% power. Participants acquired English as a second language at the mean age of 9.63 (*SD*=3.84) in school contexts. Data exclusion criteria included having accuracy below 2 SD of the sample mean on EF tasks and having response times 2 SD above or below the sample on the inference-making task. We excluded data on a trial basis.

3.1.2 Measures and Procedure

All the material in this experiment were identical to Experiment 1, except we modified the Inference-Making task so that the participants studied the sentence groups for 60 seconds instead of 90. We added 3-back after 2-back. In a separate session, we administered the LEAP questionnaire, followed by the *Peabody Picture Vocabulary Test Fourth Edition (PPVT)* by Dunn and Dunn (2007) which is a measure of receptive vocabulary for Standard American English. In this task, the experimenter utters words, and the participant chooses the picture corresponding to the word.

3.2 Results

3.2.1 Descriptive statistics[3]

Table 3.1 shows the means and standard deviations of recognition accuracy and response times. Table 3.2 shows the respective correlations of accuracy and response times. English inferred and new, DL old and new sentences were significantly

[3] Upon the preliminary analyses, we decided not to use LEAP-Q in further analyses as measures derived were negatively skewed (-1.788, *SD*= .354) with a kurtosis of 5.206 (*SD*= .695) and did not reflect variances in L2 proficiency. PPVT is used as the sole English proficiency predictor, skewness of -.726 and kurtosis of .262.

correlated with respective response times. All correlations indicated accurate recognition positively correlated with speed, indicating no tradeoff. Table 3.3 shows the means and standard deviations of working memory, inhibition, and shifting measures (2-back, 3-back, Trail Making, Letter-Number sequencing), mental imagery (SUIS), and L2 proficiency measures (PPVT).

Table 3.1: Means and standard deviations of recognition accuracy and response times.

		Recognition accuracy (%)		Recognition RT (sec)	
		Mean	Std. Deviation	Mean	Std. Deviation
Turkish	Old	83.21	10.04	4.81	1.53
	Inferred	38.77	24.67	5.44	2.15
	New	11.35	11.11	5.13	1.71
English	Old	77.53	13.71	5.10	2.14
	Inferred	40.37	20.18	5.28	2.18
	New	16.42	14.79	4.92	2.11
Dual-language	Old	75.48	12.03	4.80	1.95
	Inferred	22.46	16.37	3.93	1.30
	New	18.84	17.47	4.43	1.34

Table 3.2: Correlations of sentence recognition and response times

	TR old RT	TR inferred RT	TR new RT	ENG old RT	ENG inferred RT	ENG new RT	DL old RT	DL inferred RT	DL new RT
TR old	.099								
TR inferred		-.074							
TR new			-.006						
ENG old				.256					
ENG inferred					-.305*				
ENG new						-.411*			

DL old	.343*		
DL inferred		-.104	
DL new			-.407*

Note. *denotes p<.05.

Table 3.3: Descriptive statistics and correlations of study variables in Experiment 2.

Task	M	SD	2-back	3-back	Trail-A	Trail-B	SUIS	LNS-A	LNS-B	PPVT
2-back	3.06	1.13		**0.59***	-0.15	-0.09	0.01	0.28	**0.47***	0.06
3-back	2.07	1.24			-0.26	-0.09	0.07	0.29	**0.49***	0.00
Trail-A	21.04	5.53				**0.46***	0.11	-0.13	-0.27	**-0.38***
Trail-B	44.48	14.12					0.10	0.11	-0.09	-0.13
SUIS	3.29	0.59						0.04	0.16	0.14
LNS-A	13.42	2.46							**0.68***	0.29
LNS-B	9.83	3.45								0.24
PPVT	139.96	33.15								

Notes. SUIS= Spontaneous Use of Imagery scale, LNS=Letter Number Sequencing, PPVT=Peabody Picture Vocabulary Test.

*p<0.05.

3.2.2 Sentence Recognition

In order to test sentence recognition, we conducted a 3x3 within-subjects ANOVA with language condition (Turkish, English, DL) and sentence type (old, inferred, new) as factors. We found significant main effects of sentence type, $F(2,88)=341.074$, $p<.001$, η2=.886, and language, $F(2,88)=11.255$, $p<.001$, $\eta^2=.204$. We also found sentence type and language interaction, $F(4,176)=13.026$, $p<.001$, $\eta^2=.228$. Pairwise comparisons with Bonferroni corrections showed that Turkish yielded significantly higher recognition for old sentences compared to English and DL. Inferred sentences were (falsely) recognized significantly more in Turkish and English, compared to DL. Lastly, new sentences were (falsely) recognized as seen significantly less in Turkish than English and DL (see Figure 3.1).

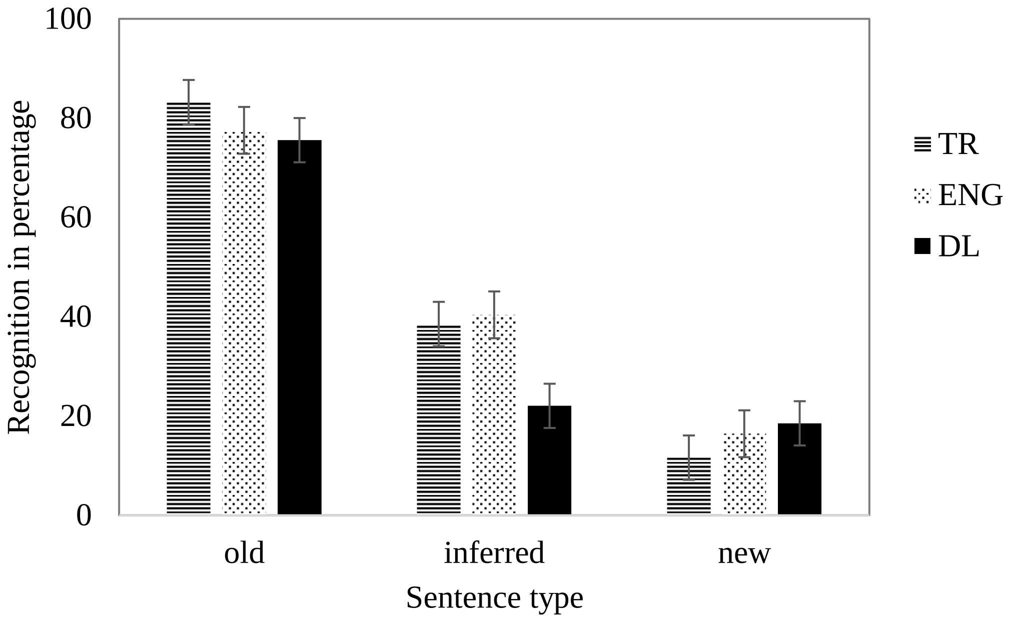


Figure 3.1: Percentage of recognized sentences.

Note. DL=dual-language.

Table 3.4 shows the correlations between sentence types for each language condition. In Turkish, old and new sentence recognitions were correlated. In English new sentences were correlated positively with inferred, and negatively with old. In DL, inferred sentences were correlated negatively with old, and positively with new sentences.

We conducted correlational analyses to determine which variables were associated with recognition (Table 3.5). To examine the effect of language condition on different sentence types, we conducted separate ANCOVAs with the significant correlates. We analyzed old sentences, using 3-back as a covariate, and found it was significantly associated, $F(1,40)=12.651$, $p=.001$, $\eta^2=.240$, with observed power .935. Having higher scores predicted higher recognition for all language conditions. The difference between Turkish and English recognition, previously found in the 3x3 ANOVA disappeared. Only recognition of DL old sentences (*M*=76.32, *SD*=.72) remained significantly lower than Turkish ones (*M*=82.80, *SD*=1.53).

Table 3.4: Correlations of true and false recognitions.

	TR inferred	TR new	ENG inferred	ENG new	DL inferred	DL new
TR old	0.206	-0.412*				
TR inferred		0.217				
ENG old			-0.176	-0.557*		
ENG inferred				0.559*		
DL old					-0.317*	-0.224
DL inferred						0.599*

Note. * denotes p<.05.

Table 3.5: Correlations of recognition performance with individual differences.

	2-back	3-back	Trail-A	Trail-B	SUIS	LNS-A	LNS-B	PPVT
TR old	0.04	**0.31***	0.05	0.01	0.00	0.03	0.11	-0.06
ENG old	-0.02	**0.32***	-0.09	-0.10	0.14	0.03	0.07	-0.12
DL old	0.20	**0.40***	-0.12	-0.14	-0.21	0.25	0.31	-0.24
TR inferred	-0.28	**-0.33***	**0.35***	**0.35***	-0.13	-0.10	-0.29	-0.18[4]
ENG inferred	-0.26	**-0.40***	0.18	**0.32***	-0.14	-0.19	**-0.34***	0.16
DL inferred	**-0.31***	-0.24	0.21	0.28	0.04	**-0.33***	**-0.34***	-0.18
TR new	-0.29	**-0.33***	0.13	0.29	-0.10	-0.17	-0.29	-0.24
ENG new	0.00	-0.30	0.15	0.25	**-0.37***	-0.12	**-0.42***	0.07
DL new	**-0.34***	-0.21	0.24	**0.36***	0.03	**-0.43***	**-0.55***	-0.19

Note. * denotes p<.05.

[4] The correlations with PPVT were significantly different at the one-tailed level.

For inferred sentences, we used 2-back, 3-back, Trail A and B, LNS-A and B, and PPVT as covariates. PPVT interacted with the language condition, $F(1,22)=6.087$, $p=.022$, $\eta^2=.217$i with observed power .655. Higher PPVT scores predicted significantly higher recognition in the English condition, $\beta_{English}=.38$, $p=.05$. The pattern for the language condition remained similar to 3x3 ANOVA; dual-language condition had lower recognition than the unilingual conditions. Probing the interaction at 1 SD below and above PPVT scores, we found for low PPVT Turkish was significantly higher than only DL, whereas, for high PPVT, both Turkish and English were higher than DL in recognition. We present the results in Figure 3.2. Trail-B was found to be a significant covariate, $F(1,22)=12.837$, $p=.002$, $\eta^2=.368$, with observed power .928. Lower Trail-B scores predicted less recognition in Turkish and English inferred sentences, $\beta_{Turkish}=.58$, $p=.010$, and $\beta_{English}=.56$, $p=.006$.

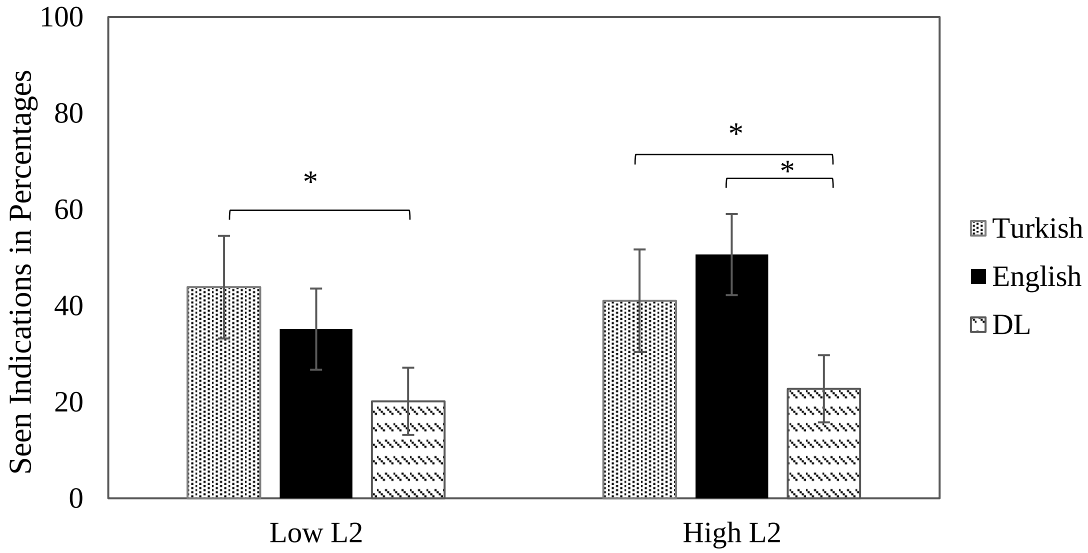


Figure 3.2: Percentage of recognized inferred sentences with L2 proficiency.

Note. L2 = second language, DL= dual-language. Low and High L2 are defined as 1 SD below and above the mean.

For new sentences, ANCOVAs with 2-back, 3-back, Trail-B, SUIS, LNS-A, LNS-B, PPVT as covariates revealed significant interaction of language with 2-back, $F(2,46)=8.065$, $p=.001$, $\eta^2=.260$ with observed power .945, and PPVT, $F(1,23)=6.55$, $p=.018$, $p=.222$, and observed power of .688. The language condition revealed similar patterns to 3x3 ANOVA, such that Turkish was significantly less recognized than English and dual-language. Parameter estimates showed that both covariates predicted higher recognition of new sentences in the English condition, but not in Turkish or dual-language, $\beta_{2back}=.531$, $p=.006$, $\beta_{PPVT}=.374$, $p=.029$.

3.2.3 Sentence Recognition Response Times

We conducted a between-subjects ANOVA with three independent variables, language, sentence type (old, inferred, new), answer given (yes/no to recognition), and subjects as the random factor, to test how response times differed for sentence recognition. Analyses revealed an interaction of sentence type and answer $F(2,120.84)=6.376$, $p=.002$, $\eta^2= .095$. For "yes" responses, there were no differences between sentence types. For "no" responses, old sentences (M=5.51, SD=.19) took significantly longer than inferred (M=4.72, SD=.12) and new sentences (M=4.71, SD=.08).

Lastly, we conducted correlational analyses between recognition and their respective response times. In the Turkish condition, we found no correlation. In the English condition, inferred and new sentences' (false) recognition were significantly and negatively correlated with their response times, $r_{\text{inferred}}(43)=-.305$, $p=.042$; $r_{\text{new}}(43)=-.411$, $p=.005$. In the DL condition, recognition of old sentences correlated positively with its response time, $r_{old}(41)=.343$, $p=.024$, whereas falsely recognizing new sentences correlated negatively with reaction time, $r_{new}(40)=-.407$, $p=.007$. For all types of sentences, higher recognition accuracy correlated with slower response times.

3.2.4 Correct Configuration

We conducted a within-subjects ANOVA to examine how configuration accuracy differed across language conditions, and found significant effect of language, $F(2,86)=16.149$, $p<.001$, $\eta^2=.273$. Pairwise comparisons revealed that accuracy was significantly higher in Turkish (*M*=42.68, *SD*=4.83, out of 48) and English (*M*=40.75, *SD*=7.04) compared to DL condition (*M*=37.55, *SD*=7.80). Turkish and English were not significantly different.

We conducted correlational analyses to examine the associations of executive function measures, mental imagery, and language proficiency with configuration accuracy (See Table 3.6). We conducted ANCOVA with language condition as predictor and added all eight correlates as covariates, and no significant covariates emerged.

Table 3.6: Means and correlations of configuration accuracy with individual difference variables.

	M** (SD)	TR	ENG	DL	2-back	3-back	Trail-A	Trail-B	SUIS	LNS-A	LNS-B	PPVT
TR	42.68 (4.84)	1	**0.60***	**0.47***	**0.38***	**0.37***	-0.15	-0.14	**0.30***	0.14	**0.49***	0.15
ENG	40.75 (7.05)		1	**0.74***	0.26	**0.43***	**-0.31***	**-0.40***	0.25	0.28	**0.57***	**0.30***
DL	37.55 (7.81)			1	0.25	0.24	-0.16	**-0.32***	0.20	**0.37***	**0.50***	**0.35***

Note. DL=dual-language, SUIS=Spontaneous Use of Imagery, LNS=Letter-Number Sequencing, PPVT=Peabody Picture Vocabulary Test.

* denotes p<.05.

**Means provided out of 48 sentences.

3.2.5 Correct Configuration Response Times

We conducted a between-subjects ANOVA to explore how response times differed based on the variables of language condition (Turkish-English-DL), sentence type (correct-incorrect), participant's answer (yes/no), and we added subject as a random factor. We found a two-way interaction of sentence type and answer, $F(1,67.96)=12.140$, $p=.001$, $\eta^2=.152$. The three-way interaction with subject was also significant. Pairwise comparisons revealed that for correct sentences, "yes" responses ($M=4.12$, $SD=.07$) took less time than "no" ($M=4.55$, $SD=.19$), while for incorrect sentences "no" ($M=4.21$, $SD=.07$) responses took less time than yes ($M=4.89$, $SD=.19$).

3.2.6 Language Tag

The tag accuracy had a mean of 19.61 ($SD=2.51$) out of 24. The Turkish tag and English tag's mean accuracy was 9.78 ($SD=1.75$) and 9.83 ($SD=3.0$) out of 12, respectively. Paired t-tests revealed no difference in tag accuracies between languages, $t(45)=-.08$, $p=.932$.

Total tag accuracy correlated with 2-back, $r(40)=.323$, $p=.037$; 3-back, $r(41)=.335$, $p=.028$; and Trail-B, $r(40)=-.392$, $p=.010$. To see how language predicted tag accuracy, we conducted ANCOVA with the corelates as covariates and found an association with Trail-B, $F(1,35)=9.379$, $p=.004$, $\eta^2=.211$, and observed power .846. Parameter estimates revealed that lower Trail-B scores (better shifting performance) predicted better English tag, $\beta=-.366$, $p=.018$, but not Turkish.

We conducted a paired t-test to examine how response time differed for tags. We averaged the sentence RTs based on the participants' indication as Turkish or English.

We found that indicating the sentence was seen in Turkish (*M*=4.15, *SD*= 1.29) took less time than English (*M*=5.07, *SD*=1.74), $t(40)=-4.302$, $p<.001$. Tag RTs did not correlate with any individual difference variables.

3.2.7 Confidence Ratings

Recognition Confidence

We analyzed the confidence ratings (on a 5-point scale) given to each recognition response. We found a significant two-way interaction of sentence type and answer, $F(2,110.64)=54.23$, $p<.001$, $\eta^2=.495$, and a four-way interaction of sentence type and answer with language and subject factors $F(137,5725)=1.655$, $p<.001$, $\eta^2=.038$. Pairwise comparisons with Bonferroni correction of the two-way interaction revealed that “yes” responses had higher confidence in old (*M*=4.04, *SD*=.02) compared to inferred (*M*=3.48, *SD*=.05) and new (*M*=3.52, *SD*=.06). “No” responses to new sentences (*M*=4.28, *SD*=.02) had significantly higher ratings compared to inferred (*M*=3.86, *SD*=.03) and old (*M*=3.53, *SD*=.05). Inferred sentences had significantly higher ratings than the old.

Configuration Confidence

We conducted 3x2x2 ANOVA to examine the confidence ratings for the configuration task, using subject as random factor. We found three significant two-way interactions of language and type, language and answer, and type and answer. We also found a significant three-way interaction of subject with type and answer, and a four-way interaction with all the predictors, $F(51,6079)=2.935$, $p<.001$, $\eta^2= .024$. We conducted pairwise comparisons with Bonferroni correction to examine the two-way interactions (see Figure 3.3). For correct sentences, DL had the lowest confidence ratings, whereas, for incorrect sentences, Turkish had the highest. Congruent responses ("yes" to correct, "no" to incorrect) had higher confidence than incongruent ones.

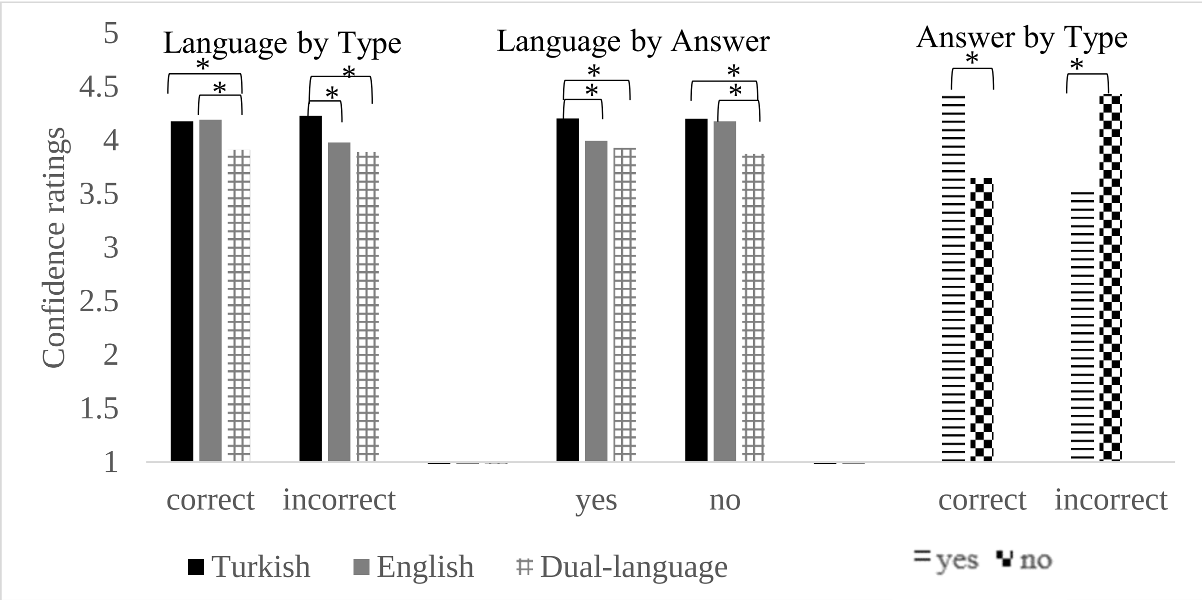


Figure 3.3: Confidence rating interactions.

3.2.8 Order effect

The counterbalanced language conditions created six orders (N=7 each). For correct configuration, the order did not interact with language, p=.268. For recognition accuracy, order interacted significantly with language and sentence type, $F(20,156)=1.893$, $p=.016$, $\eta^2=.195$. "Dual-language-Turkish-English" order had significantly higher recognition in Turkish-inferred, and English inferred and wrong sentences. We believe this to be an arbitrary result due to the low sample size in each order and potential influential cases.

3.3 Discussion

We examined the false recognition of inferences in single and dual-language contexts with reduced duration of stimuli. We confirmed our first hypothesis that sentence recognition and found high accuracy for old and new sentences, especially in the Turkish condition. Better working memory ability was associated with better accuracy for old sentences. Inferred sentences were not falsely recognized most in L1 as expected, but were least recognized in DL, suggesting the language switching increased discriminability. Another possibility is that the participants attended more in the dual-language condition. Although the desirable difficulty account had only been discussed for L2 (Bjork, 1994), we suggest the dual-language condition may have benefited similarly. We supported our hypothesis that higher L2 proficiency would predict higher false recognition for the inferred English sentences, but not when both languages were used. Aligned with the Revised Hierarchical Model, higher L2 proficiency participants processed the gist similar to L1, and this automaticity resulted in false recognition of the gist (Kroll & Stewart, 1994). However, the dual-language condition required switching between languages and hindered automatic processing. Contrary to the hypothesis, better cognitive functions was associated with less false recognition in unilingual conditions. Higher false recognition of new sentences was associated with high L2 proficiency and low shifting. For all sentence types, higher response times were associated with higher accuracy[5], indicating that more time spent increases true memories (verbatim). Confidence ratings indicated good metacognitive accuracy for recognition of old and rejection of new sentences.

[5] Accuracy refers to the recognition of old sentences, and rejection of inferred and new sentences.

Secondly, we assessed object configuration accuracy to validate our recognition task to ensure any conclusion we draw was not due to participants' inability to configure the objects. We found that correct configuration was highly accurate, albeit lowest in dual-language condition, supporting our hypothesis. We did not support our hypotheses that higher L2 proficiency, better EF skills, and mental imagery abilities would predict higher accuracy. In this level of difficulty, we expected mental imagery to become a strategy that good visualizers would employ. Mental imagery as an ability might be a more valid predictor than self-reports on frequency of imagery use to determine who would use this strategy. We saw trends of high working memory and shifting abilities being associated with better accuracy in L2 and in the dual-language condition, and L2 proficiency in the DL condition. While the trends did not reach statistical significance, they suggest the task had an appropriate difficulty level, valid for examining automaticity in the L1 condition, and EF involvement in L2 and dual-language conditions. Response times revealed that responses were faster for the right answers, demonstrating no speed-accuracy tradeoff, as desired by task design. Confidence in responses revealed higher confidence for the right answers, showing participants were metacognitively accurate about their performance. Responding to correct sentences, DL had the lowest confidence ratings. Participants had to formulate one Turkish and one English configuration, and the explicit switching demands might have reflected onto their evaluations. For incorrect sentences, Turkish had the highest confidence ratings. Since the configuration task, unlike recognition, is congruent with gist representations, the advantage of an L1 task might have created discriminability that is explicitly available to the participants.

The language tag was identified accurately and was not moderated by L2 proficiency. Low proficiency participants did not reap the benefits of desirable difficulty. Having to shift between languages in short durations may have nullified the L2 automaticity for high proficiency participants, leveling the playing field. Better shifting ability predicted better Turkish tag, and we found tag indication was faster for Turkish sentences. The results represent an L1 advantage on verbatim representation, which does not negate the L1 advantage on gist, since traces are processed separately. However, whether the traces are distinguished easily might depend on EF skills.

Chapter 4:
GENERAL DISCUSSION

We examined bilinguals' logical inference-making process in unilingual and dual-language (DL) contexts, using a spatial configuration task consisting of concrete object words in Turkish and English. Our starting assumption was that the inferring process is similar to gist representation, where the non-presented stimuli are stored as presented, if the meaning is preserved and valid. However, if the task demands verbatim information, then merely recalling inferences would result in memory errors. In both experiments, inferred sentences were recognized significantly more than new sentences, which suggested that the inference-making process created false recognition. We ensured inferring performance was not due to difficulties configuring by obtaining high accuracy configuration accuracy. Lastly, we examined if participants could report the language the sentence was presented in, and found high accuracy for this language tag. We investigated the moderating roles that L2 proficiency, working memory, inhibition, shifting, and mental imagery would play in this task. Experiment 1 did not yield such moderations; however, we used a reduced exposure time in Experiment 2 to prevent strong verbatim traces from being formed and increase the task difficulty. We found an objective measure of English vocabulary and cognitive flexibility became significant predictors in Experiment 2 with reduced exposure time. We found higher proficiency and lower EF skills were be related to higher false recognition, i.e. gist processing.

Inferences were recognized more than new sentences confirmed that false recognition occurred when information is retrieved in gist form; however, they were also discriminable from old sentences. The Fuzzy Trace Theory asserted both gist and verbatim representations are formed at the time of encoding, and true and false recognitions are independent of each other (Brainerd & Reyna, 2004). Our results are aligned with their premise, instead of Bransford and Franks (1971)'s suggesting comprehension is fully constructive. Alternatively, in our study, the time between encoding and retrieval might not have been long enough for verbatim decay. Increasing the time between studying and testing might allow more gist reliance and subsequent memory errors. Secondly, it was found that warning the participants on susceptibility to false memories predicted more verbatim-based responses, thus less errors (Carneiro & Fernandez, 2010). While we did not issue such a warning, we told the participants to watch out for changing instructions across recognition and accuracy trials. This may have hindered the automaticity in responding, given participants an insight into the study, and increased discriminability of inferred sentences from the old ones. In both experiments, false recognition was significantly lower in the DL condition compared to the unilingual conditions. This might have resulted from the integration difficulty and verbatim proneness of the dual-language condition, which created discriminability of traces and resistance against memory errors. Having to shift between languages may have created the desirable difficulty effect, since the words were not difficult to process in unilingual conditions.

While false recognition in L1 and L2 appeared similar at first, L2 proficiency created a distinction. In Experiment 2, we used a standardized task and found that more proficient participants falsely recognized L2 inferred sentences more. This result is aligned with the bilingualism models, suggesting L2 approaches L1 automaticity with increased proficiency (De Groot, 1992; Kroll & Stewart, 1994). The low susceptibility to false recognition in the dual-language suggest a level of discriminability of verbatim traces that is aligned with separate-storages. However, the inferred sentences did not substitute the translation pairs, which would be at a word-level analysis better suited for the storage accounts. We had hypothesized that higher working memory, inhibition, and shifting abilities would predict more false recognition. However, we observed the opposite patterns: they predicted less recognition of inferred sentences and generally higher accuracy in other old and new sentence types. We deduce that proficiency and cognitive functions are opposing forces acting on ease of processing. Proficiency allows us to comprehend meaning without much effort, whereas EF as a unity allows the retention of complex information in manageable bits, a skill that employs working memory capacity to hold and manipulate information. Participants with lower EF skills might prefer more automatic and less effortful strategies to process information, such as averaging, abstracting, inferring, and gist extracting. While EF abilities may be beneficial in creating meaning, they did not appear to strengthen gist traces over verbatim. Such patterns with individual differences in cognitive functions emerged in Experiment 2, when the study duration for each sentence group was shorter, rendering the experiment more demanding, and more valid measures were employed. Even though we found correlations in preliminary analyses, most did not reach statistical significance in the models. We did not sample to examine individual differences; therefore, lack of predictions can be attributed to lack of statistical power.

Response times differed for language conditions in Experiment 1. When both language lexica needed to be searched the DL condition, recognition took more time, while rejection was relatively fast. In Experiment 2, this speed was related to the accuracy of responses, such that rejection took the longest time for old sentences. Confidence ratings were lower for the dual-language condition, or when rejecting old sentences and recognizing new sentences. Both dependent variables showed patterns of good metacognition. Accuracy and response time association suggested that higher time spent was linked to responding more aligned with verbatim representation.

The configuration task was performed accurately in both experiments. The hindrance in the DL condition emerged in Experiment 2; however, it was not related to any individual difference variable. In response times, we observed indicating a sentence is inaccurate took more time, commonly known as the confirmation effect in sentence verification tasks.

In the dual-language condition, the language in which the sentences were studied was recalled very accurately by all participants; therefore, we did not confirm our hypotheses that the language tag is predicted by language proficiency. While this performance supports the separate-storage hypothesis, the results might be due to languages' discriminating properties (Bornkessel-Schlesewsky et al., 2011). Turkish and English have few cognate words, mostly borrowed words, and none were used in this study. On average, English sentences were three words longer than Turkish sentences due to the syntactic requirement of making comparisons. The length and the exposure duration might have also increased the discriminability of sentences. If the surface structure provided sufficient cues to the participant, they might not have needed to access the concepts to indicate the language. Studying in L1 and L2, and testing in L3 translations may reduce these cues and provide a clearer picture (Saegert et al., 1975). Another explication is that the experimental demands put people in a bilingual mode (Grosjean, 2001) during which people could access both cues, regardless of whether the cues would be stored commonly or separately. Contrary to hypotheses, better shifting ability predicted better tag performance, suggesting once again that EF is more linked to the ability to store and manipulate verbatim representations. This might also be related to the use of reactive mode in dual-language processing, without which we cannot process the less automatic language (Braver, 2012).

At the word level, concrete words activate mental imagery processing more compared to abstract words (Hemati & Hossein-Zadeh, 2018; West & Holcomb, 2000), and have more semantic features than abstract words (De Groot, 1992); therefore, they are shared among the languages that bilinguals know. We had hypothesized that mental imagery would facilitate gist representation, more so in the more automatic language, L1. Schrauf and Rubin (2000) on the other hand, proposed the mental imagery advantage of L2 because of novelty and effort-after-meaning. We failed to corroborate either claim. Neither in L1 nor in L2, mental imagery predicted recognition, configuration, or tag identification. While we pondered if mental imagery would be more related to nonverbal representations of relations, De Beni et al. (2005) found that spatial text processing was hindered more by concurrent spatial tasks than verbal or no interference conditions (whereas nonspatial texts were hurt by verbal interferences only), and concluded that visuospatial working memory was necessary to comprehend spatial texts and construct spatial representations. While they failed to demonstrate the interference effect with reaction times, they discussed it was due to inferences being made during mental model construction and being readily available at test time. Alternatively, the facilitatory effects of mental imagery might be revealed by the time it takes to encode information in a self-paced paradigm. The results highlight the importance of spatial processing in such tasks; however, self-report measures of imagery use should be scrutinized, and more objective measures of mental imagery or visuospatial working memory ability should be considered. Imagery ability could also be important for the representation of abstract concepts.

The reliance on verbatim structure for our low L2 proficiency participants is aligned with the lexical mediation, whereas accessing the gist appears consistent with the concept mediation path. The time measure we obtained was not as sensitive as reaction time; therefore, it does not allow us to make conclusive statements about the lexical organization. However, similar response times in the Turkish and English conditions imply that participants did not translate L2 sentences in the recognition task. Kroll et al. (2010) discussed the possibility of L2 mediation by L1 occurring *after* L2 accesses concepts, in which case, we would not observe higher RT due to the mediation. While this possibility appears in line with the findings, we should also be cautious in relating time to translate complex sentences with word translations.

We conducted the second study to examine if verbatim traces weaken when the exposure to stimuli is shortened. We could not conclude an automaticity or organization account based on the differences between studies 1 and 2. Further research underway should explore the processes with tasks of varying duration, difficulty, and object concreteness. Further studies could also incorporate difficult vocabulary that would allow us to differentiate low and high L2 proficiency participants. The inability to access the meaning or visualize image of a given word could impede the automaticity. Furthermore, the nature of the tasks should be considered when predicting the occurrence of memory errors. Comprehension tasks have been found to yield more errors, compared to immediate recognition tasks, where the perceptual cues are sufficient to perform the task, without having to access a deeper meaning structure (Harris & Monaco, 1978). There can also be differences observed for different level of discourses. Research on autobiographical memory involves complex memories such as narratives, and suggests bilinguals display a language-dependent advantage for retrieval ((Marian & Fausey, 2006; Marian & Neisser, 2000). Further research can investigate how this dependency aligns with memory errors and language tagging.

Chapter 5:
CONCLUSION

People construct meaning to manipulate information efficiently. Inferring is one of the processes that facilitates this by effortlessly extracting the gist. We found spatial inferences were falsely recognized in Turkish and English, but not when the sentences are presented in two languages. While higher proficiency predicted similar processing ease for bilinguals' two languages, those with better cognitive abilities succeeded in preserving the verbatim representations that competed with gist for recognition. This suggests our comprehension is not fully constructive, instead it is a function of decay and individual differences in strategies used. We investigated language tagging as an awareness of language membership for dual-language information. Near-perfect performance aligns with distinguishable separate components. The existing bilingualism models have not tested sentence level information, and the dual-language condition has not been extensively examined. Inferences provide insight into the bilingual information-processing and individual differences.

Bibliography

Altarriba, J., & Soltano, E. G. (1996). Repetition blindness and bilingual memory: Token individuation for translation equivalents. *Memory & Cognition, 24*(6), 700-711.

Anastasi, J., Rhodes, M., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory, 13*(8), 815-828.

Arndt, J., & Beato, M. S. (2017). The role of language proficiency in producing false memories. *Journal of Memory and Language, 95*, 146-158. https://doi.org/10.1016/j.jml.2017.03.004

Barclay, J. R. (1973). The role of comprehension in remembering sentences. *Cognitive Psychology, 4*(2), 229-254.

Bialystok, E. (2007). Cognitive Effects of Bilingualism: How Linguistic Experience Leads to Cognitive Change. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 210-223. https://doi.org/10.2167/beb441.0

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3-11. https://doi.org/10.1017/s1366728908003477

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009, Dec). Bilingual Minds. *Psychol Sci Public Interest, 10*(3), 89-129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I., & Luk, G. (2012, Apr). Bilingualism: consequences for mind and brain. *Trends Cogn Sci, 16*(4), 240-250. https://doi.org/10.1016/j.tics.2012.03.001

Bjork, R. A. (1994). Memory and metamemory considerations in the. *Metacognition: Knowing about knowing, 185*.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011, 2011/06/01/). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*(3), 133-152. https://doi.org/https://doi.org/10.1016/j.bandl.2010.09.010

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*(1), 3-47.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science, 11*(5), 164-169. https://doi.org/10.1111/1467-8721.00192

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review, 24*(4), 396-439. https://doi.org/10.1016/j.dr.2004.08.005

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive psychology, 3*(2), 193-209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive psychology, 2*(4), 331-350.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences, 16*(2), 106-113.

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5*(6), 673-678.

Carneiro, P., & Fernandez, A. (2010, Jan-Feb). Age differences in the rejection of false memories: the effects of giving warning instructions and slowing the presentation rate. *J Exp Child Psychol, 105*(1-2), 81-97. https://doi.org/10.1016/j.jecp.2009.09.004

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1057.

D'Argembeau, A., & Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and cognition, 15*(2), 342-350.

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology, 17*(1), 77-95.

De Groot, A. M. (1992). Determinants of word translation. *Journal of experimental psychology: learning, memory, and cognition, 18*(5), 1001.

Dijkstra, T. O. N., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & RekkÉ, S. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition, 22*(4), 657-679. https://doi.org/10.1017/S1366728918000287

Dornic, S. (1979). Information processing in bilinguals: Some selected issues. *Psychological Research, 40*(4), 329-348.

Dunn, L., & Dunn, D. (2007). Peabody picture vocabulary test 4th edition. *Bloomington: NCS Pearson*.

Eisen, M. L., Gomes, D. M., Lorber, W. G., Perez, C. I., & Uchishiba, H. (2013). Using an individual differences approach to examine two distinct types of suggestibility effects. *Applied cognitive psychology, 27*(1), 2-11.

Esposito, A. G., & Baker-Ward, L. (2016, 2016/08/08). Immigration, language proficiency, and autobiographical memories: Lifespan distribution and second-language access. *Memory, 24*(7), 949-960. https://doi.org/10.1080/09658211.2015.1061010

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86,* 186-204.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515-530.

Grosjean, F. (2001). The Bilingual's Language Modes. In *One mind, two languages: Bilingual language processing.* (pp. 1-22). Blackwell Publishing.

Harris, R. J. (1974). Memory and comprehension of implications and inferences of complex sentences. *Journal of Verbal Learning and Verbal Behavior, 13*(6), 626-637.

Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General, 107*(1), 1-22.

Hemati, S., & Hossein-Zadeh, G.-A. (2018). Distinct functional network connectivity for abstract and concrete mental imagery. *Frontiers in human neuroscience, 12,* 515.

Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Front Psychol, 3,* 158. https://doi.org/10.3389/fpsyg.2012.00158

Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of memory and language, 35*(2), 101-117.

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology, 98*(1), 203.

Karlsson, M. (2014). Advanced learners' L1 (Swedish) versus L2 (English) inferencing. *Australian Review of Applied Linguistics, 37*(1), 3-23.

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS One, 6*(12), e29221. https://doi.org/10.1371/journal.pone.0029221

Key-DeLyria, S. E., & Altmann, L. J. (2016, May 1). Executive Function and Ambiguous Sentence Comprehension. *Am J Speech Lang Pathol, 25*(2), 252-267. https://doi.org/10.1044/2015_AJSLP-14-0111

Kolers, P. A. (1963). Interlingual word associations. *Journal of verbal learning and verbal behavior, 2*(4), 291-300.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual review of psychology, 51*(1), 481-537.

Kroll, J. F., & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. *J Cogn Psychol (Hove), 25*(5). https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., & de Groot, A. M. (1997). *Lexical and Conceptual Memory in the Bilingual: Mapping Form to Meaning in Two Languages*. Lawrence Erlbaum Associates Publishers.

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.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373-381.

Liepmann, D., & Saegert, J. (1974). Language tagging in bilingual free recall. *Journal of Experimental Psychology, 103*(6), 1137.

Liu, H., Fan, N., Rossi, S., Yao, P., & Chen, B. (2016). The effect of cognitive flexibility on task switching and language switching. *International Journal of Bilingualism, 20*(5), 563-579.

Long, M. R., Vega-Mendoza, M., Rohde, H., Sorace, A., & Bak, T. H. (2019). Understudied factors contributing to variability in cognitive performance related to language learning. *Bilingualism: Language and Cognition*, 1-11. https://doi.org/10.1017/s1366728919000749

Luk, G., & Bialystok, E. (2013, Jan 1). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *J Cogn Psychol (Hove), 25*(5), 605-621. https://doi.org/10.1080/20445911.2013.795574

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in

Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967. https://doi.org/doi:10.1044/1092-4388(2007/067)

Marian, V., & Fausey, C. M. (2006). Language-dependent memory in bilingual learning. *Applied Cognitive Psychology, 20*(8), 1025-1047. https://doi.org/10.1002/acp.1242

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General, 129*(3), 361.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological review, 99*(3), 440.

Misdraji, E. L., & Gass, C. S. (2010). The Trail Making Test and its neurobehavioral components. *Journal of Clinical and Experimental Neuropsychology, 32*(2), 159-163.

Mortensen, L., Berntsen, D., & Bohn, O. S. (2015). Retrieval of bilingual autobiographical memories: effects of cue language and cue imageability. *Memory, 23*(2), 138-156. https://doi.org/10.1080/09658211.2013.873809

Paivio, A., Clark, J. M., & Lambert, W. E. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 163.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*(2), 261-285.

Peirce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Kastman, E. (2019). Lindelø v, J. 2019.'PsychoPy2: experiments in behavior made easy'. *Behavior research methods*.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of verbal learning and verbal behavior, 23*(1), 23-38.

Prior, A., Goldina, A., Shany, M., Geva, E., & Katzir, T. (2014, 02/16). Lexical inference in L2: Predictive roles of vocabulary knowledge and reading skill beyond reading comprehension. *Reading and Writing, 27*. https://doi.org/10.1007/s11145-014-9501-8

Reisberg, D., Pearson, D. G., & Kosslyn, S. M. (2003). Intuitions and introspections about imagery: The role of imagery experience in shaping an investigator's theoretical views. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17*(2), 147-160.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences, 7*(1), 1-75.

Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016, Mar 1). How Fuzzy-Trace Theory Predicts True and False Memories for Words, Sentences, and Narratives. *J Appl Res Mem Cogn, 5*(1), 1-9. https://doi.org/10.1016/j.jarmac.2015.12.003

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30*(2), 178.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and individual differences, 9*(2), 95-123.

Riehl, C. M. (2010, Sep). The mental representation of bilingualism. *Wiley Interdiscip Rev Cogn Sci, 1*(5), 750-758. https://doi.org/10.1002/wcs.74

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nösselt, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature, 415*(6875), 1026-1029.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition, 21*(4), 803.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*(9), 437-442.

Saegert, J., Hamayan, E., & Ahmar, H. (1975). Memory for language of input in polyglots. *Journal of Experimental Psychology: Human Learning and Memory, 1*(5), 607-613. https://doi.org/10.1037/0278-7393.1.5.607

Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence, 39*(4), 222-232.

Schrauf, R. W., & Rubin, D. C. (2000). Internal languages of retrieval: The bilingual encoding of memories for the personal past. *Memory & Cognition, 28*(4), 616-623.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne, 26*(1), 1.

Vannucci, M., Pelagatti, C., Chiorri, C., & Mazzoni, G. (2016). Visual object imagery and autobiographical memory: Object Imagers are better at remembering their personal past. *Memory, 24*(4), 455-470. https://doi.org/10.1080/09658211.2015.1018277

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Ye, Z., & Zhou, X. (2009, Sep). Executive control in language processing. *Neurosci Biobehav Rev, 33*(8), 1168-1177. https://doi.org/10.1016/j.neubiorev.2009.03.003

Appendix A: Inference-Making Study Items

Turkish condition

Çarşaf magnetin üzerinde

Fotoğraf çarşafın üzerinde

Magnet çöp torbasının üzerinde

Fotoğraf magnetin üzerinde

English condition

The leaf is above the bookmark

The flower is above the leaf

The pencil case is above the leaf

The pencil case is above the flower

Dual-language condition

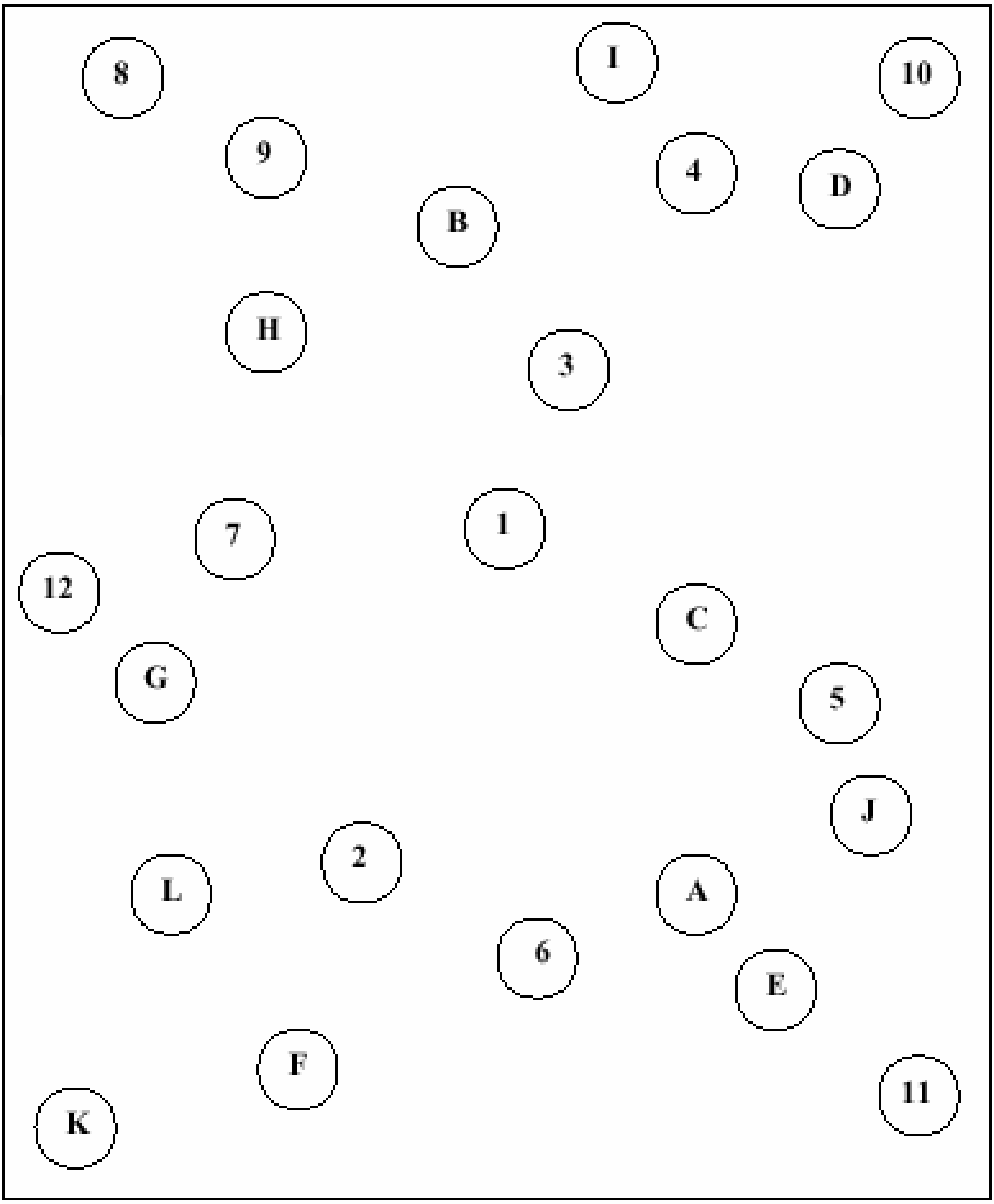
The file is above the plate

Dergi halının üzerinde

The plate is above the magazine

Dosya halının üzerinde

Appendix B: Trail Making Test Part B



Appendix C: Spontaneous Use Of Imagery Scale Turkish

1. Yeni bir yere giderken, yön bulma işareti olarak kullanılacak yerlerin isimlerine ek olarak ayrıntılı tanımlarını (örneğin benzin istasyonunun boyut, şekil ve renklerini) içeren tarifleri tercih ederim.

2. Eğer bir kısmı ağaçların arkasında kalmış bir araba görürsem, otomatik olarak onu tamamlarım ve bütün arabayı zihnimin gözünde görürüm.

3. Eğer bir mağazada yeni bir mobilya bakıyorsam, hep o mobilyanın evimin belirli yerlerinde nasıl duracağını gözümün önünde canlandırırım.

4. Hikâyedeki kişilerin nerede olduklarını ve ne yaptıklarını kafamda kolayca canlandırmamı sağlayan romanları okumayı, bu canlandırmanın zor olduğu romanları okumaya tercih ederim.

5. Bir akrabayı ziyaret etmeyi düşündüğümde, neredeyse her zaman o kişinin net bir resmi zihnimde canlanır.

6. Nispeten kolay teknik konular net bir şekilde metinde anlatıldığında, açıklayıcı şekil ve resimleri dikkat dağıtıcı bulurum çünkü benim konuyu gözümün önüne getirebilme becerime müdahale ederler.

7. Eğer biri bana kafamdan eklemek için iki basamaklı sayılar verirse (örn. 24 ve 31), onları eklemek için gözümün önüne getiririm.

8. Dışarı çıkmak için giyinmeden önce, ilk olarak farklı kıyafet kombinleri giyersem nasıl görüneceğimi gözümün önüne getiririm.

9. Yapmam gereken bir dizi işi düşündüğümde, gideceğim dükkânları kafamda canlandırırım.

10. Bir arkadaşımın sesini ilk duyduğumda, onun görsel resmi neredeyse her zaman aklımda canlanır.

11. Gerçekte hiç görmediğim bir radyo programcısı veya DJ'in sesini duyduğumda, genelde kendimi onların nasıl görünebileceğini hayal ederken bulurum.

12. Eğer bir araba kazası gördüysem, daha sonradan ayrıntıları hatırlamaya çalışırken ne olmuş olduğunu kafamda canlandırırım.