

KOÇ UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES & HUMANITIES

UNDERSTANDING GESTURE-SPEECH INTERACTION THROUGH THE LENS OF
GESTURES' SELF-ORIENTED FUNCTIONS: GESTURE PRODUCTION AND SPEECH
DISFLUENCY

BY

BURCU ARSLAN

A Dissertation Submitted for the degree of Doctor of Philosophy in Psychology

Koç University

July 2023

Koç University
Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a doctoral thesis by

BURCU ARSLAN

and have found that it is complete and satisfactory in all respects, as a thesis for
the degree of Doctor of Philosophy.

Committee Members:

Assoc. Prof. Tilbe Göksun
advisor Koç University _____

Prof. Aylin C. Küntay Koç University _____

Asst. Prof. Terry Eskenazi Koç University _____

Assoc. Prof. Reyhan Furman University of Central Lancashire _____

Assoc. Prof. Aslı Aktan-Erciyes Kadir Has University _____

Date: _____

Statement of Authorship

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Burcu Arslan

2023



Dissertation Abstract

Burcu Arslan, “Understanding gesture-speech interaction through the lens of gestures’ self-oriented functions: Gesture production and speech disfluency”

People spontaneously gesture while speaking. The gesture-for-conceptualization hypothesis argues that gestures have *self-oriented functions*. Using representational gestures (e.g., leaving a small space between two hands to mean small) decreases cognitive load and helps individuals think and speak. Healthy individuals occasionally become disfluent when they speak. Do gestures facilitate cognitive processes and pave the way for more fluent speech, particularly when speaking is difficult? In this dissertation, I investigated multimodal language production in diverse groups, with a particular focus on gesture and speech disfluency. In three studies, I examined gesture-speech interaction as a function of age and bilingualism status as well as across bilinguals’ first (L1) and second (L2) languages. In these groups, I investigated (1) speech disfluency and gesture production, (2) to which extent the two phenomena are related and temporally co-occur in spontaneous language, and (3) whether using gestures makes speech more fluent, complying with gestures’ self-oriented functions. Chapter 2 examined whether younger and older adults revealed different types and rates of disfluencies and gestures in describing paintings and how age could affect the co-occurrence between gesture and speech disfluency, which has been rarely studied in the literature. Results indicated that younger adults produced higher proportions of filled pauses while older adults produced higher proportions of repetitions. Similarly, younger adults were more likely to produce representational gestures. Moreover, the proportion of speech disfluencies that occur with a gesture was significantly higher for younger than older adults. However, the two age groups were comparable in terms of the proportion of gestures that were accompanied by a speech disfluency. These findings suggest that gestures do not necessarily compensate for weak language skills. Instead, younger adults’ language production system might be better at benefiting gesture modality to resolve temporary problems in speech. Chapter 3 investigated whether disfluency and gesture rates differed across 5- and 7-year-old Turkish monolinguals and Turkish-English bilinguals. I also examined whether children’s gesture use was associated with their speech fluency in narrative production in both languages. Results showed that bilinguals were more disfluent than monolinguals, particularly for silent and filled pauses. Monolinguals produced more iconic gestures than bilinguals. Representational gesture frequency alone was not significantly associated with the disfluency rate in Turkish narratives. Instead, children’s overall gesture frequency predicted disfluency rates only in Turkish, controlling for language proficiency and linguistic complexity. These findings together indicate that representational gestures’ self-oriented functions might not be fully reflected in an emerging language system. Chapter 4 focused on bilingualism and multimodal language production in adults. I investigated disfluency and representational gesture rates of bilinguals in their two languages and whether and when representational gestures co-occurred with disfluencies across the two languages. Results demonstrated that bilingual adults were more disfluent in L2-English even after L2 language proficiency was controlled, suggesting that disfluencies might also stem from communicative intentions. Although bilinguals produced more representational gestures in their L2, this difference disappeared after L2 proficiency was controlled. The more participants were proficient in English, the more they used representational gestures in that language. Similarly, disfluency-gesture co-occurrences were more common in English. L2 proficiency was positively correlated with the likelihood of a disfluency being accompanied by a gesture.

These findings suggest that individuals might use more representational gestures when they are disfluent if they have enough competency to detect their mistakes, which suggests a metacognitive background for gesture production. Overall, results from the three studies provide insight into typical multimodal language production through an investigation of disfluency and gesture patterns within and across populations. Despite their self-oriented functions, representational gestures do not compensate for weak language skills, but rather, they are closely linked to language proficiency. As a result, representational gestures' self-oriented functions might not be clearly reflected in children and older adults, which represent emerging and decaying language systems, respectively. In such populations, different gesture types might be orchestrated in the multimodal system, contributing to fluency. This dissertation theoretically informs the gesture-for-conceptualization hypothesis, suggesting a research avenue for representational gestures' self-oriented functions across age groups, bilingualism status, and one's L1 versus L2.

Keywords. gesture, disfluency, bilingualism, aging

Tez Özeti

Burcu Arslan, “Jest-konuşma etkileşimine jestlerin kişiye dönük fonksiyonları üzerinden bakış: Jest üretimi ve konuşma akışındaki bozulmalar”

İnsanlar konuşurken sıklıkla jest yaparlar. Kavramsallaştırma-için-just hipotezi, jestlerin öz-yönelimli işlevlere sahip olduğunu savunur. Temsilî jestler üretmek (örn., küçük anlamına gelmesi için iki el arasında küçük bir boşluk bırakmak) bilişsel yükü azaltır ve bireylerin düşünmesine ve konuşmasına yardımcı olur. Sağlıklı bireyler konuşurken zaman zaman akıcı olmayabilirler. Jestler bilişsel süreçlere yardımcı olur ve özellikle konuşma zor olduğunda daha akıcı konuşmanın yolunu açar mı? Bu tezde, çeşitli gruplarda çok modlu dil üretimini, özellikle jest ve sözlü dilin akıcılığına odaklanarak araştırdım. Üç çalışmada, jest-konuşma etkileşimini yaş, ikidillilik ve ikidilli bireylerin birinci (L1) ve ikinci (L2) dilde konuşması gibi faktörleri ele alarak inceledim. Bu gruplarda (1) konuşma akıcılığı ve jest üretimi, (2) iki olgunun ne ölçüde ilişkili olduğu ve spontan konuşmada ne zaman birlikte meydana geldiği ve (3) jestlerin öz-yönelimli işlevleriyle uyumlu olarak jest kullanmanın konuşmayı daha akıcı hale getirip getirmediği araştırıldı. İkinci bölümde, daha genç ve daha yaşlı yetişkinlerin resim betimlerken hangi türlerde ve oranlarda akıcılık ve jestler sergilediklerini ve alanyazında nadiren ele alınan bir konu olarak yaşı jest ve konuşma akıcılığı arasındaki ilişkiye olan etkisini inceledim. Sonuçlar, daha genç yetişkinlerin dolgu duraksamaları daha çok kullandığı, daha yaşlı yetişkinlerin ise daha yüksek oranlarda tekrarlamalar yaptığını gösterdi. Benzer şekilde, daha genç yetişkinlerin temsilî jest üretme eğilimleri daha yüksekti. Dahası, konuşmanın akıcı olmadığı durumlara bir jestin eşlik etmesi oranı genç yetişkinlerde daha yüksek bulundu. Bununla birlikte, bir jestin akıcılığının bozulduğu bir durumla birlikte olma oranı iki grup arasında benzerdi. Bu bulgular, jestlerin zayıf dil becerilerini zorunlu olarak telafi etmediğini göstermektedir. Ancak, daha genç yetişkinlerin dil üretim sisteminin geçici konuşma problemlerini çözmek için jest modalitesinden daha iyi faydalanyor olabilirler. Üçüncü bölümde, 5 ve 7 yaşındaki anadili Türkçe olan ve Türkçe-İngilizce ikidilli olan çocukların konuşmada akıcılık ve jest üretme oranlarını ve jest kullanımının her iki dilde de anlatılarının akıcılığı ile ilişkili olup olmadığını inceledim. Sonuçlar, ikidillilerin sessiz ve dolgu duraksamalar açısından tekdillilere göre daha az akıcı olduğunu gösterdi. Tekdilli çocuklar ikidillilere göre daha fazla temsilî jest üretti. Temsilî jest sıklığı ile Türkçe anlatıların akıcılık oranı arasında anlamlı bir ilişki bulunmadı. Bunun yerine, çocukların genel jest sıklığı, dil yeterliliği ve dilbilimsel karmaşıklık kontrol edildiğinde sadece Türkçe anlatılardaki akıcılık oranlarını yordamıştır. Tüm bu bulgular, temsilî jestlerin öz-yönelimli işlevlerinin henüz gelişmekte olan bir dil sisteminde tam olarak yansımamış olabileceğini göstermektedir. Dördüncü bölüm, yetişkinlerde ikidillilik ve çok modlu dil üretimine odaklandı. İkidilli yetişkinlerin birinci (L1) ve ikinci dillerinde (L2) akıcılık ve temsilî jest oranlarını ve temsilî jestlerin konuşmadaki akıcılığının bozulduğu durumlarla zamansal olarak bir ilişki içerisinde olup olmadığını inceledim. Sonuçlar, ikidilli yetişkinlerin ikinci dil olan İngilizce'de, dil yeterliliği kontrol edildikten sonra bile Türkçe'ye göre daha az akıcı olduklarını gösterdi. Bu durum akıcılıktaki bozulmaların iletişimsel niyetlerden de kaynaklanabileceğine işaret ediyor olabilir. İkidilliler L2'de daha fazla temsilî jest üretmiş olsalar da, bu fark L2 yeterliliği kontrol edildikten sonra ortadan kaybolmuştur. Katılımcılar İngilizce'de ne kadar yetkin olurlarsa, o kadar çok temsilî jest kullandılar. Benzer şekilde, akıcılığın bozulması ve jestlerin zamansal olarak birlikte olması durumu İngilizce'de daha yaygındı. İngilizce yeterliliği, bir akıcılık bozulmasının bir jestle birlikte olma olasılığı ile pozitif bir şekilde ilişkilendirilmiştir. Bu

bulgular, bireylerin kendi hatalarını tespit edebilecek yetkinliğe sahip olduklarında daha fazla temsilî jest kullandıklarını göstererek jestlerin üstbilişsel arka planına işaret etmektedir. Genel olarak, üç çalışmanın sonuçları, farklı popülasyonlar içinde ve arasında akıcılık ve jest desenlerini inceleyerek sağlıklı bireylerde çok modlu dil üretimi konusunda bir içgörü sağlamaktadır. Öz-yönelimli işlevlerine rağmen, temsilî jestler dil yeterliliği ile yakından ilişkilidir ve zayıf dil becerilerini her zaman telafi etmez. Sonuç olarak, temsilî jestlerin öz-yönelimli işlevleri sırasıyla gelişmekte olan ve gerileyen dil sistemlerini temsil eden çocuk ve yaşlı yetişkinlerde tam olarak kendini göstermeyebilir. Bu popülasyonlarda, farklı jest tipleri çok modlu dil sisteminde organize edilerek sözlü dilin akıcılığına katkıda bulunabilir. Bu tez, temsilî jestlerin öz-yönelimli işlevlerini anlamak için yaş grupları, ikidillilik durumu ve kişilerin birinci ve ikinci dilini kapsayan bir araştırma alanı önererek kavramsallaştırma-için-jest hipotezine teorik olarak katkıda bulunmaktadır.

Anahtar sözcükler. jest, konuşmadaki akıcılık bozulması, çiftidillilik, yaşlanma

DEDICATION



to my family...

ACKNOWLEDGEMENTS

It is difficult to believe that I am at the end of a long journey. I consider myself very lucky to have been surrounded by such smart and supportive people. Although starting a new chapter is exciting, I do not feel like leaving Language and Cognition lab, in other words, my second home.

After finishing my Bachelor's degree, it was not easy to find my path. I would like to thank my supportive and hardworking advisor, *Tilbe Göksun*. Being always present and supportive, she has guided me for more than five years and helped me to grow as a skilled researcher. Her positive energy, compassion, and dedication have been the cornerstone of my research productivity during this period. Dear *Tilbe hocam*, I cannot think of a better academic advisor than you! You have been the source of my inspiration. I am grateful that we found each other and you chose me for a five-year commitment.

I would like to thank Aslı Aktan-Erciyes for her support and guidance. I first joined the lab by involving in one of her studies. I learned a lot from her, and I hope we will continue to collaborate in the future. I also thank my dissertation jury, *Reyhan Furman*, *Terry Eskenazi*, and *Aylin Küntay*, for their valuable comments and motivating words. I also thank *Çağlar Akçay* and *Nazbanou Nozari*: you have been amazing collaborators and I learned a lot from you!

I am not sure if I can express how grateful I am to the members of *Language and Cognition* and *Language and Communication Lab*. Their presence and support have made my job easier during such a long period of time. Going to the lab to spend time together has been one of my favorite activities. Finding such a supportive and fun environment is not easy, and I feel lucky to be surrounded by such a group of people in such a challenging period. I would like to thank everyone who contributed to my journey: *Demet Özer*, *Emir Akbuğa*, *Can Avcı*, *Erim Kızıldere*, *Işıl Doğan*, *Sura Ertaş*, *Şeref Esmer*, *Hande Sodacı*, *Gyulten Hyusein*, *Ece*

Yallak, İbrahim Akkan, Sümeyye Koşulu, Salih Özdemir, Can Ceylan, Mert Kobaş, Begüm Yılmaz, and Dilay Karadöller. It is amazing to find people who are supportive and present both on good and bad days. You have been the source of my joy and laughter! I am lucky to have you. I would like to also thank all research assistants who contributed to my research at any stage. Last, I am grateful to everyone who participated in my studies.

I also thank *İsa*, who supported me with his amazing coding skills. I am excited to start a new chapter with you! I also thank *Buse*, my dark queen, for being present both on good and bad days. I am grateful to have *Sümeyye, Fatih, Beyza* for the unique connection we have been maintaining for so long. Being with you has always been full of laughter! I also thank *İlkin* and *Özgür* for their support and the amazing days we spent together.

And my family! I thank my father, *Recep*, my mother, *Makbule*, my sister, *Buket*, and my grandmother, *Aysel*. Thank you for your trust and support! You are the secret heroes of all my achievements. This dissertation would not be possible without you. Last, I thank my academic grandfather, *Ercüment Atabay*. He has contributed to my intellectual growth since I was a little kid. Talking to him about Dostoyevsky's novels and listening to his memories with historical figures was bliss. He always wanted me to become a teacher. I hope he rests in peace.

I acknowledge the financial support of the Scientific and Technological Research Council of Turkey (TÜBİTAK) during my PhD. I thank TÜBİTAK for funding my doctoral studies.

FUNDING

This work was supported by the Scholarship Programme for Ph.D. Students (2211-A) awarded by The Scientific and Technological Research Council of Turkey (TÜBİTAK) to Burcu Arslan and TÜBA-GEBİP 2018 award given (Turkish Academy of Sciences Outstanding Young Scientist Award) and a James S. McDonnell Foundation Scholar Award (Grant no: 220020510) to Tilbe Göksun.



TABLE OF CONTENTS

CHAPTER 1 General introduction.....	1
1.1. Multimodal language production.....	2
1.2. The scope and outline of the dissertation.....	5
CHAPTER 2 Aging, gesture production, and disfluency in speech: A comparison of younger and older adults.....	7
Abstract.....	8
2.1. Introduction.....	9
2.1.1. Speech and disfluency.....	10
2.1.1.1. A closer look at disfluency types.....	11
2.1.2. Gesture and disfluency.....	14
2.1.3. Aging, disfluency, and gesture.....	17
2.1.4. The present study.....	20
2.2. Methods.....	22
2.2.1. Participants.....	22
2.2.2. Materials.....	23
2.2.3. Procedure.....	24
2.2.4. Transcription and coding.....	24
2.2.4.1. Speech and disfluency.....	24
2.2.4.2. Gestures.....	25
2.2.4.3. Disfluency-gesture segments.....	26
2.3. Results.....	29
2.3.1. Speech and disfluency.....	29
2.3.2. Gesture.....	31
2.3.3. Speech disfluency and gesture co-occurrences.....	33
2.4. Discussion.....	35
CHAPTER 3 Multimodal language in bilingual and monolingual children: Gesture production and speech disfluency.....	46
Abstract.....	47
3.1. Introduction.....	48
3.1.1. Speech disfluency.....	49
3.1.1.1. Disfluency from a developmental perspective.....	50
3.1.1.2. Speech disfluency in bilingual children.....	51

3.1.2. Gesture.....	53
3.1.2.1. Gesture production from a developmental perspective.....	54
3.1.2.2. Gesture production in bilingual children.....	54
3.1.2.3. Gesture and speech disfluency.....	56
3.1.3. Present Study.....	57
3.2. Methods.....	59
3.2.1. Participants.....	59
3.2.2. Materials.....	61
3.2.3. Procedure.....	62
3.2.4. Coding.....	62
3.2.4.1. Speech and disfluency.....	62
3.2.4.2. Gesture.....	63
3.2.4.3. Syntactic complexity.....	63
3.3. Results.....	64
3.3.1. Analysis plan.....	64
3.3.2. Preliminary results.....	65
3.3.2.1. TIFALDI and PPVT-4 scores.....	65
3.3.2.2. Syntactic complexity.....	66
3.3.3. Speech disfluency.....	66
3.3.3.1. Monolingual-bilingual comparison.....	67
3.3.3.2. Bilinguals' L1 vs. L2.....	68
3.3.4. Gesture.....	69
3.3.4.1. Monolingual-bilingual comparison.....	69
3.3.4.2. Bilinguals' L1 vs. L2.....	70
3.3.5. Gesture and speech disfluency.....	70
3.3.5.1. Turkish narrative samples.....	71
3.3.5.2. English narrative samples.....	72
3.4. Discussion.....	73
CHAPTER 4 Do bilingual adults gesture when they are disfluent? Understanding gesture-	
speech interaction across first and second languages.....	81
Abstract.....	82
4.1. Introduction.....	83
4.1.1. Disfluency.....	83
4.1.2. Disfluency in L1 and L2.....	83

4.1.3. Gesture.....	85
4.1.4. Gesture in L1 and L2.....	87
4.1.5. Speech fluency and gesture.....	88
4.1.6. Present Study.....	90
4.2. Methods.....	91
4.2.1. Participants.....	91
4.2.2. Materials.....	91
4.2.3. Procedure.....	92
4.2.4. Coding.....	93
4.2.4.1. Speech and Disfluency.....	93
4.2.4.2. Gesture.....	94
4.2.4.3. Gesture-Disfluency co-occurrences.....	94
4.3. Results.....	95
4.3.1. Speech and Disfluency.....	95
4.3.2. Gesture.....	96
4.3.3. Gesture-disfluency co-occurrences.....	96
4.3.3.1. Disfluency perspective.....	96
4.3.3.2. Gesture perspective.....	97
4.4. Discussion.....	97
4.4.1. Disfluency.....	98
4.4.2. Gesture.....	100
4.4.3. Gesture-disfluency co-occurrences.....	101
4.5. Conclusion.....	104
CHAPTER 5 General discussion.....	105
5.1. Summary and the discussion of main findings.....	106
5.2. Future directions.....	111
5.3. Conclusion.....	112
APPENDICES.....	113
REFERENCES.....	114
LIST OF PUBLICATIONS.....	136
LIST OF PRESENTATIONS.....	137

LIST OF FIGURES

2.1. The filled pause use by younger and older adults.....	30
2.2. The repetition use by younger and older adults.....	30
2.3. The repair use by younger and older adults.....	31
2.4. The representational gesture use by younger and older adults.....	32
2.5. The deictic gesture use by younger and older adults.....	33
2.6. The proportion of disfluencies that occur with gesture.....	34
2.7. The proportion of gestures that occur with disfluency.....	34



LIST OF TABLES

3.1. Descriptive statistics of children's language measures.....	64
3.2. Regression analysis, the disfluency rate in Turkish as the outcome variable.....	71
3.3. Regression analysis, the disfluency rate in English as the outcome variable.....	72



CHAPTER 1

GENERAL INTRODUCTION

1.1. Multimodal language production

Language is multimodal. While speaking, people spontaneously produce hand movements called *co-speech gestures* (Goldin-Meadow, 2005). Although gesture and speech are two different modalities, research argues that the two mechanisms are closely linked (Kita & Özyürek, 2003). Gestures, particularly representational gestures (gestures that refer to objects or situations, e.g., describing a window as a square), are semantically related to the speech content (McNeill, 1992). Moreover, studies indicate that gestures temporally precede their lexical affiliates, suggesting a dynamic interplay between gesture and speech modalities (Seyfeddinipur, 2006; Seyfeddinipur & Kita, 2001; Ter Bekke et al., 2020).

Apart from their role in communication, the gesture-for-conceptualization hypothesis argues that gestures have *self-oriented functions*: using representational gestures decreases cognitive load and helps people manage and facilitate their cognitive processes (Kita et al., 2017). In line with this hypothesis, research demonstrates that people are more likely to produce gestures when they interpret complex concepts and situations (Kita & Davies, 2009; Melinger & Kita, 2007) or when they describe objects from memory (Ping & Goldin-Meadow, 2010; Morsella & Krauss, 2004). These findings suggest that gestures come into play when thinking and speaking are difficult.

How do we understand when speaking is difficult for someone? One of the prominent signs is speech disfluency, i.e., temporary interruptions in speech. While speaking, people are likely to give breaks (e.g., being silent for a while or saying *um* to fill the pause), repeat words, or revise and repair speech content (MacLay & Osgood, 1959). People are more disfluent when they describe complex objects (Bortfeld et al., 2001; Morsella & Krauss, 2005) and produce longer sentences (Oviatt, 1995). Similarly, disfluencies are frequently observed in sentence beginnings, suggesting that speech disfluencies are interwoven with planning load (Shriberg, 1996). Although research suggests that disfluent segments in speech might stem from a

cognitive load, a possible communicative role of disfluencies is also widely discussed (for a review, see Corley & Stewart, 2008). Individuals might use filled pauses (e.g., um) to signal their addressees that they have more to say and continue to talk soon (Brennan & Williams, 1995).

If gestures decrease cognitive load and facilitate one's thinking and speaking, are people more fluent when they use gestures? Previous research indicates that individuals are more disfluent when their hand use is experimentally restricted (Finlayson et al., 2003; Rauscher et al., 1996). Similarly, word retrieval is less effortful when people produce gestures (Morsella & Krauss, 2005). However, in a recent study, Avcı and colleagues (2023) suggest that experimentally restricting gesture use might yield different effects on individuals depending on how frequently they gesture in their spontaneous speech. They argued that those who frequently used abstract pointing gestures (i.e., gestures that do not have concrete referents) in spontaneous speech were more likely to be disfluent in the gesture-restricted condition. These findings indicate the role of individual differences in gesture-speech interaction, suggesting that people might differ in extracting cognitive benefits from gestures in discourse.

Research indeed argues that there are both individual and group differences in multimodal language production (for a review, see Özer & Göksun, 2020). One of the prominent variables that drive such a difference is age. Evidence suggests that, compared to younger adults, older adults are more likely to be disfluent and less likely to use complex gesture forms, i.e., representational gestures (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999). On the other side of the lifespan developmental trajectory, studies demonstrate that representational gestures are intact with language skills, and children do not commonly use representational gestures until they reach a certain language level (Iverson & Goldin-Meadow, 2005).

Another variable that drives differences in multimodal language production is bilingualism. Research on different age groups demonstrates that bilingual individuals gesture more frequently than their monolingual counterparts (e.g., Nicoladis et al., 2009; Pika et al., 2006; So, 2010). These findings suggest a role for gestures in decreasing the cognitive load of controlling two language systems and helping bilinguals manage the target vocabulary and grammar (Nicoladis, 2007). Studies also indicate that bilinguals, particularly bilingual children, are more disfluent and slower in word retrieval compared to monolinguals speaking their languages (e.g., Bedore et al., 2006; Byrd et al., 2015).

It is important to note that bilingualism is used as an umbrella term, while bilinguals differ in terms of when they acquire a second language. Speaking a second language (L2) can be a challenging task, and bilinguals are usually more disfluent in their L2 than their first language (L1) (Bergmann et al., 2015; Gollan et al., 2005). In line with gestures' self-oriented functions, Hadar and Butterworth (2001) suggests that people gesture more frequently in their L2 than L1 to compensate for weak language skills and facilitate speech production. Yet, the evidence argues that individuals with higher L2 proficiency are more likely to use complex and meaning-enhancing gesture forms (i.e., representational gestures), compared to those with lower L2 proficiency (Gregersen et al., 2009). These findings together indicate a role for individual and group differences in multimodal language production, suggesting that understanding gesture-speech interaction requires a lifespan approach where multiple characteristics are considered.

Although previous studies separately examined speech disfluency and gesture across different age and language groups, less is known about how the two modalities are orchestrated in the language system (Seyfeddinipur, 2006). In this dissertation, I study multimodal language production in diverse groups, with a particular focus on gesture and speech disfluency. Based on the gesture-for-conceptualization hypothesis (Kita et al., 2017), I examine whether gestures'

self-oriented functions are reflected in the form of fluency. Focusing on the gaps in the literature, I hold a lifespan perspective targeting different age (children, young adults, and older adults) and language groups (bilingual and monolingual) as well as within group differences (bilinguals' L1 versus L2). In these groups, I investigate (1) speech disfluency and gesture production, (2) to which extent these two phenomena are related and temporally co-occur in spontaneous language, and (3) whether using gestures makes speech more fluent, complying with gestures' self-oriented functions.

1.2. The scope and outline of the dissertation

This dissertation investigates multimodal language production, in particular gesture production and speech disfluency, through the lens of gestures' self-oriented functions. I aim to provide insight into individual and group differences in multimodal language use and the temporal and pragmatic relationship between speech disfluency and gesture. To this end, **Chapter 2** investigates whether younger and older adults reveal different types and rates of speech disfluencies and gestures in describing paintings and how age can affect the cooccurrence between gesture and speech disfluency, which has been rarely studied in the literature. **Chapter 3** focuses on the other end of the developmental trajectory: children. This chapter investigates whether disfluency and gesture rates differ across 5- and 7-year-old monolinguals and bilinguals. From a gestures' self-oriented functions perspective, I also examine whether children's gesture use is associated with their speech fluency in narrative production. Last, this chapter studies whether bilinguals reveal similar disfluency and gesture rates across their two languages. **Chapter 4** extends the previous chapter by studying bilingualism and multimodal language production in adults. In this chapter, I investigate the disfluency and representational gesture rates of bilinguals in their two languages and whether and when representational gestures co-occur with disfluencies across their two languages. In

Chapter 5, I discuss these three studies in an integrative fashion and suggest future directions for multimodal language research.



CHAPTER 2

AGING, GESTURE PRODUCTION, AND DISFLUENCY IN SPEECH: A COMPARISON OF YOUNGER AND OLDER ADULTS

This chapter is published as Arslan, B. & Göksun, T. (2022). Aging, gesture production, and disfluency in speech: A comparison of younger and older adults. *Cognitive Science*, 46(2), e13098. doi: 10.1111/cogs.13098

Abstract

Age-related changes are observed in the speech and gesture of neurotypical individuals. Older adults are more disfluent in speech and use fewer representational gestures (e.g., holding two hands close to each other to mean small) compared to younger adults. Using gestures, especially representational gestures, is common in difficult tasks to aid the conceptualization process and to facilitate lexical access. This study investigates how aging can affect gesture production and the co-occurrence between gesture and speech disfluency. We elicited speech and gesture samples from younger and older adults ($N = 60$) by using a painting description task that provided concrete and abstract contexts. Results indicated that albeit the two age groups revealed comparable overall speech disfluency and gesture rates, they differed in terms of how their disfluencies and gestures were distributed across specific categories. Moreover, the proportion of speech disfluencies that occur with a gesture was significantly higher for younger than older adults. However, the two age groups were comparable in terms of the proportion of gestures that were accompanied by a speech disfluency. These findings suggest that younger adults' language production system might be better at benefiting from other modalities, i.e., gesture, to resolve temporary problems in speech planning. However, from a gesture perspective, it might be difficult to differentiate between gestures' self-oriented and communicative functions and understand their role in speech facilitation. Focusing on specific cases where speech disfluency and gestures co-occur and considering individual differences might bring insight into multimodal communication.

2.1. Introduction

When people speak, they produce gestures that accompany their speech, called *co-speech gestures* (McNeill, 1992). There have been various attempts to understand the nature of gesture production while elaborating on how speech and gesture interact from a multimodal perspective (e.g., Alibali et al., 2000; Hostetter & Alibali, 2019; Kita et al., 2017; Kita & Özyürek, 2003). Gesture and speech are suggested to have originated from the same system (McNeill, 1992), while others have approached these two as closely linked but separate mechanisms (Kita, 2000; Kita & Özyürek, 2003). It has also been a question of whether co-speech gestures facilitate lexical access (Hadar & Butterworth, 1997; Krauss, 1998; Krauss et al., 2000; Rauscher et al., 1996) or conceptual planning (Kita, 2000) and, as a result, speech production. In line with this argument, Kita et al. (2017) suggested that gestures have self-oriented functions as they *activate*, *manipulate*, *package*, and *explore* information units. If this is the case, one might expect gesture use to occur or increase when people face temporary problems in speech, i.e., disfluencies, such as filling up pauses with ‘*um*’ or repeating a word.

Aging, on the other hand, is a critical process that is associated with decay in multimodal language skills. More specifically, previous research indicates that although overall gesture use is comparable across age groups, older adults produce fewer representational gestures (e.g., iconic gestures, deictic gestures) than younger adults (e.g., Arslan & Göksun, 2021; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999; for a review see Göksun et al., 2022). Furthermore, producing grammatically complex sentences becomes less likely as people age (Kemper et al., 2001). Older individuals also reveal more disfluent speech (Bortfeld et al., 2001; Obler & Albert, 1981; Yairi & Clifton, 1972) and face more word retrieval difficulties compared to younger adults (Brown & Nix, 1996; Burke & Shafto, 2004). Considering that younger and older adults already reveal some differences in gesture production and speech disfluency, we argued that, in a multimodal language system, the way gestures and speech are

organized might also differ across younger and older adults. This difference might be reflected in disfluency-gesture co-occurrences, providing some cues on the self-oriented functions of gestures across lifespan.

In this study, we ask whether (1) younger and older adults reveal different types and rates of speech disfluencies and gestures in describing paintings, (2) how age can affect the co-occurrence between gesture and speech disfluency, which has been rarely studied in the literature yet.

2.1.1. Speech and Disfluency

Speech is a prominent communication tool that we use in daily life. Although individuals can achieve effective communication through speech, the speech production process is not that flawless. Healthy individuals' speech is prone to reveal temporary errors, revisions, and pauses that are called *disfluencies* (Maclay & Osgood, 1959; for a review, see Wingate, 1987).

Studying disfluent speech might provide cues on how language interacts with cognitive processes. More specifically, planning and producing speech require the involvement of cognitive resources (for a review, see Acheson & Macdonald, 2009; for a review addressing the relationship among aging, working memory, and language, see Kemper, 2012; for a meta-analysis on the role of working memory in second language abilities, Linck et al., 2014). When processing and planning loads are increased by manipulating the difficulty of tasks or task items, individuals can be more disfluent due to increasing demand for cognitive resources (e.g., Bortfeld et al., 2001; Oviatt, 1995; Shriberg, 1996). Therefore, individuals' strategies and patterns while producing speech under different circumstances might shed light on how they administer and allocate their available cognitive resources to form semantically and grammatically correct sentences (Bock, 1995, 1996). Speech disfluency might have communicative implications, as well. To keep their listeners' attention, individuals might place

filler words (e.g., *well, you know*) or filled pauses (e.g., *um*) between words to signal that they are having only a temporary speech problem and will continue as soon as they resolve it (Clark, 1994; Smith & Clark, 1993).

Although previous research has suggested different ways of categorizing and studying disfluency in speech (e.g., Bortfeld et al., 2001; Cooper, 1990; Johnson, 1961; Obler & Albert, 1981; Yairi, 1981; Yairi & Clifton, 1972), the original disfluency classification goes back to Maclay and Osgood (1959) as they suggested four main speech disfluency categories: (1) *Silent pauses* are temporary discontinuities in a speech where individuals give a silent break within or between phrases, (2) *Filled pauses* are breaks in speech, filled with lexical items that do not convey meaning (e.g., *um*), (3) *Repetition* refers to repeating syllables, words, or phrases (e.g., *under under the table*), (4) *Repairs* include grammatical or semantic revisions in which grammatical structures (e.g., affixes) or words are reorganized, elaborated, and replaced with more plausible alternatives (e.g., *it is brown – I mean black*). Studies also used an aggregated single disfluency score in the literature (e.g., Oviatt, 1995). However, studying disfluency categories separately might provide a more elaborated look on the nature of disfluency in cognitive processing as well as communication strategies across age groups, sexes, and tasks with different difficulty levels (Bortfeld et al., 2001).

In sum, speech disfluency has both cognitive and communicative implications. From a cognitive perspective, disfluency patterns in speech can be considered as strategies to deal with speech planning and processing load. From a communicative perspective, some forms of disfluency might be used to maintain a *conversational floor* while talking to other(s). How disfluency is distributed across its categories might be an important point to consider its functions on cognition and communication.

2.1.1.1. A closer look at disfluency types

Disfluency rate and types are observed across different age groups (Bóna, 2011; Bortfeld et al., 2001; Duchin & Mysak, 1987; Kemper et al., 2009; Yairi & Clifton, 1972) and in different contexts (Bortfeld et al., 2001; Merlo & Mansur, 2004; Oviatt, 1995; Schachter et al., 1991; Shriberg, 1996). While some studies have focused on filled pauses (e.g., Clark & Fox Tree, 2002; Corley & Stewart, 2008; Fraundorf & Watson, 2011; Horváth, 2010), others have emphasized repetition or repair patterns in speech (e.g., Brennan & Schober, 2001; Shriberg, 1996). Parsing disfluency patterns into categories has provided a detailed understanding of the nature of disfluency from cognitive (Bock, 1995, 1996; Plevaets & Defrancq, 2018) and communicative perspectives (Brennan & Schober, 2001; Clark, 1994; Smith & Clark, 1993).

From a cognitive perspective, processing load in relation to syntactic planning might be an important point in understanding the underlying functions of disfluency types. As processing cost increases, individuals become more disfluent (Bortfeld et al., 2001; Oviatt, 1995; Shriberg, 1996). For instance, the likelihood of being disfluent is higher while dealing with longer utterances (Oviatt, 1995) and sentences (Shriberg, 1996). Furthermore, individuals are more disfluent at the beginning of their sentences due to the high amount of planning effort required (Shriberg, 1996; Boomer, 1965). Similarly, syntactically complex sentences are likely to lead to disfluencies (Haynes & Hood, 1978; Silverman & Ratner, 1997). These findings are not surprising as longer and linguistically complex sentences would be more costly to process due to their heavy demands on working memory resources.

Another way to observe the effect of processing load on disfluency rates and patterns is by manipulating task difficulty or ambiguity. For instance, Bortfeld et al. (2001) asked younger and older adults to describe pictures of tangrams and children. They found that, regardless of age, individuals were more disfluent while describing tangrams than describing pictures of children. As tasks become more cognitively demanding and ambiguous, people are more likely to face syntactic and conceptual planning problems due to increased cognitive load.

Furthermore, the likelihood of using repetitions and repairs were higher while describing tangrams as opposed to pictures of children. It was not the case for filled pauses (e.g., um) as they might not solely reflect the planning load. In such scenarios on processing cost, one could discuss what kind of cognitive and communicative processes motivates the production of different disfluency types.

Braver et al. (2007) proposes the dual mechanisms of control account (DMC), which argues for two types of cognitive control. In the case of a *proactive control*, performance is constantly monitored from the moment a goal is set until it is satisfied. On the other hand, *reactive control* is a strategy that is used in the existence of a trigger event. To put it another way, proactive control can be considered as early detection while reactive control can be considered as late corrective action in the face of an error. The presence of multiple disfluency types might provide cues into how cognitive control and behavioral correction mechanisms are organized to compensate for a problem in speech planning, depending on what the problem is and the moment it is detected (Bock, 1995, 1996). In the case of repairs, an error is already committed. However, one can consider filled pauses and repetitions as strategies when one predicts an error to occur. In other words, such disfluencies might be produced to prevent one from committing the predicted error, reflecting a proactive control strategy. Filled pauses occur mostly at the beginning of sentences, while repetitions occur after the articulation is initiated (Fraundorf & Watson, 2014). That is, filled pauses can be produced while one is still planning, reflecting a more deliberate form of early selection and planning than repetitions, which might be regarded as the last turn before committing an actual error.

For communicative implications of speech disfluency, Clark and Fox Tree (2002) discussed that filled pauses (e.g., um, uh) are as crucial as other linguistic units in speech. Although they do not carry semantic content, using them might enhance communication by keeping the conversational floor between speakers and listeners. Individuals might benefit from

filled pauses to keep listeners' attention by compensating their time loss while dealing with temporary speech problems (Clark, 1994; Schachter et al., 1991; Smith & Clark, 1993). Using filled pauses is common in the case of uncertainty, and this situation may not necessarily mean that they are used with pure communicative intentions (for a review, see Corley & Stewart, 2008). Therefore, as Bortfeld et al. (2001) argued, the function behind using filled pauses might be a combination of cognitive or communicative purposes.

In conclusion, categorizing disfluency types might be helpful to elaborate on the cognitive and communicative implications of disfluency. Using long and complex sentences and dealing with cognitively demanding tasks might result in higher disfluency rates and different disfluency patterns due to increased processing load and as a result, changing between proactive and reactive control strategies. Filled pauses might also be considered equal to words in terms of having a communicative value, but their use might not be intentional, which could be related to cognitive processes.

2.1.2. Gesture and disfluency

McNeill (1992) has considered gesture and speech originated from a single system and made a distinction among types of co-speech gestures by touching upon their possible underlying functions. For instance, iconic and metaphoric gestures, which refer to concrete and abstract notions, respectively, along with deictic gestures (i.e., pointing at something with finger(s)/hand(s)), are classified as representational co-speech gestures that convey meaning in line with speech. Beat gestures are rhythmic hand movements that do not convey meaning but might mark certain points in speech for emphasis. There are also emblems that carry meaning without an additional speech input (e.g., waving hands to say goodbye). Both beat gestures and emblems are classified as nonrepresentational gestures. Although the classification of McNeill (1992) remains as a building block of the gesture literature, the categories are open to slight changes and further divisions across studies (for a review from a developmental perspective,

see Cartmill et al., 2012; for a discussion, see Parrill & Sweetser, 2004). Additionally, in some current gesture frameworks, representational gestures only include iconic and metaphoric ones (e.g., Hostetter & Alibali, 2019; Kita et al., 2017).

Kita and colleagues (2017) suggested a framework on how representational gestures might provide benefits to speakers while retrieving, organizing, storing, and seeking information units. They focus on the self-oriented functions of representational gestures. For instance, previous research has indicated that when individuals are asked to describe objects/pictures that are not visually accessible, they are more likely to use representational gestures to reduce the cognitive load (Morsella & Krauss, 2004; Ping & Goldin-Meadow, 2010; Wesp et al., 2001). Morsella and Krauss (2004) also showed that individuals are likely to produce more representational gestures while describing abstract figures than concrete ones. That is, gesturing might enable individuals to deal with the cognitive load by saving available cognitive resources (Cook et al., 2012; Goldin-Meadow et al., 2001; Ping & Goldin-Meadow, 2010) and enhancing the conceptualization process (Alibali et al., 2000; Hostetter et al., 2007; Kita, 2000; Kita & Davies, 2009; Melinger & Kita, 2007).

If gesture use is associated with increasing cognitive load, one might expect individuals to gesture more where syntactic planning is difficult, and processing cost is high. That is, the self-oriented functions of representational gestures might facilitate the conceptual planning (Kita, 2000; Kita et al., 2017) and the lexical access (Krauss et al., 2000; Rauscher et al., 1996), and thus, speech production. In line with this argument, the Lexical Retrieval Hypothesis (Krauss et al., 2000) argues that gestures' primary role is to facilitate speech production through lexical access. When difficulties in word retrieval arise, individuals are likely to use representational gestures to facilitate the lexical access (e.g., Butterworth & Hadar, 1989; Rauscher et al., 1996; Morsella & Krauss, 2004) and the conceptual planning (Kita, 2000) by activating, manipulating, packaging, and exploring information units (Kita et al., 2017).

Research has also demonstrated that individuals have higher disfluency rates when gesture use is restricted (Finlayson et al., 2003; Morsella & Krauss, 2004; Rauscher et al., 1996).

In contrast, others suggest that individuals are more likely to gesture with fluent speech than disfluent speech (e.g., Akhavan et al., 2016; Christenfeld et al. 1991; Graziano & Gullberg, 2018; Kisa et al., 2021). People are also more prone to use gestures with disfluent speech in their second language than their first language, while these gestures used with disfluent speech seem to be associated with cognitive and communicative implications (Graziano & Gullberg, 2013; 2018). However, Gullberg (2011) argued that answering whether gestures compensate speech by saying yes or no might be misleading. There can be cases in which gesture facilitates speech (e.g., second language learning), but it does not mean that gestures are fully compensatory in nature (Kisa et al., 2021; for a review, see Özer & Göksun, 2020).

Research on individuals with speech impairments, on the other hand, suggests that using gestures might aid speech production. For instance, patients with left hemisphere brain injury are more likely to produce gestures while experiencing word retrieval problems than when they are fluent (Göksun et al., 2013; 2015). Moreover, Özer et al. (2019) obtained language samples from patients with impaired speech (e.g., individuals with temporal or left frontal lesions) and from healthy controls by asking them to describe motion events. The patient group overall produced more gestures than the control group. While the control group produced gestures with communicative intentions, patients with impaired speech used gestures to enhance communication and facilitate speech production. Similarly, research has shown that, compared to neurotypical adults, patients with aphasia produce semantically more informative gestures to support their impaired speech system (Akhavan et al., 2018; Cocks et al., 2013; Pritchard et al., 2015). Moreover, using iconic gestures might be associated with successfully resolving lexical retrieval issues. These results suggest that gestures can facilitate speech production if a speech problem is at stake.

In sum, representational gestures have self-oriented functions such as assisting lexical access, facilitating conceptual planning, and thus, lightening the cognitive load. However, whether gestures compensate for speech problems (e.g., disfluency) in neurotypical individuals and healthy aging remains controversial as there can be multiple factors that might influence gesture and speech interaction.

2.1.3. Aging, disfluency, and gesture

While aging is a critical process that influences language skills, previous studies have indicated that, compared to younger adults, older adults are more likely to be disfluent and experience word finding problems, especially in constrained tasks such as picture naming (for a review, see Mortensen et al., 2006). Different arguments are suggested to explain the nature of this relationship.

Among different theories that provide a general look at the cognitive aging process, the Transmission Deficit (TD) hypothesis (Taylor & Burke, 2002) brings an insight by specifically targeting aging and language production. It postulates that the connections from concepts to phonemes of words are weakened because of the adverse effects of the aging process. An idea can be retrieved through multiple associations, however, retrieving a specific word for that idea following a strict phonemic order might become burdensome. For example, tip-of-the-tongue states are more common in older than younger adults (Brown & Nix, 1996; Burke et al., 1991; Cohen & Faulkner, 1986) as they know the concepts but cannot reach the target lexicon due to the impaired connections.

The higher disfluency rates in older adults might also reflect processing difficulties due to an age-related decline in working memory resources (Kemper & Sumner, 2001). The degraded grammatical complexity of elderly adults' speech corroborates the working memory limitations on syntactic planning (Kemper, 2006; Kemper & Sumner, 2001). Kemper et al. (2003) demonstrated that when younger and older adults are given a dual task, older adults'

speech reveals simpler grammatical structures and higher disfluency rates as a result of increased cognitive demands. In such a scenario, not only disfluency rates but also disfluency types might provide cues on the relationship of syntactic planning with cognitive strategies.

Repetitions are the first disfluency form that appear in early language production while the use of filled pauses is not commonly observed until preschool years (Haynes & Hood, 1977; Yairi & Clifton, 1972). Similarly, the most common disfluency type used by older individuals is also repetition (Bortfeld et al., 2001; Mortensen et al., 2006). Engelhardt et al. (2010) argued that the repetition rates might follow a U-shape trend across lifespan, decreasing from childhood to young adulthood and then increasing again as individuals age. Moreover, overall, preschool children and older adults are more likely to be disfluent than younger adults (Yairi & Clifton, 1972). Research so far has shown that effectively using a proactive control strategy would require advanced executive functioning skills (Braver et al., 2007). These findings are in line with the development of the prefrontal cortex (PFC) and executive functions across lifespan (Braver et al., 2001; Salthouse, 2009). Preschool children and older adults would be more likely to rely on reactive control, which implies late correction of errors (Braver et al., 2007; Paxton et al., 2006). In other words, children's and older adults' use of their cognitive resources for conceptual and syntactic planning might not be as advanced as that of younger adults, which would end up with more errors in speech. In line with this argument, age group differences in speech disfluency rates become more pronounced when engaging with a difficult task (Bortfeld et al. 2001), in which older adults' tendency to use reactive control, and thus, late correction would be more prominent (Paxton et al., 2006).

If word retrieval and conceptual planning problems are more common in older adults, gesture use might increase in this age group. In particular, in cases of finding words, representational gestures might step in and help lexical access (Butterworth & Hadar, 1989; Rauscher et al., 1996; Morsella & Krauss, 2004) and choosing among competing concepts

(Kita, 2000; Kita & Davies, 2009). However, previous research has not fully supported these arguments with older individuals. Although overall gesture use seems comparable across younger and elderly individuals, younger adults use higher proportions of representational gestures than older adults (Cohen & Borsoi, 1996; Feyereisen & Havard, 1999). It should be noted that the Lexical Retrieval Hypothesis mainly focuses on retrieval issues in a spatial context. Therefore, tasks and contexts to obtain speech samples might become important factors to understand gesture and speech production (Arslan & Göksun, 2021; Schubotz et al., 2019). For instance, Schubotz et al. (2019) demonstrated that although younger and older adults used representational gestures at comparable rates in a narrative retelling task, only younger adults could adapt their speech and gestures based on whether knowledge is mutually shared with or novel to their addressees. Unlike younger adults, older adults gestured less frequently while conveying the novel content as opposed to the mutually shared content, suggesting that effectively narrating the novel content might be cognitively demanding for the older age group.

Even though previous research has come up with different tasks to elicit speech samples from different age groups (e.g., Arslan & Göksun, 2021; Bortfeld et al., 2001; Cooper, 1990; Kemper et al., 2010; Obler & Albert, 1981), constrained tasks such as picture naming or picture description are more likely to reveal age differences in language production (for a review, see Mortensen et al., 2006). When elderly individuals are asked about personal experiences, they are prone to be verbose (James et al., 1998) and might engage in circumlocution whenever they fail to retrieve a specific word. In a constrained task such as picture description/naming, however, the chance of circumlocution can decrease as the target lexicon is predetermined by the task itself and does not offer flexibility (Theocharopoulou et al., 2015). Similarly, the widely used narrative elicitation task, *The Cookie Theft Picture* (Goodglass & Kaplan, 1983), includes simple items and may not detect age differences in speech production (Cooper, 1990).

Moreover, Bortfeld et al. (2001) demonstrated that when describing a cognitively demanding item such as tangrams, individuals make more errors in speech.

Gesture use is also related to a task at hand. Individuals produce more gestures while talking about space, problem solving, or during complex tasks (e.g., Alibali, 2005; Arslan & Göksun, 2021; Garber & Goldin-Meadow, 2002; Morsella & Krauss, 2004). In addition to these, Kita and Davies (2009) showed that when the task is to describe geometric figures that differ in complexity, individuals use a higher proportion of representational gestures while describing the more complex ones (see also Morsella & Krauss, 2004). In a recent study, Arslan and Göksun (2021) showed that both younger and older adults produced a higher proportion of representational gestures in a spatial context (e.g., describing a route) than nonspatial ones (e.g., describing daily activities and completing stories). Only the spatial task, which was relatively more difficult than the others, revealed age differences in representational gesture use. Therefore, tasks that differ in difficulty might be helpful to understand age differences in language production.

Taken together, aging is associated with speech problems and changes in the use of gesture types. Task type and difficulty might matter while trying to understand the nature of disfluency and gesture use. Using constrained tasks, such as picture description, might be more useful to understand the word retrieval failures and whether they are associated with age. Overall, manipulating task difficulty and thus, cognitive load, might be a key to uncover the interplay among speech disfluency, gesture use, and the aging process.

2.1.4. The Present Study

The aim of this study is to understand speech disfluency and gesture use of younger and older individuals to bring an understanding of the relationship between aging and multimodal communication. We investigate how age can affect gesture production and the co-occurrence between gesture and speech disfluency, which has been rarely studied in the literature yet.

Considering previous findings on the importance of preferring constrained tasks (for a review, see Mortensen et al., 2006; Theocharopoulou et al., 2015) and manipulating task difficulty (Bortfeld et al., 2001; Kemper et al., 2003; Kita & Davies, 2009), we used a picture description task in which we presented three concrete and three abstract paintings to elicit speech and co-speech gestures. Describing abstract and complex figures would be more difficult to conceptualize (Kita, 2000; Kita & Davies, 2009) and correctly name them (Bortfeld et al., 2001; Morsella & Krauss, 2004). Therefore, we argue that the abstract paintings would be more difficult to describe than the concrete ones.

In line with previous findings (Bortfeld et al., 2001; Cooper, 1990), we expect that overall older adults would reveal higher disfluency rates than younger adults. Moreover, we hypothesize that, compared to younger adults, older adults will use a higher proportion of repetitions, which are considered an immature form of disfluency from a developmental perspective (Haynes & Hood, 1977; Yairi & Clifton, 1972). We also predict that such an age group difference on speech might be more prominent while describing abstract than concrete paintings, as abstract paintings are more difficult to describe and might better reveal age related effects on speech (cf. Bortfeld et al., 2001, for a similar design and comments). Furthermore, we expect comparable proportions of filled pauses and repairs across younger and older age groups since they are not associated with cognitive load and working memory capacity as much as repetitions are (Bortfeld et al., 2001; Yairi & Clifton, 1972).

We expect the abstract painting category to elicit more gestures than the concrete category for both age groups since an abstract context might be relatively more difficult to conceptualize (Kita & Davies, 2009; Morsella & Krauss, 2004). Moreover, consistent with previous research (Arslan & Göksun, 2021; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999), we expect younger and older age groups to have comparable gesture rates but to differ from each other in producing certain gesture types. More specifically, we hypothesize that

younger adults will produce a higher proportion of representational gestures (iconic and metaphoric gestures) than older adults while describing both concrete and abstract paintings. Our previous study on aging and gestural communication demonstrated that although there was not a statistically significant age group difference in terms of deictic gesture use, there was a salient direction that older adults were prone to produce a higher proportion of deictic gestures across different contexts (Arslan & Göksun, 2021). We interpreted this finding as there might be a hierarchy among gesture types in terms of complexity and the amount of mental imagery required to produce them. As representational gestures are likely to emerge from perceptual and motor simulations (Hostetter & Alibali, 2008, 2019), age-related decrease in one's mental imagery skills would affect the production of representational gestures (Arslan & Göksun, 2021). Then, older adults might produce gestures that are easier to produce and require less imagery, such as deictic gestures. Therefore, in this study, we similarly expect that older adults will use a higher proportion of deictic gestures than younger adults while describing both types of paintings.

Last, we hypothesize that, if gestures facilitate speech, individuals at both age groups can produce more representational gestures when they are disfluent. Considering the adverse effects of aging on language production, we then particularly expect older adults to reveal more gesture-disfluency co-occurrences than the younger adults handle the speech planning load and difficulties in lexical access. Moreover, we predict such a pattern to be more prominent while describing abstract paintings since they are cognitively more demanding and retrieving words in such a context might become more difficult.

2.2. Methods

2.2.1. Participants

We recruited 30 younger (17 females) and 30 older adults (16 females). The younger adults were between 20 and 25 years old ($M_{age} = 21.43$, $SD = 1.38$), and the older adults were

between 60 and 75 years old ($M_{age} = 65.97$, $SD = 4.45$). The participants were native Turkish speakers with no record of a neurological disorder. We also made sure that all participants had normal or corrected to normal vision and normal hearing. For the older age group, the motivation of choosing 60 as the lower age range limit was that although aging is a continuous process (Salthouse, 1994, 2009), the adverse effects of aging become more prominent around age 60 (Plassman et al., 1995; Salthouse, 2009). The motivation of choosing 75 as the upper age range limit was to recruit only neurotypically healthy older individuals, as increasing age is closely associated with the risk of developing dementia (Savva et al., 2009). The participants we recruited between this age range did not manifest any sign of cognitive impairment as measured by the Turkish version of the Standard Mini Mental Test (Güngen et al., 2002): they scored higher than 25 out of 30 whilst the suggested cut-off score is 23/24. We obtained informed consents from the participants before the experiment, which had an approval of the IRB committee of Koç University. Twelve participants from the younger age group enrolled in the sessions through the subject pool of Koç University and received 1 course credit for their participation. The rest of the participants, 18 younger and 30 older adults, were recruited based on convenience, and they did not receive any reward for participating.

2.2.2. Materials

We used a picture description task to elicit speech and gesture samples. The task consists of the pictures of three concrete and three abstract paintings (Appendix A). The paintings were randomly chosen through online research. An online norming was done with 35 independent young adults to make sure that we correctly classified the paintings as abstract or concrete. The paintings were presented in a mixed and randomized order and people were asked to rate them from 1(completely abstract) to 5(completely concrete). In line with our expectations, the ratings of the abstract paintings we selected were close to the abstract end ($M = 1.65$, $SD = .51$) and the ratings of the concrete paintings we selected were close to the

concrete end ($M = 4.49$, $SD = .59$). During the experiment, all pictures were presented in A4 size. The pictures were placed in a holder to ensure that participants' hands are free, and they have enough space for gesturing.

2.2.3. Procedure

The sessions were carried out in a silent room where the participants were seated in front of a table. We started the session by collecting demographic information from the participants. Then, the experimenter presented a total of six pictures of paintings (three concrete and three abstract) in an enclosed file. The participants were asked to describe each of the pictures as if they were describing them to someone who had not seen the pictures before. We did not set an order of description or mention anything related to the painting types. Rather, participants randomly chose among the six paintings that were in a mixed order, regardless of the painting types. The paintings were visible as participants described them. There was no time limitation, and when participants finished describing one painting, they could move on to another. We also warned the participants not to hold the pictures with their hands but to place them in a holder and leave them in front of the table while describing. The session took approximately 15 minutes, and it was videotaped. Participants knew that they were being recorded, as indicated in the consent form. At the end of each session, we asked participants whether they had any questions regarding the experiment, they could guess what we studied or suspected of anything. None of them realized that we study gestures or speech disfluency as they were not familiar with the field.

2.2.4. Transcription and coding

2.2.4.1. *Speech and disfluency*

We transcribed speech and coded disfluencies in the ELAN software (Version 5.8, 2019). We classified disfluency into three categories as filled pauses (e.g., Şapka giyen *um* bir kadın var. “There is a woman *um* wearing a hat”), repetitions (e.g., Farklı renkler *renkler* var.

“There are different colors *colors*”), and repairs (e.g., Sandal- *masanın yanında ayakta duruyor*. “He stands next to the chair- *table*”). Although the classification of Maclay and Osgood (1959) also includes silent pauses, we excluded them since their identification is quite subjective, especially for comparisons between younger and older adults (see Bortfeld et al., 2001, for similar comments). A single coder, who was trained by the first author, coded all disfluencies. Moreover, another trained assistant independently identified and coded 20% of the participants to test reliability. Two coding attempts revealed a significant high correlation for disfluency frequency ($r = 0.89, p < .001$) with a high interrater agreement for categorizing disfluencies ($\kappa = .87, p < .001$). The first author checked all coding and resolved any discrepancies between coders.

For each participant, we calculated the total number of words separately for the concrete and the abstract painting categories. To calculate the disfluency rate for each painting category, we divided the total number of disfluencies observed for a painting category by the total word count for that category. Moreover, we calculated the relative use of each disfluency type among all disfluency types in proportions. In other words, we calculated the filled pause use, the repetition use, and the repair use separately for the concrete and the abstract painting categories: we divided the total number of occurrences of each disfluency type (filled pause, repetition, repair) for a painting category by dividing the total number of disfluencies for that category.

2.2.4.2. Gestures

Together with speech disfluencies, we coded co-speech gestures in the ELAN software (Version 5.8, 2019). Our coding schema was based on the classification of McNeill (1992) as we categorized gestures into iconic (e.g., drawing a circle on the air with fingers to mean round), metaphoric (e.g., raising two hands close to chest as holding something on the air to refer to a feeling), deictic (e.g., pointing with an index finger at an image within a picture), beat (e.g., rhythmically moving hands while speaking), and emblem (e.g., using a thumbs up sign

to mean okay) gestures. In line with the current gesture frameworks (Hostetter & Alibali, 2019; Kita et al., 2017), we included only iconic and metaphoric gestures under the representational gesture category. It is important to note that there was a limited number of metaphorical gestures in the dataset. The metaphorical gestures were only 10% of the gestures produced for the description of abstract paintings and 6% of the gestures produced for the description of concrete paintings as participants were more likely to produce iconic forms of representational gesture in either context. The first author and a trained assistant coded gestures for all participants. The two coding attempts revealed a significant high correlation for gesture frequency ($r = 0.88, p < .001$) and a high interrater agreement in categorizing gestures ($\kappa = .85, p < .001$). After the independent coding, the coders discussed the differences and resolved the conflicts.

For each participant, we calculated overall gesture frequency separately for concrete and abstract painting categories as we divided the total number of gestures for a painting category by dividing the total word count for that category. We calculated the relative use of representational gestures for the concrete and the abstract painting categories: we divided the total number of representational gestures (iconic and metaphoric gestures) for a painting category by dividing the total number of gestures for that category. To obtain a measure for deictic gesture use for the concrete and the abstract painting categories, we divided the total number of deictic gestures for each painting category by the total number of gestures for that category.

2.2.4.3. *Disfluency-Gesture Segments*

We focused solely on the representational gestures in the case of disfluency-gesture segments as we made predictions based on the theories regarding the self-oriented functions of representational gestures (Kita et al., 2017). Therefore, we coded disfluency-gesture co-occurrences by focusing only on the stroke onset and offset of representational gestures. The

stroke phase is commonly identified and coded in terms of kinematic movement properties, which requires a frame-by-frame coding in transitions between clear and blurred images (e.g., Seyfeddinipur, 2006). While coding the strokes, we mainly took the onset as the moment we first observed the shape of an object or action in hands, complying with the kinematic movement properties. Similarly, we clearly determined the onset and offset of each disfluency type. For filled pauses, the onset was the start of e.g., um sound, and the offset was the end of it. For the repetition category, the onset of disfluency was the start of the repeated section, and the offset was the end of it (e.g., white *white* shirt). For the repair category, the onset of a repair was the moment where an alternative word was suggested, and the offset was the end of that word. For instance, in the case of an example “He stands next to the chai- *table*”, the repair part was “table” where the onset was the beginning, and the offset was the end of this word. An independent trained assistant coded half of the participants’ representational gestures and speech disfluency to check the onset and offset time of the gestures and disfluency. The intraclass correlation coefficients (ICC) obtained from the two coding attempts revealed a high interrater reliability for the onset (ICC = .94, $p < .001$) and the offset (ICC = .95, $p < .001$) of gestures as well as for the onset (ICC = .97, $p < .001$) and the offset (ICC = .96, $p < .001$) of speech disfluencies.

To determine co-occurrences of gestures and disfluency in spontaneous speech, we relied on the Delayed-Interruption-for-Planning hypothesis, which suggests that individuals do not experience a problem in speech as soon as they detect an error (Blacfkmer & Mitton, 1991). Rather, there is a gap between error detection and the arrival of disfluency in which a resolution process is already initiated (Seyfeddinipur, 2006). Similarly, previous literature suggested that gestures temporarily precede their lexical counterparts (McNeill, 1992; Morrel-Samuels & Krauss, 1992; Seyfeddinipur, 2006; Seyfeddinipur & Kita, 2001; Ter Bekke et al., 2020). According to Seyfeddinipur and Kita (2001), gestures are likely to suspend around 240

milliseconds before a speech disfluency occurs. However, this finding was based on a very small sample, which only consisted of young adults, while our study targets both younger and older adults. Therefore, considering cognitive and motor slowing as a result of the aging process (for a review, see Myerson & Hale, 1993), we followed a more conservative method to determine the co-occurrences of gestures and speech disfluency. That is, we included the co-occurrences where the offset of the stroke of a representational gesture was slightly before (up to 500 milliseconds) than the onset of a speech disfluency along with the co-occurrences where the onset of the stroke of a representational gesture was slightly later (up to 500 milliseconds) than the offset of a speech disfluency. It is important to note that, in each case, representational gestures were semantically/deictically related to the content in speech.

We followed two separate strategies while calculating disfluency-gesture co-occurrences as proportions: one from a gesture and one from a disfluency perspective. The rationale behind such an approach is that one might gesture quite a few, but those gestures might occur with a disfluency. Similarly, one might reveal only few speech disfluencies but those might intersect with a gesture. Therefore, calculating proportions separately from gesture use perspective and speech disfluency perspective might prevent bias that might arise from individual differences.

In line with such a strategy, we first counted the number of (representational) gestures that occur with disfluencies. That is, for each gesture, we coded whether it overlapped with a disfluency, applying the 500 milliseconds rule that we mentioned earlier. Then, we divided the total number of these co-occurrences by the total number of representational gestures produced for each painting category. These proportions were used as a dependent variable in the analyses. Second, we counted the number of disfluencies that occur with (representational) gestures. In other words, for each disfluency, we coded whether it overlapped with a gesture, again applying the 500 milliseconds rule. Then, we divided the total number of these co-

occurrences by the total number of speech disfluencies produced for each painting category. We used these proportions as a dependent variable in the analyses.

2.3. Results

2.3.1. Speech and disfluency

We found that, overall, younger adults ($M = 506.57$, $SD = 240.44$) produced more words than older adults ($M = 235.87$, $SD = 154.87$) while describing paintings, $t(58) = 5.18$, $p < .001$. To investigate whether disfluency rates varied between the two age groups for the two painting categories, we conducted a 2x2 repeated measures ANOVA. The results demonstrated that the main effects of painting category and age groups were not significant, $F(1,58) = .98$, $p = .325$, $\eta p^2 = .017$ and $F(1,58) = 2.65$, $p = .109$, $\eta p^2 = .044$, respectively. Younger and older adults revealed comparable disfluency rates for both painting categories. The interaction between painting category and age group was not significant, either, $F(1,58) = .01$, $p = .915$, $\eta p^2 = .000$.

We repeated similar analyses to examine the use of each disfluency type across age groups and tasks/contexts. For the proportion of filled pauses, the results indicated a significant main effect of age, $F(1,56) = 8.92$, $p = .004$, $\eta p^2 = .137$. The younger adults ($M = .71$, $SD = .21$) produced a higher proportion of filled pauses compared to the older age group ($M = .52$, $SD = .32$). However, the main effect of painting category, $F(1,56) = .80$, $p = .375$, $\eta p^2 = .014$, and the interaction between painting category and age group, $F(1,56) = 2.35$, $p = .131$, $\eta p^2 = .040$, were not significant (Figure 2.1).

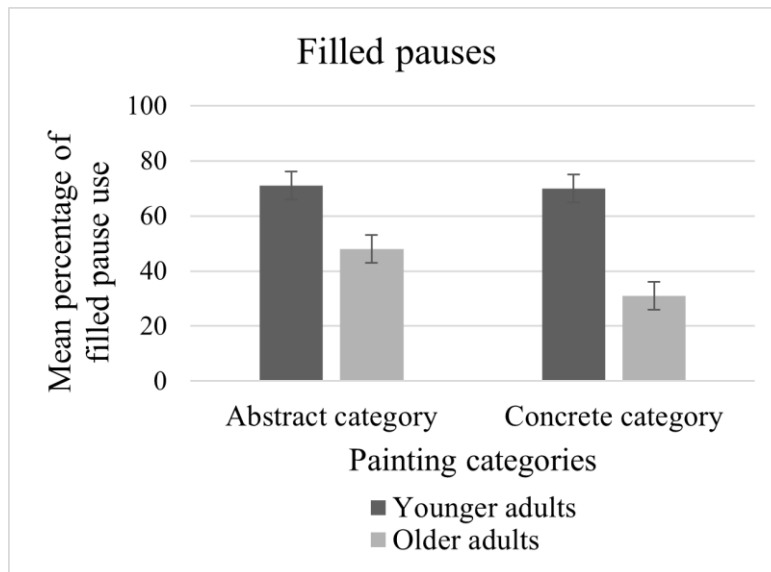


Figure 2.1. The mean percentage of the filled pause use by younger and older adults across abstract and concrete painting categories (filled pause use for each participant was calculated by dividing total number of filled pauses by total number of disfluencies).

For the proportion of repetitions, the repeated measures ANOVA yielded a significant main effect of age, $F(1,56) = 8.15$, $p = .006$, $\eta^2 = .127$, as older adults ($M = .23$, $SD = .26$) produced higher proportion of repetitions than younger adults ($M = .12$, $SD = .10$). However, the results did not reveal a significant main effect of painting category, $F(1,56) = .61$, $p = .440$, $\eta^2 = .011$, or a significant interaction between age group and painting category, $F(1,56) = 1.49$, $p = .228$, $\eta^2 = .026$ (Figure 2.2).

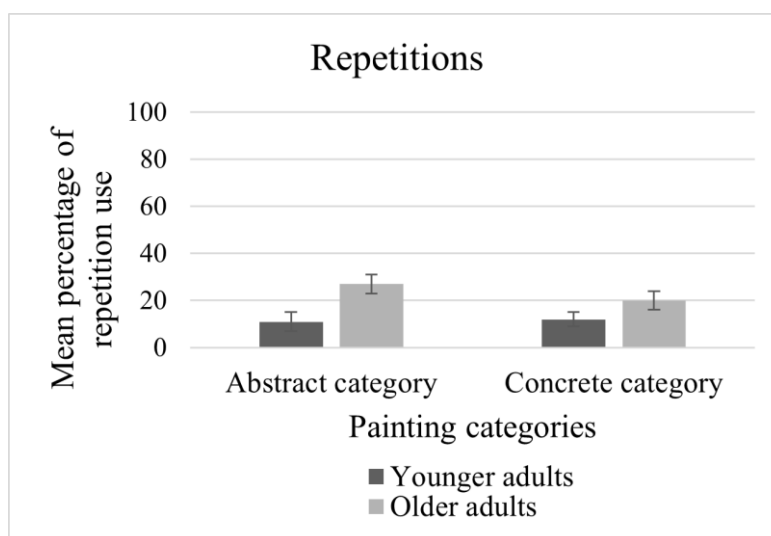


Figure 2.2. The mean percentage of the repetition use by younger and older adults across abstract and concrete painting categories (repetition use for each participant was calculated by dividing total number of repetitions by total number of disfluencies).

For the repair use, the analysis revealed a nonsignificant main effect of age, $F(1,56) = 2.92$, $p = .093$, $\eta p^2 = .050$, and a nonsignificant main effect of painting category, $F(1,56) = .01$, $p = .941$, $\eta p^2 = .000$. Also, the interaction between these variables was not significant, $F(1,56) = .07$, $p = .786$, $\eta p^2 = .001$ (Figure 2.3).

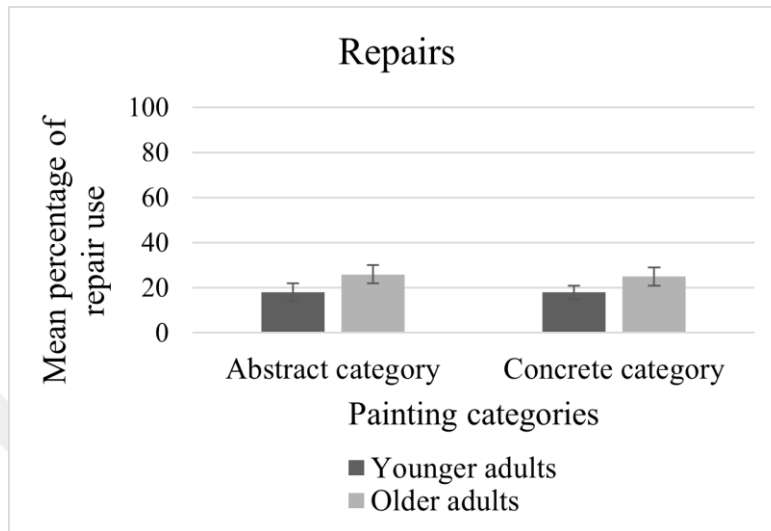


Figure 2.3. The mean percentage of the repair use by younger and older adults across abstract and concrete painting categories (repair use for each participant was calculated by dividing total number of repairs by total number of disfluencies).

2.3.2. Gesture

To examine gesture use across age groups and paintings, we conducted a repeated measures ANOVA, taking age group as the between-subjects factor and the two painting categories as the within-subject factor. The dependent variable was gesture frequency. We did not find significant main effects of age, $F(1,58) = .11$, $p = .743$, $\eta p^2 = .002$, or painting category, $F(1,58) = .43$, $p = .513$, $\eta p^2 = .007$, on gesture frequency. Similarly, the interaction between age group and painting category was not significant, $F(1,56) = .01$, $p = .966$, $\eta p^2 = .001$.

We carried out similar analyses to examine the use of each gesture type across age groups and painting categories. For the proportion of representational gestures, a 2x2 repeated measures ANOVA revealed a significant main effect of age, $F(1,51) = 38.31$, $p < .001$, $\eta p^2 = .429$. In particular, the proportion of representational gestures was significantly higher for the younger ($M = .61$, $SD = .22$) than the older age group ($M = .26$, $SD = .24$). However, the main

effect of painting category, $F(1,51) = 1.40$, $p = .242$, $\eta p^2 = .027$, and the interaction between age group and painting category were not significant, $F(1,51) = .11$, $p = .741$, $\eta p^2 = .002$ (Figure 2.4).

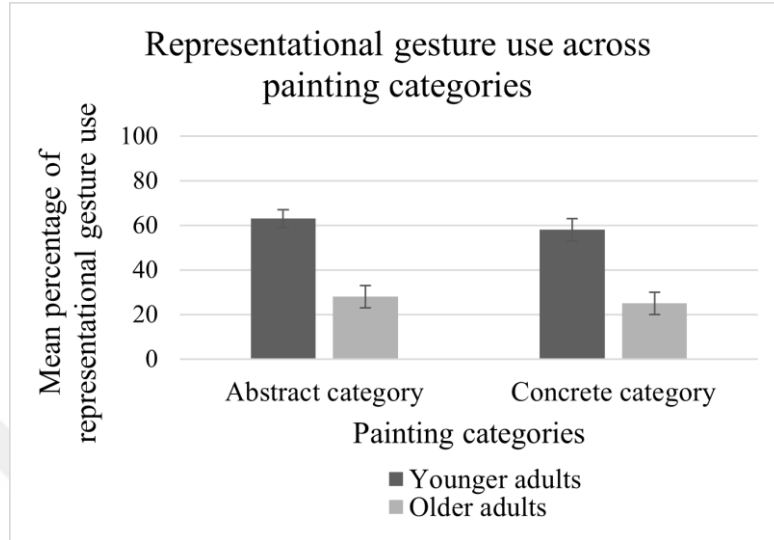


Figure 2.4. The mean percentage of the representational gesture use by younger and older adults across abstract and concrete painting categories (representational gesture use for each participant was calculated by dividing total number of representational gestures by total number of gestures).

For the proportion of deictic gestures, we again found a significant main effect of age group, $F(1,51) = 8.25$, $p = .006$, $\eta p^2 = .139$. Older adults ($M = .29$, $SD = .28$) used a higher proportion of deictic gestures compared to younger adults ($M = .12$, $SD = .14$). The main effect of painting category was not significant, $F(1,51) = .04$, $p = .839$, $\eta p^2 = .001$, however, there was a significant interaction between age group and painting category, $F(1,51) = 4.48$, $p = .039$, $\eta p^2 = .081$. Post-hoc analyses indicated that while younger adults produced higher proportion of deictic gestures in the concrete category ($M = .16$, $SD = .22$) than the abstract category ($M = .09$, $SD = .10$), older adults produced higher proportion of deictic gestures in the abstract category ($M = .32$, $SD = .30$) than the concrete category ($M = .25$, $SD = .27$) (Figure 2.5).

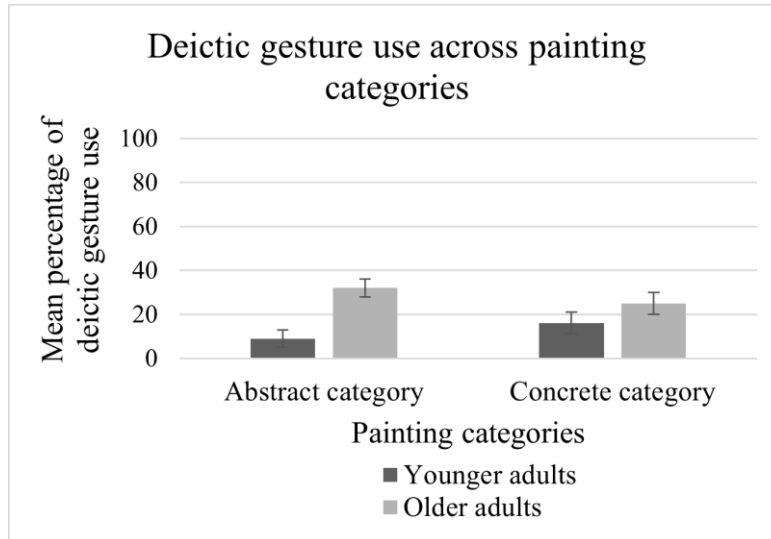


Figure 2.5. The mean percentage of the deictic gesture use by younger and older adults across abstract and concrete painting categories (deictic gesture use for each participant was calculated by dividing total number of deictic gestures by total number of gestures).

2.3.3. Speech disfluency and gesture co-occurrences

To examine to what extent speech disfluencies occurred with gestures across painting categories and age groups, we carried out a 2x2 repeated measures ANOVA. The dependent variable was the proportion of disfluencies that occurred with representational gestures¹. We found a significant main effect of age, $F(1,56) = 16.60$, $p < .001$, $\eta p^2 = .229$. Overall, the proportion of disfluencies that occurred with representational gestures was significantly higher for younger ($M = .25$, $SD = .18$) than older adults ($M = .09$, $SD = .13$) (Figure 2.6). However, the main effect of painting category, $F(1,56) = 3.29$, $p = .075$, $\eta p^2 = .055$, and the interaction between age group and painting category, $F(1,56) = .00$, $p = .969$, $\eta p^2 = .000$, were not significant.

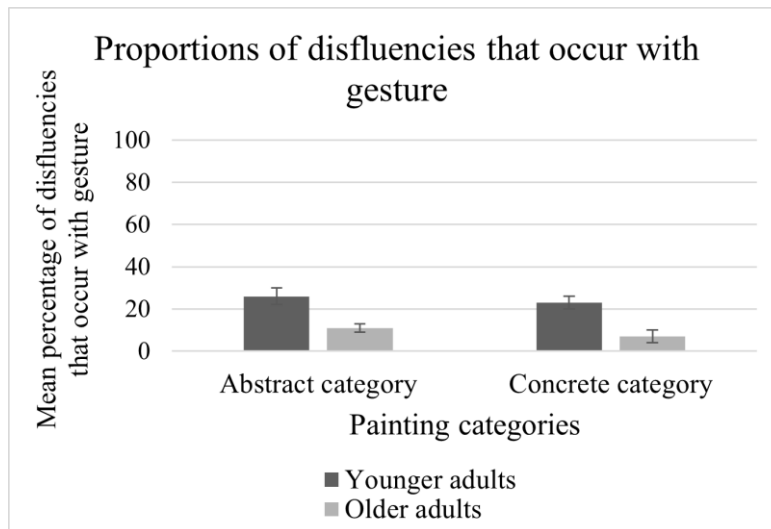


Figure 2.6. The mean percentage of the proportion of disfluencies that occur with gesture across abstract and concrete painting categories.

We also conducted a 2x2 repeated measures ANOVA to investigate to what extent representational gestures² occurred with speech disfluencies across painting categories and age groups. The dependent variable was the proportion of representational gestures that occurred with speech disfluencies. The main effect of age, $F(1,42) = 2.62$, $p = .113$, $\eta p^2 = .059$, and the main effect of painting category, $F(1,42) = .00$, $p = .955$, $\eta p^2 = .037$, were not significant. Similarly, the results did not yield a significant interaction between painting category and age group variables, $F(1,42) = 1.52$, $p = .224$, $\eta p^2 = .035$ (Figure 2.7).

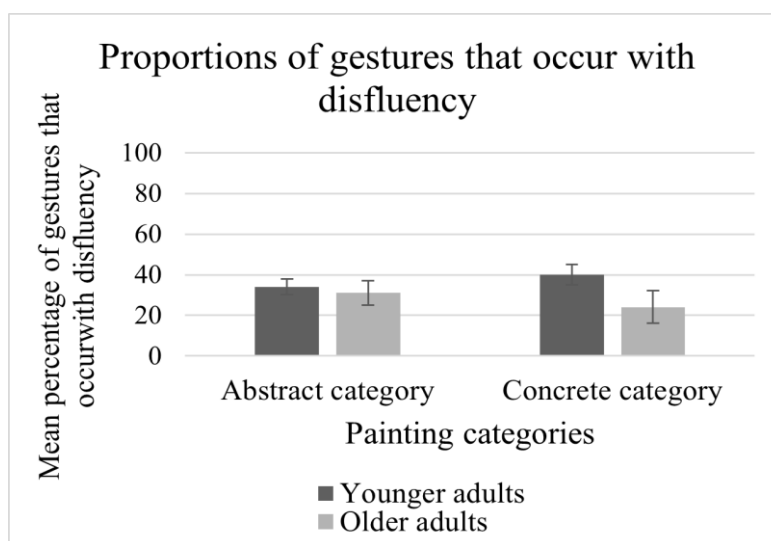


Figure 2.7. The mean percentage of the proportion of gestures that occur with disfluency across abstract and concrete painting categories.

2.4. Discussion

This study examined speech disfluency and co-speech gestures of younger and older adults through a concrete and abstract painting description task. We asked how age could affect the co-occurrence between gesture and speech disfluency, which has been rarely studied in the literature yet. First, we showed that younger adults produced a higher amount of speech in both types of paintings compared to older adults. Second, although overall disfluency rates were comparable between the two age groups and two contexts, we found age group differences in terms of how the disfluencies were distributed across categories. In particular, the younger adults produced higher proportions of filled pauses than the older adults for both paintings. Regardless of the painting type, the older adults produced higher proportions of repetitions than the younger adults, whereas the proportion of repairs were comparable between the two contexts and groups. Third, although the overall gesture rates revealed similar trends between the two age groups and contexts, we found age group differences in terms of how the gestures were distributed across categories. Regardless of the context, the proportion of representational gestures was higher for the younger than the older age group. In contrast, the proportion of deictic gestures was higher for older than younger adults. Moreover, the concrete and abstract paintings elicited comparable proportions of deictic gestures. However, the proportion of deictic gestures was higher in the concrete than the abstract context for the older age group, whereas it was the other way around for younger adults. Finally, we demonstrated that the proportion of representational gestures that occurred with a speech disfluency was comparable across the two age groups and the two contexts. However, the proportion of speech disfluencies that occurred with a representational gesture was significantly higher for the younger than the older age group, regardless of the context.

Although our results do not indicate an age group difference in terms of overall disfluency rates, we found that older adults produced almost half times fewer words than

younger adults. This situation can be associated not only with the nature of the task that prevents circumlocution to some extent (Theocharopoulou et al., 2015) but also with the effect of aging on risk-taking behavior (for a review, see Shao & Lee, 2014). Older adults in our study might have preferred talking less than the younger ones while describing paintings so as not to make mistakes in a context in which they are thematically constrained. We can also argue that producing a minimum amount of speech not to make any mistakes might have been an effective strategy that can preclude older adults from revealing higher disfluency rates than younger adults. The fact that older participants in our study were motivated to perform well in the descriptions further corroborates this argument.

The two age groups, however, differed in terms of how their speech disfluencies were distributed across disfluency types. As we expected, the proportion of repetitions among all disfluency types was significantly higher for older than younger individuals. This finding corroborates the suggestion that repetition is a common disfluency form when task demands exceed available cognitive resources (Bortfeld et al., 2001). Engaging with a thematically constrained task and explaining concrete and abstract art figures might have been relatively more difficult for the older age group than the younger ones. Although producing fewer words might be a strategy for older adults to avoid making mistakes, the types of disfluencies individuals dominantly use might be a key to understanding the cognitive implications of disfluency within and across age groups. Our finding is also consistent with previous research that repetition use is more common in the case of young children and elderly individuals (Yairi & Clifton, 1972), who are more likely to rely on automatic forms of thinking (Braver et al., 2001) as their cognitive skills are limited compared to young adults (Salthouse, 1994). That is, the dominant use of repetitions observed in the older age group might signal that the speech planning process is not well-executed. Older adults might be more likely to detect planning related problems while speaking. In such cases, one uses the repeating strategy, reflecting a

last-minute intervention of proactive control before committing an error. If the proactive control strategies were effectively involved in speech planning of older adults from the very beginning, they would be less likely to predict errors after initiating speech and as a result, less likely to use repetitions to prevent those errors. This pattern might suggest that older individuals are less likely to rely on deliberate forms of cognitive control (e.g., proactive control) in speech planning. Therefore, producing repetitions at higher proportions, in the case of older than younger adults, might reflect their difficulty in speech planning and interpreting novel stimuli with relatively limited cognitive resources and tendency to dominantly use automatic forms of thinking.

As predicted, we found comparable proportions of repairs across the two age groups, but the use of filled pauses differed across age groups. Filled pauses were the dominant disfluency type for the younger than the older age group. Considering that filled pauses might not only have cognitive but also communicative implications (for a review, see Corley & Stewart, 2008), one would expect this disfluency type to occur at comparable rates across the two age groups. From a cognitive perspective, filled pauses are associated with conceptual difficulties (Smith & Clark, 1993) and are commonly used at the beginning of a new sentence, where speech planning occurs (Fraundorf & Watson, 2014; Shriberg, 1996). Filled pauses might be used while one is still planning, unlike repetitions that come into play after speech is already initiated. Then, filled pauses might be the byproducts of one's conceptual and syntactic planning processes, suggesting that the person is engaging with effortful control strategies as a non-automatic form of thinking. Therefore, observing higher proportions of filled pauses in younger adults' speech corroborates with the argument that this age group more effectively uses proactive control strategies compared to older adults (Braver et al., 2007).

From a communicative perspective, younger adults' more frequent use of filled pauses might be related to the amount of speech produced by younger and older adults and possibly

their motivation to describe paintings. Using filled pauses also helps maintain the conversational floor between speaker and listener(s) by making pauses more natural (Clark, 1994; Smith & Clark, 1993). In our study, younger adults could be more eager to describe paintings and, therefore, produced more words compared to the older age group. Thus, using a higher proportion of filled pauses in the younger age group is not surprising since the motivation to describe paintings might be associated with keeping the conversational floor to obtain their listener's attention during speech planning.

For the gesture use, in line with the earlier findings (Arslan & Göksun, 2021; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999), the two age groups did not differ in terms of overall gesture frequency but how their gestures were distributed across specific gesture categories. Overall, younger adults produced higher proportions of representational gestures than older adults, whereas deictic gesture use was more prominent for the older age group. This finding corroborates the argument of Arslan and Göksun (2021) that, as individuals age, their representational gestures might be replaced by deictic gestures, which are considered as a simpler gesture form compared to representational gestures (Özçalışkan et al., 2016). Similarly, Arslan and Göksun (2021) found an age group difference in representational gesture use only in a spatial context where individuals described a route. Although it is not purely spatial, the painting description task we used in this study requires describing not only characteristics of objects, subjects, and shapes but also their relative position in relation to each other to convey a comprehensive picture in listener(s)' mind. Therefore, the painting description task in this study can be similar to the route description task of Arslan and Göksun (2021) in terms of including spatial components and, thus, revealing age group differences in gesture production.

Moreover, describing pictures is considered a constrained task that reveals age group differences in language production (for a review, see, Mortensen et al., 2006). We included both concrete and abstract paintings to provide contexts that vary in difficulty. However, we

could not find differential effects of the two contexts, neither on disfluency nor on gesture frequencies. Our results did not yield a difference between the concrete and the abstract contexts in terms of the co-occurrence of speech disfluency and representational gestures, either. One could expect disfluency rates to be higher or age group differences in disfluency to be more prominent in an abstract context (Bortfeld et al., 2001). Describing tangrams as opposed to baby pictures elicited higher disfluency rates and presented age group differences in disfluency. In a similar way, as representational gestures facilitate the conceptualization process and lighten the cognitive load (Kita & Davies, 2009; Morsella & Krauss, 2004), individuals are more likely to produce representational gestures in an abstract context, which might be difficult to conceptualize. One possible reason for this inconsistent finding can be related to the paintings themselves. In Bortfeld et al. (2001), there was a prominent gap between describing baby pictures versus tangram pictures in terms of difficulty and the level of abstractness. Additionally, Morsella and Krauss (2004) presented figures that were quite easy to label (e.g., picture of a flower) and figures that looked like tangrams and were difficult to code. Describing complex objects was associated with higher representational gesture use. The abstract paintings in our study might not necessarily be more difficult than the concrete ones. Although the concrete paintings were relatively easier to label than the ones of the abstract ones, they were impressionist in nature. The concrete paintings we chose are not as simple as the Cookie Theft Picture (Goodglass & Kaplan, 1983) or drug store/bus stop pictures used in previous research (Cooper, 1990). Interpreting artwork might have complicated the process, regardless of the paintings' level of abstractness.

Moreover, although we did not ask participants their opinions on using hand gestures and body language in formal and semi-formal settings, we cannot exclude the possibility that older adults might come from a generation that was instructed not to gesture in such settings. Schubotz et al. (2019) demonstrated that in a context where participants communicated with

another participant rather than a confederate or an experimenter, older and younger adults were comparable in their overall representational gesture use. However, while conveying information, which is unknown to their addressees, older adults used fewer representational gestures compared to younger adults' narrating such content. Even though age group differences in overall representational gesture use might not be observed in relatively more informal settings, it might appear as the task becomes more difficult. As in the case of Schubotz et al. (2019), narrating a content, which is novel to their addressee might be cognitively more costly for older adults. Therefore, regardless of whether individuals communicate in a formal or an informal setting, age group differences in representational gesture production is likely to appear when effective communication becomes cognitively more demanding.

Moreover, we adopted a slightly different methodology than previous research on different age groups (Bortfeld et al., 2001; Cooper, 1990) as we used proportions to measure the relative use of each disfluency type across age groups and contexts. That is, we examined how specific disfluency types were distributed across categories, whereas, for instance, Bortfeld et al. (2001) measured the use of each disfluency type as per 100 words. We used the relative use of each disfluency type because one might reveal only a few cases of disfluency as a part of their speech, but most of these disfluencies might be related to one disfluency category. In such a case, dominantly using a certain disfluency type among all disfluency cases might be a better measure to understand individual differences, particularly when studying age group differences in language. Since such a point of view also holds for gesture production, we used the same methodology to calculate the relative use of each gesture type, among others. Previous research on gesture production of different age groups corroborates this argument. Although younger and older adults are comparable in their overall gesture use, the groups differ in terms of how their gestures are distributed across categories (Arslan & Göksun, 2021; Cohen & Borsoi, 1996; Feyereisen & Havard, 1999). Our study, therefore, brings a new insight into

multimodal communication of different age groups by combining speech disfluency and gesture use under a single methodology.

Last, we found that the proportion of speech disfluencies that occurred with representational gestures was higher for younger than older adults. That is, speech disfluencies of younger adults were more likely to be accompanied by representational gestures. However, the proportion of representational gestures that occurred with speech disfluencies was comparable between the two age groups. That is, the likelihood of representational gestures to be accompanied by speech disfluencies was similar for younger and older adults. Then, a simple response in one direction for gestures' facilitative role is difficult to obtain. Observing age group differences for the proportion of speech disfluencies that occur with a gesture suggests that younger adults are more likely to benefit from representational gestures to resolve temporary problems in speech planning. Research demonstrated that although older individuals can comprehend information conveyed through either speech or gesture channels, they are likely to ignore gestures and focus on speech when these modalities are presented together (Cocks et al., 2011). Similarly, unlike younger adults, older adults are less likely to benefit from gestural information in difficult listening conditions (Thompson, 1995; Thompson & Guzman, 1999). These findings suggest that older adults may rely on speech as their main communication tool since integrating information coming from multiple modalities might be costly for them. As a result of this integration issue, younger adults' language production system might be better at benefiting from other modalities, i.e., gesture, and integrating information coming from different modalities to resolve temporary problems, and thus, facilitate the speech production. However, from a gesture perspective, it might not be the whole story as the proportion of gestures that occur with speech disfluencies were found to be comparable between the two age groups. Gesture production is associated with not only self-oriented but also communicative functions (Kita et al., 2017). Therefore, from a gesture

perspective, it is difficult to differentiate between these motivations and understand the facilitative roles of gestures. Then, it is not surprising that the facilitative roles of gestures are more prominent from a speech disfluency perspective, when a temporary problem regarding speech planning is at stake. It is also possible that gestures facilitate speech in the absence of an overt error (e.g., speech disfluency), and we might see no trace of such help in a properly functioning language system. Therefore, observing speech disfluency rates and patterns in a context where gesture use is restricted might provide more precise answers on the link between gesture's facilitative roles and speech disfluency. Moreover, although previous research has mainly focused on between-group variations in gesture use (e.g., Akhavan et al., 2018; Azar et al., 2019; Cohen & Borsoi, 1996; Özçalışkan & Goldin-Meadow, 2010), addressing within-group variations on gesture production might be critical (e.g., Hostetter & Alibali, 2007, 2011). Individual differences within a group based on cognitive, social, or contextual domains might be required to draw a more comprehensive picture of to what extent gestures might become compensatory (for a review, see Özer & Göksun, 2020).

Gestures' facilitative roles may be more specific to some cases of speech production. For instance, research on second language acquisition indicates facilitative roles of gestures as individuals are more likely to gesture with disfluent speech in their non-dominant language (Graziano & Gullberg, 2018). Similarly, people with aphasia (Akhavan et al., 2018) and brain injury (Göksun et al., 2013; 2015) are prone to gesture when they have difficulties in speech. Although patients with aphasia are more likely to use gestures, especially iconic gestures, than healthy individuals (Cocks et al., 2013; Sekine et al., 2013; Sekine & Rose, 2013; Van Nispen et al., 2013; but for severe aphasia cases see Mol et al., 2013), they might use gestures not only during word retrieval problems to ease lexical access (Cocks et al., 2013) but also with communicative intentions (Ruiter & Beer, 2013; Sekine & Rose, 2013; Van Nispen et al., 2013). In other words, gestures produced by patients with aphasia are likely to be *essential* to

convey a message, which is understandable by others (Van Nispen et al., 2017). Then, the frequent use of iconic gestures in such populations does not mean that gestures' primary role is to facilitate speech production (Ruiter & Beer, 2013). Additionally, the facilitative role of gestures is not limited to lexical access (Kita et al., 2017). Consistent with these arguments, Christenfeld et al. (1991) indicated that filled pauses and gestures are not likely to co-occur in spontaneous speech, suggesting the possibility that an overt error (e.g., speech disfluency) might not be required to claim gestures as having facilitative roles in human cognition. Therefore, rather than trying to come up with a global answer to the facilitative roles of gestures, taking individual differences into account and focusing on various cases and contexts where gestures aid speech production might be a more effective strategy to understand the nature of the multimodal communication and how it is affected by the aging process (Göksun et al., 2022).

In conclusion, this study examined speech disfluency and co-speech gestures produced by younger and older adults across the concrete and abstract painting description tasks. We addressed whether and to what extent gestures had a facilitative role in speech disfluencies across age groups and contexts. It is among the first studies that combine speech disfluency and gesture use of younger and older age groups under a single methodology. Moreover, the painting description task we used in our study brings novelty to the classic picture description tasks used in the literature. Our findings suggest that focusing on how speech disfluency and gestures are distributed across categories might be central to understanding the effects of aging on multimodal communication. Gesture production is associated with not only self-oriented but also communicative functions (Kita et al., 2017). As a result, from a gesture perspective, it is difficult to differentiate between these motivations and understand gestures' role in speech facilitation. The facilitative roles of gestures across age groups might be more prominent from a speech disfluency perspective, when a temporary problem regarding speech planning is at

stake. Such a perspective reveals that younger adults are more likely to benefit from representational gestures to resolve temporary problems in speech planning. Understanding the facilitative roles of gestures might require further attention on specific cases and contexts in which speech disfluency and gestures co-occur.



Footnotes

^{1,2} When we merged deictic gestures into the representational gesture category, the results did not change.



CHAPTER 3

MULTIMODAL LANGUAGE IN BILINGUAL AND MONOLINGUAL CHILDREN: GESTURE PRODUCTION AND SPEECH DISFLUENCY

This chapter is published as Arslan, B., Aktan-Erciyes, A., & Göksun, T. (2023). Multimodal language in bilingual and monolingual children: Gesture production and speech disfluency.

Bilingualism: Language and Cognition, 1-13. doi: 10.1017/S1366728923000196

Abstract

Bilingual and monolingual children might have different styles of using multimodal language. This study investigates speech disfluency and gesture production of 5- and 7-year-old Turkish monolingual (N=61) and Turkish-English bilingual children (N=51). We examined monolinguals' Turkish narratives and bilinguals' Turkish and English narratives. Results indicated that bilinguals were more disfluent than monolinguals, particularly for silent and filled (e.g., umm) pauses. Bilinguals used silent pauses and repetitions (e.g., cat cat) more frequently in English than in Turkish. Gesture use was comparable across language and age groups, except for iconic gestures. Monolinguals produced more iconic gestures than bilinguals. Children's overall gesture frequency predicted disfluency rates only in Turkish. Different gesture types might be orchestrated in the multimodal system, contributing to narrative fluency. The use of disfluency and gesture types might provide insight into bilingual and monolingual children's language development and communication strategies.

3.1. Introduction

When people speak, they are likely to pause, repeat, or revise their message. This temporary disruption in speech is called disfluency (MacKay & Osgood, 1959). Putting their role in communication aside (Corley & Stewart, 2008), speech disfluencies are likely to stem from increased cognitive load and word retrieval issues (Bortfeld et al., 2001). People are more disfluent when they produce longer words (Oviatt, 1995) or sentences (Shriberg, 1996). Similarly, disfluent segments in speech are more likely to be observed in sentence beginnings as a result of high planning load (Shriberg, 1996). Being disfluent is a part of a healthy speech production, yet the likelihood of being disfluent might differ across populations and individuals. Although age has been considered an important variable while studying disfluency (e.g., Arslan & Göksun, 2022; Bortfeld et al., 2001; Cooper, 1990), there is relatively limited focus on bilingualism to understand disfluent speech, especially from a developmental perspective (e.g., Brundage & Rowe, 2018; Dumont, 2010; Kaur et al., 2011). Previous research on bilingual children's language development suggests that bilinguals may reach language competency later than monolinguals (Bialystok et al., 2010; Oller et al., 2007). Bilingual children are comparable with monolingual children in receptive vocabulary, yet they are more likely to experience difficulties in producing the target words (Yan & Nicoladis, 2009). Disfluent segments in speech might reflect individuals' speech planning process and communication strategies (Fraundorf & Watson, 2014). Focusing on whether bilingual and monolingual children's disfluency rates and patterns differ might provide insight into the language development of these groups, particularly for language production.

Individuals' speech can be accompanied by spontaneous co-speech gestures (McNeill, 1992). As a nonverbal form of language, gesture interacts with the speech production mechanism (Kita & Özyürek, 2003). Gestures might help speech production by facilitating lexical retrieval (Krauss et al., 2000) and conceptualizing information (Kita et al., 2017). When

individuals use gestures, they are more fluent compared to when their gesture use is experimentally restricted (Morsella & Krauss, 2004; Rauscher et al., 1996). The gesture-for-conceptualization hypothesis suggests that gestures have self-oriented functions. In particular, using gestures can reduce cognitive load by enabling gesturers to activate, manipulate, package, and explore information units (Kita et al., 2017). In general, regardless of age, bilinguals gesture more frequently in either of their languages compared to monolinguals speaking those languages (e.g., Nicoladis et al., 2009; Pika et al., 2006; So, 2010). Through the lens of gestures' self-oriented functions, gesturing might enable bilingual children to deal with the cognitive load of having two separate language systems by helping them to reach the correct lexicon from the target language (Nicoladis et al., 2009).

Considering the link between cognitive load and disfluency rates in speech, bilingual children are more disfluent than their monolingual peers due to the load of managing two vocabulary systems (Yan & Nicoladis, 2009). Yet, gestures might come into play to decrease the cognitive load in language production, particularly for bilinguals in the form of frequent gesture use. From a developmental perspective, gestures help children complement and build on their verbal messages, playing a role in language acquisition (Iverson & Goldin-Meadow, 2005). Therefore, we focus on bilingual and monolingual children to understand the interplay within and between gesture and speech modalities, which would shed light on individual and group differences in the course of language development. This study asks whether (1) disfluency and gesture rates differ across 5- and 7-year-old monolinguals and bilinguals, (2) bilinguals reveal similar disfluency and gesture rates across their two languages, and (3) children's gesture use is associated with their speech fluency in narrative production.

3.1.1. Speech disfluency

The main classification that is proposed by Maclay and Osgood (1959) has four disfluency types: (1) *Silent pauses* are temporary silent moments within or between phrases,

(2) *Filled pauses* are breaks in one's speech, filled with nonword sounds (e.g., um), (3) *Repetition* means repeating sentence units, such as syllables or words, or sentences themselves (e.g., the child *child* woke up), (4) *Repairs* are revisions in speech that refer to replacing already uttered words or grammatical structures with their more plausible alternatives (e.g., the child – *the dog* barked). It is important to note that disfluency categories might slightly differ in the literature (e.g., Brundage & Rowe, 2018; Graziano & Gullberg, 2018; for the stuttering population, see Yaruss, 1997; Yaruss et al., 1999). Yet, the classification of Maclay and Osgood (1959) encompasses most of the disfluency classifications, which are frequently used while studying typical populations (e.g., Arslan & Göksun, 2022; Avcı et al., 2022; Fraundorf & Watson, 2014).

3.1.1.1. Disfluency from a developmental perspective

Infants start producing simple speech sounds at around 6 months and can construct one or two-word phrases by 18 months (Parish-Morris et al., 2013). Earlier studies on disfluency target as young as 2-year-old children through interviews (Yairi, 1981, 1982), suggesting individual differences in their disfluent speech. Additionally, children's disfluency patterns slightly change as they move from age two to three. More specifically, repetitions are more likely to occur at the phrase level than at the word level while the proportion of revisions in total disfluency also increases. Similarly, DeJoy and Gregory (1985) obtained speech samples from 3.5- and 5-year-old children using a picture story book. They found that although the two age groups did not differ in terms of their total disfluency rates, they varied in using certain disfluency types. More specifically, 3.5-year-olds' disfluent speech was dominated by repetitions and incomplete phrases while 5-year-old children's disfluencies occurred in the form of grammatical pauses (e.g., silent pauses). These findings are in line with children's increasing syntactic complexity due to their growing vocabulary and grammar knowledge (Haynes & Hood, 1977, 1978). Moreover, using filled pauses over repetitions in general

conversation increases with age (Haynes & Hood, 1977), which can be interpreted as filled pauses being more socially acceptable than repeating words in daily communication. Thus, there is a great variability in children's disfluency patterns that are prone to reveal rapid changes year by year due to their growing vocabulary and grammar knowledge.

In sum, yearly or even monthly changes in disfluency rates and patterns until the first years of primary school might be informative in understanding language development, particularly language production (Nettelbladt & Hansson, 1999). Only a few and earlier studies targeted speech disfluency among typically developing nonstuttering children (e.g., DeJoy & Gregory, 1985; Haynes & Hood, 1977, 1978; Yairi, 1981, 1982). Moreover, these studies have not differentiated between monolingual and bilingual children, which might provide a greater insight into the link between speech disfluency and language development.

3.1.1.2. Speech disfluency in bilingual children

Since bilingual children simultaneously master two languages, they might have more trouble in language production compared to their monolingual peers (Bialystok, 2009), which might be reflected in the form of disfluency. Bilinguals are also sensitive to communication (Gampe et al, 2019). Their use of disfluencies might reflect not only their language development but also communication strategies. Therefore, comparing bilingual children with their monolingual peers would give further insight into the nature of speech disfluencies. Although some studies focus on bilingualism and second language (L2) in the case of stuttering children (for a review, see Van Borsel et al., 2001), there is relatively less focus on typically developing young children's disfluency patterns with mixed findings. Some of these studies that recruit typically developing children either have a small sample size (e.g., Bedore et al., 2006; Byrd et al., 2015; Lee et al., 2007) or a wide age range with a limited sample size (e.g., Carias & Ingram, 2006).

Developmental studies on the relationship between bilingualism and speech disfluency are mainly conducted with Spanish-English bilingual children. For instance, Bedore et al. (2006) obtained narrative retellings from L1-Spanish-L2-English bilinguals (N=22) as well as Spanish (N=22) and English monolinguals (N=22) between 4 and 6 years of age. They found that, in both languages, bilinguals were more disfluent than monolinguals. Additionally, grammatical repairs were common while speaking Spanish both for bilinguals and monolinguals. However, those grammatical repairs were not common for bilinguals and monolinguals while speaking English. This finding, however, was attributed to phonological and grammatical differences between the two languages. Moreover, using a narrative task, Byrd et al. (2015) found that L1-Spanish-L2-English bilingual children between 5 and 7 years of age were more disfluent when their disfluency rates were compared with that of monolinguals reported in the literature.

In contrast to the above findings, Brundage and Rowe (2018) found that 30-month-old L1-Spanish-L2-English bilinguals (N=53) revealed lower disfluency rates compared to monolingual disfluency rates reported in the literature. This study, however, did not have an age-matched monolingual control group. Moreover, those monolingual disfluency rates obtained from the literature as a comparison were not representative for 30-month-olds. The authors also suggested that this finding might be the result of using a spontaneous speech task instead of a narrative task, which might be cognitively more demanding (see Bedore et al., 2006; Byrd et al., 2015).

Furthermore, Fiestas et al. (2005) obtained narrative retellings from L1-Spanish-L2-English bilingual as well as Spanish monolingual and English monolingual children between 4 and 7 years of age. They found that there was a direction that bilinguals were slightly more disfluent. Compared to monolingual children, the number of repetitions was higher in the bilingual group. The study also asked whether bilinguals revealed different disfluency rates

and patterns in their L1 and L2. Results showed that the proportion of disfluent segments in children's speech was comparable across their two languages. However, in L1-Spanish, bilingual children commonly used grammatical repairs, indicating a high awareness of grammatical rules as a result of being more frequently exposed to Spanish. When they retold the story L2-English, in which they were less proficient, disfluencies mostly occurred in the form of word-finding failures. These findings suggest that language proficiency might be crucial in understanding disfluency in a bilingual context. Children's knowledge in grammar and vocabulary might be linked to disfluency rates and patterns.

In sum, there are limited studies in the literature that target bilingualism in relation to disfluency in children. Not only disfluency rates but also disfluency patterns might differ between monolingual and bilingual children. However, speech disfluency alone might not be enough to portray the difference between the two groups in using language. Language is multimodal: children also produce gestures while speaking. Considering gestures' self-oriented functions (Kita et al., 2017), examining gesture production among bilingual and monolingual children might shed light on gestures' role in decreasing cognitive load and enhancing speech fluency.

3.1.2. Gesture

McNeill (1992) classified gestures into five categories as iconic, metaphoric, deictic, beat, and emblem gestures. Iconic (e.g., marking a space in the air with two hands to refer to the size of an object) and metaphoric gestures (e.g., moving one hand forward while referring to the future) depict shapes and relations in space regarding concrete and abstract concepts, respectively. A deictic gesture is pointing at something with fingers or hands. Beat gestures, on the other hand, are rhythmic hand movements that lack propositional content. Last, emblems (e.g., waving hands to say bye) are gestures that are understood without using their lexical affiliates. The current gesture frameworks mainly focus on iconic and metaphoric gestures by

using the umbrella term *representational gesture* (Hostetter & Alibali, 2019; Kita et al., 2017; Krauss et al., 2000).

3.1.2.1. Gesture production from a developmental perspective

Children start producing gestures even before they utter their first words (Iverson & Goldin-Meadow, 2005). These gestures are generally simple in form such as pointing at objects (Bates et al., 1979) and help children produce sentence level messages later when they are combined with words (e.g., Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005). Gesture-word combinations are likely to predict the onset of children's two-word combinations (e.g., Iverson & Goldin-Meadow, 2005; Iverson et al., 2008; Özçalışkan & Goldin-Meadow, 2005). Therefore, one could consider early pointing gestures as being a precursor to children's language development, although these gestures might reflect weak language proficiency by the time they are produced (Iverson & Goldin-Meadow, 2005). On the other hand, iconic and beat gestures are relatively more complex compared to deictics, and they appear mostly when children become more proficient in language (Goldin-Meadow, 1998; Nicoladis et al., 1999). As children start constructing more complex ideas, deictic gestures mostly get replaced by iconic gestures (Mayberry & Nicoladis, 2000; Nicoladis et al., 1999) that are intended to convey complex messages (McNeill, 1992). Similarly, as beats are associated with temporal word relations and stress patterns in speech (McClave, 1994), at least sentence level information and multimorphemic utterances should be at stake for the production of beat gestures (Nicoladis et al., 1999). These findings together suggest that gesture development is closely linked with language development (Bates et al., 1979; Goldin-Meadow, 1998; Mayberry & Nicoladis, 2000).

3.1.2.2. Gesture production in bilingual children

Bilinguals gesture more frequently in both of their languages compared to monolinguals speaking those languages (e.g., Nicoladis et al., 2009; Pika et al., 2006). Such a

pattern might stem from bilinguals' controlling of two language systems. Gesture and speech are closely associated mechanisms and gestures comply with the verbal packaging system in speech (Kita & Özyürek, 2003). Considering that bilinguals select among competing representations as a result of having two language systems, and thus, two verbal packaging styles, gestures might facilitate the process by helping individuals to opt for the most appropriate packaging method (Nicoladis, 2007; Nicoladis et al., 2009). This situation might reflect gestures' self-oriented functions as suggested by the gesture-for-conceptualization hypothesis (Kita et al., 2017). Additionally, bilinguals' frequent use of gestures might also stem from speaking a high gesture frequency language (So, 2010). That is, gesture frequency observed in a high gesture frequency language (e.g., Spanish, French) might be transferred to the low gesture frequency language (e.g., English), called *gesture transfer* (Pika et al., 2006).

Findings from developmental research is in line with the former argument. For instance, Nicoladis et al. (2009) compared French-English bilingual preschoolers (between age 4 and 6) with their French and English monolingual counterparts. Narrative retellings of both language groups indicated that bilingual children gestured more frequently than monolingual children both in French and English. However, French and English monolinguals gestured at comparable rates, suggesting that frequent gesture production might be a result of bilinguals' challenge to choose between the competing vocabulary systems.

Understanding the use of specific gesture types might provide further insight into bilingual children's language development. Research has suggested that the use of iconic and beat gestures goes hand in hand with language development while deictics or emblems are considered prelinguistic gestures, appearing quite early in language development (Iverson & Goldin-Meadow, 2005; Nicoladis et al., 1999). Research on bilinguals' gesture production has indicated similar findings. In either of their languages, bilingual children start using iconic and beat gestures by the time they can construct sentence-like utterances (Mayberry & Nicoladis,

2000; Nicoladis et al., 1999). In line with these findings, studies have indicated that bilinguals use fewer iconic gestures in the language they are far less proficient in (Gullberg, 1998). Therefore, one could suggest that iconic gestures do not compensate for weak language skills (Gullberg, 2013; Nicoladis, 2007).

Considering that the language one speaks influences the way gestures are produced (Kita & Özyürek, 2003), it is not surprising that the development of iconic gestures is closely linked to language development. From such a perspective, iconic gestures can be interpreted as an advanced gesture form, its final shape and referent is determined by linguistic factors. Therefore, it would be plausible to expect more iconic gestures where one's language skills are advanced (e.g., the dominant language of bilinguals, see Nicoladis, 2002). On the other hand, bilinguals produce more deictic gestures in their weak language compared to monolinguals of that language (Azar et al., 2020; Gullberg, 2013). The critical point here is that most of those studies were carried out by using story retelling tasks. In such narrative contexts, deictic gestures would be abstract deictics as they do not directly point at concrete objects (Nicoladis, 2007). Azar et al. (2020) found that Turkish-Dutch bilinguals used more deictic gestures in either of their languages compared to monolinguals, suggesting that those (abstract) deictics might be a part of the bilingual strategy that uses space to organize notions and helps speakers to deal with cognitive load.

3.1.2.3. Gesture and speech disfluency

Using gestures might decrease cognitive load by enabling speakers to organize information units (Kita et al., 2017). People are more likely to gesture when they engage with cognitively demanding tasks (Kita & Davies, 2009; Melinger & Kita, 2007; Morsella & Krauss, 2004) that are particularly spatial in nature (Alibali, 2005; Arslan & Göksun, 2021). Moreover, individuals are more fluent when they spontaneously gesture compared to when their hand use is experimentally restricted (Morsella & Krauss, 2004; Rauscher et al., 1996). Then, gesturing

might pave the way for a more fluent speech by decreasing the cognitive load of the speech production process. Therefore, people can frequently produce gestures when speech is difficult.

Individuals might differ in terms of benefiting from using gestures to facilitate cognitive processes, including speech production. Since disfluent segments in one's speech might reflect individuals' speech planning process (Fraundorf & Watson, 2014), studying disfluencies of different groups in a multimodal context might provide insight into how cognitive load and communicative strategies interact (Arslan & Göksun, 2022). In the case of bilingual children, the frequent use of gestures might reduce the cognitive load of having two separate language systems by helping them to reach the correct lexicon from the target language (Nicoladis et al., 2009). Observing the relationship between gesture production and speech fluency in bilingual and monolingual children would give insight into the facilitative roles of gestures in the speech production process.

3.1.3. Present Study

The aim of this study is to examine speech disfluency and gesture production of 5- and 7-year-old typically developing Turkish monolingual and Turkish-English bilingual children. We examine monolingual's Turkish language samples and bilinguals' Turkish and English language samples. Focusing on two typologically different languages, English and Turkish, would bring diversity to the existing literature that is mainly built on the disfluency rates and patterns of Spanish-English bilingual children. Moreover, gesturing frequency and the use of specific gesture types revealed by bilingual and monolingual children might inform the literature about the language development of these groups. The association of speech disfluency with gesture in these groups might shed light on gestures' facilitative roles in speech.

We hypothesize that, compared to Turkish monolingual children, bilinguals would be more disfluent in L1-Turkish, as the challenge of being bilingual is to master and control two languages instead of one. To eliminate a possible bias, we aim to control children's syntactic

(grammatical) complexity scores. It is because compared to simple sentences, producing complex sentences are cognitively more demanding, resulting in more errors and higher disfluency rates in children's speech (Gordon et al., 1986; Haynes & Hood, 1978; Yaruss et al., 1999). Considering that the use of pauses are linked to speech planning (Fraundorf & Watson, 2014), bilingual children might be more likely to use silent pauses and filled pauses to make sure that they stick to the vocabulary of the target language. The use of repetitions and repairs, however, might be comparable between bilingual and monolingual groups.

In line with the previous literature (Nicoladis et al., 2009; Pika et al., 2006), we hypothesize that bilingual children would gesture more frequently than their monolingual peers. Similarly, since bilinguals' challenge is to master two languages instead of one, they might slightly lag behind their monolingual peers in language development (Bialystok, 2009) and iconic gestures are linked to spoken language skills (Nicoladis et al., 1999). Therefore, we expect monolingual Turkish children to use more iconic gestures than the same age bilinguals while speaking Turkish. Conversely, compared to Turkish monolinguals, bilingual children might be more likely to produce deictic gestures in Turkish, which might be a part of their bilingual strategy (Azar et al., 2020). Last, we expect Turkish-English bilingual children of both age groups to reveal similar overall gesture frequency across their two languages as they are proficient in both. We also explore whether bilingual children differ in terms of the use of specific gesture types in L1-Turkish and L2-English.

Moreover, we hypothesize that 5- and 7-year-old children reveal comparable disfluency rates in L1-Turkish and in L2-English. However, we expect disfluency types to differ between 5- and 7- year-old children due to increasing language competency and starting primary school. Compared to 7-year-olds, we expect 5-year-old children to produce more repetitions, which are developmentally less advanced than other disfluency forms (e.g., filled pause, repairs) (Bortfeld et al., 2001). In contrast, 7-year-old children would produce more filled pauses than

5-year-olds, reflecting their planning processes (Fraundorf & Watson, 2014) to express more complex ideas boosted by primary school education. As iconic and beat gestures are closely associated with language development (Nicoladis et al., 1999), we expect 7-year-old children to produce more iconic and beat gestures compared to 5-year-olds, regardless of being bilingual or monolingual.

Last, as gestures help one to think and speak (Kita et al., 2017), we expect that gesture frequency in Turkish, particularly iconic gesture use, would be negatively associated with speech disfluency rates, regardless of age group (5- and 7-year-old) and language group (bilingual and monolingual) by also controlling for participants' syntactic complexity and language proficiency scores in Turkish. Similarly, regardless of age, we expect bilinguals' iconic gesture frequency in L2-English to negatively predict their speech disfluency rate in this language, controlling for their syntactic complexity and L2-English proficiency.

3.2. Methods

3.2.1. Participants

Fifty-one Turkish-English bilingual (25 females) and 61 Turkish monolingual (31 females) children participated in the study as a part of a larger project that investigated the effects of early and intense exposure of L2-English on L1-Turkish narrative skills and motion event conceptualization (see Aktan-Erciyes, 2019). For the original study the data came from, G*Power tool (Version 3.0) (Faul et al., 2007) was used to estimate sample size with .80 power and .30 effect size. The calculated sample size was 118 and data was collected from 112 children. We also performed post-hoc sensitivity analyses for the present study again using the G*Power software package (Faul et al., 2007) to determine the effect size. The sample size of the present study 112 was used in the sensitivity analysis for the regression analyses. The alpha level set to .05 and power was taken as .80. The sensitivity analysis revealed an effect size of 0.13, which around the same level we obtained in our regression analyses ($R^2=0.12$)

We recruited 5-year-old bilinguals ($M=68.95$ months, $SD=4.02$, $N=26$) and monolinguals ($M=69.24$ months, $SD=3.40$, $N=28$) as well as 7-year-old bilinguals ($M=92.37$ months, $SD=3.50$, $N=25$) and monolinguals ($M=89.98$ months, $SD=4.62$, $N=33$). Bilingual children's L1 was Turkish and L2 was English, who were first exposed to English around age three or earlier. The motivation of recruiting 5-year-olds was that the bilinguals of this age had been exposed to early intense exposure to L2-English before starting L1-Turkish dominant primary school. The reason behind recruiting 7-year-olds was that the bilinguals of this age had been exposed to full-time L2-English instruction at the preschool and then started L1-Turkish dominant primary school at the age of 6. As a result, the study could also investigate the effects of this shift.

Bilingual participants were recruited from eligible private schools for intense L2 exposure and monolingual participants were recruited from SES comparable private schools. Our bilingual participants were enrolled in a preschool in which native English speaker teachers and staff communicated with children only in English (8 hours/day). All bilingual children were exposed to American English for instruction. By the time we tested 5-year-old bilinguals, they were already exposed to English in this preschool for three years. On the other hand, the 7-year-old bilingual group included children who had completed three years of education in the same English immersion preschool and continued their primary school education in Turkish for two years. By the time we tested 7-year-old bilinguals, they were about to finish the second grade. Although 7-year-old bilinguals started the L1-Turkish dominant curriculum at the age of 6, they still received English instructions for 10 hours per week and gained L2-English literacy, which further supported their L2-English.

We asked parents to fill a demographic form to better understand children's language background. For both age groups, parents reported that the use of L2 was not limited to the school context, they were also exposed to L2 at home from time to time. Children were not

exposed to another language other than Turkish and English. In the demographic form, parents were also asked to indicate their income level (in monetary terms) in a 5-point Likert scale and give information about their education level on an ordinal scale (1 to 6; from primary school to Ph.D. degree). Both bilingual and monolingual children were raised by Turkish native speaker parents who identified themselves as upper-middle to high socioeconomic status. The mothers of bilingual and monolingual children did not differ in terms of their education levels, $t(110)=-1.64, p=.103$, as most of the mothers from the two groups had completed at least an undergraduate degree. It is important to note that mothers were the only ones who responded to demographic forms and that's why only maternal education level was taken into account. Yet, research suggests a high correlation between maternal and paternal education (Jeong et al., 2018).

The project was reviewed and approved by Koç University Ethics Committee Board with acceptance number 2016/22. Informed consents were obtained from the parents prior to data collection. Parents were given their children's vocabulary scores and children were given stickers as compensation.

3.2.2. Materials

The Frog Story (Mayer, 1969), a wordless book depicting the story through a series of pictures, was used to elicit narrative retellings from the participants. A video camera was used to record the sessions. We used ELAN software (Version 6.2) (Lausberg & Sloetjes, 2009) to transcribe speech and code speech disfluency and gesture.

To measure vocabulary knowledge in Turkish, we used the Turkish Expressive and Receptive Language Test (TIFALDI)–Receptive subtest (Berument & Güven, 2010). As a norm referenced test, TIFALDI includes 159 items and targets children between 2 and 12 years old. In each trial, a participant is presented with a word along with four pictures and asked to

indicate the picture, which best corresponds to the given word. A total score is calculated based on the number of correct answers.

To examine vocabulary knowledge in English, the Peabody Picture Vocabulary Test (PPVT-4) (Dunn & Dunn, 2015) was used. As a norm referenced test, PPVT 4 includes 228 items and aims to measure individuals' receptive vocabulary in English. In each trial, a participant is presented with a word along with four pictures and asked to indicate the picture, which best corresponds to the given word. As in TIFALDI, a total score is calculated based on the number of correct answers.

3.2.3. Procedure

Each child participated in a session where they were alone with the experimenter in a silent room of their school. First, children were presented with The Frog Story book. They were asked to examine the pictures while the experimenter was turning the pages one by one. While turning the pages, the experimenter did not comment or tell the story. After children examined all pictures, the experimenter asked them to retell the story by looking at the pages. The experimenter sat in front of the participants and turned the pages one by one while children were retelling the story. For bilinguals, another session with the same task was carried out in English. The English session took place one week after the Turkish session, and it was carried out by a different experimenter who ran all sessions in English. The same standardized instructions were used both in Turkish and English sessions to prevent possible confounding factors. All sessions were videotaped for further transcription and coding. Additionally, all children completed the TIFALDI receptive vocabulary subtest by the end of the first session, which was held in Turkish. Bilingual children were also asked to complete PPVT-4 by the end of the English session.

3.2.4. Coding

3.2.4.1. Speech and disfluency

Speech samples obtained from children were transcribed including all disfluent segments. The Turkish sessions were transcribed by Turkish native speakers while the English sessions were transcribed by a native English speaker. Four types of speech disfluency as indicated by the classification of Maclay & Osgood (1959) were targeted. Silent pauses, where children were silent for more than 200 milliseconds within their sentences, filled pauses (e.g., *Eee kurbağa kaçtı. “Um the frog escaped.”*), repetitions (e.g., *Köpek köpek çocuğu takip etti. “The dog dog followed the boy.”*), and repairs (e.g., *Köpek kurbağa- arılardan kaçıyordu. “The dog was running away from the frog- the bees.”*) were coded. For the Turkish sessions, a trained assistant coded all participants’ disfluencies. For reliability, another trained assistant coded 38% of participants. The interrater agreement was high between the two coders ($\kappa=.92$, $p<.001$). Similarly, for the English sessions, a trained research assistant coded all participants’ disfluencies while a second assistant coded 33% of participants for reliability, yielding a high interrater agreement ($\kappa=.90$, $p<.001$). For each participant, we counted the total number of words produced in each session. All disfluency rates were calculated per word.

3.2.4.2. Gesture

In line with the classification of McNeill (1992), iconic, metaphoric, deictic, beat and emblem gestures were coded. As children did not produce metaphoric gestures at all, we mainly focused on iconic gestures as representational gestures. For the Turkish sessions, a trained assistant coded all participants’ gestures. For reliability, a second assistant coded 38% of participants. The interrater agreement was found to be high between the two coders ($\kappa=.88$, $p<.001$). Similarly, for the English sessions, a trained research assistant coded all participants’ gestures while another trained assistant coded 33% of participants for reliability, yielding a high interrater agreement ($\kappa=.87$, $p<.001$). All gesture frequencies were calculated per word.

3.2.4.3. Syntactic Complexity

For each participant, we transcribed Turkish and English speech samples in Microsoft Excel files. We used the coding schema of Berman and Slobin (1994) by parsing discourse into *verbed clauses*, “...expressing a single situation (activity, event, or state)” (Berman & Slobin, 1994, p.660) and placing them sequentially per line. A clause consists of at least one predicate. If a clause included a single predicate, it was coded as a simple clause (e.g., Bir çocuk vardı. “There was a boy.”). If, under a single clause, two or more predicates were linked with conjunctions (e.g., and, or, but), adverbials (e.g., while, when), relative clauses, reported speech, or if-then statements, it was considered a complex clause (for a detailed coding scheme for both languages, see Aktan-Erciyes, 2019; Kızıldere et al., 2020). For each participant, the total number of complex clauses was divided by the total number of clauses to calculate the syntactic complexity score, separately for Turkish and English narrative samples. The coding was done by two trained assistants. We found a high interrater reliability between the coders ($\kappa=.92, p<.001$).

3. 3. Results

3. 3.1. Analysis plan

We will present two sets of results, namely *preliminary results* and *main results*. The preliminary results section includes analyses that do not specifically target our hypotheses, but provide initial information with regard to the sample characteristics and participants’ overall performance in the given tasks (Table 3.1). The main results section, however, presents analyses, which are directly motivated to test our hypotheses.

Table 3.1

Descriptive statistics of children’s language measures.

	Language group				Age group			
	Monolingual		Bilingual		5-year		7-year	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Narrative length (word count)								
Turkish narrative length	195.80	73.10	162.90	105.86	160.74	116.15	199.52	51.98
English narrative length	-	-	237.59	105.92	214.88	130.83	261.25	66.41

Speech disfluency in L1-Turkish

Overall disfluency rate	.19	.07	.24	.07	.21	.08	.21	.07
Silent pause rate	.11	.05	.14	.05	.15	.05	.10	.04
Filled pause rate	.03	.03	.06	.05	.03	.03	.06	.05
Repetition rate	.03	.02	.02	.02	.02	.02	.03	.02
Repair rate	.02	.01	.02	.01	.01	.01	.02	.01

Speech disfluency in L2-English

Overall disfluency rate	-	-	.25	.10	.25	.11	.26	.10
Silent pause rate	-	-	.16	.07	.18	.08	.15	.06
Filled pause rate	-	-	.05	.05	.03	.03	.07	.05
Repetition rate	-	-	.03	.02	.03	.03	.03	.02
Repair rate	-	-	.01	.01	.01	.01	.02	.01

Gesture use in L1-Turkish

Overall gesture frequency	.074	.082	.077	.076	.073	.083	.078	.075
Iconic gesture frequency	.010	.013	.003	.007	.004	.009	.009	.012
Deictic gesture frequency	.056	.075	.068	.073	.062	.081	.061	.068
Beat gesture frequency	.004	.007	.002	.007	.001	.004	.004	.009

Gesture use in L2-English

Overall gesture frequency	-	-	.066	.066	.072	.064	.060	.068
Iconic gesture frequency	-	-	.007	.009	.009	.009	.005	.008
Deictic gesture frequency	-	-	.054	.059	.058	.057	.051	.062
Beat gesture frequency	-	-	.001	.003	.001	.003	.001	.003

Language proficiency

TIFALDI score (out of 159)	81.67	11.70	79.84	11.60	72.70	11.25	88.41	5.01
PPVT-4 score (out of 228)	-	-	86.33	25.23	72.54	20.12	99.56	22.66

Syntactic Complexity

Syntactic complexity score (L1-Turkish)	.54	.16	.39	.20	.41	.22	.51	.15
Syntactic complexity score (L2-English)	-	-	.24	.18	.17	.14	.30	.19

Note. All disfluency and gesture rates/frequencies were calculated per word. For gesture frequencies, we reported three decimals as there was a limited number of gestures per word. Syntactic complexity was calculated as proportions by dividing the total number of complex clauses by the total number of clauses.

3.3.2. Preliminary results

3.3.2.1. TIFALDI and PPVT-4 scores

We carried out a two-way ANOVA to examine the effect of language group (bilingual or monolingual) and age group (5- or 7-year-old) on receptive vocabulary scores measured by TIFALDI (Berument & Güven, 2010). Results indicated a main effect of age group, $F(1,108)=91.10$, $p<.001$, $\eta p^2=.458$. The receptive vocabulary scores were higher in 7-year-old

than 5-year-old children (Table 1). However, the main effect of language group, $F(1,108)=0.41$, $p=.525$, $\eta p^2=.004$, and the interaction between age group and language group, $F(1,108)=0.10$, $p=.747$, $\eta p^2=.001$, was not significant.

We also conducted an independent samples *t*-test to examine whether 5- and 7-year old bilingual children differed in terms of the English vocabulary scores measured by PPVT-4 (Dunn & Dunn, 2015). Results suggested that 7-year-old children were more advanced in English vocabulary knowledge than 5-year-old children, $t(47)=-4.41$, $p<.001$, $d=1.26$ (Table 1). Moreover, as a norm-referenced test, the PPVT-4 results indicated that 5-year-old bilingual children's vocabulary performance corresponded to 4;2 age level for monolingual English speakers. Similarly, the 7-year-old bilingual children's English vocabulary skills were found to be around the 6;1 age level for monolingual English speaker children.

3.3.2.2. Syntactic Complexity

We conducted a two-way ANOVA to examine the effect of language group (bilingual or monolingual) and age group (5- or 7-year-old) on syntactic complexity scores. We found that the main effect of age group, $F(1,107)=8.06$, $p=.005$, $\eta p^2=.070$, was significant. The syntactic complexity scores were higher in the 7-year-old than the 5-year-old group. Similarly, we found a significant main effect of the language group, $F(1,107)=17.36$, $p<.001$, $\eta p^2=.140$. In other words, monolinguals' speech was syntactically more complex than bilinguals' speech (Table 1). However, results did not yield a significant interaction between these variables, $F(1,107)=2.56$, $p=.113$, $\eta p^2=.023$.

We also carried out a paired samples *t*-test to examine whether bilinguals' English and Turkish narratives were comparable in terms of syntactic complexity. We found that bilingual children produced syntactically more complex speech in Turkish ($M=0.40$, $SD=0.20$) than English ($M=0.24$, $SD=0.18$), $t(45)=5.00$, $p<.001$, $d=0.73$.

3.3.3. Speech disfluency

3.3.3.1. Monolingual-bilingual comparison

We carried out a two-way ANCOVA to examine the effect of language group (bilingual or monolingual) and age group (5- or 7-year-old) on children's overall disfluency rates in Turkish, controlling for the (Turkish) syntactic complexity scores. The main effect of the language group was significant, $F(1,106)=12.42$, $p<.001$, $\eta p^2=.105$. Bilingual children were more disfluent than monolingual children in Turkish (Table 1). The main effect of age group, $F(1,106)=0.15$, $p=.702$, $\eta p^2=.001$, and the interaction between age group and language group were not significant, $F(1,106)=0.23$, $p=.634$, $\eta p^2=.002$. The syntactic complexity score was not a significant covariate, $F(1,106)=0.70$, $p=.792$, $\eta p^2=.001$.

We then repeated the same two-way ANCOVA for the specific disfluency categories. For the silent pause rate, we found a significant main effect of the language group, $F(1,106)=12.03$, $p<.001$, $\eta p^2=.102$. Overall, bilingual children used more silent pauses than monolingual children (Table 1). Similarly, the main effect of the age group was significant, $F(1,106)=25.25$, $p<.001$, $\eta p^2=.192$, suggesting that the 5-year old group produced more silent pauses than the 7-year-old group. However, the interaction between these variables was not significant, $F(1,106)=0.04$, $p=.847$, $\eta p^2=.000$. The syntactic complexity score was not a significant covariate, either, $F(1,106)=0.14$, $p=.712$, $\eta p^2=.001$.

For the filled pause rate in children's Turkish narrations, we again found a main effect of language group, $F(1,106)=11.89$, $p<.001$, $\eta p^2=.101$, suggesting that bilingual children used more filled pauses than monolingual children (Table 1). Similarly, the main effect of age was significant, $F(1,106)=24.12$, $p<.001$, $\eta p^2=.185$, suggesting that 7-year old children produced more filled pauses than 5-year-old children. However, the interaction between language group and age group was not significant, $F(1,106)=0.93$, $p=.337$, $\eta p^2=.009$. The syntactic complexity score was not a significant covariate, either, $F(1,106)=0.23$, $p=.637$, $\eta p^2=.002$.

For the repetition rate in Turkish narrations, we found that the main effect of language group, $F(1,106)=2.38$, $p=.126$, $\eta p^2=.022$, the main effect of age group, $F(1,106)=0.01$, $p=.908$, $\eta p^2=.000$, and interaction between these variables, $F(1,106)=3.44$, $p=.067$, $\eta p^2=.031$, were not significant. The syntactic complexity score was not a significant covariate, $F(1,106)=0.06$, $p=.800$, $\eta p^2=.001$. For the repair rate, results suggested that the main effect of language group, $F(1,106)=0.33$, $p=.564$, $\eta p^2=.003$, the main effect of age group, $F(1,106)=3.56$, $p=.062$, $\eta p^2=.032$, and the interaction between these variables were not significant, $F(1,106)=3.75$, $p=.056$, $\eta p^2=.034$. The syntactic complexity score was not a significant covariate, $F(1,106)=1.55$, $p=.216$, $\eta p^2=.014$.

3.3.3.2. Bilinguals' L1 vs. L2

To examine whether overall disfluency differed between bilinguals' two languages (Turkish and English) and across age groups (5- and 7-year-old), we conducted a repeated measures ANOVA. Results showed that the main effect of language, $F(1,45)=1.01$, $p=.319$, $\eta p^2=.022$, the main effect of age group, $F(1,45)=0.01$, $p=.915$, $\eta p^2=.000$, and the interaction between these variables, $F(1,45)=0.69$, $p=.412$, $\eta p^2=.015$, were nonsignificant.

We then repeated the same repeated measures ANOVA for the use of specific disfluency categories. The dependent variables were silent pause rate, filled pause rate, repetition rate, and repair rate. For the silent pause rate, we found a significant main effect of language, $F(1,45)=4.06$, $p=.050$, $\eta p^2=.083$, suggesting that, regardless of age, bilingual children produced more silent pauses in English ($M=0.16$, $SD=0.07$) than Turkish ($M=0.14$, $SD=0.05$). The main effect of age group, $F(1,45)=6.82$, $p=.012$, $\eta p^2=.132$, and interaction between age group and language were not significant, $F(1,45)=0.33$, $p=.567$, $\eta p^2=.007$. For the filled pause rate, results yielded a significant main effect of age group, $F(1,46)=13.26$, $p<.001$, $\eta p^2=.224$. Regardless of language, 7-year-old bilinguals produced more filled pauses than 5-year-old bilinguals (Table 1). The main effect of language, $F(1,46)=2.74$, $p=.104$, $\eta p^2=.056$,

and the interaction between age group and language, $F(1,46)=0.19$, $p=.665$, $\eta p^2=.004$, were not significant. For the repetition rate, results suggested a significant main effect of language, $F(1,46)=7.15$, $p=.010$, $\eta p^2=.134$. Regardless of age, bilingual children produced more repetitions in English ($M=0.03$, $SD=0.02$) than Turkish ($M=0.02$, $SD=0.02$). The main effect of age group, $F(1,46)=0.84$, $p=.366$, $\eta p^2=.018$, and the interaction between age group and language, $F(1,46)=0.44$, $p=.513$, $\eta p^2=.009$, were not significant. Last, for the repair use, results showed that the main effect of language, $F(1,46)=1.39$, $p=.245$, $\eta p^2=.029$, the main effect of age group, $F(1,46)=2.98$, $p=.091$, $\eta p^2=.061$, and the interaction between these variables, $F(1,46)=3.68$, $p=.061$, $\eta p^2=.074$, were all nonsignificant.

3.3.4. Gesture

3.3.4.1. Monolingual-bilingual comparison

We carried out a two-way ANOVA to examine the effect of language group (bilingual or monolingual) and age group (5- or 7-year-old) on children's overall gesture frequency in Turkish. We found that the main effect of language group, $F(1,108)=0.04$, $p=.841$, $\eta p^2=.000$, and the main effect of age group, $F(1,108)=0.12$, $p=.732$, $\eta p^2=.001$, were not significant. Similarly, the interaction between these variables was nonsignificant, $F(1,108)=0.82$, $p=.369$, $\eta p^2=.007$.

We then repeated the same two-way ANOVA for the specific gesture categories. For the iconic gesture frequency, we found a significant main effect of language group, $F(1,108)=10.64$, $p=.001$, $\eta p^2=.090$. Overall, monolingual children used more iconic gestures than bilingual children (Table 1). However, the main effect of age, $F(1,108)=2.90$, $p=.091$, $\eta p^2=.026$, and the interaction between language group and age group were not significant, $F(1,108)=1.50$, $p=.223$, $\eta p^2=.014$. For the deictic gesture frequency, results demonstrated that the main effect of language group, $F(1,108)=0.82$, $p=.367$, $\eta p^2=.008$, and the main effect of age group, $F(1,108)=0.02$, $p=.967$, $\eta p^2=.000$, were not significant. Similarly, the interaction

between these variables was nonsignificant, $F(1,108)=0.40$, $p=.528$, $\eta p^2=.004$. Last, for the beat gesture frequency, the main effect of language group, $F(1,108)=1.26$, $p=.265$, $\eta p^2=.011$, the main effect of age group, $F(1,108)=3.09$, $p=.082$, $\eta p^2=.028$, and the interaction between these variables, $F(1,108)=0.33$, $p=.569$, $\eta p^2=.003$, were not significant.

3.3.4.2. Bilinguals L1 vs. L2

To examine whether overall gesture frequency differed between bilinguals' two languages (Turkish and English) and across age groups (5- and 7-year-old), we carried out a repeated measures ANOVA. Results indicated that the main effect of age, $F(1,46)=0.52$, $p=.476$, $\eta p^2=.011$, and the main effect of language was not significant, $F(1,46)=1.76$, $p=.191$, $\eta p^2=.037$. We also did not find an interaction between these two variables, $F(1,46)=0.04$, $p=.848$, $\eta p^2=.001$.

We then repeated the same repeated measures ANOVA for the use of specific gesture categories. The dependent variables were iconic gesture frequency, deictic gesture frequency, and the beat gesture frequency. For the iconic gesture frequency, we found a significant main effect of language, $F(1,46)=7.17$, $p=.010$, $\eta p^2=.135$, suggesting that, although few in number, bilingual children produced more iconic gestures in English ($M=0.01$, $SD=0.01$) than Turkish ($M<0.001$, $SD=0.01$). The main effect of age group, $F(1,46)=0.72$, $p=.400$, $\eta p^2=.015$, and interaction between age and language were not significant, $F(1,46)=3.00$, $p=.090$, $\eta p^2=.061$. For the deictic gesture frequency, the main effect of language, $F(1,46)=3.46$, $p=.069$, $\eta p^2=.070$, the main effect of age group, $F(1,46)=0.34$, $p=.564$, $\eta p^2=.007$, and the interaction between these variables, $F(1,46)=0.17$, $p=.685$, $\eta p^2=.004$, were not significant. Similarly, for the beat gesture frequency, the main effect of language, $F(1,46)=0.03$, $p=.866$, $\eta p^2=.001$, the main effect of age group, $F(1,46)=0.16$, $p=.688$, $\eta p^2=.004$, and the interaction between these variables, $F(1,46)=0.02$, $p=.899$, $\eta p^2=.000$, were nonsignificant.

3.3.5. Gesture and speech disfluency

3.3.5.1. Turkish narrative samples

We carried out two separate linear regression analyses to predict the total disfluency rate in children's Turkish narratives. For the first regression analysis, the predictor variables were iconic gesture frequency, language group (monolingual or bilingual), age group (5 or 7 years), TIFALDI score, and syntactic complexity score (in Turkish) (Model 1). The dependent variable was the total disfluency rate in children's Turkish narratives. Results yielded a significant regression equation explaining 10% of the total variance ($R^2=.10$, $F(5,105)=3.45$, $p=.006$). Only language group was a significant predictor of disfluency rate (Table 2). That is, being bilingual was associated with higher disfluency rates.

We then repeated the same linear regression analysis by replacing the iconic gesture frequency with the overall gesture frequency. The dependent variable was again the total disfluency rate in children's Turkish narratives. The predictor variables were overall gesture frequency, language group (monolingual or bilingual), age group (5 or 7 years), TIFALDI score, and syntactic complexity score (in Turkish) (Model 2). The regression equation was significant explaining 12% of the variance ($R^2=.12$, $F(5,105)=3.88$, $p=.003$). This time, not only language group but also gesture frequency significantly predicted the total disfluency rate in Turkish. Being bilingual was associated with a higher disfluency rate. Moreover, the frequency of using gestures negatively predicted the total disfluency rate; as the children produced more gestures, they were less disfluent. Age group, TIFALDI score, or syntactic complexity score did not significantly contribute to the model (Table 3.2).

Table 3.2

Regression analysis, the total disfluency rate in Turkish narratives as the outcome variable.

Outcome: The total disfluency rate in children's Turkish narratives.

Predictors	Model 1			Predictors	Model 2		
	β	p	R^2		β	p	R^2
Model 1		.006**	.100	Model 2		.003**	.116

Iconic gesture frequency	-.171	.080	Overall gesture frequency	-.210	.027*
Language group (bilingual or monolingual)	.289	.006**	Language group (bilingual or monolingual)	.325	.001**
Age group (5 or 7 years)	.020	.875	Age group (5 or 7 years)	.034	.789
TIFALDI score	-.036	.774	TIFALDI score	-.070	.573
Syntactic complexity score	.017	.870	Syntactic complexity score	-.028	.784

Note. * $p < .05$, ** $p < .01$, $N = 112$.

3.3.5.2. English narrative samples

We carried out two separate linear regression analyses to predict the total disfluency rate in bilingual children's English narratives. For the first regression analysis, the predictor variables were iconic gesture frequency, age group (5 or 7 years), PPVT score, and syntactic complexity score (in English) (Model 3). The dependent variable was the total disfluency rate in children's English narratives. The regression equation was not significant, $F(4,40)=1.26$, $p=.303$, with an R^2 of .023, and none of the variables significantly predicted the total disfluency rate in English (Table 3.3).

When we conducted the same regression analysis by replacing the iconic gesture frequency with the overall gesture frequency, the model was still nonsignificant, $F(4,40)=1.83$, $p=.142$, with an R^2 of .070 (Model 4). Neither overall gesture frequency nor other predictors (age group, PPVT score, and syntactic complexity score) significantly contributed to the model (Table 3.3).

Table 3.3

Regression analysis, the total disfluency rate in English narratives as the outcome variable.

Outcome: The total disfluency rate in children's English narratives.

Predictors	Model 3			Predictors	Model 4		
	β	p	R^2		β	p	R^2
Model 1		.303	.023	Model 2		.142	.070

Iconic gesture frequency	-.176	.268	Overall gesture frequency	-.297	.074
Age group (5 or 7 years)	.203	.265	Age group (5 or 7 years)	.285	.110
PPVT score	-.305	.116	PPVT score	-.395	.050
Syntactic complexity score	-.082	.640	Syntactic complexity score	-.148	.391

Note. *N* = 51.

3.4. Discussion

This study investigated the speech disfluency and gesture production of 5- and 7-year-old typically developing Turkish monolingual and Turkish-English bilingual children. We examined monolinguals' Turkish narratives and bilinguals' both Turkish and English narratives. We asked whether children's gesture use was associated with their speech fluency in narrative production, regardless of being bilingual or monolingual. Our results indicated that, overall, bilinguals were more disfluent than monolinguals. There were also some age and language group differences in terms of the frequency of using specific disfluency types. Bilinguals used silent pauses and filled pauses more frequently than monolinguals. Regardless of being monolingual or bilingual, the frequency of using silent pauses was higher in the 5-year-old group while the filled pause frequency was higher in the 7-year-old group. However, there was neither an age group nor a language group difference in terms of the frequency of using repetitions and repairs. We also found that bilinguals used silent pauses and repetitions more frequently in English than Turkish. For gesture use, we demonstrated that all gesture frequencies were comparable across the two language groups and the two age groups, except the iconic gesture frequency. Monolingual children produced more iconic gestures than bilinguals. Additionally, bilingual children produced iconic gestures more frequently in English than Turkish. Last, our results suggested that along with the overall gesture frequency, being bilingual or monolingual were significant predictors of overall disfluency rate in Turkish. Iconic gesture frequency, however, was not associated with the disfluency rate. For the English

narrative samples, we found that neither overall gesture frequency nor iconic gesture frequency were significant predictors of the overall disfluency rate.

As we expected, compared to Turkish monolinguals, bilingual children were more disfluent in Turkish. The use of specific disfluency types might give cues regarding the nature of disfluent speech in a bilingual context. For instance, the higher frequency of silent pauses and filled pauses observed in bilinguals' than monolinguals' speech might reflect bilinguals' effort to control two lexicons and maintain the conversational floor at the same time. Considering that bilinguals' two language systems are simultaneously active (Bialystok, 2009), they might have frequently preferred giving silent breaks while planning their speech to make sure that they select from the vocabulary of one language and ignore the other. Those silent breaks might indeed prevent bilinguals from switching between two languages and help them harmonize with their listeners. Using filled pauses might be a different form of maintaining the conversational floor as filled pauses signal speakers' aim to continue speaking, which in turn keeps listeners engaged (Bortfeld et al., 2001; Corley & Stewart, 2008). Bilinguals might strategically use filled pauses to keep listeners' attention while resolving the competition between their two vocabulary systems. Then, bilinguals' frequent use of filled pauses might result from their heightened communicative sensitivity to maintain the conversational floor.

Similarly, as we predicted, silent pauses were more commonly observed in the 5-year-old group, while the use of filled pauses was more frequent in the 7-year-old group. This finding is in line with children's enhanced language and communication skills (Nettelbladt & Hansson, 1999) as well as increased pragmatic skills and narrative competence (Hickmann, 2003). Perspective-taking and meta-representational abilities, which might further strengthen communication, also increase with age (Astington & Baird, 2005). As children's language and social skills develop, they might prefer filled pauses over silent pauses to create a stronger communicative ground in which they better manage listeners' point of view and turn takings

(Haynes & Hood, 1977). Moreover, unlike what we expected, we found that the use of repetitions was comparable between the two age groups. We argue that as the decrease in repetitions are usually observed before children reach preschool age (DeJoy & Gregory, 1985; Haynes & Hood, 1977), there might not be a prominent difference in repetition rates between 5- and 7-year-old children's speech.

Bedore et al. (2006) demonstrated that Spanish-English bilingual children used more repetitions and grammatical repairs in Spanish than in English. The authors suggested that as Spanish is more complex than English in terms of morphosyntactic elements, bilinguals might be likely to repeat and revise their language output to comply with the specific rules such as gender-noun agreement in Spanish. Our study focused on the two typologically different languages, Turkish and English. Although bilinguals' overall disfluency rates were comparable between Turkish and English, our results indicated differences in terms of the frequency of using silent pauses and repetitions. The interpretation of this finding might be open to alternative explanations, but the results may stem from the following factors which might be future questions to be asked. Turkish is a gender-neutral language, meaning that it does not contain grammatical genders or gender pronouns. In English, however, there are gender pronouns. Turkish is a gender-neutral language, meaning that it does not contain grammatical genders or gender pronouns. In English, however, there are gender pronouns. Moreover, the two languages differ in terms of determinants as Turkish does not have a definite article such as *the* in English. Therefore, Turkish-English bilinguals might give silent breaks more frequently in English to evaluate whether they comply with the specific and relatively complex grammatical rules that do not exist in Turkish. Similarly, in some cases, after passing the internal evaluation, overtly articulating some parts of speech in the form of repetitions might reflect bilinguals' hesitations, and they need to double check whether everything sounds good regarding the grammatical rules in English.

Unlike what we expected, we found that bilingual and monolingual children's overall gesture frequencies were comparable in Turkish. This finding is in line with Azar et al. (2020) as Turkish-Dutch bilingual adults' gesture frequencies were comparable with those of monolinguals both in Turkish and Dutch. Similarly, Cavicchio and Kita (2013) obtained narratives from Italian-English bilingual and English and Italian monolingual adults. They found Italian monolinguals gesturing more frequently and using a larger gesture space than English monolingual speakers. However, bilinguals' gestures were similar to the baseline rates in either of their languages. Although our sample did not include English monolinguals as a control group, observing similar overall gesture rates in Turkish across the two groups partially corroborates with these findings. One explanation is that since the bilingual children in our study were mostly dominant in Turkish with limited native and cultural contact to English, this finding might have stemmed from a dominance effect. They might be using the gestures of their dominant language in both languages. Yet, similar to bilingual children in this study who used Turkish mainly at home and English mainly at school, the bilingual adults of Azar et al. (2020) and Cavicchio and Kita (2013) used one of their languages mostly at home with friends and family. In these both studies, bilinguals were highly proficient in both of their languages. Our replication with children supports the argument that cultural and experience-related differences might explain contrasting results regarding the gesture frequency of bilingual speakers.

Our results indicated a higher frequency of iconic gestures in monolinguals' than bilinguals' narratives in Turkish. Since bilinguals acquire two languages, they might practice less in each of their languages compared to monolinguals speaking those languages. As a result, they control a smaller vocabulary in either of their languages (Bialystok & Feng, 2009) and they are disadvantaged in language related tasks (for a review, see Bialystok, 2009). Language is multimodal, and the use of iconic gestures is positively related to spoken language skills

among children (Nicoladis, 2002). As a result, monolingual children might also be ahead of their bilingual peers in terms of using complex gesture forms such as iconic gestures (Nicoladis et al., 1999). On the other hand, although we expected higher deictic gesture frequency in the bilingual group than the monolingual group, our results did not yield such a difference. Deictic gestures are considered a part of the bilinguals' strategy of decreasing the cognitive load; however, those deictic gestures are abstract in nature (i.e., pointing in the absence of referents). In our study, as the story book was open in front of the participants, the deictic gestures they produced were mostly concrete deictics (i.e., pointing at visually present objects). Unlike abstract deictics (Nicoladis, 2007), concrete deictics are not associated with the bilingual strategy of decreasing the cognitive load. Then, it is not surprising to observe similar deictic gesture rates across the narratives of bilingual and monolingual groups. Moreover, although the two age groups were comparable in iconic and beat gesture frequencies, there was a tendency that the 7-year-old group used these gestures more frequently than the 5-year-old group. We suggest that the two years of age difference between these groups might not be sufficient to yield a significant difference in terms of the use of iconic and beat gestures. Rather, such a difference might be more prominent when 5-year-olds are compared with children older than 7 years of age. In line with this argument, Colletta (2009) suggests that around 9 years of age, children are able to spontaneously produce narratives similar to that of adults in the sense that they effectively use co-speech gestures. Future work can examine 9-year-old Turkish-English bilinguals' gesture use.

Although we expected bilinguals to reveal comparable gesture frequencies across their two languages, we found that iconic gesture frequency was higher in bilinguals' English than Turkish narratives. Having Turkish native speaker parents and living in Turkey, bilingual children has been predominantly exposed to Turkish since birth. They are sequential bilinguals and English is their L2. As a result, overall, bilingual children might have relatively less

practice in English than Turkish, although they are proficient in both languages. Levelt et al. (1999) suggests that lexical selection and phonological form retrieval are linked to each other, and together they form the lexical retrieval process. As a result of less practice, bilinguals' lemma to word form connections might be weaker in English. Such weak connections might be associated with slower lexical retrieval (Levelt et al., 1999). Considering iconic gestures' role in lexical retrieval (Morsella & Krauss, 2004; Rauscher et al., 1996), bilinguals might have used more iconic gestures in English than Turkish to enhance the word retrieval process.

Our results indicated that although not significant, there was a tendency for iconic gesture use to be negatively associated with children's disfluency rates both in Turkish (controlling for language group) and English narrative samples. Producing gestures might decrease cognitive load and enhance cognitive processes (Kita et al., 2017). Yet, using gestures in a way that facilitates cognitive processes, particularly speech production might be more prominent in a more mature language system. It is because iconic gestures are complex in nature, which are intact with advanced language skills (Iverson & Goldin-Meadow, 2005). Yet, it is important to note that even though a significant negative association of iconic gestures with speech disfluency was observed in adults, it would only be an indirect evidence of gestures' facilitative roles in speech production. Disfluency is a complex concept to study as disfluent segments in speech might also stem from communicative intentions (Fraundorf & Watson, 2014).

We demonstrated that rather than iconic gesture frequency, overall gesture frequency was negatively associated with disfluency rates in Turkish narratives, regardless of being bilingual or monolingual. Although iconic gestures are extensively investigated in terms of having self-oriented functions (for a review, see Kita et al., 2017; Kita & Davies, 2009; Melinger & Kita, 2007), recently there is an increased focus on the role of different gesture types in relation to cognitive processes, such as deictic gestures (e.g., Arslan & Göksun, 2022;

Avcı et al., 2022; Azar et al., 2020) and beat gestures (e.g., Llanes-Coromina et al., 2018; Vilà-Giménez et al., 2019; Vilà-Giménez & Prieto, 2020). Research suggests that not only using iconic gestures (Demir et al., 2014; Stites & Özçalışkan, 2017) but also using deictic gestures (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005) and beat gestures (Vilà-Giménez et al., 2021) can predict children's language development and narrative performance. Moreover, the use of deictics are observed in different populations, including older adults (Arslan & Göksun, 2021, 2022) and bilinguals (Azar et al., 2020; Nicoladis, 2007), suggesting that pointing at abstract referents might decrease the cognitive load in these groups. Similarly, Vilà-Giménez and Prieto (2020), demonstrated that children produce narratives more fluently when they were encouraged to produce beat gestures, suggesting that beats might be pragmatically meaningful gestures.

These findings together indicate that different gesture types might be orchestrated in a coordinated multimodal system, which contribute to the narrative performance. Focusing solely on iconic gestures might provide a limited perspective, considering the role of different gesture types in terms of decreasing cognitive load and enhancing fluency. Our results are partially in line with this argument as we found a significant negative association of overall gesture frequency with disfluency rates only in children's Turkish narratives. One possible explanation of not observing such an association in the English narratives might be related to the sample size. We ran the regression analysis for the Turkish narratives with a total of 112 children, 61 of them being monolingual and 51 of them being bilingual. However, as our sample did not include English monolingual children, we carried out the regression analysis for the English narratives with a total of 51 bilinguals. The multiple linear regression analysis is sensitive to the number of observations and the number of independent variables (Cohen et al., 2003). Then, it would not be surprising to demonstrate such association in Turkish but not in English narratives. In line with this argument, although not significant, there is still a salient

tendency that overall gesture frequency and disfluency rate in children's English narratives were negatively associated. Future research should investigate and compare children's multimodal language samples in different contexts such as free talk or play talk, which might further shed light on typical language development as a result of being cognitively less demanding than narrative production.

In conclusion, the current study suggests that speech disfluency cannot be fully understood without incorporating the gesture modality into the picture, in line with the close link between gesture and speech (Kita & Özyürek, 2003). Our findings highlight the importance of overall gesture use rather than iconic gesture use alone in portraying gesture-speech interaction with regard to speech fluency. We also found that bilingual and monolingual children differed in the use of specific disfluency and gesture types, which might shed light on their language development and communication strategies. Overall, our study suggests that each gesture type might carry a function in a coordinated multimodal system, which might, in turn, influence speech quality.

CHAPTER 4

DO BILINGUAL ADULTS GESTURE WHEN THEY ARE DISFLUENT?: UNDERSTANDING GESTURE-SPEECH INTERACTION ACROSS FIRST AND SECOND LANGUAGES

This chapter is based on Arslan, B., Avcı, C., Yılmaztekin, A., & Göksun, T. (under review).

Do bilingual adults gesture when they are disfluent?: Understanding gesture-speech interaction across first and second languages.

Abstract

People are more disfluent in their second language (L2) than their first language (L1). Using gestures facilitates cognitive processes, including speech production. This study investigates speech disfluency and representational gesture production across Turkish-English bilingual adults' L1 (Turkish) and L2 (English) through a narrative retelling task (N=27). Results showed that people were more disfluent and used more representational gestures in English. Controlling for L2 proficiency, people were still more disfluent in English. The more people were proficient in L2, the more they used representational gestures in that language. Similarly, disfluency-gesture co-occurrences were more common in English. L2 proficiency was positively correlated with the likelihood of a disfluency being accompanied by a gesture. These findings suggest that gestures do not only compensate for weak language skills. Rather, people might use more representational gestures during disfluent moments if they have the knowledge to detect their own errors, signaling a metacognitive motivation for gesture production.

4.1. Introduction

People spontaneously gesture while speaking. Research indicates that using representational gestures (e.g., describing a window as a square) might facilitate speakers' speech production (Morsella & Krauss, 2004; Rauscher et al., 1996). However, there are individual and population-related differences in gesturing (e.g., Arslan & Göksun, 2021; 2022; Avcı et al., 2022; for a review, Özer & Göksun, 2020). Bilingual individuals frequently gesture (e.g., Nicoladis et al., 2009; Pika et al., 2006; So, 2010), suggesting that gesturing helps them pick out the target lexicon among the two competing language systems (Nicoladis, 2007). Research also demonstrates that speech disfluency rates are higher in adult bilinguals' second language (L2) than in their first language (L1) (Bergmann et al., 2015; Gollan et al., 2005). Yet, when they are disfluent, do bilingual adults differ in using representational gestures across the two languages? In this study, we investigate (1) disfluency and representational gesture rates of bilinguals in their two languages, and (2) whether and when representational gestures co-occur with disfluencies across their two languages.

4.1.1. Disfluency

Disfluencies are temporary interruptions in speech (Maclay & Osgood, 1959). These interruptions can occur in different forms, such as silent breaks, filled pauses (e.g., um), repetitions (e.g., the right *right* corner), or repairs (e.g., it is a blue - *green* hat). Disfluencies are considered part of language and might reflect speakers' metacognitive processes and communication-related intentions (Brennan & Williams, 1995; Corley & Stewart, 2008). Research also argues that disfluencies might stem from cognitive load (Bock, 1996; Bortfeld et al., 2001; Oviatt, 1995). People are more disfluent when they describe complex than easy items (Bortfeld et al., 2001). Similarly, disfluencies are frequently observed at the beginning of sentences, which might be a byproduct of the cognitive load (Fraundorf & Watson, 2014).

4.1.2. Disfluency in L1 and L2

People are more disfluent in certain situations where speaking is difficult (Corley & Stewart, 2008). For instance, speaking a second language (L2) can be a difficult task, and bilinguals are usually more disfluent in their L2 than their first language (L1) (Bergmann et al., 2015; Gollan et al., 2005). One of the most intuitive and experimentally supported factors explaining differential disfluency rates across one's L1 and L2 is language proficiency (Bergmann et al., 2015; Kosmala & Crible, 2022). Speech fluency depends on being knowledgeable about the language, such as having a large lexicon and being able to rapidly access and use the stored knowledge (Gollan et al., 2005). Such factors which make speech fluent are absent in low language proficiency (Gollan et al., 2005; Ortega, 2003; Sabourin & Stowe, 2008). While improving language proficiency can pave the way to a more fluent speech in L2, research suggests that bilinguals, given that most learn their L2 later in life, are quite unlikely to reach the same level of fluency as they have in their L1 (Ortega, 2003; Sabourin & Stowe, 2008). Many people, in a way, inevitably possess at least some degree of disfluency in their L2, and different factors might account for that. Language acquisition from birth is characterized by the better ability to match concepts and their linguistic representations (Appel, 1996), while the presence of a socially supporting condition also allows for faster language acquisition (Zhang & Wang, 2007). In sum, while higher language proficiency can lead to less disfluent speech, there seem to be fundamental barriers standing in the way of equating L1 and L2 fluency rates since L2 proficiency is likely to be lower than L1 proficiency.

It is still possible to be more disfluent in L2 than L1, despite being equally proficient in both languages (Bergmann et al., 2015; but also see Arslan et al., 2023). One of the primary reasons behind this counter-intuitive pattern is believed to be the perpetual co-activation of L1 and L2 (Bergmann et al., 2015; Hermans et al., 1998). The activation of one's L1 might interfere with the activation of their L2 since representations used in L1 are strong and automatically activated, which makes it difficult to switch to L2. This co-activation of both

language systems causes the task-irrelevant language to interfere with the workings of the other and most likely causes L2-related task performance to drop (Costa et al., 1999; Hermans et al., 1998).

Similarly, speech conceptualization might affect speech fluency differentially for L1 and L2. Felker and colleagues (2019) employed a network description task where participants were tasked with describing paths leading to target stimuli. Speech conceptualization was made more difficult/cognitively taxing by forcing participants to change their path description unexpectedly, requiring them to abandon their already constructed speech conceptualizations and regenerate them accordingly. Even though there was no difference in overall disfluency rates across languages, participants still differed in one measure of disfluency, speech speed, evident by speech in L2 being significantly slower than in L1. Thus, it seems that conceptualization can also be a factor that influences speech disfluency differently depending on the language.

Taken together, different factors make people more disfluent at speaking in L2 than in L1. Low language proficiency leads to more disfluent speech because it fails to provide the speaker with the crystallized information and linguistic automaticity needed to speak fluently. L1 and L2 co-activation might also make people more disfluent because L1 interferes with the efficient working of L2. Last, conceptualization failures damage fluency by forcing one to reconceptualize the intended utterance, which is performed slower in L2.

4.1.3. Gesture

Gestures have been studied in a variety of paradigms, and as a result, new knowledge regarding their interaction with a variety of cognitive and communicational functions has been unearthed. As in this study, most current frameworks use the term gesture to refer to representational gestures (e.g., Kita et al., 2017; Krauss, 1998). Perhaps the most prominent facilitations representational gestures bring about is helping with the process of lexical retrieval

and, consequently, speech production (Krauss, 1998; Rauscher et al., 1996). The Lexical Retrieval Hypothesis (Krauss, 1998) argues that gesturing activates sensorimotor features of words to be retrieved, making their retrieval and production easier. Morsella and Krauss (2004) conducted a study where the participants were tasked with describing objects that are easy (codable) or difficult (non-codable) to remember, either from memory (absent condition) or readily presented in front of them (present condition). Participants gestured more when describing objects from memory or describing difficult objects instead of easy ones, both results suggesting that gesturing kicks in to enhance speech production when spatial working memory is taxed. Yet, as studies examining the relationship between gestures and cognitive functions increased in number, the Lexical Retrieval Hypothesis started to be seen as incomplete, shedding light on only one possible function of representational gestures.

Currently, there has been an increased focus on gestures' self-oriented functions (Kita et al., 2017). As opposed to the Lexical Retrieval Hypothesis, the Gesture-for-conceptualization hypothesis argues that representational gestures (e.g., leaving a small space between two hands to mean close) facilitates not only speech but also cognitive processes in general, in which speech production is one of the functions (Kita et al., 2017). Individuals gesture more frequently when they interpret complex stimuli, which is difficult to conceptualize (Kita & Davies, 2009; Melinger & Kita, 2007). Individuals' frequent use of gestures while describing objects from memory is also in line with the cognitive load exerted by a memory task (Ping & Goldin-Meadow, 2010; Morsella & Krauss, 2004). These findings suggest that gesturing, particularly using representational gestures might help people lighten their cognitive load. Yet, people differ in the extent to which they benefit from gestures both at an individual and group level (Özer & Göksun, 2020). In the scope of this paper, we focus on bilingual adults' gesture production to understand how gestures' self-oriented functions are reflected across their L1 and L2.

4.1.4. Gesture in L1 and L2

Individuals gesture more frequently in their L2 (Gullberg, 1998; Hadar et al., 2001; Nicoladis et al., 2007). Hadar and colleagues (2001) demonstrated that Hebrew-English bilinguals produced more representational gestures in their L2-English compared to their L1-Hebrew, suggesting the role of gesture in the word retrieval process. In a recent study, Arslan and colleagues (2023) demonstrated that Turkish-English bilingual children were more likely to produce representational gestures in their L2-English. Children were proficient in both languages, but they had less practice in L2 than L1. Therefore, children might prefer using more representational gestures in L2 to decrease the cognitive load of speaking L2 and facilitate the speech process. Yet, Gullberg (2013) argued that albeit individuals produce more gestures in their L2, they might not be necessarily iconic gestures but more like pragmatic gestures as a part of the bilingual strategy (e.g., pointing gestures that do not have concrete referents).

Research argues that when the mapping a word to a concept gets difficult such as when expressing spatial metaphors (Akbuğa & Göksun, 2023) or describing emotions (Özder et al., 2022), individuals are more likely to produce representational gestures in their L2. Özder and colleagues (2022) demonstrated that bilingual adults frequently gesture when they talk about emotions, particularly in their L2. These findings suggest that bilinguals might not only get help from representational gestures to compensate for low language proficiency but also that gesturing facilitates cognitive processes in L2 where cognitive load and mapping cost might be high.

Gestures might also help communication in L1 and L2. In a study conducted by Zhao (2006), Chinese bilinguals' gesture frequencies were recorded in a task where the task was to simply hold a conversation with native Chinese speakers in their L1 (Chinese) and native English speakers in their L2 (English). The results supported the argument that individuals gesture more in their L2 (Hadar et al., 2001). Follow-up interviews revealed participants'

concerns over the quality of their communication with native English speakers and that participants sometimes gestured more purposefully while speaking to them. Participants also voiced their observation, which they gained through living in an English-speaking environment, that by comparison to themselves, English speakers gestured more.

It should, however, be noted that gesturing more frequently or more efficiently is not employed by all groups of L2 speakers equally well to facilitate communication. In a study conducted by Gregersen et al. (2009), participants were tasked with playing a role-playing game in dyads, where one participant played the role of a crime witness, and the other, a police officer, interviewed the witness. The conversation dyads were held in Spanish (L1) and English (L2). Higher L2 proficiency speakers used gestures that enhanced the meaning of their speech more frequently, whereas low proficiency L2 speakers opted more for speech-unrelated gestures. This pattern of results, in return, supports the view that gestures do not always come about as a compensation strategy in L2, where access to lexical items is challenging. There might be more to the functions of gestures than what the lexical retrieval hypothesis predicts.

To sum up, gesturing might help bilinguals facilitate cognitive processes and speak, particularly in their L2. Similarly, communicative intentions might motivate bilinguals' gesture use, and it might be challenging to differentiate gestures' self-oriented and communicative functions. As gestures are closely related to speech (Kita & Özyürek, 2003), L2 proficiency might play a role in the way that gestures are produced across L1 and L2.

4.1.5. Speech fluency and gesture

Speaking a second language is characterized by lexical, phonological, and grammatical difficulties. As a result, gestures might come into play to compensate for such difficulties and pave the way for more fluent speech. Gullberg (2011) argues that gestures are good candidates for making up for speech difficulties for three main reasons. First, gestures have expressive power, and they might fill semantic gaps in the language (McNeill, 1992). Moreover, to be able

to compensate for speech difficulties, gestures should be closely related to speech, and the current frameworks already indicate such a close relationship between the two modalities (e.g., Kita et al., 2017; Krauss, 1998; McNeill, 1992). Similarly, a gesture is a tool that is relevant to addressees. Regardless of whether they are intentionally produced by speakers, research indicated that individuals process both verbal and gestural information as a part of language comprehension (Kelly et al., 2010).

In line with these arguments, research indicated that people with aphasia produce semantically rich gestures compared to neurotypical individuals, suggesting that gestures might compensate for speech difficulties when needed (Akhavan et al., 2018; Cocks et al., 2013). Similarly, people who suffer left hemisphere brain injury are found to produce more gestures when they face a word-retrieval issue compared to when they are fluent (Göksun et al., 2013, 2015). Studies have also shown that people are more disfluent when their hand use is experimentally restricted (Morsella & Krauss, 2004; Rauscher et al., 1996; but also see Avcı et al., 2023).

Research also argues for a temporal relation between gesture and speech (McNeill, 1992; Seyfeddinipur, 2006; Seyfeddinipur & Kita, 2001). However, studies that focus on disfluency and gesture co-occurrences are limited in number (Akhavan et al., 2016; Arslan & Göksun, 2022; Graziano & Gullberg, 2018; Kısa et al., 2022; Seyfeddinipur, 2006). In a recent study, Arslan and Göksun (2022) compared younger and older Turkish-speaking adults' speech disfluency and representational gesture co-occurrences. While determining co-occurrences, they held a conservative method by including cases where gestures were slightly before or after a speech disfluency. They found that the proportion of disfluencies that co-occurred with gestures among all disfluencies was higher in the younger age group. Yet, the proportion of gestures that co-occurred with a disfluency among all gestures was similar between younger and older adults. The authors argued that although people might get help from gestures when

they have speech problems, gestures' self-oriented functions are not limited to facilitating speech. Additionally, even when gestures accompany disfluent speech, they may not necessarily be performed to make up for speech difficulties, as it is challenging to differentiate gestures' cognitive and communicative functions in neurotypical adults.

Yet, more frequent gesture use in fluent than disfluent speech is mainly based on L1 (Arslan & Göksun, 2022; Graziano & Gullberg, 2018). In a recent study, Graziano and Gullberg (2018) demonstrated that compared to L1 speakers, adults speaking L2 were more likely to produce representational gestures when their speech is disfluent. This finding suggests that various factors such as language proficiency and practice might come into play when it comes to understanding the role of gestures as a compensation mechanism for disfluent speech.

4.1.6. Present Study

This study investigates gesture production and speech disfluency across Turkish-English bilinguals' L1-Turkish and L2-English. By using a narrative retelling task, we examine the extent to which bilinguals produce representational gestures when they are disfluent across their two languages. We also measure L2-English proficiency as a control variable since insufficient knowledge of L2 is a prominent source of speech disfluency (Bergmann et al., 2015).

In line with the cognitive load of speaking L2 (Bergmann et al., 2015; Felker et al., 2019), we expect bilinguals to be more disfluent in L2 than in L1. Moreover, we expect participants to use more representational gestures in their L2 compared to L1 to lighten the cognitive load (Gullberg, 2011) and facilitate the word retrieval process (Hadar et al., 2001).

Last, based on the coding method used in Arslan and Göksun (2022), we hypothesize that representational gestures are more likely to co-occur with disfluencies in L2 than in L1 to ease the word retrieval and the speech production process, even when L2 proficiency is controlled.

4.2. Methods

4.2.1. Participants

We determined the sample size as 27 using the G*power tool (Faul et al., 2009), with an alpha level of 0.05, a medium effect size of (d) 0.50, and a power of 0.80. A total of 30 Turkish-English bilinguals participated in the study through online Zoom Video Conferencing. We excluded three participants whose hands were not fully visible during the meeting, and we carried out the main analyses based on remaining 27 participants (15 female). Participants were young adults with similar education levels between 20 and 33 years old ($M_{age} = 23.19$, $SD = 3.31$). Their L1 was Turkish, and their L2 was English. Participants received 1-course credit for their participation. The study was approved by the Koç University Institutional Review Board committee, with the approval number 2020.462.IRB3.181. We obtained participants' informed consent prior to the study.

Our participants' age of L2-English acquisition ranged from 4 to 12 ($M = 8.44$, $SD = 2.28$). When asked about their age of fluently speaking L2-English, participants' answers ranged from age 9 to 21 ($M = 14.89$, $SD = 3.31$). All participants indicated that they had an L2 experience in a family environment, a country, or a workplace, where English was frequently spoken. They also reported their English speaking competence as on average around 80 out of 100 ($M = 80.33$, $SD = 13.98$). The results of the Oxford Quick Placement Test indicated that participants' L2 proficiency scores ranged from 37 to 57 out of 60 ($M = 46.27$, $SD = 6.78$), suggesting that they have good L2 knowledge.

4.2.2. Materials

Oxford Quick Placement Test (OQRT). The Oxford Quick Placement Test (OQRT) (Syndicate, 2001) was applied to assess the English skills of the participants. All questions

were in English and weighted equally. There were 60 questions in total, and all of them were multiple-choice. The maximum score on the test was 60.

The Language Experience and Proficiency Questionnaire (LEAP-Q). We used LEAP-Q (Marian et al., 2007) to obtain participants' information about the number of spoken languages, their age of acquisition, exposure to the language, and relative proficiency levels. We eliminated the questions about accent and immigration history since they were not relevant to the current study.

Cartoon videos. Six short videos of Tom and Jerry cartoons were used to elicit narrative samples. The videos were mainly about cartoon characters chasing each other, moving from one location to another, and encountering some objects or other characters. There was no verbal content in the videos, which were similar in content. The duration of each cartoon video was around 1 minute.

4.2.3. Procedure

Upon their enrollment in the experiment, a Qualtrics survey link was sent to the participants' email addresses. The survey link contained the informed consent form, demographics form, LEAP-Q, and the Oxford Quick Replacement Test. Participants who completed the survey were contacted in three days to set a date for the second session.

The second session was conducted in an online setting using Zoom Video Conferencing. The experimenter first introduced the task and asked participants if they had any questions. She also mentioned that it was not an exam and participants should not have concerned themselves with making mistakes. The experimenter made sure that participants were seated in the view of the camera and that their upper bodies could be seen. While doing so, the experimenter did not mention anything related to hands or gestures. Each participant watched six Tom and Jerry cartoon clips. After each clip, they were instructed to narrate what

they had watched. They retold half of the cartoons in Turkish and the other half in English. Participants were randomly assigned to a condition where they narrated the first three cartoons either in Turkish or in English. The order of narrative language (i.e., Turkish or English), the order of the cartoon clips, and the combination of cartoon clips and narrative language were all counterbalanced. All sessions were carried out by a single experimenter and recorded with the consent of the participants for transcription and coding. At the end of the experiment, the experimenter asked participants whether they had any guesses about what we intended to measure and then provided a debriefing to participants regarding the purpose of the study.

The data and the materials are available on the Open Science Framework (OSF): [<https://osf.io/d6ngm/>].

4.2.4. Coding

ELAN software (version 6.2) (Lausberg & Sloetjes, 2009) was used to transcribe speech and code gestures, disfluencies, and their co-occurrences.

4.2.4.2. *Speech and Disfluency*

Bilingual research assistants transcribed speech, including all disfluencies for both languages (Turkish and English). As forms of disfluency, a trained bilingual research assistant identified and coded silent pauses (silent breaks within sentences, which are longer than 200 milliseconds), filled pauses (e.g., “ee”, uhm”), repetitions (e.g., “the mouse-*mouse* escaped him”, “fare-*fare* ondan kaçtı.”), and repairs (e.g., “the cat-*dog* was running.”, “kedi-köpek koşuyordu.”) (Arslan & Göksun, 2022; Avcı et al., 2023; Bortfeld et al., 2001; MacLay & Osgood, 1959). Another trained bilingual research assistant coded 22% of the participants for reliability. We found high interrater agreement between the two coders for disfluency coding both for Turkish ($\kappa=.93$, $p < .001$) and in English ($\kappa=.89$, $p < .001$) narratives. The total disfluency rate was calculated by summing all disfluencies. The disfluency rate was calculated

by dividing the total number of disfluencies by the total word count separately for Turkish and English narratives.

4.2.4.2. *Gesture*

We coded iconic (e.g., flapping hands to describe a bird) and metaphoric gestures (e.g., leaving small space between two hands to refer to little time) under the representational gesture category, complying with the gesture-for-conceptualization hypothesis (Kita et al., 2017). A trained research assistant coded gestures for Turkish and English retelling conditions with the authors' supervision. Another trained assistant coded 22% of the participants for reliability purposes. The inter-rater agreement between the coders was high for both the Turkish ($\kappa=.89$, $p < .001$) and the English condition ($\kappa=.85$, $p < .001$). Gesture frequency was calculated by dividing the total number of gestures by the total word count separately for Turkish and English narratives.

4.2.4.3. *Gesture-Disfluency co-occurrences*

Since we are interested in gestures' self-oriented functions, we only included representational gestures in our gesture-disfluency co-occurrence analysis (Kita et al., 2017). Gesture-disfluency co-occurrences were coded by targeting only the stroke onset and offset of the representational gestures (Arslan & Göksun, 2022). In line with kinematic movement properties (Seyfeddinipur, 2006), we identified stroke onset as the initial starting point of gestures at which the object or action shape is observed in participants' hands. In a similar vein, we identified the onset and offset of each disfluency type. The onset of filled pauses was the starting point of the sound and the offset was the endpoint (e.g., for “umm”, starting point = “u” sound, endpoint = second “m” sound). For the repetitions, the second word, which is the repeated word (e.g., cat *cat* jump), was taken as the stroke, with the starting and ending points

as the onset and offset, respectively. We followed a similar pattern in repairs, that is, the second fixing word was taken as a stroke. As an example, for the sentence “The bir-cat jumped from the fence.” the beginning of the word “cat” was taken as the onset and the ending point as the offset of a repair. A trained assistant and researcher coded the strokes of disfluencies and gestures for all conditions. We acquired intraclass correlation coefficients (ICC) for both gesture and disfluency from two coding attempts. Results indicated that there was high interrater reliability for the onset (ICC = 0.91, $p < .001$) and the offset (ICC = 0.89, $p < .001$) of gestures. Similar results were also found for speech disfluencies at onset (ICC = 0.89, $p < .001$) and at offset (ICC = 0.90, $p < .001$) as well.

As we focus on Turkish-English bilinguals, we identified and coded co-occurrences based on Arslan and Göksun (2022), which investigated gesture-disfluency co-occurrence in native Turkish speakers. As co-occurrences, we considered instances where a representational gesture stroke offset was timed up to 500 ms before the start of a speech disfluency, as well as instances where the start of the representational gesture stroke onset was up to 500 ms after the end of a speech disfluency offset. We calculated the co-occurrence rate both from disfluency and gesture perspectives, complying with Arslan and Göksun (2022). From the disfluency perspective, we divided the total number of disfluencies that co-occurred with a gesture by the total number of disfluencies. Similarly, from the gesture perspective, we divided the total number of gestures that co-occurred with a speech disfluency by the total number of gestures.

4.3. Results

4.3.1. Speech and Disfluency

A paired samples *t*-test indicated that individuals produced more words in English ($M = 420.59$, $SD = 223.89$) than in Turkish ($M = 310.96$, $SD = 161.09$), $t(26) = -4.16$, $p < .001$, $d = 0.80$. We found that participants were more disfluent in English ($M = 0.36 = 0.12$) than in

Turkish ($M = 0.20$, $SD = 0.05$), $t(26) = -7.14$, $p < .001$, $d = 1.37$. We carried out repeated measures ANCOVA by taking language (Turkish and English) as the within-subject variable, and the disfluency rates as the dependent variable, controlling for English proficiency. Results indicated a main effect of language, $F(1,25) = 24.90$, $p < .001$, $\eta^2 = .499$, suggesting that individuals were still more disfluent in their L2-English than L1-Turkish. The interaction between language type and English proficiency was also significant, $F(1,25) = 12.53$, $p = .002$, $\eta^2 = .334$. We found that English proficiency was a significant covariate, $F(1,25) = 7.68$, $p = .010$, $\eta^2 = .235$. Bivariate correlation analysis indicated that English proficiency score was significantly negatively correlated with disfluency rate in English, $r(27) = -0.57$, $p = .002$.

4.3.2. Gesture

Results of a paired samples t -test indicated that representational gesture frequency was higher in English ($M = 0.14$, $SD = 0.07$) than in Turkish ($M = 0.10$, $SD = 0.06$), $t(26) = -3.36$, $p = .001$, $d = 0.65$. When we controlled for English proficiency, results of repeated measures ANCOVA indicated that English proficiency score was a significant covariate of representational gesture use, $F(1,25) = 8.03$, $p = .009$, $\eta^2 = .243$. Moreover, representational gesture frequency turned out to be comparable across two languages, as indicated by a nonsignificant main effect of language, $F(1,25) = .23$, $p = .634$, $\eta^2 = .009$. The interaction between language type and English proficiency was not significant, either, $F(1,25) = 1.19$, $p = .286$, $\eta^2 = .045$. Bivariate correlation analysis showed that English proficiency score was significantly positively correlated with representational gesture frequency in English, $r(27) = 0.50$, $p = .009$.

4.3.3. Gesture-disfluency co-occurrences

4.3.3.1. Disfluency perspective

We carried out a paired samples t -test to examine whether gesture-disfluency co-occurrence from the disfluency perspective. We found that the proportion of disfluencies that

co-occurred with a gesture was higher in English ($M = 0.45$, $SD = 0.21$) than in Turkish ($M = 0.27$, $SD = 0.17$), $t(26) = -4.50$, $p < .001$, $d = 0.87$. When we controlled for English proficiency, the results of ANCOVA showed that the main effect of language type disappeared, $F(1,25) = 0.75$, $p = .396$, $\eta^2 = .029$. English proficiency score was a significant covariate, $F(1,25) = 8.00$, $p = .009$, $\eta^2 = .242$. There was no interaction between the language type and the English proficiency score, $F(1,25) = 2.92$, $p = .100$, $\eta^2 = .105$. Bivariate correlation analysis demonstrated that the English proficiency score was significantly positively correlated with the proportion of disfluencies that co-occur with a gesture in English, $r(27) = 0.54$, $p = .004$.

4.3.3.2. Gesture perspective

We repeated the same analyses for the proportion of gestures that occurred with a disfluency. Results showed that the proportion of gestures that occurred with a disfluency was significantly higher in English ($M = .88$, $SD = .16$) than in Turkish ($M = .51$, $SD = .20$), $t(25) = -7.95$, $p < .001$, $d = 1.56$. When we controlled for English proficiency, results of ANCOVA showed that the main effect of language was still significant, $F(1,24) = 5.33$, $p = .030$, $\eta^2 = .182$. English proficiency score was not a significant covariate, $F(1,24) = 1.34$, $p = .258$, $\eta^2 = .053$. The interaction between the language type and the English proficiency score was not significant, $F(1,24) = 1.40$, $p = .249$, $\eta^2 = .055$.

4.4. Discussion

The current study investigated co-speech gestures and speech disfluency in Turkish-English bilingual adults' L1-Turkish and L2-English, using a narrative retelling task. We examined to which extent speech disfluencies and gestures co-occur in L1 and L2. Our results indicated that the overall disfluency rate was higher in English than in Turkish, even after L2 proficiency was controlled. Results also showed that although representational gesture frequency was higher in English than in Turkish, when L2 proficiency was controlled, it was comparable across L1 and L2. We found that L2 proficiency was positively correlated with

representational gesture frequency in L2. Our results also yielded that the proportion of disfluencies that co-occurred with a representational gesture was higher in English than in Turkish. Those who have higher L2 proficiency were more likely to use gestures when they were disfluent. Last, from a gesture perspective, the proportion of representational gestures that co-occurred with a disfluency was higher in English regardless of L2 proficiency.

4.4.1. Disfluency

As we expected, individuals were more likely to be disfluent in L2-English than L1-Turkish. Although L2 proficiency was positively correlated with fluency in the given language, the fact that this difference still existed after controlling the English proficiency score suggests that language proficiency might not be the only factor paving the way for disfluent speech. Research indeed indicates that people might be more disfluent in L2 than in L1 even when they are equally proficient in both languages (Bergman et al., 2015). In one study conducted by Bergmann and colleagues (2015), to be tasked with re-telling movies scenes in German, German-speaking monolinguals were compared to two groups of bilinguals: German-learners (with their L1 being English) and German-attriters (whose L1 was German, yet haven't gotten exposed to it for a long time, due to immigrating to an L2-English environment). The results yielded that both bilingual groups were more disfluent than the monolingual group. However, the difference in disfluency rates was attributed to the co-activation of languages instead of language proficiency because all three groups were equally proficient in the language the task was performed in (German). Rather, interference caused by the co-activation of languages seemed to be the explanatory factor since both bilingual groups had two highly active and uninhabitable languages activated at the same time (in the case of attriters, it was their L2, English; and for German-learners, it was their L1, English). These results suggest that language proficiency alone is not the sole factor explaining bilingual disfluency and that language co-activation is also effective.

Arslan and colleagues (2023) demonstrated that 5- and 7- year-old bilingual children revealed similar overall disfluency rates across their L1-Turkish and L2-English. However, they were more disfluent in Turkish compared to their Turkish monolingual peers. It is important to note that bilingual children in this study were similarly proficient in both of their languages, and they were exposed to L2-English latest by age 3. Therefore, it might be tricky to compare bilingual children's disfluency rate with the one of adult bilinguals, as children have not completed their acquisition process in both languages, whereas healthy adults are supposed to be fully proficient, at least in their L1.

Similar to Bergmann et al. (2015), albeit our adult participants seemed to have good L2 proficiency, they started learning English after being exposed to Turkish (i.e., between 4 and 12 years of age). Since L1 representations are stronger in nature, they are automatically activated and might interfere with L2 performance (Costa et al., 1999). Our results are in line with Bergmann et al. (2015) in the sense that our participants were still more disfluent in L2-English even when the effect of L2 proficiency was eliminated. Therefore, we argue that the higher disfluency rate we observed in English might be explained by the perpetual co-activation of L1 and L2, as L1-Turkish might have interfered with L2-English.

It is also important to acknowledge that disfluencies might stem from communicative intentions (Bortfeld et al., 2001; Brennan & Williams, 1995; Corley & Stewart, 2008). For instance, filled pauses used by a speaker might signal their intention to continue speaking and might aid turn-taking between the interlocutors (Corley & Stewart, 2008). In that sense, disfluencies might reflect speakers' metacognitive states (Brennan & Williams, 1995) and might be interpreted as communication-related signals (Brennan & Schober, 2001). Therefore, the higher disfluency rate in English than in Turkish might be at least partially accounted for by participants' intention to maintain the conversational floor in their L2, where they might be struggling with inhibiting L1 representations.

4.4.2. Gesture

As we expected, our results yielded that adults produced more representational gestures in their L2 than L1. This finding is in line with Hadar et al. (2001), which found more frequent representational gesture production in L2. Similarly, Arslan and colleagues (2023) suggested that representational gesture frequency was higher in L2 than L1 narrative retelling, even in the case of children who had not mastered their language skills. Five- and 7-year-old children produced more representational gestures in English than in Turkish, although English was their second language. In the present study, the representational gesture frequency was comparable between the two languages after we controlled for L2 proficiency. At first look, our findings might support the Lexical Retrieval Hypothesis and imply that gestures compensate for weak language skills by facilitating word retrieval and the speech production process. Yet, we found that L2 proficiency was positively associated with the representational gesture frequency in L2-English, contrasting with the argument that gestures mainly come into play to compensate for weak language skills. That is, the more proficient our participants were in English, the more representational gestures they used in that language.

This finding points out the link between gesture and speech mechanisms (Kita & Özyürek, 2003), suggesting that the two systems evolve in an interrelated fashion. Representational gestures are complex forms of gestures that appear later in language development (Gullberg, 2013; Nicoladis, 2007). In line with this argument, developmental research indicates that the more complex sentences children form, the more representational gestures they produce (e.g., Capirci et al., 2005; for a review, see Gullberg, 2013). In another study with adults, Gregersen and colleagues (2009) also found that the gestures of proficient L2-Spanish speakers were more likely to be meaning-enhancing compared to L2-Spanish speakers who had lower levels of proficiency. These findings comply with our results with

regard to the positive correlation between L2-English proficiency and representational gesture use in English narratives.

One could then ask why participants used more representational gestures in English than in Turkish in the first place if it was not to compensate for weak L2 skills. There might be alternative explanations for this pattern. First, speaking L2 is a cognitively demanding task, particularly for late bilinguals (Gollan et al., 2005; Ortega, 2003). Although people might benefit from gestures to decrease such load, there are individual and group differences in benefiting from gestures (for a review, Özer & Göksun, 2020). For instance, Avcı and colleagues (2023) demonstrated that individuals' likelihood of being disfluent when their hand gestures were restricted depended on to which extent they gestured in spontaneous speech. It is possible that some people do not frequently gesture and asking them to gesture might create an extra cognitive load on them. In the case of our study, using representational gestures more in English than in Turkish is in line with the cognitive load of speaking L2. Yet, producing complex gesture forms, i.e., representational gestures, in L2 might be cognitively burdensome for those who have low L2 proficiency. On the other hand, those who have high proficiency might be more likely to use representational gestures in a way that complements speech and facilitates cognitive processes. Second and less plausible, the experimenter was a native Turkish speaker who provided instructions in Turkish and English. Participants did not exactly know the experimenter's English level. Therefore, overall, participants might have used more representational gestures in English to make sure that they are being understood and on the same page with the experimenter.

4.4.3. Gesture-disfluency co-occurrences

Our results yielded that the proportion of disfluencies that co-occurred with a gesture among all disfluencies was higher in participants' English than in Turkish narratives. Yet, when we controlled for L2-English proficiency, the results turned out to be comparable. The more

people were proficient in English, the more they gestured during disfluent speech segments. These findings are in line with our previous argument that those who have high L2 proficiency might be more likely to use representational gestures in a way that complements speech and facilitates cognitive processes. We argue that the fact that bilinguals' likelihood of disfluencies to be accompanied by gestures is higher in L2, particularly for those who have higher L2 proficiency, might stem from a combination of cognitive and communicative factors.

As suggested by the conflict-based theory (Nozari et al., 2011), errors in speech, observed in the form of disfluencies, are detected through a monitoring mechanism and repaired by responsible mechanisms upon receiving signals from the monitor. The anterior cingulate cortex (ACC) is closely associated with monitoring skills, suggesting that it monitors performance and evaluates whether top-down control is required (Botvinick et al., 2001). Detecting a high conflict where bottom-up processes cannot handle the situation, the ACC signals the prefrontal cortex (PFC) to exert top-down control. Therefore, the monitor functions as a bridge between automatic and controlled processes (Badre & Wagner, 2002; Braver et al., 2007). We suggest that speaking L2 might have required our participants to engage in a more controlled process to monitor whether they correctly express themselves. As a result of recruiting such proactive control strategies, people might detect their mistakes faster and produce gestures during such periods to complement speech and keep their listeners' attention. Using representational gestures at disfluent segments might signal that the speaker already knows the concept but has difficulty in putting it into words (Schwartz, 2006). The fact that this tendency is higher in those who have high L2 proficiency might suggest that in order for such a controlled mechanism to come into play, people first should have enough L2 proficiency to detect their own errors and have metacognitive awareness in that language.

We also examined co-occurrences from a gesture perspective. We found that the proportion of gestures that co-occurred with a disfluency among all gestures was higher in

English than in Turkish narratives. Unlike our findings from the disfluency perspective, we found that this difference remained even when the L2 proficiency score was controlled. One explanation for this situation would be that although gestures are closely related to speech (Kita & Özyürek, 2003), the L2 proficiency test primarily measures participants' proficiency in L2 speech, suggesting an indirect relation to gestures. Therefore, we might not have seen an effect of L2 proficiency on co-occurrences from the gesture perspective. Yet, although we compare L1 and L2, it is also important to focus on tendencies within a language. In Turkish narratives, on average, 51% of the gestures were accompanied by a disfluency, higher than Arslan and Göksun (2022) suggested (38% for young adults). There might be two reasons for this pattern. First, Arslan and Göksun (2022) did not code silent pauses since they argued that identifying silent breaks was subjective when comparing younger and older adults. Second, they used a picture description task where each picture was physically present in front of participants. Yet, in our study, we used a cartoon retelling task where participants were required to recall the scenario and produce a narrative accordingly. In line with such cognitive load argument, we found that this proportion drastically increased in English narratives, being 88% on average, suggesting that speaking L2 might have created an additional cognitive load on speakers.

Disfluencies are frequently observed when people describe difficult stimuli (e.g., Bortfeld et al., 2001), and representational gestures come into play to decrease cognitive load (Arslan & Göksun, 2021; Kita & Davies, 2009; Melinger & Kita, 2007). Therefore, the fact that gestures and disfluencies co-occur in speech does not mean that gestures are produced to resolve disfluencies. Rather, there might be a spurious relationship between the two phenomena in the sense that the cognitive load or communicative needs individuals face paves the way for both disfluency and gesture production. Arslan and Göksun (2022) also acknowledged that the temporal co-occurrence of gestures and disfluencies would be only indirect evidence of gestures' self-oriented functions. Future research should investigate disfluency-gesture co-

occurrences via neural oscillations to understand the temporal relationship between gestures, speech disfluency, and error detection.

4.5. Conclusion

Our study suggests that language proficiency might be an important factor in explaining a higher disfluency rate in L2 than in L1. Yet, our results indicate that L2 proficiency alone is not sufficient to explain the disfluency rate in speech. It is because disfluencies might stem from a combination of cognitive and communicative factors. Contrasting with the Lexical Retrieval Hypothesis, our findings suggest that representational gestures do not compensate for weak language skills. Rather, as Gullberg (2013) proposes, the more proficient people are in a language, the more representational gestures they produce. We also argue that people should have enough L2 proficiency to detect their own errors and thus, have metacognitive awareness to use gestures as a complementary tool when they are disfluent. From a gesture perspective, gestures might come into play where there is cognitive load, particularly in L2. Yet, gesture-disfluency co-occurrences in discourse are only indirect evidence of gestures' self-oriented functions and do not rule out the possibility of both phenomena resulting from a third factor.

CHAPTER 5

GENERAL DISCUSSION

5.1. Summary and the discussion of the main findings

This dissertation investigated multimodal language production within and across different groups, with a specific focus on gesture production and speech disfluency. I demonstrated that people might differ in terms of the likelihood of gesturing and being disfluent depending on specific characteristics such as age, bilingualism status, or whether one speaks in their first or second language. The findings revealed by the three studies indicated a close relationship between speech and gesture mechanisms, suggesting that although representational gestures have self-oriented functions, they might not always be to compensate for weak language skills. However, taking the communicative aspects of speech disfluency and gesture into consideration makes the gesture-speech interaction process more complicated than it is thought.

In **Chapter 2**, I studied speech disfluency and co-speech gestures produced by younger and older adults through a painting description task. Results indicated that although the two age groups revealed comparable overall speech disfluency and gesture rates, they differed in terms of how their disfluencies and gestures were distributed across specific categories. In particular, younger adults produced higher proportions of filled pauses, while older adults produced higher proportions of repetitions. This finding is in line with the argument that repetition is commonly observed where task demands exceed available cognitive resources (Bortfeld et al., 2001), and it is frequently observed in young children's speech (Haynes & Hood, 1978). Older participants may have faced relatively greater challenges in engaging with a task that required them to analyze and describe both concrete and abstract art compared to the younger individuals.

I also found that the proportion of representational gestures was higher for the younger than the older age group. In contrast, the proportion of deictic gestures was higher for older than younger adults. This finding supports the argument proposed by Arslan and Göksun

(2021) that as people get older, they start producing fewer representational gestures, complying with decreasing mental imagery skills. Instead, they may start using more deictic gestures, which is a simpler gesture form compared to representational gestures, as observed in children's discourse (Özçalışkan et al., 2016). A deictic gesture is the first gesture form to appear in children's language, while representational gesture use becomes common when children reach a certain level of language proficiency (Iverson & Goldin-Meadow, 2005).

This chapter also addressed whether and to what extent representational gestures temporally co-occur with disfluencies across the two age groups. Results showed that the proportion of speech disfluencies that occur with a gesture was significantly higher for younger than older adults. However, the two age groups were comparable in terms of the proportion of gestures that were accompanied by a speech disfluency. These findings suggest that gestures do not compensate for weak language skills. Instead, younger adults' language production system might be better at benefiting gesture modality to resolve temporary problems in speech. Yet, gestures serve both self-oriented and communicative purposes, as indicated by Kita et al. (2017). As a result, from a gesture perspective, it might become challenging to distinguish between these motivations and understand gestures' facilitative roles in speech. We acknowledge that it is only indirect evidence of gestures' self-oriented functions. Moreover, we cannot rule out the possibility that gestures aid speech even in the absence of overt errors (such as speech disfluencies), and we may not observe any evidence of such assistance in a normally functioning language system.

Chapter 2 indicated that older adults might reveal similar trends with children in terms of speech disfluency and gesture use. While older adults' language skills decay as they age (Kemper et al., 2001; Burke & Shafto, 2004), children represent an emerging language system (Iverson & Goldin-Meadow, 2005). Moreover, bilingualism has become a prominent feature of today's world, and research argues its effects on multimodal language development (for a

review, see Nicoladis, 2007). Yet, although research separately targeted speech disfluency (e.g., Bedore et al., 2006; Byrd et al., 2015; Fiestas et al., 2005) and gesture production (Mayberry & Nicoladis, 2000; Nicoladis et al., 1999; Yan & Nicoladis, 2009) among children, there is a gap in the literature on how these two modalities interact in this age group. Therefore, as a next step, I investigated bilingual and monolingual children's multimodal language production based on a lifespan development perspective.

In **Chapter 3**, I focused on 5- and 7-year-old Turkish-English bilingual and Turkish monolingual children's speech disfluency and gesture production and to which extent the two modalities are related in narrative production. Results indicated that bilinguals were more disfluent than monolinguals, particularly for silent and filled pauses. Gesture use was comparable across language and age groups, except for iconic gestures. Monolinguals produced more iconic gestures than bilinguals. Based on these results, I argued that disfluency and gesture patterns in children's speech reflect their language development process. These findings are in line with previous research that repetitions are immature forms of disfluency (Bortfeld et al., 2001), and they are replaced by filled pauses later in development as children's metacognitive skills increase (Haynes & Hood, 1978). Similarly, the fact that I observed iconic gestures more frequently in older children and monolinguals is in line with the findings of the previous chapter, suggesting that representational gestures do not compensate for weak language skills. Children should reach certain language competency to be able to produce such complex gesture forms (Goldin-Meadow, 1998; Iverson & Goldin-Meadow, 2005). Moreover, representational gesture frequency alone was not significantly associated with the disfluency rate in Turkish narratives. Instead, children's overall gesture frequency predicted disfluency rates only in Turkish, controlling for language proficiency and linguistic complexity. These findings together indicate that representational gestures' self-oriented functions might not be fully reflected in an emerging language system. In children's discourse, different gesture types

might be orchestrated in the multimodal system, which contributes to the narrative performance.

Chapter 3 also investigated and compared bilingual children's narrative production across their two languages. Results revealed that bilinguals used silent pauses and repetitions more frequently in English than in Turkish. Moreover, bilingual children produced more iconic gestures in English than in Turkish. It is important to note that the bilingual participants of this study were proficient both in Turkish and English, although English was their L2. Therefore, I argue that bilinguals' use of more iconic gestures in their L2 does not stem from incompetency in this language but rather from less practice. Based on these findings, as a next step, I focused on bilingual adults' multimodal language production across their L1 and L2 to dissect the disfluency-gesture interaction in a relatively mature language system.

In **Chapter 4**, I inquired disfluency and representational gesture rates of Turkish-English bilingual adults and whether representational gestures temporally co-occurred with disfluencies across their two languages. Results showed that bilingual adults were more disfluent in L2-English even after L2 language proficiency was controlled. This finding is in line with the argument that disfluencies might be part of communication (Corley & Stewart, 2008). As a result, it might be tricky to dissect disfluencies' cognitive and communicative motivations. Although bilinguals produced more representational gestures in their L2, this difference disappeared after L2 proficiency was controlled. The more participants were proficient in English, the more they used representational gestures in that language. Similarly, disfluency-gesture co-occurrences were more common in English. L2 proficiency was positively correlated with the likelihood of a disfluency being accompanied by a gesture. These findings comply with the argument that representational gestures do not compensate for weak language skills (Gullberg, 2013). Instead, people might use more representational gestures when they are disfluent if they have enough competency to detect their mistakes, which

suggests a metacognitive background for gesture production. A similar situation is also frequently observed in tip-of-the-tongue states, as individuals are likely to gesture when they know a word but cannot retrieve it at that moment (Schwartz, 2006).

This dissertation contributes to the field by demonstrating the necessity of incorporating population characteristics into the picture to better understand multimodal language production. Most gesture frameworks mainly focus on young adults to explain multimodal language production, and they do not discuss the possible effects of bilingualism status (Hostetter & Alibali, 2019; Kita et al., 2017; Krauss et al., 2000). Yet, the findings of this dissertation indicate that such focus might be limited while studying populations with emerging or decaying language systems. Moreover, this thesis draws attention to the use of deictic gestures and verbal repetitions across the lifespan, indicating a reverse U-shaped trend in using these forms from childhood to late adulthood. I also showed that bilingual and monolingual children differed in the use of specific disfluency and gesture types, which might shed light on their language development and communication strategies.

It is likely that gestures' self-oriented functions are reflected differently across and within populations, and representational gestures are less likely to be produced by those who have weak language skills. As a result, this dissertation theoretically informs the gesture-for-conceptualization hypothesis (Kita et al., 2017), suggesting a research avenue for representational gestures' self-oriented functions across age groups, bilingualism status and one's L1 versus L2. Although gestures do not necessarily compensate for speech difficulties, there are still cases where gestures help individuals to resolve disfluencies. Individual differences in cognitive resources might interact with gestures' role as a compensation tool (for a review, see Özer & Göksun, 2020), pointing out the complex nature of gesture-speech interaction. Focusing on specific cases where speech disfluency and gestures co-occur and

considering individual differences might bring insight into the circumstances under which gestures come forward as compensation tools.

Moreover, understanding disfluency and gesture patterns within and across populations brings pragmatic implications in diagnosing atypical language development across the lifespan. Taking individual and population related differences in language into account would yield improved diagnosis of those with speech problems. These findings also inform the intervention techniques designed for individuals with speech impairment.

5.2. Future directions

There are questions that remained unanswered in this dissertation. First, putting their self-oriented functions aside, people also gesture to communicate (Kita et al., 2017). The findings of this dissertation provide only indirect evidence of gestures' self-oriented functions in relation to speech disfluency. Individuals might also gesture during disfluent segments to keep their listeners' attention and communicate (Kısa et al., 2020). How would research dissect gestures' self- and other-oriented motivations? Research already indicates that experimentally restricting gesture use might yield different effects on individuals depending on how frequently they gesture in spontaneous speech (Avcı et al., 2022). Future research should experimentally manipulate hand use and the presence of an addressee in combination to investigate whether restricting gesture use yields different effects on disfluency rate depending on an addressee. The same question holds for speech disfluency. It is often argued that disfluencies stem from a combination of cognitive load and communicative intentions (Corley & Stewart, 2008). Such a design would provide insight into whether certain disfluency types play a more prominent role in a context where the presence of an addressee is emphasized.

Second, it is argued that abstract deictics might also have self-oriented functions when one is not confident with his/her language skills (Azar et al., 2020; Nicoladis, 2007). These gestures are different from concrete deictics because they are mostly used in a narrative context,

and they are not produced to point at something physically present at that moment (Avcı et al., 2023). Therefore, further research is needed to differentiate between concrete and abstract deictics to understand whether abstract deictics are similar to iconic gestures in terms of having self-oriented functions.

Last, as mentioned earlier, the fact that gestures and disfluencies co-occur in speech does not mean that gestures are produced to resolve disfluencies. Do people gesture right upon detecting an error in their speech? Such errors might not always be observed in overt speech, as the inner loop detects some errors in the planning phase before it is fully articulated (Nozari et al., 2011). As a result, people can still benefit from gestures to think and speak even though they do not reveal overt speech errors. Future research should investigate disfluency-gesture co-occurrences via neural oscillations to understand the temporal relationship among gesture, speech disfluency, and error detection.

5.3. Conclusion

Individuals produce gestures when they speak. The speech production process is not always flawless, and people become disfluent when they speak. The current dissertation investigated speech disfluency and gesture production and examined to which extent these two phenomena were related across different age groups, bilingualism status, and one's L1 versus L2. This dissertation provides insight into typical multimodal language production through an investigation of disfluency and gesture patterns within and across populations. Despite their self-oriented functions, representational gestures do not necessarily compensate for weak language skills, but rather, they are closely linked to language proficiency. As a result, representational gestures' self-oriented functions might not be clearly reflected in children and older adults, which represent emerging and decaying language systems, respectively. In such populations, different gesture types might be orchestrated in the multimodal system, contributing to fluency.

APPENDICES

APPENDIX A

Concrete paintings



Abstract paintings



REFERENCES

- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135(1), 50-68.
- Akbuğa, E., & Göksun, T. (2023). EXPRESS: The role of spatial words in the spatialization of time. *Quarterly Journal of Experimental Psychology*, 0(0).
- Akhavan, N., Goksun, T., & Nozari, N. (2016). Disfluency production in speech and gesture. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the Cognitive Science Society* (pp. 716–721). Austin, TX: Cognitive Science Society.
- Akhavan, N., Göksun, T., & Nozari, N. (2018). Integrity and function of gestures in aphasia. *Aphasiology*, 32(11), 1310-1335.
- Aktan-Erciyes, A. (2019). İkinci Dil Ediniminin Okul Öncesi ve Okul Çağı Çocuklarında Anlatı Becerilerinin Kurgusal ve Dilbilgisel Süreçlerine Olan Etkisi. *Psikoloji Çalışmaları*, 39(2), 369-399.
- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4), 307-331.
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15(6), 593-613.
- Appel, R. (1996). The lexicon in second language acquisition. In P. Jordens, & J. A. Lalleman (Eds.), *Investigating second language acquisition* (pp. 381–404). Berlin: Mouton de Gruyter.
- Arslan, B., Aktan-Erciyes, A., & Göksun, T. (2023). Multimodal language in bilingual and monolingual children: Gesture production and speech disfluency. *Bilingualism: Language and Cognition*, 1-13. <https://doi.org/10.1017/S1366728923000196>

- Arslan, B., & Göksun, T. (2021). Ageing, working memory, and mental imagery: Understanding gestural communication in younger and older adults. *Quarterly Journal of Experimental Psychology*, 74(1), 29-44.
- Arslan, B., & Göksun, T. (2022). Aging, Gesture Production, and Disfluency in Speech: A Comparison of Younger and Older Adults. *Cognitive Science*, 46(2), e13098.
- Astington, J. W., & Baird, J. A. (2005). *Why language matters for theory of mind*. New York: Oxford University Press.
- Avcı, C., Arslan, B., & Göksun, T. (2022). Gesture and Speech Disfluency in Narrative Context: Disfluency Rates in Spontaneous, Restricted, and Encouraged Gesture Conditions. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 1912-1917). Austin, TX: Cognitive Science Society.
- Azar, Z., Backus, A., & Özyürek, A. (2019). General-and language-specific factors influence reference tracking in speech and gesture in discourse. *Discourse Processes*, 56(7), 553-574.
- Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13), 2883-2901.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and Communication in Infancy*. New York: Academic Press.
- Bedore, L. M., Fiestas, C. E., Peña, E. D., & Nagy, V. J. (2006). Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. *Bilingualism: Language and Cognition*, 9(3), 233-247.
- Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, 161, 25-35.
<https://doi.org/10.1016/j.actpsy.2015.07.015>

- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: L. Erlbaum.
- Berument, S. K., & Güven, A. G. (2010). Turkish Expressive and Receptive Language Test: Receptive Vocabulary Sub-Scale (TIFALDI-RT). Turkish Psychological Society.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12(1), 3-11.
- Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language*, 109(2-3), 93-100.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525-531.
- Blacfkmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39(3), 173-194.
- Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition. Vol. 11: Speech, Language, and Communication* (pp. 181–216). Orlando, FL: Academic Press.
- Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, 3, 395–421. doi:10.3758/BF03214545
- Bóna, J. (2011). Disfluencies in the spontaneous speech of various age groups: Data from Hungarian. *Govor*, 28(2), 95-115.
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8(3), 148-158.

- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123-147.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624.
- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., ... & Reed, B. R. (2001). Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, 130(4), 746-763.
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in working memory*, 75, 76-106.
- Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2), 274-296.
- Brennan, S. E., & Williams, M. (1995). The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*, 34(3), 383-398. <https://doi.org/10.1006/jmla.1995.1017>
- Brown, A. S., & Nix, L. A. (1996). Age-related changes in the tip-of-the-tongue experience. *The American Journal of Psychology*, 109, 79-91.
- Brundage, S. B., & Rowe, H. (2018). Rates of typical disfluency in the conversational speech of 30-month-old Spanish-English simultaneous bilinguals. *American Journal of Speech-Language Pathology*, 27(3S), 1287-1298.
- Burke, D. M., & College, P. (1997). Language, aging, and inhibitory deficits: Evaluation of a theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(6), P254-P264.

- Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults?. *Journal of Memory and Language*, 30(5), 542-579.
- Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science*, 13(1), 21-24.
- Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96, 168-174.
- Byrd, C. T., Bedore, L. M., & Ramos, D. (2015). The disfluent speech of bilingual Spanish–English children: Considerations for differential diagnosis of stuttering. *Language, Speech, and Hearing Services in Schools*, 46(1), 30-43.
- Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, 5(1-2), 155-177.
- Carias, S., & Ingram, D. (2006). Language and disfluency: Four case studies on Spanish–English bilingual children. *Journal of Multilingual Communication Disorders*, 4(2), 149-157.
- Cartmill, E. A., Demir, O. E., & Goldin-Meadow, S. (2012). Studying gesture. In E. Hoff (Ed.), *Research methods in child language: A practical guide*. Blackwell Publishing.
- Cavicchio, F., & Kita, S. (2013). Bilinguals switch gesture production parameters when they switch languages. In *Proceedings Tilburg Gesture Research Meeting (TIGeR) 2013*. Retrieved from <http://www.researchgate.net/publication/236899431>
- Christenfeld, N., Schachter, S., & Bilous, F. (1991). Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research*, 20(1), 1-10.
- Clark, H. H. (1994). Managing problems in speaking. *Speech Communication*, 15(3-4), 243-250.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*,

84(1), 73-111.

- Cocks, N., Dipper, L., Pritchard, M., & Morgan, G. (2013). The impact of impaired semantic knowledge on spontaneous iconic gesture production. *Aphasiology*, 27(9), 1050-1069.
- Cocks, N., Morgan, G., & Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture*, 11(1), 24-39.
- Cohen, R. L., & Borsoi, D. (1996). The role of gestures in description-communication: A cross-sectional study of aging. *Journal of Nonverbal Behavior*, 20(1), 45-63.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression /correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Cohen, G., & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, 4, 187-197.
- Colletta, J. M. (2009). Comparative analysis of children's narratives at different ages: A multimodal approach. *Gesture*, 9(1), 61-96.
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27(4), 594-610.
- Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology*, 45(5), P210-P214.
- Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, 2(4), 589-602.
- Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language & Cognitive Processes*, 14(4), 381-391. <http://doi.org/10.1080/016909699386275>
- De Ruiter, J. P., & de Beer, C. (2013). A critical evaluation of models of gesture and speech production for understanding gesture in aphasia. *Aphasiology*, 27(9), 1015-1030.

- DeJoy, D. A., & Gregory, H. H. (1985). The relationship between age and frequency of disfluency in preschool children. *Journal of Fluency Disorders*, 10(2), 107-122.
- Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: seeing gesture matters. *Developmental Psychology*, 50(3), 815.
- Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, 20(3), 245-257.
- Dumont, J. (2010). Testing the cognitive load hypothesis: Repair rates and usage in a bilingual community. *Studies in Hispanic and Lusophone Linguistics*, 3(2), 329-352.
- Dunn, L. M., & Dunn, D. M. (2015). *Peabody picture vocabulary test: PPVT 4*. Pearson.
- ELAN (Version 5.8) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition*, 38(5), 617-628.
- 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.
- Felker, E. R., Klockmann, H. E., & De Jong, N. H. (2019). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, 40(1), 111-136. doi:10.1017/S0142716418000474
- Feyereisen, P. (1987). Gestures and speech, interactions and separations: A reply to McNeill (1985). *Psychological Review*, 94, 493-498.
- Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, 23(2), 153-171.

- Fiestas, C. E., Bedore, L. M., Peña, E. D., Nagy, V. J., Cohen, J., & McAlister, K. T. (2005). Use of mazes in the narrative language samples of bilingual and monolingual 4-to 7-year old children. In *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 730-740). Somerville, MA: Cascadilla Press.
- Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. In *Proceedings of DISS, Gothenburg Papers in Theoretical Linguistics, Göteborg University, Sweden, 90*, 21–24.
- Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, 65(2), 161-175.
- Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, 29(9), 1083-1096.
- Gampe, A., Wermelinger, S., & Daum, M. M. (2019). Bilingual children adapt to the needs of their communication partners, monolinguals do not. *Child Development*, 90(1), 98-107.
- Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, 26(6), 817-831.
- Goldin-Meadow, S. (1998). The development of gesture and speech as an integrated system. *New Directions for Child Development*, (79), 29-42.
- Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press.
- Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In *Pointing* (pp. 93-116). Psychology Press.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12(6), 516-522.

- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33(7), 1220–1234. <https://doi.org/10.3758/BF03193224>
- Goodglass, H., & Kaplan, E. (1983). *The Boston Naming Test*. Philadelphia, PA: Lea and Febiger.
- Gordon, P. A., Luper, H. L., & Peterson, H. A. (1986). The effects of syntactic complexity on the occurrence of disfluencies in 5 year old nonstutterers. *Journal of Fluency Disorders*, 11(2), 151-164.
- Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2013). Naming and gesturing spatial relations: Evidence from focal brain-injured individuals. *Neuropsychologia*, 51(8), 1518–1527.
- Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (2015). Spontaneous gesture and spatial language: Evidence from focal brain-injury. *Brain and Language*, 150, 11-13.
- Göksun, T., Özer, D., & Akbıyık, S. (2022). Gesture in the aging brain. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 269–293). De Gruyter Mouton; American Psychological Association.
- Graziano, M., & Gullberg, M. (2013). Gesture production and speech fluency in competent speakers and language learners. In *Tilburg Gesture Research Meeting (TiGeR) 2013*. Tilburg University.
- Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, 9, 879.
- Gregersen, T., Olivares-Cuhat, G., & Storm, J. (2009). An examination of L1 and L2 gesture use: What role does proficiency play?. *The Modern Language Journal*, 93, 195–208.
- Gullberg, M. (1998). *Gesture as a communication Strategy in Second Language Discourse. A Study of Learners of French and Swedish*. Lund : Lund University Press.

- Gullberg, M. (2009). Reconstructing verb meaning in a second language: How English speakers of L2 Dutch talk and gesture about placement. *Annual Review of Cognitive Linguistics*, 7(1), 221-244. <https://doi.org/10.1075/arcl.7.09gul>
- Gullberg, M. (2011). Multilingual multimodality: Communicative difficulties and their solution in second-language use. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 137–151). Cambridge, UK: Cambridge University Press.
- Gullberg, M. (2013). Bilingualism and gesture. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 417–437). Malden, MA: Wiley-Blackwell.
- Güngen, C., Ertan, T., Eker, E., Yaşar, R., & Engin, F. (2002). Standardize mini mental test'in Türk toplumunda hafif demans tanısında geçerlik ve güvenilirliği. *Türk Psikiyatri Dergisi*, 13(4), 273-281.
- Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, 115(1-2), 147-172.
- Hadar, U., Dar, R., & Teitelman, A. (2001). Gesture during speech in first and second language: Implications for lexical retrieval. *Gesture*, 1(2), 151-165.
- Haynes, W. O., & Hood, S. B. (1977). Language and disfluency variables in normal speaking children from discrete chronological age groups. *Journal of Fluency Disorders*, 2(1), 57-74.
- Haynes, W. O., & Hood, S. B. (1978). Disfluency changes in children as a function of the systematic modification of linguistic complexity. *Journal of Communication Disorders*, 11(1), 79-93.

- Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213–229.
- Hickmann, M. (2003). *Children's discourse: Person, space, and time across languages*. Cambridge: Cambridge University Press.
- Horváth, V. (2010). Filled pauses in Hungarian: Their phonetic form and function. *Acta Linguistica Hungarica (Since 2017 Acta Linguistica Academica)*, 57(2-3), 288-306.
- Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, 7(1), 73-95.
- Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy?. *Gesture*, 11(1), 40-60.
- Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26(3), 721-752.
- Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3), 313-336.
- Iverson, J. M., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language*, 28(2), 164-181.
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367-371.
- James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of verbosity in younger and older adults. *Psychology and Aging*, 13(3), 355-367.
- Jeong, J., Kim, R., & Subramanian, S. V. (2018). How consistent are associations between maternal and paternal education and child growth and development outcomes across 39

- low-income and middle-income countries?. *Journal of Epidemiol Community Health*, 72(5), 434-441.
- Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *Journal of Speech and Hearing Disorders*, 7, 1-20.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260-267.
<https://doi.org/10.1177/0956797609357327>
- Kemper, S. (2006). Language in adulthood. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 223-238). New York: Oxford University Press.
- Kemper, S. (2012). The interaction of linguistic constraints, working memory, and aging on language production and comprehension. In M.N. Benjamin & N. Ohata (Eds.), *Memory and Aging: Current Issues and Further Directions* (pp. 43-60). Psychology Press.
- Kemper, S., Herman, R. E., & Lian, C. H. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech of noise. *Psychology and Aging*, 18(2), 181-192.
- Kemper, S., Schmalzried, R., Herman, R., Leedahl, S., & Mohankumar, D. (2009). The effects of aging and dual task demands on language production. *Aging, Neuropsychology, and Cognition*, 16(3), 241-259.
- Kemper, S., Schmalzried, R., Hoffman, L., & Herman, R. (2010). Aging and the vulnerability of speech to dual task demands. *Psychology and Aging*, 25(4), 949-962.
- Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, 16(2), 312-322.
- Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production:

- effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, 16(4), 600-614.
- Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2021). Do gestures really facilitate speech production?. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0001135>.
- Kita, S. (2000). How representational gestures help speaking. *Language and Gesture*, 1, 162-185.
- Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, 124(3), 245-266.
- Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, 24(5), 761-775.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16-32.
- Kizildere, E., Aktan-Erciyes, A., Tahiroğlu, D., & Göksun, T. (2020). A multidimensional investigation of pretend play and language competence: Concurrent and longitudinal relations in preschoolers. *Cognitive Development*, 54, 100870.
- Kosmala, L., & Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech*, 65(1), 216-239.
- Krauss, R. M. (1998). Why do we gesture when we speak?. *Current Directions in Psychological Science*, 7(2), 54-60.
- Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and Gesture: Window into thought and action*, pp. 261–283. Cambridge, UK: Cambridge University Press.

- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41(3), 841-849.
- Lee, S. B., Sim, H. S., & Shin, M. J. (2007). Disfluency characteristics in preschool bilingual children. *Communication Sciences & Disorders*, 12(2), 296-316.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.
- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861-883.
- Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*, 172, 168-188.
- MacLay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15(1), 19-44.
- 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, 940-967.
- Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science*, 9(6), 192-196.
- Mayer, M. (1969). *Frog, Where are you?* New York: Dial Press.
- McClave, E. (1994). Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23(1), 45-66.
- McNeill, D. (1985). So you think gestures are nonverbal?. *Psychological Review*, 92, 350-371.
- McNeill, D. (1987). So you do think gestures are nonverbal?. Reply to Feyereisen (1987). *Psychological Review*, 94, 499-504.

- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
- Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, 22(4), 473-500.
- Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: topic familiarity and disfluencies. *Journal of Communication Disorders*, 37(6), 489-503.
- Mol, L., Krahmer, E., & van de Sandt-Koenderman, M. (2013). Gesturing by speakers with aphasia: How does it compare? *Journal of Speech, Language, and Hearing Research*, 56, 1224–1236.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3), 615-623.
- Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 411-424.
- Morsella, E., & Krauss, R. M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories. *Journal of Psycholinguistic Research*, 34(4), 415-427.
- Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, 21(1-3), 238-290.
- Myerson, J., & Hale, S. (1993). General slowing and age invariance in cognitive processing: The other side of the coin. In J. Cerella, J. M. Rybash, W. J. Hoyer, & M. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 115–141). San Diego: Academic Press.
- Nettelbladt, U., & Hansson, K. (1999). Mazes in Swedish pre-school children with specific language impairment. *Clinical Linguistics and Phonetics*, 13(6), 483–497.

- Nicoladis, E. (2002). Some gestures develop in conjunction with spoken language development and others don't: Evidence from bilingual preschoolers. *Journal of Nonverbal Behavior*, 26(4), 241-266.
- Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics*, 28(3), 441.
- Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, 35(2), 514.
- Nicoladis, E., Pika, S., & Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children?. *Journal of Psycholinguistic Research*, 38(6), 573-585.
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1-33. <https://doi.org/10.1016/j.cogpsych.2011.05.001>
- Obler, L. K., & Albert, M. (1981). Language and aging: A neurobiological analysis. In D. S. Beasley & G. A. Davis (Eds.), *Aging: Communication processes and disorders* (pp. 107-121). New York: Grune & Stratton.
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(2), 191-230.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.
- Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, 9(1), 19-36.
- Özçalışkan, Ş., Adamson, L. B., & Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism*, 20(6), 754-763.

- Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3), B101-B113.
- Özçalışkan, Ş., & Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Developmental Science*, 13(5), 752-760.
- Özder, L. E., Özer, D., & Göksun, T. (2022). Gesture use in L1-Turkish and L2-English: Evidence from emotional narrative retellings. *Quarterly Journal of Experimental Psychology*, 0(0). <https://doi.org/10.1177/17470218221126685>
- Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology*, 11. doi: 10.3389/fpsyg.2020.573555.
- Özer, D., Göksun, T., & Chatterjee, A. (2019). Differential roles of gestures on spatial language in neurotypical elderly adults and individuals with focal brain injury. *Cognitive Neuropsychology*, 36(5-6), 282-299.
- Parish-Morris, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). From coo to code: a brief story of language development. In P.D. Zelazo (Ed.), *The Oxford handbook of developmental psychology* (vol. 1, pp. 867–908). New York: Oxford University Press.
- Parrill, F., & Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, 4(2), 197-219.
- Paxton, J. L., Barch, D. M., Storandt, M., & Braver, T. S. (2006). Effects of environmental support and strategy training on older adults' use of context. *Psychology and Aging*, 21(3), 499-509.
- Pika, S., Nicoladis, E., & Marentette, P. F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer?. *Bilingualism*, 9(3), 319.
- Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34(4), 602-619.

- Plassman, B.L., Welsh, K.A., Helms, M., Brandt, J., Page, W.F., & Breitner, J.C.S. (1995). Intelligence and education as predictors of cognitive state in late life: A 50-year follow-up. *Neurology*, 45, 1446–1450.
- Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh (m). *Interpreting*, 20(1), 1-28.
- Pritchard, M., Dipper, L., Morgan, G., & Cocks, N. (2015). Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology*, 29(7), 826-844.
- Ratner, N. B., & Sih, C. C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders*, 52(3), 278-287.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7(4), 226-231.
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, 24(3), 397–430. <http://doi.org/10.1177/0267658308090186>
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535-543.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin?. *Neurobiology of Aging*, 30(4), 507-514.
- Savva, G. M., Wharton, S. B., Ince, P. G., Forster, G., Matthews, F. E., & Brayne, C. (2009). Age, neuropathology, and dementia. *New England Journal of Medicine*, 360(22), 2302-2309.
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60(3), 362-367.

- Schubotz, L., Özyürek, A., & Holler, J. (2019). Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. *Language, Cognition and Neuroscience*, 34(2), 254-271.
- Schwartz, B. L. (2006). Tip-of-the-tongue states as metacognition. *Metacognition and Learning*, 1(2), 149-158. <https://doi.org/10.1007/s11409-006-9583-z>
- Sekine K., & Rose M. (2013). The relationship of aphasia type and severity to gesture productions in people with aphasia. *American Journal of Speech-Language Pathology*, 22, 662–672.
- Sekine, K., Rose, M. L., Foster, A. M., Attard, M. C., & Lanyon, L. E. (2013). Gesture production patterns in aphasic discourse: In-depth description and preliminary predictions. *Aphasiology*, 27(9), 1031-1049.
- Seyfeddinipur, M. (2006). *Disfluency: interrupting speech and gesture* (Ph.D. thesis), Radboud University, Nijmegen, Netherlands.
- Seyfeddinipur, M., & Kita, S. (2001). Gesture as an indicator of early error detection in self-monitoring of speech. In *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.
- Shao, R., & Lee, T. M. (2014). Aging and risk taking: toward an integration of cognitive, emotional, and neurobiological perspectives. *Neuroscience and Neuroeconomics*, 3, 47-62.
- Shriberg, E. E. (1996). Disfluencies in SWITCHBOARD. In *Proceedings of International Conference on Spoken Language Processing*, Addendum (pp. 11–14), Philadelphia, PA.
- Silverman, S. W., & Ratner, N. B. (1997). Syntactic complexity, fluency, and accuracy of sentence imitation in adolescents. *Journal of Speech, Language, and Hearing Research*, 40(1), 95-106.

- Slobin, D. (1987). Thinking for speaking. *Proceedings of the Berkeley Linguistic Society*, 13, 435–445.
- Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, 32(1), 25-38.
- So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes*, 25(10), 1335-1353.
- Stites, L. J., & Özçalışkan, Ş. (2017). Who did what to whom? Children track story referents first in gesture. *Journal of Psycholinguistic Research*, 46(4), 1019-1032.
- Syndicate, U. C. L. E. (2001). *Quick placement test*. Oxford University Press.
- Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and Aging*, 17(4), 662-676.
- Ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech. In *The 7th Gesture and Speech in Interaction (GeSpIn) Conference*. Stockholm.
- Theocharopoulou, F., Cocks, N., Pring, T., & Dipper, L. T. (2015). TOT phenomena: Gesture production in younger and older adults. *Psychology and Aging*, 30(2), 245-252.
- Thompson, L. A., & Guzman, F. A. (1999). Some limits on encoding visible speech and gestures using a dichotic shadowing task. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 54(6), 347-349.
- Van Borsel, J., Maes, E., & Foulon, S. (2001). Stuttering and bilingualism: A review. *Journal of Fluency Disorders*, 26(3), 179-205.
- Van Nispen, K., van de Sandt-Koenderman, M., Sekine, K., Krahmer, E., & Rose, M. L. (2017). Part of the message comes in gesture: how people with aphasia convey

- information in different gesture types as compared with information in their speech. *Aphasiology*, 31(9), 1078-1103.
- Vilà-Giménez, I., Dowling, N., Demir-Lira, Ö. E., Prieto, P., & Goldin-Meadow, S. (2021). The Predictive Value of Non-Referential Beat Gestures: Early Use in Parent–Child Interactions Predicts Narrative Abilities at 5 Years of Age. *Child Development*, 92(6), 2335-2355.
- Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children’s narrative discourse performance. *Developmental Psychology*, 55(2), 250.
- Vilà-Giménez, I., & Prieto, P. (2020). Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Developmental Science*, 23(6), e12967.
- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, 114(4), 591-600.
- Wingate, M. E. (1987). Fluency and disfluency; illusion and identification. *Journal of Fluency Disorders*, 12(2), 79-101.
- Yairi, E. (1981). Disfluencies of normally speaking two-year-old children. *Journal of Speech, Language, and Hearing Research*, 24(4), 490-495.
- Yairi, E. (1982). Longitudinal studies of disfluencies in two-year-old children. *Journal of Speech, Language, and Hearing Research*, 25(1), 155-160.
- Yairi, E., & Clifton Jr, N. F. (1972). Disfluent speech behavior of preschool children, high school seniors, and geriatric persons. *Journal of Speech and Hearing Research*, 15(4), 714-719.
- Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, 12(3), 323-335.

- Yaruss, J. S. (1997). Clinical measurement of stuttering behaviours. *Contemporary Issues in Communication Science and Disorders*, 24, 33-44.
- Yaruss, J. S., Newman, R. M., & Flora, T. (1999). Language and disfluency in nonstuttering children's conversational speech. *Journal of Fluency Disorders*, 24(3), 185-207.
- Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10(2), 147–160.
- Zhao, J. (2006). The communicative functions of gestures in L2 speech. *Journal of Second Language Acquisition and Teaching*, 13, 1-17.



LIST OF PUBLICATIONS

Peer-Reviewed Journal Articles

Arslan, B., Ng, F., Göksun, T., & Nozari, N. (2023). Trust my gesture or my word: How do listeners choose the information channel during communication?. *Journal of Experimental Psychology: Language, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0001253>

Arslan, B., Aktan-Erciyes, A., & Göksun, T. (2023). Multimodal language in bilingual and monolingual children: Gesture production and speech disfluency. *Bilingualism: Language and Cognition*. 1-13. <https://doi.org/10.1017/S1366728923000196>

Arslan, B., & Göksun, T. (2022). Aging, Gesture Production, and Disfluency in Speech: A comparison of younger and older adults. *Cognitive Science*, 47(2), e13098. <https://doi.org/10.1111/cogs.13098>

Arslan, B., & Göksun, T. (2021). Aging, Working Memory, and Mental Imagery: Understanding gestural communication in younger and older adults. *Quarterly Journal of Experimental Psychology*, 74(1), 29-44. <https://doi.org/10.1177/1747021820944696>

Manuscripts under review

Arslan, B., Göksun, T., & Akçay, Ç. (under revision). Does source reliability moderate the survival processing effect?: The role of linguistic markers as reliability cues.

Arslan, B., Avcı, C., Yılmaztekin, A., & Göksun, T. (under review). Do bilingual adults gesture when they are disfluent?: Understanding gesture-speech interaction across first and second languages.

Peer-Reviewed Conference Proceedings

Avcı, C., **Arslan, B.,** & Göksun, T. (2022). Gesture and Speech Disfluency in Narrative Context: Disfluency Rates in Spontaneous, Restricted, and Encouraged Gesture Conditions. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society* (pp. 1912-1917). Austin, TX: Cognitive Science Society.

Arslan, B., & Göksun, T. (2020). Understanding multimodal communication: Gesture production and disfluency in speech. *Proceedings of the 7th Gesture and Speech in Interaction (GeSpIn) Conference*. Stockholm, Sweden.

LIST OF PRESENTATIONS

Oral presentations

Arslan, B., Avcı, C., & Göksun, T. (2022, July). *Do gestures facilitate speech production?: Understanding contextual and individual differences.* The talk given at the 9th Conference of the International Society for Gesture Studies (ISGS), Chicago, The United States.

Arslan, B., & Göksun, T. (2021, September). *Cognitive processes associated with gesture production and speech disfluency.* The talk given at the 6th National Experimental-Cognitive Psychology Symposium (Virtual Symposium), Ankara, Turkey.

Arslan, B., & Göksun, T. (2020, September). *Understanding multimodal communication: Gesture production and disfluency in speech.* The talk given at the 7th Gesture and Speech in Interaction (GeSpIn) Conference (Virtual Conference), Stockholm, Sweden.

Poster presentations

Arslan, B., Avcı, C., Yilmaztekin, A., Yilmaz, F., Yürekli, M., Kaya, M., & Göksun, T. (2023, May). *Do bilinguals use gestures when they are disfluent?.* Poster presented at the 9th International Symposium on Brain and Cognitive Science, hosted by Özyeğin University, Istanbul, Turkey.

Avcı, C., **Arslan, B.,** Ceylan, C., & Göksun, T. (2023, May). *How Does Gesture Restriction Affect Grammatical Complexity? Individual differences in Gesture and Speech Production.* Poster presented at the 9th International Symposium on Brain and Cognitive Science, hosted by Özyeğin University, Istanbul, Turkey.

Arslan, B., Aktan-Erciyes, A., & Göksun, T. (2022, September). *Tekdilli ve ikidilli çocuklarda multimodal iletişim: Konuşmanın akıcılığının bozulması ve jest üretimi.* Poster presented at the 3rd National Developmental Psychology Symposium, hosted by Koç University, Istanbul, Turkey.

- Arslan, B.,** Aktan-Erciyes, A., & Göksun, T. (2022, May). *Multimodal communication in bilingual and monolingual children: Speech disfluency and co-speech gestures*. Poster presented at the 8th International Symposium on Brain and Cognitive Science (Virtual Symposium), hosted by Koç University, Istanbul, Turkey.
- Arslan, B.,** Avcı, C., & Göksun, T. (2022, July). *Multimodal Communication in Virtual and Face-to-Face Settings: Gesture Production and Speech Disfluency*. Poster presented at the 44th Annual Conference of Cognitive Science Society, Toronto, Canada.
- Ng, F., **Arslan, B.,** Göksun, T., Nozari, N. (2022, June). *Trust my gesture or my word? How do listeners choose the information channel during communication?* Poster presented at the 12th International Conference on Language Production (IWOLP). Pittsburgh.
- Arslan, B.,** Göksun, T., & Akçay, Ç. (2021, July). *The effect of evidentiality markers on the survival processing effect*. Poster presented at the 43rd Annual Meeting of the Cognitive Science Society (Virtual Conference).
- Arslan, B.,** Aktan-Erciyes, A., Avcı, C., Tezel, S., & Göksun, T. (2021, April). *Bilingualism and Disfluency in Children: Do gestures help?*. Poster presented at Society for Research in Child Development (SRCD) Biennial Meeting (Virtual Symposium).
- Arslan, B.,** Avcı, C., Karbuz, C., Şentürk, Y., & Göksun, T. (2020, May). *Aging, Gesture Production, and Disfluency in Speech: A comparison of younger and older adults*. Poster presented at the 7th International Symposium on Brain and Cognitive Science (Virtual Symposium), hosted by Bilkent University, Ankara, Turkey.
- Arslan, B.,** Caba, B., & Göksun, T. (2019, September). *Aging, Working Memory, and Mental Imagery: Understanding gestural communication in younger and older adults*. Poster presented at the 6th Gesture and Speech in Interaction (GeSpIn) Conference, Paderborn, Germany.
- Arslan, B.,** Caba, B., & Göksun, T. (2019, September). *Yaşlanma, İşler Bellek, ve Zihinsel İmgeleme: Farklı yaş gruplarında jest kullanımını anlamak*. Poster presented at the 5th National Experimental-Cognitive Psychology Symposium, Istanbul, Turkey.