Multi-modal Expression of Events by Bilinguals: Do Proficiency, Language, Discourse Context, and Event Type Matter?

Multi-modal Expression of Events by Bilinguals:

Do Proficiency, Language, Discourse Context, and Event Type Matter?

by

Armita Ghobadi

Under the Direction of Şeyda Özçalışkan, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2025

ABSTRACT

Gesture and speech form tightly integrated systems in first language (L1) production, with gestures either augmenting or supplementing the information conveyed in speech. We know less about the gesture-speech system in second language (L2) production, particularly with respect to speaker proficiency, language typology and task demands imposed by the discourse context. In this study, we focus on the speech and gestures produced by two groups of adult Persian (L1)-English (L2) bilinguals, as compared to adult English monolinguals. In Study 1, we focus on productions in English and ask whether speaker proficiency (native, high, low) and discourse context (narrative, explanation) influence the overall amount, diversity, and complexity of speech and gesture production. In study 2, we focus on productions in both English and Persian and ask whether speaker proficiency (native, high, low) and event (self-motion, caused- motion) have an effect on language-specific patterns of speech and gesture production. Understanding key factors that influence the expression of motion events in both L1 and L2 production contexts not only advances our knowledge of bilingualism but also carries significant implications for language education and cross-cultural communication in second-language learning contexts.

INDEX WORDS: Bilingualism, Motion events, Gestures, Speech, Second language acquisition, First language production

Multi-modal Expression of Events by Bilinguals:

Do Proficiency, Discourse Context, and Event Type Matter?

by

Armita Ghobadi

Committee Chair: Şeyda Özçalışkan

Committee: Rose Sevcik

Elizabeth Tighe

Michael Beran

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2025

DEDICATION

To my ride-or-die, my dad, Alireza, who never stopped believing in me and my dreams; to my mom, Masoumeh, whose positivity carried me through every challenge; to my brother, Parsa, my best friend and confidant; and to Eaman, who kept me grounded and reminded me to keep going when the road felt too long. To my incredible research assistants—Kimberly, Chris, Angela, Jacqueline, Cindy, Isabelle, and Taylor—your dedication and work ethic made this dissertation possible, and I am endlessly grateful. To my Persian community in Atlanta, who became my family when I needed one the most—thank you for your love and support. This journey would not have been the same without you.

And finally, to my mentor, advisor, and dare I say, friend, Şeyda —thank you for believing in my vision, giving me the freedom and space to pursue the research that meant so much to me, and for being more than a mentor—not just in academia, but in life. I could not have done this without you.

ACKNOWLEDGEMENTS

I am deeply grateful to all my committee members for their invaluable insights, dedication, guidance, and support throughout this project. First and foremost, my deepest thanks go to Dr. Şeyda Özçalışkan for helping me shape this research in a way that not only aligned with my academic interests but also allowed me to bring recognition and awareness to the Persian language—something that holds immense significance for me. I am incredibly thankful to Dr. Rose Sevcik for their unwavering support throughout my PhD journey, for fostering a safe and welcoming space within the Developmental program, and for being the steady pillar I could always rely on whenever I faced challenges. To Dr. Elizabeth Tighe, thank you for being an inspiring role model and for encouraging me to embrace statistics and data analysis—to truly *"be the captain of my own ship!"* I still remember this from our Statistics II class. And to Dr. Beran, I am deeply grateful for the way you always recognized something in my research that I could not see myself. Your intellect and insights have been invaluable, enriching my work in ways I could not have imagined.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Adult native speakers of structurally different languages systematically differ in how they express events, particularly those involving spatial motion, following language-specific patterns in both speech and co-speech gesture in their first language (L1) (e.g., Kita & Özyürek, 2003; Özçalıskan, Lucero, Goldin-Meadow, 2016a, 2018). However, relatively little is known about how these patterns manifest in second language (L2) production, and existing research presents mixed results (Brown & Gullberg, 2008; Cadierno, 2010; Özçalışkan, 2016; Özyürek, 2002; Nicoladis et al., 2009, 2018; Pika et al., 2006). In this project, we focused on adult bilingual speakers of Persian (L1)-English (L2) – a group of adult learners that has received minimal attention in previous research—as compared to English monolinguals. We examined their speech and gesture production using a variety of tasks across two studies. In Study 1, we focused solely on productions in English, examining how levels of speaker proficiency (native, high, low) and discourse context (narrative, explanation) contributed to differences in gesture and speech production. In Study 2, we focused on patterns in both Persian and English, examining how levels of speaker proficiency (native, high, low) and event context (self-motion, caused-motion) predicted language-specific patterns of speech and gesture production. The project, with one study focusing on *overall patterns of production* and the other focusing on *language-specific patterns of production* in adult bilinguals, provides a comprehensive assessment of multi-modal communication strategies bilingual speakers employ across a variety of contexts, and, in turn, has important implications for instructional strategies in second language learning contexts.

1 INTRODUCTION

Previous research in first language (L1) production has shown a strong connection between speech and gesture use in adults (Kita, Özyürek, Allen, Brown, Furman, & Ishizuka, 2007). Adults use gesture when they speak, with gestures either reinforcing what they express in speech (e.g., "chair" + point at chair) or adding new information not found in speech (e.g., "sit" + point at chair; Goldin-Meadow, 2007; Hostetter, 2011). However, we know less about how speech and gesture are linked in second language (L2) production; and the existing studies provide inconclusive results (Brown & Gullberg, 2008; Özçalışkan, 2016; Özyürek, 2002). In this study, we focused on two groups of adult Persian (L1)-English (L2) bilinguals with either high or low proficiency in their L2 and compared them to monolingual English speakers. Our goal was to determine whether there were systematic differences in the amount, diversity, and complexity of speech and gesture production among these three groups of speakers in different contexts of language use, with a focus on narratives and explanations. One possibility is that gesture and speech follow the same patterns, with greater speech production outcomes coupled with greater gesture production outcomes across all groups. Another possibility however is that gesture might compensate for the difficulties in speech production – particularly for bilinguals with low L2 proficiency—resulting in greater gesture production outcomes coupled with lower speech production outcomes.

The study, focusing on multiple facets of speech and gesture production in adult bilinguals with different L2 proficiencies, provided a comprehensive assessment of multi-modal communication strategies bilingual speakers employ across a variety of contexts, which has important implications for instructional strategies in second language learning contexts.

1.1 Patterns of Speech and Production in Bilinguals

There are numerous ways to convey a message in a given language (Nicoladis, 2024). For example, when narrating a cartoon scene where a *woodpecker is knitting*, one speaker might say, *"The mother bird is knitting diligently while wearing her glasses,"* whereas another might only say, *"She is doing something."* According to models of language production, conveying a message involves a series of sequential steps: first, speakers conceptualize what they want to say; next, they select the appropriate words and syntax to structure the message; and finally, they produce speech by activating the phonemes corresponding to the selected words (Costa, Roelstraete, & Hartsuiker, 2006; Levelt, Roelofs, & Meyer, 1999; Nicoladis, 2024).

Lexical access—the ability to retrieve and produce words—plays a crucial role in speech production and is influenced by a speaker's level of experience with a language (Nicoladis, 2024). Since bilinguals divide their exposure between two languages, they may experience difficulties in lexical access, particularly in the weaker of their two languages. This challenge does not typically arise for monolinguals, who have continuous and consistent exposure to a single language. Research suggests that bilinguals generally have less exposure to one language (the weaker language) compared to their dominant language (Cattani et al., 2014). Consequently, this pattern of exposure influences speech production, introducing variability in how bilinguals communicate compared to monolinguals. That being said, bilinguals show variability in speech production in their L1 and L2 and as compared to monolingual speakers. Beginning with the *amount of speech production*, research has shown that bilingual speakers talk more, particularly compared to the monolingual speakers of their L2. Studies that examined amount of speech production across a variety of contexts—from descriptions of animated stimuli to narratives—showed that bilinguals produce more speech in their L2 compared to monolingual speakers of their L2—a pattern that has

been shown across different bilingual groups, including of Mandarin (L1)-English (L2), Hindi (L1)-English (L2), French (L1)-English (L2), Spanish (L1)-English (L2): Nicoladis, Nagpal, Marentette, & Hauer, 2018; Spanish (L1)-English (L2): Cruz (2021); Turkish (L1)-English (L2): Özçalışkan (2016); and Korean (L1)-English (L2): Park (2020). One possible explanation for this difference is the relative L2 proficiency of bilingual speakers. For example, in a discourse context, bilinguals might lack the vocabulary to express the different entities in speech (e.g., characters, objects; Bosch, 1983; Garrod, 2001; Lyons, 1977; So, Kita, & Goldin-Meadow, 2009), and might instead rely on other linguistic tools, such as more extended descriptions (Johns, Sheppard, Jones, & Taler, 2016). This, in turn, might result in greater speech production in bilinguals, particularly compared to monolinguals.

This difference becomes particularly pronounced for bilinguals with lower proficiency in L2. An earlier study by So, Kita, & Goldin-Meadow (2013) that examined Mandarin (L1)-English (L2) bilinguals with either high or low proficiency in L2 showed that bilinguals with low—but not high proficiency—produced greater amount of speech than bilinguals with high proficiency. The greater amount of speech production among less proficient bilinguals was attributed to their tendency to over specify referents, by including greater amounts of descriptive speech to characterize them (Gullberg, 2003, 2006; So et al., 2013; Yoshioka, 2008)—a pattern that becomes particularly pronounced in more complex speech production tasks (i.e., tasks that place higher cognitive demands such as reasoning, argumentation; Yoshioka, 2008).

Some other studies also suggested that the amount of speech production could show variability based on the type of L1 and L2 (Nicoladis et al., 2018). Languages like English tend to employ a more chronicle-style narrative, focusing on actions (e.g., how events happened); languages like Greek, on the other hand, employ a more descriptive type of narrative, focusing on

why events happened (Ryan, 1993; Tannen, 1980). Accordingly, the greater emphasis placed on providing a chronicle of the story with more details results in longer narratives in a language like English, which has been the L2 in much of the earlier work with bilinguals (Tannen, 1982). Most of the earlier work also primarily used either narrative tasks (e.g., So et al., 2013) or description of animated scenes (e.g., Nicoladis et al., 2018; Pika, Nicoladis, & Marentette, 2006), leaving patterns of speech production in other relatively more demanding speech contexts (e.g., explanations) unexamined, highlighting this as an important area in the need of further research.

Turning to the *diversity of speech* production, we know relatively less about the variability in the range of meanings bilinguals convey in speech. We know from developmental work that bilingual children convey a narrower range of meanings in speech in each of their languages compared to their monolingual peers (Bialystok, 2009; Oller and Eilers, 2002), suggesting early differences. Research with adult bilinguals also showed that the proficiency level of the speaker might be one of the key factors in determining the diversity of spoken vocabulary in bilinguals (Johns et al., 2016). As shown in earlier work with English (L1)-French (L2) bilinguals, bilinguals with high proficiency were more likely to exhibit greater semantic diversity (i.e., greater variety of words) in their speech than the ones with low proficiency (Johns et al., 2016). A study conducted on three groups of French (L1)-English (L2) bilinguals with varying levels of proficiency in a picture description task found that bilinguals with high proficiency displayed higher levels of lexical diversity, particularly for verbs (Treffers-Daller, 2009). Similarly, another study with Korean (L1)-English (L2) bilinguals with different levels of proficiency showed that bilinguals with high proficiency produced a more diverse set of verb types than the ones with low proficiency in L2 (Park, 2020).

Research also suggests that discourse context is an important variable in determining speech diversity in bilinguals. An earlier study with Persian (L1)-English (L2) bilinguals—all with high proficiency—examined lexical diversity in three types of contexts, including argumentation (e.g., argue why money might bring happiness), description (e.g., describe a good time you had with a friend), and narrative (e.g., narrate a story based on pictures). The study showed that the argumentation and description resulted in higher levels of lexical diversity compared to the narratives (Bayazidi, Ansarin, & Mohammadnia (2019), thus suggesting greater speech diversity in discourse contexts that impose greater production demands.

Turning to the *complexity of speech* production, the existing relatively sparse research suggests that proficiency also serves as an important factor for speech complexity. As shown in earlier work, (e.g., Bayazidi et al., 2019; Johns et al., 2016), bilinguals with high L2 proficiency produced more complex speech in more demanding tasks (e.g., argumentation), further highlighting discourse context as an important contributor to speech complexity as well.

In summary, research on amount, diversity, and complexity of speech production in bilinguals suggests that proficiency and discourse context serve as important factors in explaining variability in production. Bilinguals with high L2 proficiency show lower amount, but greater diversity and complexity of speech production as compared to bilinguals with low L2 proficiency. Discourse context and language type might also affect speech production, with certain types of tasks (e.g., argumentation) and types of languages (e.g., languages with chronicle style of reporting) resulting in higher speech production outcomes in bilinguals compared to their monolingual peers. Even though existing research underscores the significant influence of both proficiency and context demands on speech production in bilinguals, there is no research to date that examined systematic variability in speech production in bilinguals based on both proficiency

and discourse context across multiple measures, including amount, diversity, and complexity of speech production.

1.2 Patterns of Gesture Production in Bilinguals

When we communicate, we not only move our mouths to produce speech but also use our hands, heads, and bodies (Stam, 2022). The hand movements that accompany speech are known as co-speech gestures (Kendon, 2004; McNeill, 1992, 2005; Stam, 2022). Gesture and speech form a tightly integrated system in monolinguals (e.g., de Ruiter, 2000; Kita & Özyürek, 2003; McNeill, 1992). However, relatively less is known about gesture production in bilinguals, with most studies focusing on overall rates of gesture production. Beginning with the *amount of gesture production*, studies suggest that bilinguals' gesture more, particularly when speaking their L2, compared to monolinguals (Nicoladis, 2007; Nicoladis, Pika & Marentette, 2009; Nicoladis et al., 2018; So, 2010; Özçalışkan, 2016; Pika et al., 2006; So et al., 2013). An earlier study by Pika et al., (2006) which involved two groups of advanced French (L1)-English (L2) and English (L1)-Spanish (L2) bilinguals compared to English monolinguals, revealed that bilinguals produced more gestures in their L2 compared to their monolingual counterparts. This pattern was also evident in a study conducted by Özçalışkan (2016), in which Turkish (L1)-English (L2) bilinguals with high proficiency produced more gestures in their L2 than monolingual English speakers. This pattern was even evident when comparing bilinguals with different proficiency levels to monolingual speakers. Two studies, one examining four groups of bilinguals with Mandarin, Hindi, French, or Spanish as L1 and English as L2 (Nicoladis et al., 2018) and the other studying Korean (L1)-English (L2) bilinguals—both with various levels of proficiency (Park, 2020)—showed that bilinguals gestured more in L2 English than monolingual English speakers.

Some studies (e.g., So et al., 2013) examined amount of gesture use in relation to proficiency level and showed that bilinguals might differ in their relative use of different gesture types based on their proficiency level. So et al. (2013), in a picture description study with Mandarin (L1)-English (L2) bilinguals, showed that bilinguals with lower L2 proficiency produced more deictic gestures that indicated objects (e.g., pointing to a car), while bilinguals with higher L2 proficiency used more iconic gestures that were symbolic (e.g., flapping arms to convey flying). This was a finding that was also evident at the early ages in children (Nicoladis, Mayberry, & Genesee, 1999), suggesting that not only overall gesture production but also relative use of each gesture type might differ based on the proficiency level of the bilinguals.

Much of the earlier research on gesture production has focused on narrative or picture/video description tasks, leaving open the question of whether these patterns extend to other discourse contexts, such as explanations, and how discourse context might interact with speakers' level of L2 proficiency. Even studies that have examined discourse context (e.g., Tabensky, 2008; Stam, 2016) primarily focused on L2 learners in classroom settings. For example, Tabensky (2008) investigated L2 learners' gesture use in an expository task that included two sections: (1) a presentation and (2) answering follow-up questions. The study found that adult L2 learners used more metaphorical gestures during the presentation task but relied more on iconic gestures when answering follow-up questions. Similarly, a longitudinal case study by Stam (2016) supported these findings. Examining a Spanish L1, L2 English learner over 14 years, Stam analyzed gesture use across two tasks—a narrative and an oral proficiency interview. The study found that the learner used more metaphorical gestures when explaining the interview process but employed more iconic gestures during narrative tasks.

Turning to the *diversity of gesture production* (i.e., the different meanings conveyed in gesture) in bilinguals, the research remains non-existent. We do not yet know whether diversity of meanings conveyed in gesture varies either by proficiency level or discourse context. However, we know from earlier work on speech diversity that bilinguals with high proficiency produce more diverse speech (i.e., use more different word types) than the ones with low proficiency. If gesture further augments what is conveyed in speech, then we would expect the same pattern in gesture, namely that bilinguals with low proficiency would convey a narrower range of meanings in their gesture akin to their speech. If, on the other hand, gesture compensates for the difficulties in speech production, we would then expect that bilinguals with low proficiency would use a greater diversity of meanings in gesture to convey meanings that they cannot convey in speech – a possibility that might also be more pronounced in discourse contexts that impose greater demands in production. The paucity of research in this domain thus renders it as an important area for future research.

Turning last to the *complexity of gesture production*, there is relatively limited research as well. The few studies that focused on the informational relation gesture has in relation to speech showed that bilinguals with high proficiency were more likely to use gesture to supplement what they conveyed in speech than bilinguals with low proficiency (Gullberg, 2003, 2006; So et al., 2013). They might, for example use a deictic gesture to clarify an ambiguous referent in speech (e.g., “The lady asked him a question” + point at picture of tall man) or produce an iconic gesture to add further descriptive detail to a spoken description (e.g., “The lady asked the man to take a taxi” + spread palms away from each other vertically to convey how tall the man is). Apart from these two earlier studies that focused on narratives produced by

bilinguals with high L2 proficiency, there is no work that has yet examined how discourse context or speaker proficiency might impact complexity of gesture production.

In summary, the limited research on bilinguals suggests that proficiency might serve as an important factor in explaining variability in gesture production—with some studies suggesting an advantage for bilinguals with higher proficiency in the amount and complexity of gesture production. However, no study has systematically examined how gesture production varies across different discourse contexts, how it differs in bilinguals with lower proficiency, or how proficiency and discourse context interact. These gaps highlight important directions for future research.

1.3 Current Study

In this study, we focused on the gestures and speech produced by Persian (L1)-English (L2) bilinguals with either high or low L2 proficiency in two different language production tasks (narratives, explanations) and compared them to the gestures and speech produced by a group of monolingual English speakers—all focusing on productions only in English. We asked two questions:

(1) We first asked whether language proficiency (native, high proficiency, low proficiency) and discourse context (narrative, explanation) would have an effect on the amount, diversity, and complexity of speech production in English. We expected an effect of proficiency. Specifically, we predicted that bilinguals with low L2 proficiency would produce a greater amount of speech than bilinguals with high L2 proficiency, who, in turn, would use more speech than monolinguals, a pattern that we expected to be reversed for diversity and complexity of speech production. These predictions stemmed from prior research, which indicated that bilinguals with low proficiency talked more but used less diverse and less complex speech when

speaking their L2 (Gulberg, 2003, 2006; Nicoladis et al., 2018; So et al., 2013; Johns et al., 2016). We also expected an effect of discourse context, namely that the differences based on proficiency would be more pronounced in the explanation task as compared to the narrative task, based on earlier research that underscored the significance of task demands on L2 speech production (e.g., Bayazidi et al., 2019).

(2) We next asked whether language proficiency (native, high proficiency, low proficiency) and discourse context (narrative, explanation) would have an effect on the amount, diversity, and complexity of gesture production in English. We predicted one of two possibilities, based on the relatively scarce and inconclusive findings. One possibility is that gestures would mirror the patterns observed in speech, based on earlier work that suggested a strong coupling between gesture and speech production in adults across a variety of tasks in first language production contexts (Chu & Kita, 2011; Hostetter, Alibali, & Kita, 2007; Theocharopoulou et al., 2015). Another possibility is that gestures would show the opposite pattern as speech, compensating for the difficulties in speech production. That is, we expected bilinguals with low proficiency to compensate for the limited diversity and complexity in their speech by producing a greater number of gestures that are both complex and diverse compared to the other groups. This prediction was based on earlier work that suggested a supplementary role for gesture in bilingual children and adults (Arslan, Aktan-Erciyes, & Göksun, 2023; Nicoladis, 2007; Smithson & Nicoladis, 2013). We expected these differences to be more prevalent in the explanation task than in the narrative task, following earlier work which showed that more difficult tasks (e.g., argumentation, intentional reasoning) result in greater gesture production than the ones with less difficulty (i.e., narratives; Nicoladis, 2007).

2 METHODS

2.1 Participants

The sample consisted of 44 Persian (L1)-English (L2) adult bilingual speakers—22 with high L2 proficiency (8 Females; M_{age}= 37, SD = 16;3), 22 with low L2 proficiency (13 Females; M_{age}= 31.5, SD = 7;6)—and 22 adult monolingual English L1 speakers (13 Females; M_{age}= 27;6, SD = 7;4) who had very limited or no knowledge of other languages. The sample size was based on a power analysis (G*Power 3.1; Faul et al., 2007), which showed that N=22 per group would provide a power of 85% at an alpha level of 0.05 with medium effect sizes (R^2 = 0.20).

Building on prior research (Olson, 2024; Portocarrero, Burright, & Donovick, 2007) and given the quantitative nature of this study, proficiency was assessed using three distinct metrics: (1) the Test of English as a Foreign Language (TOEFL) score, (2) length of residency in the United States, and (3) performance on a fluency test.

The bilinguals were classified into high vs. low proficiency groups based on their Test of English as a Foreign Language (TOEFL; Educational Testing Service, 2024) score, extent of residency in the United States, and their score on a verbal fluency task that assesses vocabulary knowledge in English (Portocarrero et al., 2007; Spreen & Strauss, 1998).[1] Bilinguals

[1] For the verbal fluency task, the participant was first asked to generate as many words as possible related to a specific category, such as animals and fruits, in one minute assessing their category (i.e., semantic) fluency. The participant was next asked to produce words starting with a given letter, such as "A" ,"S" or "F" in one minute, assessing their letter (i.e., phonemic) fluency. The mean fluency score (as reported here)—in line with earlier work (Portocarrero et al., 2007)— provides an average across both semantic and letter fluency scores.

with low English L2 proficiency included speakers who had a B2 level on the Common European Framework of Reference for Languages (CEFR; M_{TOEFL}=92.8 [SD=5.47], range=72-94, $M_{FLUENCY}$ =13.72 [SD= 4.12] and resided in the United States for less than four years ($M_{year\ of\ residency}$= 2;1, SD= 2;2). Bilinguals with high English L2 proficiency consisted of speakers who had a CEFR level of C1 (M_{TOEFL}=106.82 [SD=7.01], range = 95-113, $M_{FLUENCY\ SCORE}$ =15.23 [SD= 2.86] and who have lived in the United States for more than four years ($M_{year\ of\ residency}$= 17;6, SD= 13;4). Five of the participants from the high L2 proficiency group did not have TOEFL scores or any other equivalent English proficiency tests as they immigrated to the United States at an early age ($M_{age\ of\ immigration}$= 7;2, SD=2;5) and completed their education in English. Their fluency score ($M_{Fluency}$= 16.54, SD= 1.97) was in the range of fluency scores of bilinguals in the high proficiency group; their extent of residency was also within the range of the high proficiency group; we therefore included these 5 participants in the high L2 proficiency group. All monolingual English speakers were born in English-speaking households, had minimum knowledge of any languages other than English and completed all their education in English.

Bilingual participants were recruited through the Georgia Iranian Student Organization and the international offices at Georgia State University and Georgia Institute of Technology. Monolingual participants were recruited through student organizations at Georgia State University and via word-of-mouth referrals.

The participants in each group were either college students or recent college graduates and were comparable in gender and education (see Table 1 for sample characteristics). All participants received a small gift for their participation. The research was approved by the institutional review board of Georgia State University and was conducted in accordance with the Code of Ethics for the protection of human research participants.

Table 1. Summary of the Sample Characteristics by Group

	MONOLINGUAL	BILINGUAL	
	Native English Proficiency n=22	High English Proficiency n=22	Low English Proficiency n=22
M_{AGE} (SD)	27;6 (8;4)	37 (16;3)	31;5 (7;6)
GENDER	13 females, 9 males	8 females, 14 males	13 females, 9 males
M_{TOEFL} (SD)	N/A	106.82 (7.01)	92.8 (5.5)
$M_{RESIDENCY\ IN\ US}$ (SD)	N/A	17;7 (13;4)	2;1 (1;2)

M=mean (years; months); SD=Standard Deviation (years; months)

2.2 Procedure for data collection

Each participant was interviewed in English by a native English speaker individually in one session in a quiet place. All participants first completed the consent form and video release forms, then instructed to fill out demographic questionnaire that provided information about their age, gender, education, and language background. Bilingual participants also completed the short form of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian, Blumenfeld & Kaushanskaya, 2007), which provides information about language dominance, attitudes towards second language learning, and self-assessment of fluency in speaking, listening, reading, and writing in English. Upon completion of both questionnaires each participant was asked to participate in two language production tasks—one eliciting narratives and one eliciting explanations—in counterbalanced order. Each participant completed both the narrative and explanation tasks, with each task lasting approximately five minutes, resulting in a total duration

of ten minutes. They also completed a fluency task in English at the end of the session (see footnote 1, p. 12 for description of the task). All responses were videorecorded.

Narration task: Each participant was shown two cartoons, each approximately 2.5 minutes long, one at a time for an overall of 5 minutes. They were then asked to narrate what happened in the cartoon to the experimenter (i.e., "Tell me what happened in the cartoon in as much detail as you can remember"). Each cartoon depicted adventures of Tom and Jerry, featuring dynamic, temporally organized event scenes that were well-suited for eliciting gestures in a narrative context (see Figure 1 for stills from one of the cartoons used for the narrative task; Hannah & Barbera, 2014). The order of the two cartoons was counterbalanced across participants.

Figure 1. Screenshots from a Sample Task Eliciting Narrative

Explanation task: Each participant was presented with five short cartoons, each lasting approximately 60 seconds, thus resulting in a total of 5 minutes of video exposure, similar to the narrative task (the length of both tasks was equal). After each cartoon they were asked to explain what happened in the cartoon to the experimenter (i.e., "Explain to me what the problem in the cartoon was and how it was solved in as much detail as you can remember"). Each cartoon depicted a mouse and an elephant facing a particular problem and then solving it individually or together, making them suitable for eliciting speech and gestures within an explanatory

framework (see Figure 2 for stills from a sample cartoon used for the explanation task; Schmidt, 1971). The order of the five cartoons was counterbalanced across participants.

Figure 2. Screenshots from a Sample Task Eliciting Explanation

2.2.1 Speech

All speech produced by each participant was transcribed from video recordings by native speakers, using the Codes for the Human Analysis of Transcripts (CHAT) systems (MacWhinney, 2000). Speech transcripts were divided into utterances, defined as a sequence of words that were preceded and followed by a change in conversational turn, intonation, or pause, following the CHAT system guidelines.

Following previous research (Özçalışkan et al., 2017), we considered sounds that were used to refer to concrete and abstract entities (e.g., cat, idea, tree), events (e.g., run, eat), properties of entities or events (e.g., pretty, fast), onomatopoeic sounds (e.g., meow), conventionalized evaluative sounds (e.g., wow), and function words (e.g., and, that, with) as words. Speech responses were further coded for the amount, diversity, and complexity of production. Specifically, we used the number of words as a measure of *speech amount*, the number of different types of words or lexical diversity (e.g., "cat" vs. "bird") as a measure of *speech diversity*, and the Mean

Length of Utterance[2] in words (MLU) as a measure of *speech complexity*, following earlier work (Ozturk et al., 2021). We treated words with the same stem but with derivational morphemes as different words (e.g., *"run"* vs. *"runner"* as two different word types), while words with the same stem but with inflectional morphemes (e.g., *"run"* vs. *"running"*) were considered the same word type. Repetition of the same sentence, false start (if a speaker starts a sentence but abandoned the idea without finishing the task), nonverbal vocalization (e.g., laughter, throat cleaning), and backchannel responses (e.g., serving as a listener to experimenter's instruction, for example saying: *"yes," "sure," "of course,"* etc.) were excluded.

2.3.2 Gesture

All gestures produced by each participant were also coded. Gesture was defined as a communicative hand or body movement that did not involve direct manipulation of an object (e.g., twisting open a jar; Özçalışkan & Goldin-Meadow, 2005a). Each gesture was categorized into different types, including deictic gestures (e.g., pointing at empty space to indicate the imaginary location of a cat), emblematic gestures (e.g., thumbs up as if OK), iconic gestures (e.g., flapping arms to convey flying, placing a cupped-shaped hand in the air as if holding a bird's nest), and beats (e.g., rhythmic hand gestures marking speech boundaries; McNeill, 1992). Each gesture was also coded in terms of the informational relation it held to the accompanying speech. These included reinforcing (i.e., combinations where the gesture conveyed the same information as speech; e.g., saying "flying" while flapping arms), emphasizing (combinations where gesture marked speech boundaries; e.g., flicking fingers at the end of each speech segment),

[2] MLU was used as a metric for speech complexity to examine how gestures interact with speech in multimodal communication.

disambiguating (i.e., combinations in which gesture clarified a pronominal referent; e.g., saying "this one" while pointing to the character on screen), and supplementary (i.e., combinations where the gesture added new information to speech; e.g., saying "he ran to that door" while moving right index finger in a circle to indicate that the door was a revolving door), following the guidelines outlined in earlier work (Özçalışkan & Goldin-Meadow, 2005a, 2005b). The reinforcing and emphasizing gestures were classified as simple as they did not add any new semantic information to speech; the disambiguating and supplementary gestures were categorized as complex as gesture added new semantic information to speech, following earlier work (Ozturk et al., 2021).

Gesture responses were further coded for the amount, diversity, and complexity of production. Specifically, we used the number of gestures (beat, iconic, deictic and emblem) as a measure of *gesture amount*, the number of different referents conveyed in gesture (e.g., flapping arms for bird vs. tracing whiskers under nose for cat) that was only available in iconic and deictic gestures[3] as a measure of *gesture diversity*, and the number gestures that disambiguated or supplemented speech that was only available in iconic and deictic gestures as a measure of gesture complexity, following earlier work (Ozturk et al., 2021).

2.3.3 Reliability

We assessed reliability for gesture coding by a trained independent coder, naïve to the hypotheses of the study, who coded a randomly selected 30% of the video records. The

[3]Only iconic and deictic gestures have referents, as they explicitly represent or point to something. For example, a deictic gesture might involve pointing at a dog rather than a cat to specify a particular object, while an iconic gesture could involve moving the right hand horizontally from right to left to depict direction. In contrast, beats and emblematic gestures do not function in the same way—beats are primarily rhythmic, and emblematic gestures follow conventional meanings.

agreement between coders was 97% for identifying gestures (i.e., presence vs. absence of gesture in an utterance), 98% for assigning meaning to gesture (i.e., referents conveyed in gesture), 93% for coding gesture type (as deictic, iconic, conventional, beat) and 92% for coding gesture's informational relation to speech (reinforcing, emphasizing, disambiguating, supplementing).

2.3 Statistical Analysis

We analyzed speech and gesture separately, using a set of Generalized Linear Mixed Models (GLMMs). The use of GLMMs allowed us to address for the non-normal distribution of the data and effectively model the count-based outcomes associated with speech and gesture production measures. We conducted three separate GLMMs for speech, one for each of the three outcome measures (speech amount, speech diversity, speech complexity). Similarly, we conducted three separate GLMMs for gesture, one for each of the three outcome measures (gesture amount, gesture diversity, gesture complexity) Each model had the same structure: Outcome = Proficiency * Context + (1| Subject_ID), for each of the 6 different outcome measures[4]. Proficiency (native, high, low) and context (narrative, explanation) were included as fixed effects, and participant was treated as a random effect due to the repeated-measures design. Dummy coding was applied to each of the two categorical independent variables: native group served as the baseline for proficiency variable (native, high, low) and the narrative task served as the baseline for the discourse context (narrative, explanation). Data manipulation was conducted using Python (Python Software Foundation, 2023), with the Pandas (McKinney, 2010) and NumPy (Harris et al., 2020) libraries facilitating data management. Data visualization was performed using Matplotlib (Hunter,

[4] Each dependent variable was uncorrelated with the others, and each GLMM was conducted separately for its respective dependent variable. As a result, the power for each GLMM remained consistent at 85% with an alpha level of 0.05, assuming a medium effect size ($R^2 = 0.20$).

2007) and Seaborn (Waskom, 2021). Scikit-learn (Pedregosa et al., 2011) was employed to standardize the data and ensure consistency in scale. For multiple regression model analyses, we utilized Statsmodels package (Seabold & Perktold, 2010).

3 RESULTS

3.1 Patterns of speech production

Overall, participants spent more time completing the narrative tasks, despite each task having the same allotted duration of five minutes. We first examined whether production of speech showed variability by speaker proficiency (native, high, low) and discourse context (narrative, explanation). Beginning with *speech amount*, we found no main effect of speaker proficiency ($\beta = 0.29$, $z = 0.30$, $p = 0.33$), but a main effect of context ($\beta = 0.42$, $z = 2.33$, $p = 0.01$). Overall, speakers produced greater amount of speech in the narrative than in the explanation context. The interaction between proficiency and discourse context approached, but not reached significance ($\beta = 0.47$, $z = 1.87$, $p = 0.06$) (See Fig.3, panel A_1).

Turning next to *speech diversity* (i.e., number of different words), we found no main effect of speaker proficiency ($\beta = 0.17$, $z = 0.58$, $p = 0.56$), but there was a marginal main effect of context ($\beta = 0.35$, $z = 1.88$, $p = 0.06$) as all speakers produce more lexically diverse components only in narrative contexts. There was also no interaction between proficiency and discourse context ($\beta = 0.06$, $z = 0.23$, $p = 0.82$; see Fig.3, panel B_1).

Turning last *to speech complexity* (i.e., MLU), we found a main effect of proficiency: native speakers produced more complex speech (longer sentences) compared to bilinguals, both with high proficiency ($\beta = 0.70$, $z = 2.61$, $p = 0.01$) and low proficiency ($\beta = 1.22$, $z = 4.53$, $p < 0.001$). However, there was neither an effect of context in the complexity of speech ($\beta = 0.03$, $z =$

0.19, $p = 0.85$) nor an interaction between speaker proficiency and discourse context ($\beta = 0.19$, $z = 0.99$, $p = 0.32$; see Fig.3, panel C_1).

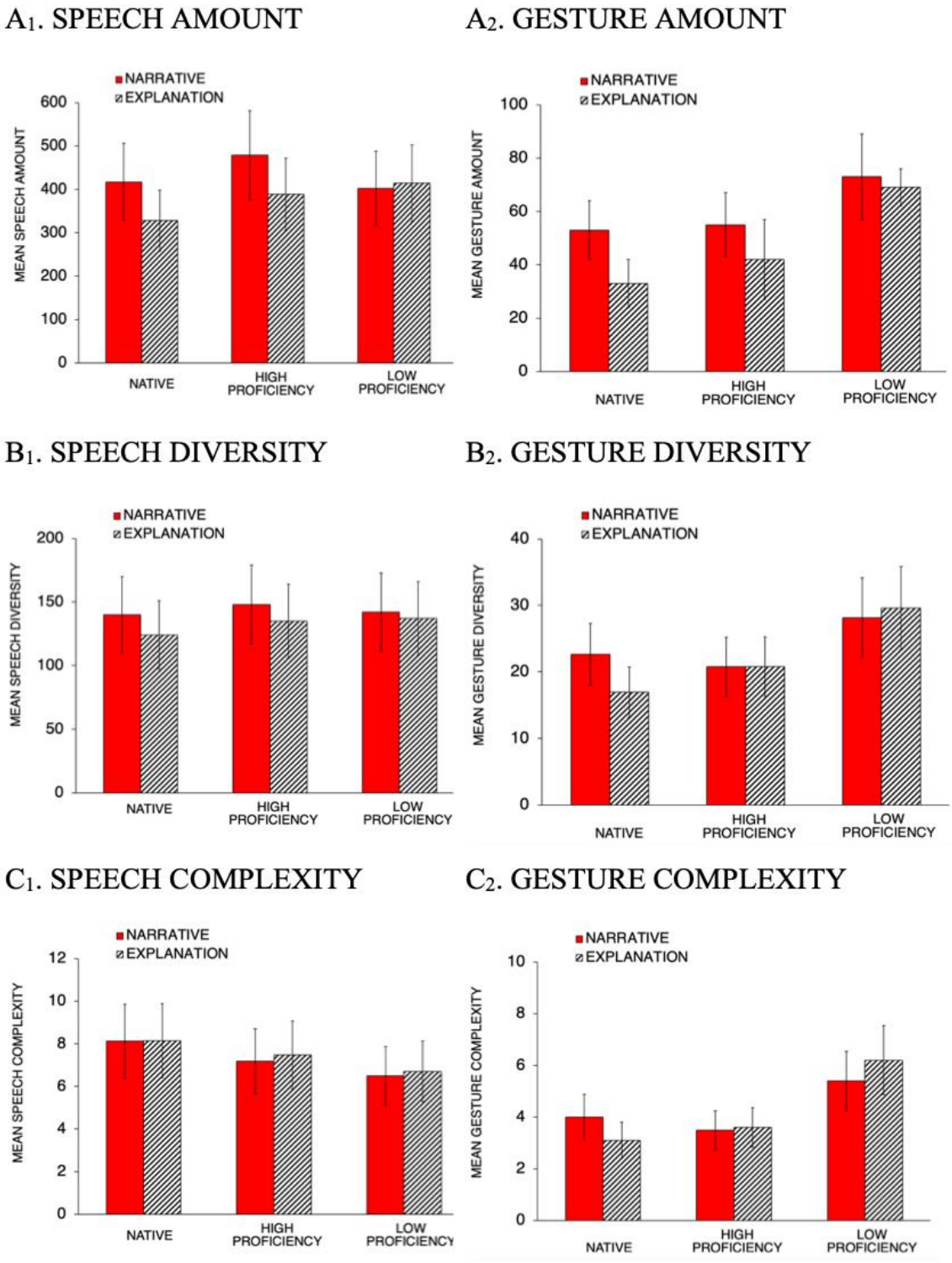


Figure 3. Mean amount (A_1-A_2), diversity (B_1-B_2), complexity (C_1-C_2) of speech and gesture production by bilinguals with low proficient bilinguals, bilinguals with high proficient, and monolingual English speakers with native proficiency in narrative (red bars) and explanation tasks (striped bars). Error bars represent standard error; also note that the scales of the bars in panels A and B are different.

3.2 Patterns of gesture production

We next examined patterns of gesture production by speaker proficiency and discourse context. Beginning with *gesture amount* (i.e., number of gestures), our results showed no main effect of proficiency ($\beta = 0.06$, $z = 0.13$, $p = 0.89$), but a main effect of context ($\beta = 0.43$, $z = 2.18$, $p = 0.02$). Speakers produced a greater number of gestures in the narrative than in the explanation context–a finding consistent with the patterns of speech production. There was no interaction between proficiency and discourse context ($\beta = 0.06$, $z = 0.23$, $p = 0.82$) (see Fig.3, panel A_2).

Speakers—across proficiency levels and contexts— produced iconic gestures (e.g., moving index finger in circles to convey a revolving door; moving the index and middle finger forward repeatedly to convey a bird's pecking; *M*=24.02, *SD*=18.33) and beat gestures (*M*=21.59, *SD*=18.12) the most, followed by deictic gestures (e.g., pointing to the screen to indicate a character; *M*=7.34, *SD*=6.12) and emblematic gestures (e.g., thumbing up as if OK; *M*=3.02, *SD*=3.62). As can be seen in Table 2, these patterns were evident in each proficiency level and discourse context as well.

Turning next to *gesture diversity* (i.e., number of different referents conveyed in gesture), we found no main effect of either speaker proficiency ($\beta = 0.17$, $z = 0.38$, $p = 0.70$), context ($\beta =$

0.16, z = 0.88, p = 0.38), or an interaction between proficiency and discourse context (β = 0.16, z = 0.64, p = 0.52; see Fig.3, panel B_2).

Table 2. Mean (SD) Production of Gesture Types by Speaker Proficiency and Context

	MONOLINGUAL		**BILINGUAL**			
	Native		High Proficiency		Low Proficiency	
	Narrative	**Explanation**	**Narrative**	**Explanation**	**Narrative**	**Explanation**
GESTURE TYPE						
Deictic	6.59 (5.37)	3.50 (3.06)	7.14 (6.52)	5.45 (4.84)	12.41 (10.53)	8.95 (6.41)
Iconic	24.91 (23.10)	18.77 (17.68)	21.77 (21.72)	18.14 (12.19)	28.32 (19.69)	32.23 (15.85)
Emblematic	2.59 (3.65)	3.09 (3.85)	3.00 (3.54)	2.64 (4.25)	3.91 (3.74)	2.91 (2.70)
Beat	19.09 (19.30)	9.45 (10.07)	27.77 (28.93)	15.41 (12.27)	30.77 (19.52)	27.05 (18.65)

SD: standard deviation

Turning last to *gesture complexity*, we found no main effect of context (β = 0.29, z = 1.50, p = 0.14), but a main effect of proficiency (β= 1.15, z= 2.37, p =0.02): bilinguals with low proficiency produced more complex gestures that added further information to speech, showing an opposite pattern to the one observed for speech complexity. Speaker proficiency also interacted with context (β= 0.88, z= 3.23, p =0.001), with bilinguals with low proficiency producing more complex gestures, but only in the explanation context (see Fig.3, panel C_2).

Speakers—across proficiency levels and discourse contexts— primarily used gesture to reinforce (e.g., "revolving door" + moving index finger in circles to convey revolving; M=26.66, SD=21.78) or emphasize (i.e., beats; M=14.68, SD=13.07) their speech. As can be seen in Table

3, using gesture to supplement speech (e.g., "the mouse uses his tail to exit the hole" + spinning index finger upward to convey spinning motion; *M*=4.43, *SD*=3.43) or disambiguate speech (e.g., "the mouse looks for something sweet inside that" + point to sugar bowl on screen; *M*=3.73, *SD*=0.58) was overall less frequent across groups and contexts. At the same time however, bilinguals with low proficiency used gesture more to add further information to their speech than the other two groups (see Figs. 3 and 4 for descriptions of a sample scene in speech and gesture by speakers with different levels of proficiency in the narrative and the explanation context).

Table 3. Mean (SD) Production of Gesture + Speech Combinations by Proficiency and Context

	MONOLINGUAL		**BILINGUAL**			
	Native		High Proficiency		Low Proficiency	
	Narrative	**Explanation**	**Narrative**	**Explanation**	**Narrative**	**Explanation**
GESTURE COMPLEXITY						
Reinforce	35.10 (34.19)	22.82 (18.35)	39.59 (28.39)	15.18 (10.33)	24.00 (17.00)	23.32 (12.40)
Emphasize	14.18 (16.67)	8.91 (9.13)	19.82 (20.24)	12.82 (8.64)	16.77 (11.47)	14.95 (12.28)
Supplement	4.48 (4.25)	3.36 (3.30)	3.55 (3.33)	3.64 (3.20)	5.32 (3.25)	6.18 (3.21)
Disambiguate	1.00 (0.47)	1.00 (0.71)	1.20 (0.40)	0.17 (0.37)	0.23 (0.42)	0.77 (1.08)

SD: Standard Deviation

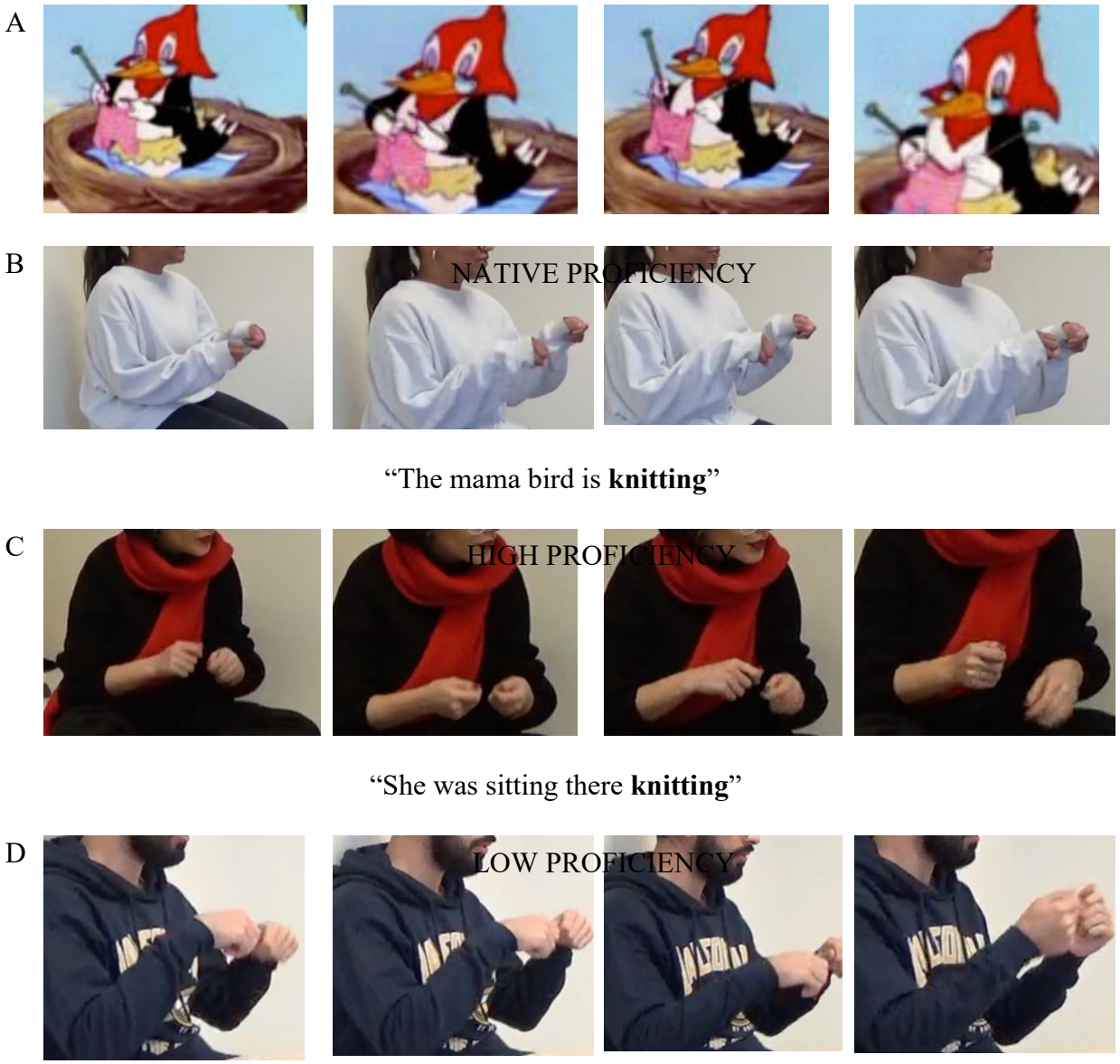
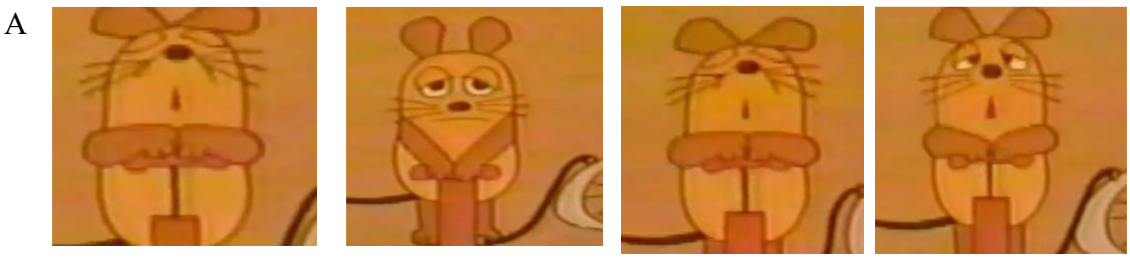


Figure 3. The sample scene of mother bird knitting (A) and its depiction in gesture by a monolingual speaker with native fluency (B) and by bilingual speakers with either high (C) or low (D) English proficiency. The speakers with native and high proficiency both used gestures to reinforce the information conveyed in their speech ("the mama bird is knitting" + moving fisted hands in opposite circular directions to convey knitting); the speaker with low English proficiency used gesture to supplement the information

expressed in speech ("she <u>is doing something</u>" + moving fisted hands in opposite circular directions to convey knitting).

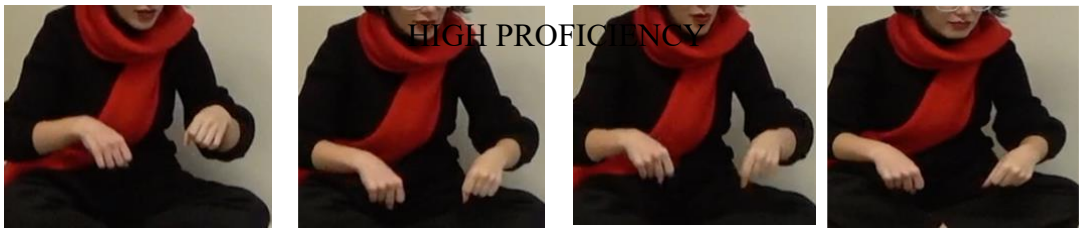


A

B

NATIVE PROFICIENCY

"He **pumped** the bicycle"



C

HIGH PROFICIENCY

"He tried to resolve it by **pumping** air"

D

LOW PROFICIENCY

"She is doing ***this***"

Figure 4. The sample explanation scene of mouse pumping a tire (A) and its depiction In gesture by a monolingual speaker with native fluency (B) and by bilingual speakers with either high (C) or low (D) English proficiency. The speakers with native and high proficiency both used gestures to reinforce the information conveyed in their speech ("the mouse is pumping the tire" + moving cupped hands up and down rapidly to convey

pumping); the speaker with low English proficiency used gesture to supplement the information expressed in speech ("she is doing this" + moving cupped hands up and down rapidly to convey pumping).

4 CONCLUSION

Gesture and speech form an integrated system in first language (L1) production as gestures either reinforce or supplement the speech (Kita, Alibali, & Chu, 2017). In this study, we asked whether this pattern extends to second language (L2) production across different proficiency levels (native, high, low) and discourse contexts (narratives, explanations), using data from a sample of 44 Persian (L1)-English (L2) bilinguals –22 with high and 22 with low L2 proficiency and 22 native English speakers. Our analysis showed an effect of context in the amount of speech and gesture production, suggesting close coupling between the two systems in L2 production.

We also found a strong effect of proficiency in the complexity of speech and gesture production—with lower complexity in speech coupled with higher complexity in gesture production by bilinguals with low L2 proficiency, compared to monolinguals and bilinguals with high proficiency—also suggesting a compensatory for gesture in the communications of bilinguals with lower proficiency in L2.

4.1 Does discourse context matter in patterns of speech and gesture production of monolinguals and bilinguals?

The participants in our study—monolinguals and bilinguals—produced more speech in their narratives than in their explanations. This finding aligns well with previous research, which also showed an effect of discourse context on speech production, with an advantage for narratives (Masson-Carro, Goudbeek, & Krahmer, 2017; Nicoladis, Pika, Yin, & Marentette, 2007; Yoshioka, 2008). One possible explanation for the greater amount of speech production in narratives could be the nature of the narrative practice in English. Several studies have shown that languages like English emphasize chronicle-style narratives focusing on actions (e.g., how

events happened), which, in turn, result in more detailed descriptions and greater speech production in the narrative context (Nicoladis et al., 2018; Ryan, 1993; Tannen, 1980).

More importantly, participants also produced more gestures in their narratives than in their explanations, suggesting a strong connection between the two modalities in production. The close coupling between speech and gesture use has been shown in earlier work in both L1 (de Ruiter, 2000; Kita & Özyürek, 2003; McNeill, 1992) and L2 production (Cruz, 2021; Nicoladis et al., 2018; Nicoladis, 2007; Özçalışkan, 2016; So et al., 2013), and our study furthers these findings, showing a similar pattern in the amount of speech and gesture production in a new group of bilinguals, namely Persian (L1)-English (L2) bilinguals. Notably, all participants, regardless of proficiency level, took longer to complete the narrative task. This suggests that producing a narrative required greater cognitive effort and verbal elaboration, leading to increased speech production, which was closely linked to higher gesture rates.

The context-based variability in production was not quite as evident when we looked at the diversity of meanings conveyed in speech and gesture production. Speakers did not produce reliably more diverse speech (i.e., word types) in their narratives than in their explanations—a pattern that was also observable in the diversity of meanings conveyed in gesture (i.e., number of different gesture referents). This finding differs from earlier work (Bayazidi et al., 2019), which showed lower diversity in speech production in the narrative context than in contexts that elicit descriptive and argumentative discourse among Persian (L1)-English (L2) bilinguals. However, different from our study, Bayazidi et al. (2019) did not include speaker proficiency as a variable in their study, thus differing in its design from our study. But this still leaves open the question about lack of reliable differences in the diversity of speech and gesture production in the two discourse contexts –even if the amount of production showed an advantage for narratives. One

possibility could be nature of the tasks. The stimuli—both narratives and explanations—centered around the same characters, with specific verbal prompts in elicitation (i.e., Narrative: "Can you tell me *what happened* in the video; Explanation: "Can you tell me *what the problem in the video was* and *how the character resolved it*), which in turn, might have limited the range of words or gesture referents participants could employ in their production. In fact, the participants frequently repeated the same character in their speech and gesture (e.g., the mouse; point at the mouse on the screen; the bird; pecking gesture for woodpecker) multiple times in the production of a single narrative or explanation. This, in turn might have affected the diversity of the meanings conveyed in speech and gesture equally in both discourse contexts. One way to address this possibility might be to ask participants to produce narratives and explanations by using cartoons with a greater range of characters, settings, and events —a possibility that could be explored in future work.

The lack of context-based variability was also observed in the complexity of speech and gesture production. Research on the complexity of speech and gestures is quite limited. One of the few studies focusing on bilinguals with high proficiency found that they produced more complex speech in more demanding tasks, such as argumentation or picture descriptions (Bayazidi et al., 2019; Johns et al., 2016). The relatively small body of research on gesture complexity has focused primarily on narrative tasks and only on bilinguals, with no comparison to other discourse contexts. Our study as the first to examine how discourse context might impact the complexity of gesture production, did not show any main effect of context.

More importantly, the context-based variability (or lack thereof) in speech and gesture production was tightly integrated. Narratives elicited a greater amount of speech than explanations, a pattern that was also mirrored in gesture production. Conversely, narratives and explanations did not differ in the diversity or complexity of speech production—a pattern that was also reflected in

gesture production. This close coupling between the two modes of communication suggests that gesture and speech form tightly integrated systems across different production contexts across proficiency levels.

4.2 Does speaker proficiency matter in patterns of speech and gesture production of monolinguals and bilinguals?

The three groups of participants in our study did not show differences in the amount of speech or gesture production—a pattern that diverged from our predictions. This finding also contrasts with earlier work that showed greater speech production in bilinguals with low proficiency as compared to those with high proficiency (e.g., Gullberg, 2003, 2006; So et al., 2013; Yoshioka, 2008).

Several factors could explain the lack of differences between proficiency levels in our study. One possibility could be the *plateau effect in language proficiency*. After reaching an intermediate level, learners may experience a plateau where further increases in proficiency do not significantly impact language production in their L2. The participants in our study had all passed the English proficiency test required for admission into colleges in the United States and had lived in an English-speaking environment for an extended period. Therefore, even though the low proficiency group was lower in proficiency than the high proficiency group, they both were still using English frequently in their everyday lives. As such, they may have developed enough experience to manage tasks like narratives and explanations at levels comparable to native speakers or bilingual speakers with high proficiency. On a related point, Han (2004) discusses the concept of fossilization, where language development slows down or stops, potentially leading to similar speech patterns among intermediate and advanced learners. This

plateau might explain why our participants, despite varying proficiency levels, exhibited comparable production in both speech and gesture.

The comparability of overall speech (and consequently gesture) production might also be a result of more formulaic language use (e.g., chunks of memorized phrases) among bilinguals with low proficiency in our study. The use of such formulaic speech might have allowed them to produce enough speech, making them comparable to the other two groups in terms of the sheer amount of speech output (Pawlak, 2011). This, in fact, aligns well with Grosjean's (2010) notion of "functional bilingualism," where bilinguals develop coping strategies that enable them to adapt their language use effectively, thus reducing variability across proficiency levels. Future studies that include bilinguals at lower proficiency levels (e.g., beginners) or that analyze speech content in terms of formulaic speech might offer more insight into the lack of proficiency-based differences in the relative amount of speech and gesture production.

The lack of proficiency-based variability was also evident in the diversity of speech and gesture production—a pattern that also differed from our predictions. Earlier work focusing on the diversity of speech production has shown that bilinguals with high proficiency used a more diverse lexicon in speech than those with low proficiency (Johns et al., 2016; Treffers-Daller, 2009). For example, Treffers-Daller (2009) conducted a study on French (L1)-English (L2) bilinguals with varying levels of proficiency in a picture description task and found that bilinguals with high proficiency displayed higher levels of lexical diversity, particularly for verbs. Our findings, however, did not replicate the previously observed differences in speech diversity and further extended this lack of difference to gesture diversity. The only marginal effect we found for speech diversity was a higher lexical diversity in the narrative context. One possible explanation for this could be the task design. In the narrative task, we prompted

participants to "tell the story in as much detail as possible," which encouraged them to use a more diverse range of vocabulary to describe the characters' distinctive features and elaborate on how the events progressed. In contrast, for the explanation context, we provided a more specific prompt: "Can you tell me what the problem was and how the main character solved it?" This framing inherently constrained elaboration by directing participants toward specific information rather than encouraging expansive storytelling. While prompting participants in this distinct way was necessary to examine the effect of discourse context, it may have directly influenced the diversity, breadth, and depth of speech production.

Tasks that are too simple or too structured can limit the range of linguistic and gestural expressions participants could use. Robinson (2001) suggests that task complexity influences language output in a way that less complex tasks may produce more uniform results across proficiency levels. Therefore, the simplicity of our narrative and explanation tasks might have constrained participants' expressive possibilities, leading to similar levels of diversity across all proficiency groups.

Another explanation could be the plateau effect in language proficiency, as we have suggested above for amount of speech and gesture production. The relatively smaller proficiency gap between the two bilingual groups in our study might have affected not only overall amount but also the diversity of and gesture production, making the groups look more similar (De Bot, Lowie, & Verspoor, 2007).

Importantly, however, we observed that the patterns of speech and gesture production remained closely aligned across all proficiency levels. Specifically, the three groups did not show differences in either the amount or the diversity of their speech production; this was also the case for their gesture production. This suggests that the integration of speech and gesture

operates similarly across proficiency levels, reinforcing the idea that both modalities are tightly linked in communication regardless of language proficiency. These findings also further support earlier models of gesture-speech integration, where the two modalities form a tightly integrated system from conceptualization to articulation (Kita & Özyürek, 2003; McNeill, 1992).

The lack of a group difference in the amount and diversity of speech and gesture production, however, co-occurred with a strong effect of proficiency on the complexity of speech and gesture production. Bilinguals with low proficiency produced less complex speech than both bilinguals with high proficiency and monolinguals with native proficiency—a pattern consistent with earlier work suggesting that higher proficiency correlates with greater linguistic complexity (Bayazidi et al., 2019; Johns et al., 2016). More importantly, this pattern was reversed for the complexity of gestures. Bilinguals with low proficiency produced more complex gestures that either disambiguated their speech or added information not already mentioned verbally, suggesting a compensatory role for gesture.

This finding aligns with some of the earlier theories of gesture production, starting with McNeill (1992), who considered speech and gestures as a unified system emerging from a single unit of thought called the "Growth Point" (GP). GP is defined as an initial unit of thinking, combining imagery (gestures) and linguistic components into a cohesive entity during communication. Gestures reflect internal cognitive states and actively shape the formulation and articulation of speech. Putting GP in the context of speech deficiencies (e.g., aphasia or developmental speech delays), gestures can facilitate verbal expression by tapping into non-verbal cognitive resources, helping speakers communicate more effectively. In other words, when the speaker struggles with verbal communication, gestures could provide an alternative or supplementary way of communication for speech development.

Other theories, such as those proposed by Krauss, Morrel-Samuels, and Colasante (1991) and Butterworth and Hadar (1989), suggest that gestures occur more frequently in situations where speech failure arises. The lexical retrieval theory argues that gestures, particularly iconic gestures, facilitate lexical access during speech production and enhance fluency (Krauss, Chen, & Gottesman, 2000; Rauscher, Krauss, & Chen, 1996; Stam, 2022). However, other studies, such as those by Beattie and Coughlan (1998, 1999), indicate that iconic gestures are not the only type of gestures involved in lexical retrieval. Despite these differences, both research perspectives converge on the idea that bilinguals—especially those with lower proficiency—use gestures more frequently when struggling to retrieve words. Starting with Rauscher et al. (1996), who found that using gestures facilitated speech, especially in discourse contexts requiring complex spatial descriptions. Addressing this compensatory function, de Ruiter (2000) also proposed the Sketch Model, which considers gesture and speech as two separate but interrelated systems. As he argues, when one system does not function fully, the other can compensate. We see evidence of a similar phenomenon in adults with aphasia, where speech production is belabored. Not surprisingly, adults with aphasia frequently rely on gesture to either replace speech (i.e., pantomimes) or add information to their speech (e.g., Cocks, Morgan, & Kita, 2011; Mol, Krahmer, & van de Sandt-Koenderman, 2013; Sekine, Rose, Foster, Attard, & Lanyon, 2013; Ozturk & Özçalışkan, 2024). The participants in our low proficiency group also showed word retrieval difficulties and frequently used gesture to convey meanings that they could not do with their L2 speech (He is doing this + *moving cupped hands up and down rapidly to convey pumping*; She is doing something + *moving fisted hands in opposite circular directions to convey knitting*). The Sketch Model (de Ruiter, 2000) aptly explains this phenomenon in bilinguals with low proficiency. Gestures appear to compensate for difficulties in L2 production and assist

speech production by providing visual cues that support lexical retrieval—as suggested by Krauss et al., (2000)— also by expanding the complexity of the meaning conveyed in speech. This was particularly evident in the explanation task, where problem-solving and reasoning were required, and bilinguals with low proficiency produced more complex gestures. This illustrates how proficiency interacts with context, with the low-proficiency group producing more complex gestures in tasks demanding higher cognitive resources.

The compensatory role of gesture in language production of adult L2 speakers with low proficiency in our study also aligns well with earlier studies with children on L1 development. Children frequently use gesture to supplement their speech across multiple milestones in their L1 development, from early sentences (e.g., Goldin-Meadow & Butcher, 2003; Özçalışkan & Goldin-Meadow, 2005a) to early narratives and explanations (Demir, Levine, & Goldin-Meadow, 2015; Stites & Özçalışkan, 2017; 2021; Özçalışkan, 2007). Moreover, gestures may reduce production demands, facilitating communication for speakers with limited linguistic proficiency (Goldin-Meadow & Alibali, 2013). Similarly, in second language learning contexts, gestures might serve as a compensatory tool, helping low-proficiency bilinguals convey complex ideas they cannot express verbally.

In summary, our analysis of speech and gestures produced by speakers with respect to discourse context and speaker proficiency showed that gesture and speech form tightly integrated systems. Gesture both mirrors the patterns observed in speech (e.g., greater amount of speech and gesture in narratives than explanations) and compensates for difficulties in speech production (i.e., greater use of more complex gestures by speakers with low proficiency in the more challenging explanation context). These findings contribute to our understanding of the intricate relationship between speech and gesture in bilingual communication, highlighting that

this integration becomes even more pronounced in speakers with lower proficiency. By revealing that low-proficiency bilinguals employ more complex gestures to overcome linguistic challenges, our study extends existing theories on gesture's compensatory role in language production. This underscores the adaptive strategies speakers use to maintain effective communication, emphasizing gesture as a crucial resource in conveying complex ideas when speech alone is insufficient. Furthermore, our research emphasizes the influence of discourse context on gesture use, suggesting that task complexity can modulate the reliance on gestural communication, especially among those with limited language proficiency. These insights have practical implications for language education and therapy, where encouraging gesture use could facilitate language learning and fluency in second-language speakers. Future research could build on these findings by exploring other communicative contexts or by examining how gestures interact with other non-verbal modalities to support communication in bilinguals, particularly focusing on low-proficiency groups.

STUDY 2: HOW LANGUAGE PROFICIENCY, TYPOLOGY AND EVENT TYPE INFLUENCE THE PATTERN OF MOTION EXPRESSION IN BILINGUAL SPEAKERS?

5 INTRODUCTION

Gesture and speech form a tightly integrated system in first language production contexts (Kita, 2000; McNeill, 1992, 2000). The two modalities work in tandem from conceptualization to articulation of a message—with patterns of gesture largely mirroring patterns of speech production (i.e., Kelly, Özyürek, & Maris, 2010). In this study, we focus on the domain of motion that shows strong crosslinguistic variability in speech and gesture production in first language production (Özçalışkan, Lucero & Goldin-Meadow, 2016a, Slobin, 2004, Talmy, 2000). Our goal is to extend the applicability of these patterns to the multi-modal expression of motion in second language production contexts. We aim to provide a comprehensive account of the multi-modal expression of motion events by focusing on a relatively understudied language (i.e., Persian) as compared to English and by including speaker proficiency, along with event and language type, as possible factors in explaining variability in patterns of speech and gesture production among bilinguals.

5.1 Talking about motion events

5.1.1 Expression of motion in speech by monolinguals

Languages show systematic cross-linguistic variation in their encoding of motion events, particularly in the way they map Path and Manner (Talmy, 1985, 2000). According to Talmy, the world's languages can be divided into two types based on how speakers semantically package the two key components of motion: namely path (direction of movement) and manner (style of movement). In describing motion events, speakers of satellite-framed languages (S-languages,

e.g., English, German, Polish) use a conflated strategy, expressing manner in the main verb and path outside the verb, typically in a path particle or a preposition, within a single clause (e.g., “He RUNS INTO the house”). In contrast, speakers of verb-framed languages (V-languages, e.g., Spanish, Turkish, Japanese) use a separated strategy, expressing path in the main verb and manner in an additional subordinate clause to describe the same event (e.g., "Eve GIRER KOŞARAK " = house-to ENTER RUNNING; Özçalışkan & Slobin, 1999; Slobin, 2004). The preference for encoding path inside or outside the verb influences how speakers of each language type express manner or path components of motion. S-language speakers often express both manner and path in a single clause in their descriptions of motion. In contrast, V-language speakers prefer to exclude manner and express only the path of motion, likely an outcome of the extra processing load imposed by the necessity to use a subordinate construction (Özçalışkan & Slobin, 1999, 2003).

These differences in the semantic organization of event components have been shown in the relative production of the different packaging strategies speakers employ in speech across different languages of the world. Speakers of S-languages use a greater number of descriptions with conflated strategies (e.g., English: Ibarretxe-Antuñano, 2004; Özçalışkan, 2006, Dutch: Bowerman et al., 2002, Polish: Lewandowski & Özçalışkan, 2019, German: Lewandowski & Özçalışkan, 2018). Speakers of V-languages, on the other hand, rely on a greater number of descriptions with separated strategies (Basque: Ibarretxe-Antuñano, 2004; French: Gullberg et

al., 2008, Japanese: Allen et al., 2007, Kita & Özyurek, 2003; Korean: Choi & Bowerman, 1991, Spanish: Naigles & Terrazas, 1998; Turkish: Özçalışkan et al., 2016a, b, 2018)[5].

Most of this earlier work focused on self-motion events (i.e., events that involve voluntary motion of a figure on its own, e.g., man crawls over a carpet), leaving caused-motion events (i.e., events that involve motion of an entity caused by an agent; e.g., man throws ball across a net) relatively unexamined. A few studies that focused on caused-motion events across different languages suggested both similarities and differences in the expression of motion in speech. One set of studies showed that the expression of caused-motion events largely follows the same patterns but includes a greater amount of manner expression, resulting in more conflated descriptions (Hendricks & Hickmann, 2015; Özçalışkan, 2000, 2005). For instance, prior research on self vs. caused-motion of animate entities (e.g., “boy enters the house” vs. “boy is pushed into the house”; Özçalışkan, 2000) and metaphorical motion of abstract entities (e.g., "economy enters into recession" vs. "unemployment pushes the economy into recession"; Özçalışkan, 2005) in English and Turkish has provided some preliminary evidence for this phenomenon—with greater manner expression in caused-motion events. This pattern has also

[5] The binary classification into S- versus V-languages was also later expanded to include a third group of languages, namely equipollently-framed languages (e.g., Chinese, Slobin, 2004). Most of these languages use serial verb constructions (e.g., RUN-ENTER: one for manner and one for path), both expressed within a single clause, thus also express both components within the bounds of a single clause (Chen & Guo, 2009; Paul, Emerson, Özçalışkan, 2022). Similar to S-languages, E-language speakers also use primarily conflated descriptions, but using serial verb constructions (e.g., PǍOJÌN fángzi= Adam RUN-ENTER house; Tütüncü et al., 2023).

been extended to some other V-languages. Hendriks and Hickmann (2015) examined caused- and self-motion descriptions produced by adult speakers of English (S-framed) and French (V-framed). The two groups differed in describing self-motion events with English and French speakers producing more conflated and separated responses, respectively. This crosslinguistic difference was not evident in the description of caused-motion events, however, where speakers of both languages relied primarily on conflated strategies in speech. A study by Lewandowski and Özçalışkan (2018) extended these results to another set of S- and V-languages, namely German (S-framed) and Spanish (V-framed). German speakers differed from Spanish speakers producing more conflated and fewer separated responses in their speech about self-motion. The crosslinguistic differences diminished however when speakers described caused-motion events, with both groups relying primarily on conflated responses. These findings thus suggest that event type might be an important factor in the expression of motion events in speech.

In this study, we focused on a relatively understudied language, namely Persian, with the goal to extend the applicability of the findings from earlier studies that focused primarily on languages spoken in continental Europe and America. Persian, also known as Farsi, is an Indo-European language widely spoken in Iran, and Afghanistan, along with a variation of it in Tajikistan and with significant communities of its speakers residing in the Gulf region (Akhavan et al., 2017; Feiz, 2011). There are only three studies that have examined patterns of motion expression in Persian – all focusing on first-language production. Feiz (2011) examined the expression of manner and path in the narratives produced by 30 native Persian and 30 native English speakers, focusing on self-motion events. They found that Persian speakers showed characteristics of both S- and V-languages in their expression of motion. More specifically, Persian speakers displayed predominantly V-language patterns, using primarily path verbs (i.e.,

the frog entered the treehouse). However, when manner was a necessary part of the action, they used manner along with path, using a conflated construction (i.e., *the frog ran into the treehouse*). In a more recent study, Khorvash and Lotfi (2022) examined motion expression in Persian monolinguals and Persian (L1)-English (L2) learners, using an animated narrative elicitation task, also with a focus on self-motion events. They found that monolingual Persian speakers showed more V-language patterns in describing self-motion events—a pattern consistent with Feiz's findings. In contrast, another study by Akhavan et al. (2017), using an animated motion description task with adult native Persian speakers, also with a focus on self-motion events, showed the opposite pattern –with descriptions following the S-language pattern (i.e., more conflated)—thus differing from earlier studies. The inconclusive results in the placement of Persian along the typological continuum thus highlights the need for additional research.

5.1.2 Expression of motion in speech by bilinguals

There is considerable research on motion expression in bilingual speakers of structurally different languages but with an almost exclusive focus on self-motion events and also with largely inconclusive results. Some studies focusing on self-motion have shown proficiency as an important element in achieving target L2 patterns. For example, Özyürek (2002) examined Turkish (L1)-English (L2) bilinguals, who acquired English later in life, including advanced learners who had substantial experience with L2 by living in an English-speaking environment. She found that advanced L2 learners consistently achieved L2 patterns, while beginners and intermediate learners largely followed L1 patterns when speaking their L2. Similarly, other work, either comparing advanced Danish (L1)-Spanish (L2) bilinguals (Cadierno & Ruiz, 2006) to monolingual Spanish and Danish speakers or advanced Turkish (L1)-English (L2) bilinguals

to monolingual Turkish speakers (Özçalışkan, 2016), showed no differences in patterns of speech production of bilinguals in their L2 compared to the monolingual speakers of each target language (i.e., Spanish, Danish, or Turkish).

In some other studies, focusing mostly on intermediate L2 speakers, bilinguals showed an effect of L1 on their motion descriptions in their L2 speech. For example, an earlier study by Negueruela et al. (2004) examined English (L1)-Spanish (L2) and Spanish (L1)-English (L2) bilinguals—all intermediate learners—as compared to monolingual English and Spanish speakers, using a picture description task depicting self-motion. The results showed that Spanish (L1)-English (L2) bilinguals faced challenges attaining L2 speech patterns, often producing more separated clauses—a pattern consistent with their L1. Similarly, other studies (e.g., Japanese (L1)-English (L2) bilinguals by Brown & Gullberg, (2008); Spanish (L1)-Danish (L2) bilinguals by Cadierno (2010)—all with intermediate proficiency level in L2) showed the same challenge in the attainment of the L2 speech pattern, with bilinguals opting to produce more separated responses in their speech about motion, thus showing an effect of their L1. Similarly, English (L1)-Spanish (L2) bilinguals tended to use more manner verbs and conflated them with path satellites in their L2 Spanish, also showing an effect of L1 on L2 speech production. Stam (2006, 2010), examining motion descriptions of Spanish (L1)-English (L2) bilinguals with intermediate proficiency in L2 compared to Spanish and English monolinguals (N=5/each), showed that bilinguals largely resorted to L1 patterns in their L2 speech about self-motion—a pattern consistent with earlier work.

Lastly, turning to bilinguals with low proficiency in L2, a recent study by Aktan-Erciyes et al. (2023) on Turkish (L1)-English (L2) bilinguals with low proficiency corroborated Özyürek (2002), showing strong L1 patterns when speaking about self-motion in L2 when describing

animations depicting motion. Similarly, Özçalışkan (2016) conducted a study with advanced Turkish (L1)-English (L2) bilinguals who had resided in an English-speaking environment for a prolonged period, using a set of animated motion scenes involving self-motion. She found that bilinguals were able to adopt L2 patterns in speech, comparable to monolingual English speakers. This pattern was also observed in the case of advanced Danish (L1) -Spanish (L2) bilinguals, where native Danish speakers (S-framed) successfully followed Spanish (V-framed) speech patterns by producing more separated clauses in their L2 Spanish when describing a picture story book (Cadierno & Ruiz, 2006). Similar results were observed in a study with Russian (L1)-English (L2) learners (both S-languages) with high intermediate and advanced proficiency in L2 English: bilinguals followed L2 speech patterns at rates comparable to monolingual English speakers (Iakovleva, 2012). While there is no research on Persian (L1)-English (L2) bilinguals' motion expression, one study on Persian (L1)-English (L2) learners with either low or high proficiency residing in Iran, showed that the bilinguals with high proficiency were able to attain target L2 patterns in their speech about motion when describing self -motion events (Khorvash & Lotfi, 2022).

Yet another set of studies on bilingual speech production in describing self-motion events shows bidirectional effects between L1 and L2 resulting in a speech pattern that shows similarities to both L1 and L2 (Brown & Gullberg, 2008; Park, 2019; see Cadierno, 2010 for a review). For instance, Brown and Gullberg (2008) investigated intermediate Japanese (L1)-English (L2) bilinguals residing in both Japan and the United States as compared to English and Japanese monolinguals. Their study revealed a strong bidirectional relationship between L1 and L2, with bilinguals' speech production in L2 showing properties akin to both L1 and L2, but different from monolingual speakers of each language. Further evidence for bidirectional effects

comes from studies comparing Spanish (L1)-Danish (L2) bilinguals as well as Spanish (L1)-English (L2) bilinguals (Alonso, 2020; Aveledo & Athanasopoulos, 2023; Park, 2019). For example an earlier study by Aveledo and Athanasopoulos (2023) found that while advanced Spanish (L1)-English (L2) bilinguals show the target L2 conflated pattern when speaking L2, they still relied on a fewer manner verbs but greater number of path verbs compared to monolingual English speakers. As suggested by Aveledo and Athanasopoulos, these patterns are largely driven by both proficiency and age of acquisition effects, namely that bilinguals who acquired L2 English earlier in life and had higher proficiency levels not only used more conflated descriptions, but also produced more manner verbs in their English L2.

Similar results were observed in an earlier study by Lai, Rodriguez, and Narasimhan (2014), which examined advanced Spanish (L1)-English (L2) adult bilinguals, categorizing them into two groups: early bilinguals who learned English before the age of six (0–5 years) and late bilinguals who acquired English between ages of 6 and 15. They found that while both groups of adults demonstrated target L2 patterns on the surface, closer analysis revealed subtle differences. Early bilinguals showed similar patterns when speaking either L1 or L2, expressing either path of motion (L1 Spanish pattern) or both manner and path simultaneously in a compact description (L2 English pattern). In contrast, late bilinguals exhibited different patterns when speaking each language: they predominantly expressed manner in English, frequently conflating it with path, while they encoded manner less frequently in their L1 Spanish.

Overall, studies that examine expression of self-motion in speech among bilinguals suggest that bilingual speakers are more likely to adhere to the patterns of the target L2 language if they are advanced learners of L2, but either resort to L1 patterns or show a pattern between L1 and L2 if they are beginner or intermediate learners.

Studies focusing on the expression of *caused-motion events* among bilinguals received relatively less attention. One of the few studies that delved into the distinction between caused and self-motion events was conducted by Choi and Lantolf (2008) on advanced speakers of Korean (L1)-English (L2) and English (L1)-Korean (L2). Their findings indicated that, even at high levels of proficiency, both groups of bilinguals followed the patterns of their L1 in their description of both self and caused motion events. That is, English (L1)-Korean (L2) speakers largely followed the patterns of their L1 English, encoding manner at a higher rate, and Korean (L1)-English (L2) bilinguals followed patterns of their L1 Korean, encoding path at a higher rate. Another study by Lewandowski and Özçalışkan (2021b), examined caused-motion events, with a particular focus on object-placement events (e.g., put stone in pocket) produced by Polish (L1)-German (L2) and Polish (L1)-Spanish (L2) bilinguals with intermediate proficiency in comparison to Polish, German and Spanish monolinguals on a set of animated tasks. They found that speakers showed greater L1 effects on L2 speech production if the L2 required more specific encoding of event structure than their L1, thus suggesting an effect of language type in expressing caused-motion events. These results thus suggest that both speaker proficiency and target language type influence patterns of L2 speech production in describing caused motion events among bilinguals.

5.2 Gesturing about motion

5.2.1 Expression of motion in gestures by monolinguals

Speakers frequently use hand gestures to accompany their verbal descriptions of motion events, known as co-speech gestures (McNeill, 2000). The *expression of self-motion* in co-speech gestures produced by adult S- and V-language speakers largely follows the language-specific patterns observed in their speech, as demonstrated across several different studies. S-

language speakers typically use the same conflated pattern in gestures, synthesizing manner, and path components into a single gesture (e.g., wiggling index and middle fingers while moving the hand from right to left to convey running from right to left). In contrast, V-language speakers follow the separated pattern that they use in speech, producing one gesture for path (e.g., tracing a line from left to right to convey a left-to-right trajectory) and one for manner (e.g., wiggling fingers in the same place to convey running; Kita and Özyurek, 2003; Lewandowski & Özçalışkan, 2018; Özçalışkan et al., 2016a, 2018). Similar to the patterns observed in speech, V-language speakers frequently express only the path of motion in their gestures, leaving out manner from their gestures (see Özçalışkan, 2016 for Turkish; Stam, 2001, 2006 for Spanish).

Studies comparing the gestural description of self-motion in English (S-language), Turkish (V-language), or Japanese (V-language) in monolinguals, using animated motion scenes, found patterns similar to speech (e.g., Kita & Özyürek; 2003; Özçalışkan et al, 2016a, 2018). English speakers used more conflated gestures (e.g., running fingers left to right to convey running toward the house), while Turkish and Japanese speakers relied more on separated gestures (e.g., tracing a line left to right to convey the trajectory of motion towards the house). Similar patterns were observed in other studies. Stam (2006a, 2006b, 2008, 2010) comparing monolinguals speaking Spanish vs. English found more separated gestures encoding path in Spanish and more conflated gestures encoding both manner and path in English. The same patterns were also evident in studies with Korean vs English monolinguals, with greater use of separated path gestures by Korean speakers and conflated gestures by English speakers (Choi & Lantolf, 2008; Park, 2019).

Research on patterns of gesture production for caused-motion events remains sparse. The only existing study by Lewandowski & Özçalışkan (2018) examined patterns of caused-motion

expression in the gestures produced by German (S-framed) vs. Spanish (V-framed) monolinguals in describing animated scenes from a silent movie. Even though the speakers showed different gesture patterns in describing self-motion events (more conflated in German, more separated in Spanish), they showed similarities when describing caused-motion events with both groups using more conflated gestures that synthesized manner and path components of motion. Even though some of the earlier work on motion event expression in Persian included both self-motion and caused-motion stimuli in their design, these studies did not examine motion type variability in Persian monolinguals' response strategies in gesture (Akhavan et al., 2015, 2017).

Overall, the existing gesture research on self-motion and the relatively sparse research on caused-motion descriptions in L1 production has shown strong crosslinguistic differences in the gestures that accompany self-motion and crosslinguistic similarities in the gestures that accompany caused-motion event descriptions.

5.3 Expression of motion in gestures by bilinguals

Compared to monolingual speakers, relatively less is known about the expression of motion in the gestures of bilingual speakers, particularly with different proficiency levels; and the existing findings are largely inconclusive. One set of studies, focusing on advanced bilinguals showed evidence of L2 patterns in gesture when describing self-motion events in L2. For example, Özyürek (2002) in her study with advanced Turkish (L1)-English (L2) bilinguals found gesture patterns similar to monolingual English speakers when they described self-motion events in L2 English—a pattern that did not extend to Turkish (L1)-English (L2) bilinguals with beginner level proficiency in English. This pattern however was not evident in another study with advanced Turkish (L1)-English (L2) bilinguals, where the bilinguals largely followed L1 patterns (more separated) in their gestures when describing events in L2 English (Özçalışkan,

2016). The L1 effect on gesture was also shown in other work focusing on Spanish (L1)-English (L2) bilinguals with varying levels of proficiency from intermediate to advanced (Stam, 2001): even with advanced L2 proficiency, bilinguals still adhered to their L1 patterns in their gestures about motion events.

Yet some other studies suggested bidirectional effects on L1 and L2 gesture production among bilinguals. Negueruela et al. (2004) studied Spanish (L1)-English (L2) and English (L1) Spanish (L2) advanced bilinguals and showed that bilinguals followed gestural patterns that were distinct from both their L1 and L2. For example, Spanish (L1)-English (L2) bilinguals expressed manner more in their L2 (similar to English monolinguals), but without combining it with path, thus differing from English L1 speakers. Similar patterns were observed in Korean (L1)-English (L2) bilinguals with different levels of proficiency (i.e., intermediate to advanced) residing in Korea, who displayed gestural patterns reflecting a strong influence of both L1 and L2 (Park, 2020). These findings were further corroborated in Gullberg and Brown's (2008) study on intermediate Japanese (L1)-English (L2) speakers, where gestures showed a strong bidirectional effect and interaction between L1 and L2. More specifically, Gullberg (2008, 2011) found that Japanese (L1)-English (L2) speakers' L2 gestures changed over time. While bilinguals initially relied more on separated gestures, as they immersed more in L2 environments, they began to add more manner to their L1 gestures and showed a gestural pattern that was not similar to either L1 or L2. As such, bilinguals developed a hybrid gesture pattern that is influenced by both languages—a phenomenon referred as "convergence" (Brown, 2015), namely that bilinguals' gesture use—particularly for features such as manner—gravitated toward a midpoint that is distinct from both their L1 and L2.

Similar to speech, most earlier studies focused on gestures depicting self-motion events, leaving gestural depictions of *caused-motion events* relatively unexplored. One of the few studies was with advanced Korean (L1)-English (L2) and advanced English (L1)-Korean (L2) bilinguals, as compared to monolingual Korean speakers (Choi & Lantolf, 2008). The study found that the patterns of motion expression in gesture remained similar in the description of self- vs. caused-motion events in both groups of bilinguals with closer alignment to L1 patterns, namely more gestures encoding manner in L2 Korean and more gestures encoding path in L2 English. A later study also showed adherence to L1 gesture patterns in Korean (L1)-English (L2) bilinguals with varying levels of proficiency in both self- and caused-motion events (Park, 2020). The few studies that compared self- vs. caused-motion events, with either small sample sizes (n=2 per group; Choi & Lantolf) or with designs that did not include proficiency as a variable (Park, 2020, suggest that bilingual speakers tend to adhere to L1 patterns in gesture regardless of event type—even though this evidence remains anecdotal in nature, highlighting the need for future work.

5.4 Current Study

Earlier research with monolinguals has shown strong cross-linguistic differences in motion expression patterns in speech and gesture. These differences are also largely evident in bilinguals, with proficiency emerging as a significant factor that explains variability in the attainment of L2 patterns in bilinguals. In this study, we focused on a relatively understudied language, namely Persian, which shows systematic variability in its typological classification with respect to different event types (Feiz, 2011). More specifically, we focused on the speech and gestures produced by adult Persian (L1)-English (L2) bilinguals with either high or low L2 proficiency as compared to English L1 speakers. Our goal was to understand how language

proficiency (native, high, low) and event type (self-motion, caused-motion) contribute to patterns of speech and gesture production in Persian (L1)-English (L2) bilinguals and English monolinguals. We asked two questions:

1. We first focused on productions in English. We asked whether the expression of motion events in speech and gesture would show an effect of *speaker proficiency* (native, high, low), *event type* (self-motion, caused-motion), and *packaging type* (conflated, separated) in attaining target L2 patterns. Starting with speech, we predicted an interaction between proficiency and packaging, namely those bilinguals with high—but not low—proficiency would be more likely to produce more conflated responses (i.e., the monolingual English pattern) in speech, based on earlier work that showed greater alignment to L2 patterns at higher proficiency levels (Cadierno & Ruiz, 2006; Özçalışkan, 2016; Özyürek, 2002, Stam 2010, 2015). We also expected a 3-way interaction between proficiency, event type, and packaging, namely that monolinguals and bilinguals with high—but not low proficiency—would be more likely to produce more conflated responses when describing caused-motion events than self-motion events in speech, based on earlier research showing greater manner expression and more conflated packaging in caused motion events (Choi & Lantolf, 2008; Lewandowski & Özçalışkan, 2021b; Khorvash & Lotfi, 2022).

Turning to gestures, and based on the mixed findings in the literature, we predicted two possibilities. One possibility was an interaction between proficiency and packaging, with bilinguals with high—but not low—proficiency producing more conflated gesture responses, similar to monolingual English speakers—based on earlier work (e.g., Özyürek, 2002). We also expected this pattern to be more pronounced in caused events across proficiency levels, based on earlier work that suggested greater conflated packaging in caused events (Choi & Lantolf, 2008;

Lewandowski & Özçalışkan, 2021b). Another possibility was a bidirectional effect between bilinguals' L1 and L2 (Negueruela et al., 2004; Park, 2019), as bilinguals may follow distinct gestural patterns that do not fully align with those of monolingual English speakers and instead fall somewhere between their first and second languages in their patterns of gesture production—a pattern that we expected to be more pronounced in the expression of self-motion events.

2. We next focused only on productions by bilingual speakers in the two languages (English and Persian); we asked whether the speech and gestures produced by the two groups of Persian (L1)-English (L2) bilinguals in their L1 Persian and L2 English would show an effect of L2 proficiency (high, low), event type (self-motion, caused-motion), language (Persian, English), and packaging type (conflated, separated) in attaining target L1 and L2 patterns.

We predicted an interaction between proficiency and packaging, namely that bilinguals with high—but not low—proficiency would be more likely to produce more conflated responses in their L1 and L2 speech, showing an effect of their proficiency in their both first and second language (Cadierno & Ruiz, 2006; Özyürek, 2002; Stam, 2001, 2006). We also expected a 3-way interaction between proficiency, event type, and packaging type namely that bilinguals with high—but not low proficiency—would be more likely to produce more conflated responses, particularly when describing caused-motion events (Choi & Lantolf, 2008; Lewandowski & Özçalışkan, 2021b).

Given the inconclusive findings with regards to gesture production by proficiency levels (Gullberg, 2008, 2011; Negueruela et al., 2004; Stam, 2001; Özçalışkan, 2016; Choi & Lantolf, 2008; Park, 2020), we formulated two possible predictions. The first possibility was that gesture production in L1 and L2 would mirror speech patterns: bilinguals with higher proficiency would use more conflated packaging responses—a pattern that we expected to be more pronounced in

the description of caused-motion events. An alternative possibility was that gestures would show bidirectional effects in both L1 and L2. In this scenario, both groups of bilinguals would demonstrate patterns that show an in-between pattern that does not fully align with either L1 or L2 gesture production, but still with greater expression of conflated responses in the expression of caused-motion events.

6 METHODS

6.1 Participants

The sample was same as the study 1, consisting of 44 Persian (L1)-English (L2) adult bilingual speakers—22 with high L2 proficiency, 22 with low L2 proficiency—and 22 adult monolingual English L1 speakers who had very limited or no knowledge of other languages. The sample size was based on a power analysis (G*Power 3.1; Faul et al., 2007), which showed that N=22 per group would provide a power of 0.81 at an alpha level of 0.05 with medium effect sizes ($f^2 = 0.25$). The bilinguals were classified into high vs. low proficiency groups based on their Test of English as a Foreign Language score (TOEFL; Educational Testing Service, 2024), extent of residency in the United States, and their score on a verbal fluency task that assesses vocabulary knowledge in English (Portocarrero et al., 2007; Spreen & Strauss, 1998).[6] Bilinguals with low English L2 proficiency included speakers who had a B2 level on the Common European Framework of Reference for Languages (CEFR; M_{TOEFL}=92.8 [SD=5.47], range=72-94, $M_{FLUENCY}$ =13.72 [SD= 4.12] and resided in the United States for less than four years ($M_{year\ of\ residency}$= 2;1, SD= 2;2). Bilinguals with high English L2 proficiency consisted of speakers who

[6] For the verbal fluency task in L2 English, the participant was first asked to generate as many words as possible related to a specific category, such as animals and fruits, in one minute assessing their category (i.e., semantic) fluency. The participant was next asked to produce words starting with a given letter, such as "A","S" or "F" in one minute, assessing their letter (i.e., phonemic) fluency. The mean fluency score (as reported here)—in line with earlier work (Protocarreo et al., 2007)— provides an average across both semantic and letter fluency scores.

had a CEFR level of C1 (M_{TOEFL}=106.82 [SD=7.01], range = 95-113, $M_{FLUENCY\ SCORE}$ =15.23 [SD= 2.86] and who have lived in the United States for more than four years ($M_{year\ of\ residency}$= 17;6, SD= 13;4). Five of the participants from the high L2 proficiency group did not have TOEFL scores or any other equivalent English proficiency tests as they immigrated to the United States at an early age ($M_{age\ of\ immigration}$= 7;2, SD=2;5) and completed their education in English. Their fluency scores ($M_{Fluency}$= 16.54, SD= 1.97) were in the range of fluency scores of bilinguals in the high proficiency group; their extent of residency was also within the range of the high proficiency group; we therefore included these 5 participants in the high L2 proficiency group. All monolingual English speakers were born in English-speaking households, had minimum knowledge of languages other than English and completed all their education in English. Both samples were recruited through student organizations at an urban research university in the United States. As such, the participants in each group were either college students or recent college graduates and were comparable in gender and education (see Table 1 for sample characteristics). All participants received a small gift for their participation. The research was approved by the institutional review board of a research university in the United States and was conducted by the Code of Ethics for the protection of human research participants.

6.2 Procedure for data collection

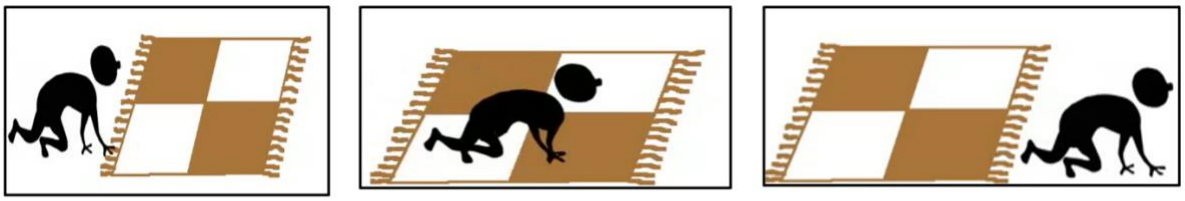
Upon completion of Study 1 and an extended break, each participant was interviewed individually again, once in English and once in Persian, by a native speaker of each language during a single session in a quiet location. All participants first completed the consent form and then were instructed to fill out a demographic questionnaire that recorded details about their age, gender, education, language background, and handedness. Bilingual participants also filled out

the short form of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), which provides information about language dominance, attitudes toward second language learning, and self-assessed fluency in speaking, listening, reading, and writing in English. After the completion of the questionnaires, participants engaged in two experimental language production tasks involving 18 short-animated videos (2–3 seconds each) that elicited speech and gestures; they also completed a fluency task, separately in English and in Persian, at the end of the session[2]. All responses were videorecorded.

Experimental task: Participants were introduced to a cartoon character (named Adam in English and Arya in Persian), who performed actions depicted in the animations. Each participant viewed 18 animated events—9 depicting self-motion (e.g., crawling across carpet) and 9 depicting caused-motion (e.g., throwing the ball across the net)—each with a distinct manner and path. The 18 scenes were organized into 3 sets with six items in each (3 self-motion events, 3 caused-motion events), and the presentation order of the sets was counterbalanced across participants. We also randomized the presentation order of the six items within each set across participants (see Figure 1 for two sample events and Table 2 for a full list of events). The participants were asked to describe the scenes while using their hands naturally (i.e., *Tell me what is happening in this video while also using your hands as naturally as possible*). Our decision to ask participants to use their hands was based on earlier work that instructing speakers to gesture during a task increases the number of gestures but does not change the nature of those gestures (Cook et al., 2010, Özçalışkan et al., 2016a, 2018). This allowed us to elicit comparable amounts of gesture production in the two languages across all speakers. Monolingual participants described the scenes only once in English. Bilingual participants described the scenes twice, once in English and once in Persian. Half of the bilingual participants completed the task in

English first and the other half in Persian first to account for possible order effects in patterns of speech and gesture production in the two languages

5A. Self-motion



5B. Caused-motion

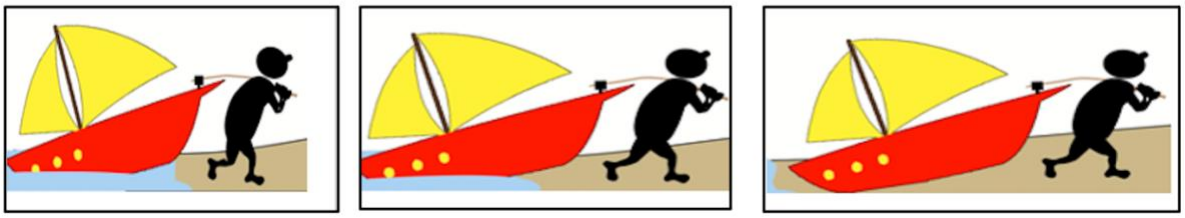


Figure 5. Snapshots from sample scenes depicting self-motion events (Crawl over the carpet; 5A panels) and caused-motion events (e.g., Dragging the boat out of the water; 5B).

Table 4. List of stimulus events

Self-motion	**Caused-motion**
Set 1	**Set 1**
Run into the house	Hammer nail into the wall
Fly out of the trashcan	Pull yellow flower out of pot
Flip over log	Throw the ball over the net
Set 2	**Set 2**
Tumble into the trashcan	Push stroller into the store
Crawl out of the house	Dig treasure out of the ground
Jump over hurdle	Sprinkle confetti over the cake
Set 3	**Set 3**
Tumble out of treehouse	Drag boat out of lake
Climb into the treehouse	Submerge the boat into the tub
Crawl over carpet	Slide Frisbee across the fence

6.3 Transcription, coding, reliability

All speech produced by the participants was transcribed and segmented into sentence units by native speakers of each language. A sentence unit was defined as a segment of speech that contained at least one verb and its associated arguments and subordinate clauses (e.g., Adam CRAWLS over the carpet", "Arya bar rooye farsh MIKHAZAD= Arya above carpet CRAWLS), following earlier work (Özçalışkan et al., 2016a, 2018). All gestures that accompanied each sentence unit were also coded. Gesture was defined as a communicative hand or body movement that conveyed information about the stimulus events (e.g., wiggling fingers left to right to convey running, placing a cupped hand on the left of the body to convey house; Özçalışkan, 2016; Özçalışkan et al., 2016a, 2018).

Each sentence unit was further coded for the packaging of semantic elements in speech and co-speech gestures, following guidelines outlined in earlier work (Özçalışkan et al., 2016b,

2018). For packaging of semantic elements, each sentence-unit was categorized as either conflated (i.e., both manner and path described within a single clause or gesture) or separated (i.e., manner and path described in different clauses or gestures). A sentence unit was classified as separated when it contained manner-only, path-only, or when path and manner were conveyed in two separate clauses or two sequential gestures (see Figure 2 and Figure 3 for sample gestures produced by participants in each group when describing self vs. caused-motion events).

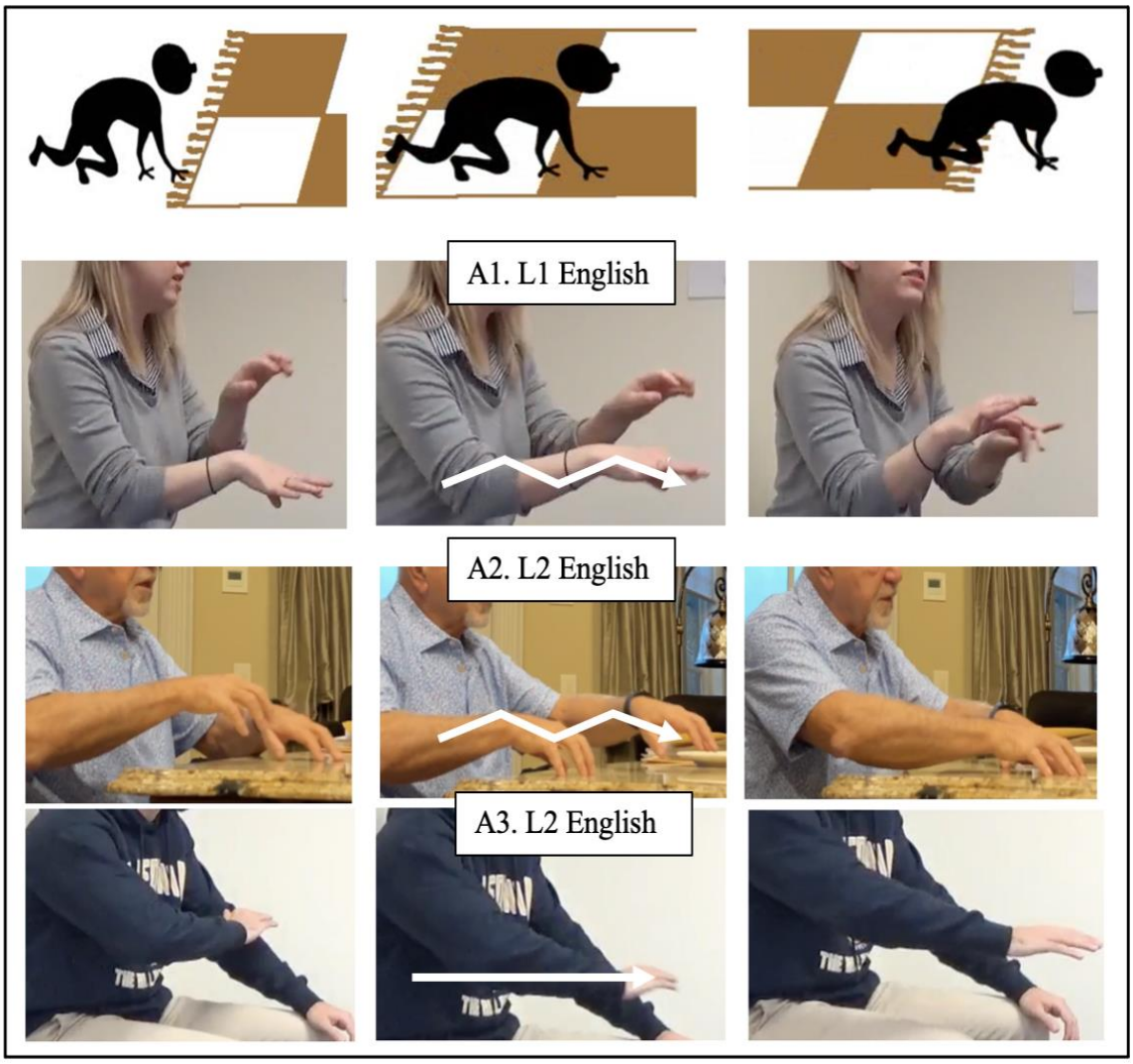


Figure 6. Screenshots from the sample scene of "crawling over the carpet" (top panel)

for self-motion events and its description in gesture by a monolingual English speaker in L1 English (A1), by a bilingual with high proficiency in L2 English (A2), and by a bilingual with low proficiency in L2 English (A3). English monolinguals and bilinguals with high proficiency expressed manner and path simultaneously in one conflated clause in speech and gesture ("[He] is crawling over the carpet" + walking hands (manner) in forward trajectory (path). Bilinguals with low proficiency expressed the same using a separated expression in both speech and gesture ("He passes from this carpet," + moving right hand in rightward trajectory (path).

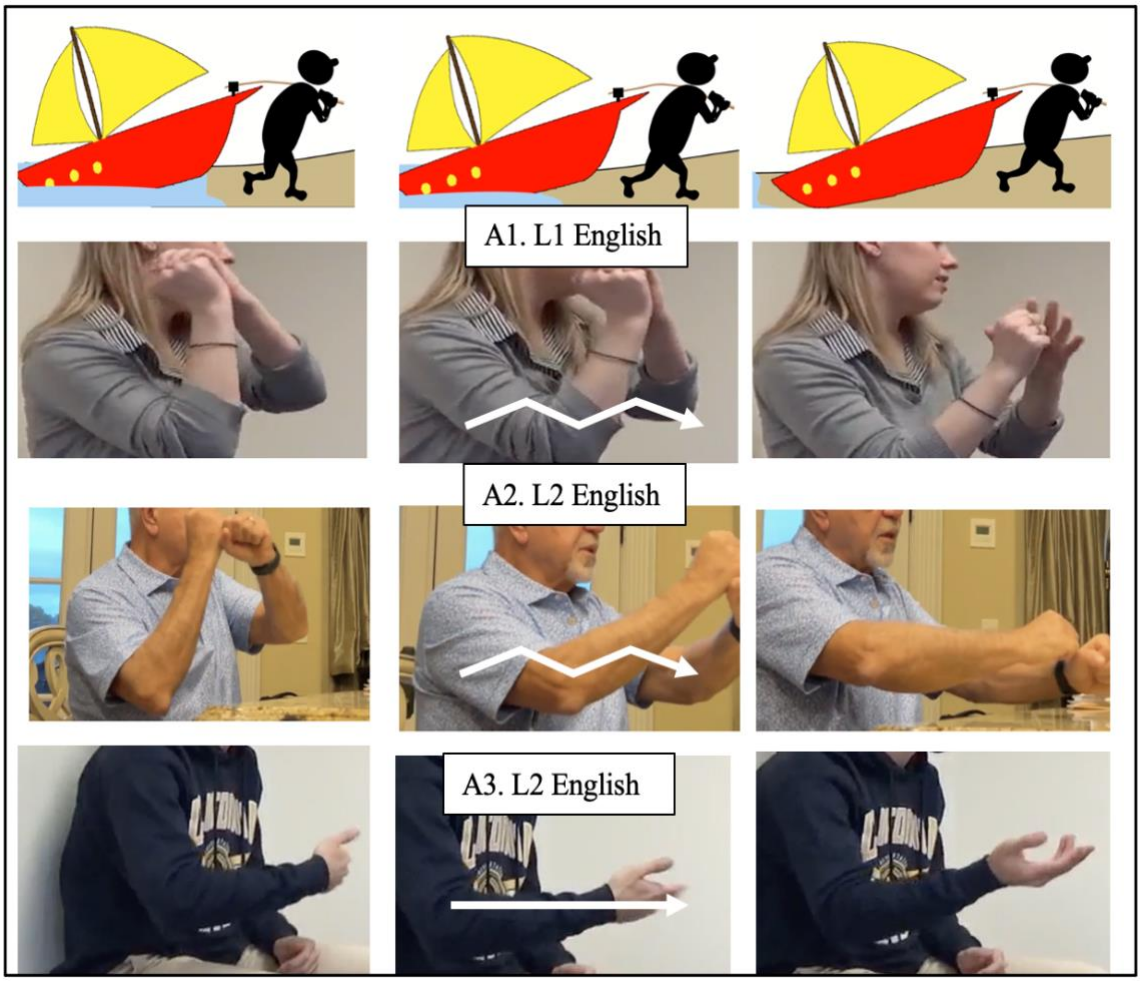


Figure 7. Screenshots from the sample scene of dragging the boat out of water (top panel) for the caused-motion events and its description in gesture by a monolingual English speaker in L1 English (A1), by a bilingual with high proficiency in L2 English (A2) and by a bilingual with low proficiency in L2 English (A3). English monolinguals and bilinguals with high proficiency expressed manner and path simultaneously in the same gesture by moving their hands backward away or from left to right(path), dragging the imaginary boat forcefully(manner); however, bilinguals with low proficiency expressed the same scene by two separate gestures: one by dragging their hands forcefully (manner), then pause and move B hands from left to right (path).

6.4 Reliability

All gestures and speech produced by speakers in each language were coded by native speakers of the language, who were trained for speech and gesture coding. Intercoder reliability was established by additional coders in each language—also native speakers—who coded a randomly selected subsample of 20% of the speech, and co-speech gestures. Agreement between coders was 98% for identifying gestures, 97% for describing gesture form (i.e., hand shape of gesture), and 98% and 96% for coding motion elements in speech and gesture, respectively.

6.5 Statistical Analysis

Each participant's responses were tabulated for the packaging of semantic elements as separated vs. conflated, separately for speech and gesture. In some cases, the participants produced both a separated and a conflated gesture within the same sentence-unit (i.e., a mixed pattern), which accounted for roughly 0.02% of sentence-units across proficiency levels; all of these instances were excluded from our analysis for packaging as they could not be classified as separated or conflated. The tabulated raw scores were then analyzed using two separate sets of Generalized Linear Mixed Models (GLMMs). The nature of our data, namely nested count data, rendered GLMMs as the appropriate statistical method for analysis as it allowed us to control random effects (i.e., subject variability) as well as potential non-normality in the distribution of our dependent variables. GLMMs also allowed us to account for the hierarchical structure of our data by modeling both fixed effects (i.e., event type, language type, proficiency level) and random effects (e.g., individual differences), while also handling overdispersion and accommodating different distributions suited for count data, such as Poisson or Negative Binomial distributions (Baayen et al., 2008; Bolker et al., 2009; Jaeger, 2008).

We first analyzed productions only in English, comparing L1 English speakers, L2 English speakers with high proficiency, and L2 English speakers with low proficiency. Proficiency (native, high, low), event type (self-motion, caused-motion), and packaging type (conflated, separated) were entered as fixed effects, with participant ID included as a random intercept to account for individual variability. Separate GLMMs were conducted for speech and gesture data. Dummy coding was applied to each categorical predictor: native speakers served as the reference level for the proficiency variable, self-motion served as the baseline for the event type variable and separated clauses served as the baseline for packaging. Data preprocessing and transformation were conducted using Python (Python Software Foundation, 2023), with Pandas (McKinney, 2010) and NumPy (Harris et al., 2020) used for data manipulation. Visualization of descriptive and model-related trends was performed using Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021). Scikit-learn (Pedregosa et al., 2011) was used to standardize predictors where necessary to improve model convergence and interpretability. All GLMMs were implemented using the Statsmodels package (Seabold & Perktold, 2010).

We next focused only on the two groups of bilinguals, comparing patterns in L1 (Persian) and L2 (English) to determine the effect of proficiency and event type on patterns of speech and gesture production in speakers' first vs. second language. We ran another set of GLMMs, including proficiency (high, low), packaging (conflated, separated), language (Persian, English), and event type (self, caused) as fixed variables, with subject (individual variability) as a random effect. We did dummy coding, considering one level of each variable as a baseline running our regression. We later reported the main effect for each variable as well as the interaction between them using coefficients, standard errors, z-scores, and p-value for each main effect or interaction term. We used z-scores in our reporting of the statistical results as it allowed for a more readily

available comparison of results across different GLMMs, especially when the variables or scales included in the analysis differed between the models. The z-scores standardized the effect sizes, allowing us to compare the effects of different fixed variables in a standardized manner.

7 RESULTS

7.1 Patterns of speech and gesture production in L2 English

We first examined *speech production* in English in three groups of speakers: monolingual English speakers, Persian (L1)-English (L2) bilinguals with high proficiency, and Persian (L1)-English (L2) bilinguals with low proficiency (see Table 4, top portion, for a summary of statistical results for patterns of speech production). Our analysis showed no main effect of proficiency (native, high, low; $\beta = 0.14$, $z = 0.18$, $p = 0.85$) or event type (self-motion, caused-motion; $\beta = 0.10$, $z = 0.17$, $p = 0.86$), but a significant main effect of packaging (separated, conflated; $\beta = 0.14$, $z = 0.22$, $p = 0.02$). As can be seen in Figure 2 (panels A_1- A_2), across groups, speakers produced more conflated than separated responses in their speech about motion when speaking English. At the same time, this pattern tended to be more pronounced among speakers with either high English proficiency or native English proficiency, accounting for 64% of the responses in both groups. Speakers with low English proficiency, on the other hand, produced comparable amounts of conflated (49%) and separated (51%) responses.

Turning next to *gesture production* in English in the three groups of speakers, we found patterns similar to speech (see Table 4, bottom portion, for a summary of statistical results for patterns of gesture production). Our analysis showed no main effect of proficiency ($\beta = 0.10$, $z = 0.17$, $p = 0.86$) or event type ($\beta = 0.10$, $z = 0.17$, $p = 0.86$), but a marginal main effect of packaging ($\beta = 0.50$, $z = 1.86$, $p = 0.06$). As Figure 2 (panels B_1- B_2) illustrates, speakers produced more conflated than separated responses in their gestures about motion when speaking English. However, as in speech, this pattern tended to be more evident in the high proficiency (63% conflated responses) and native proficiency (70% conflated responses) groups and less so in the low proficiency group (51% conflated responses). Our analysis also showed a marginal

interaction between event type and packaging ($\beta = 0.86$, $z = 1.86$, $p = 0.07$), highlighting a more pronounced preference for conflated responses when describing caused-motion events.

Overall, the patterns of speech and gesture production when speaking English remained similar, with a preference for the conflated strategy in both speech and gesture. However, this preference tended to be stronger among speakers with native or high proficiency as compared to speakers with low proficiency in English, along with a greater tendency for conflated responses when gesturing about caused-motion.

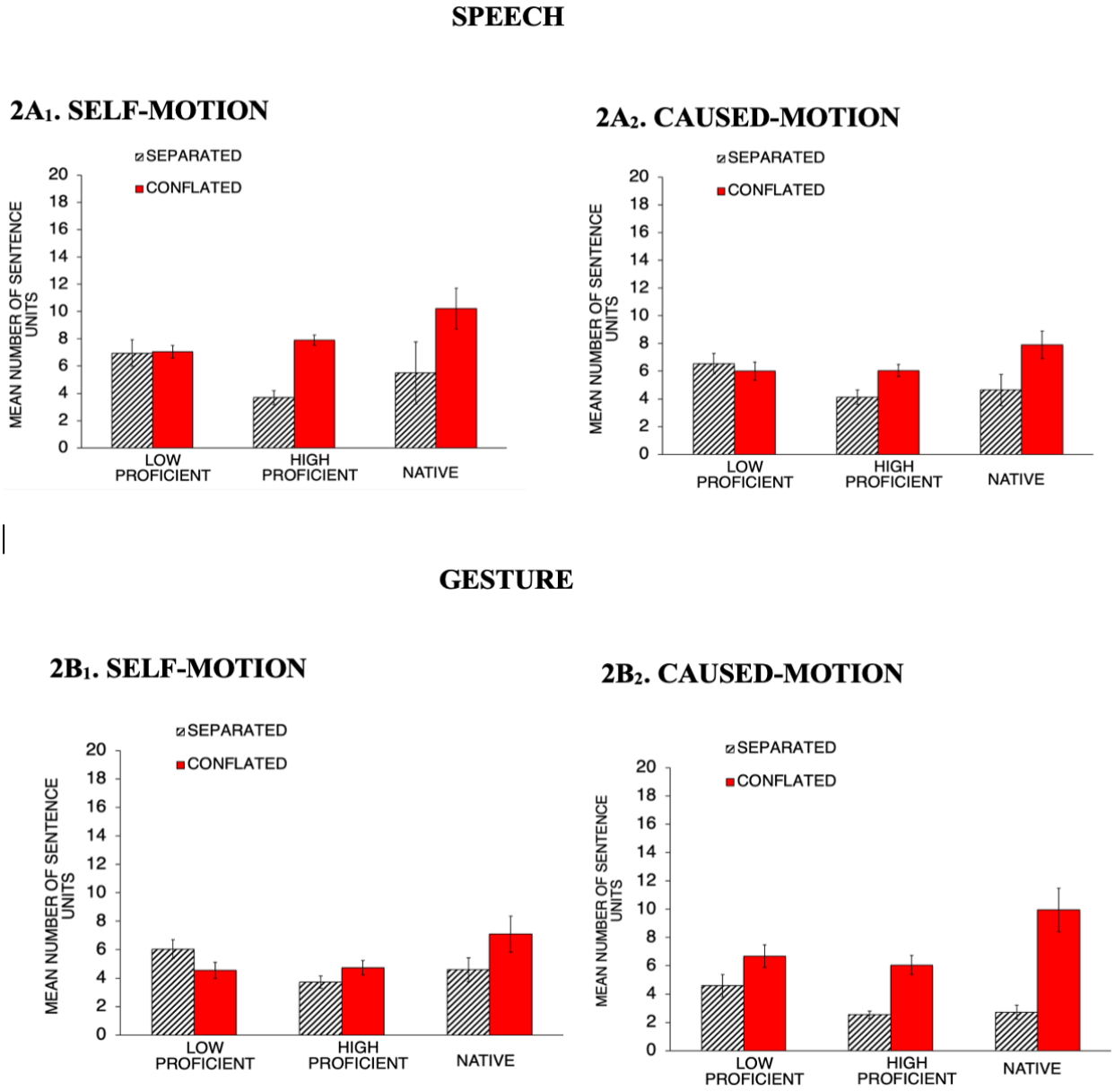


Figure 8. Mean number of sentence units with separated (striped bars) or conflated (solid red bars) packaging in speech (A panels) and gesture (B panels) produced by speakers with native, high, or low proficiency in English when describing self-motion (panels on the left) or caused-motion (panels on the right) events (error bars represent standard error).

7.2 Patterns of speech and gesture production of bilinguals in L1 Persian and L2 English

We next focused only on Persian (L1)-English (L2) bilinguals that differed in their English proficiency and asked whether their patterns of speech and gesture production would differ in their two languages based on their proficiency level (low, high) and the type of event that they described (self-motion, caused-motion).

Beginning with *speech production*, we found differences (see Table 5, top portion for a summary of statistical results for patterns of speech production). Our analysis showed no effect of language, event type, or packaging type, but a main effect of proficiency. As can been seen in Figure 3 (panels A1-A2), speakers with low proficiency produced significantly more motion descriptions than speakers with high proficiency across their productions in L1 and L2 (M_{HIGH}=43.00, SD=5.84 vs. M_{LOW}=48.96, SD=7.91). We also found 2 two-way interactions. One was an interaction between event type and packaging type: bilinguals used more conflated strategies when describing caused-motion than self-motion events ($\beta = 0.22$, $z = 2.41$, $p = 0.01$). the other was an interaction between proficiency and packaging type: Bilinguals with high but *not* low proficiency produced more conflated packaging in speech ($\beta = 0.23$, $z = 0.09$, $p = 0.006$). In addition, there was a three-way interaction between language, proficiency, and packaging ($\beta = 0.71$, $z = 0.24$, $p = 0.004$). Bilinguals with high proficiency produced more conflated responses in their L2 English than bilinguals with low proficiency—a pattern that was more evident in the description of caused-motion events.

Turning to gesture production, we also found differences (see Table 5, bottom portion for a summary of statistical results for patterns of gesture production). Similar to speech, our analysis showed no effect of language, event type, or packaging type, but a main effect of proficiency. As can been seen in Figure 3 (panels B1-B2), speakers with low proficiency

produced more gestures than speakers with high proficiency across their productions in L1 and L2 (M_{HIGH}=34.01, SD=4.82 vs. M_{LOW}=41.22, SD=7.03). We also found 2 two-way interactions: One was a marginal interaction between proficiency and packaging: bilinguals with high but not low proficiency tended to produce greater conflated responses ($\beta = 0.91$, $z = 1.66$, $p = 0.08$). The other was a reliable interaction between proficiency and language: bilinguals with high proficiency producing more gestures in their L2 English than bilinguals with low proficiency ($\beta = 1.01$, $z = 3.28$, $p = 0.001$). In addition, there was a three-way interaction between language, proficiency, and packaging ($\beta = 0.72$, $z = 2.58$, $p = 0.01$) showing that bilinguals with high proficiency produced more conflated gestures in their L2 English than bilinguals with low proficiency, a pattern that mirrored the patterns observed in speech.

Overall, the patterns of speech and gesture production among bilinguals with high vs. low proficiency showed that bilinguals with lower proficiency produced a greater amount of speech and gestures in their L2 English. There was also an interaction between language type, proficiency level, and packaging type: bilinguals with higher proficiency produced more conflated clauses in their L2 English—a pattern that was reflected in their gestures as well.

Table 6. Summary of statistical results for speech and gesture production of bilinguals in their L1 English and L2 Persian

	Coefficient	Standard Error	z	*p*
SPEECH				
Proficiency	**0.13**	**0.4**	**2.91**	**0.004**
Event Type	0.08	0.19	0.43	0.66
Packaging	0.12	0.19	0.61	0.54
Language	0.07	0.19	0.37	0.71
Proficiency * Event Type	0.13	0.09	1.53	0.12
Event Type * Packaging	**0.22**	**0.09**	**2.41**	**0.01**
Proficiency *Packaging	**0.23**	**0.09**	**2.77**	**0.006**
Language * Proficiency	0.13	0.09	1.54	0.12
Language * Packaging	0.01	0.09	0.16	0.86
Language * Event Type	0.14	0.09	1.53	0.13
Proficiency * Event Type * Packaging	0.08	0.19	0.43	0.66
Language * Proficiency * Packaging	**0.71**	**0.24**	**2.91**	**0.004**
Language * Packaging * Event Type	0.72	0.60	1.20	0.22
Language * Packaging * Event Type * Proficiency	0.02	0.12	0.22	0.82
GESTURE				
Proficiency	**1.38**	**0.64**	**2.15**	**0.03**
Event Type	0.79	0.67	1.19	0.23
Packaging	1.13	0.97	1.17	0.24
Language	0.61	0.66	0.9	0.36
Proficiency * Event Type	0.11	0.31	0.35	0.73
Event Type * Packaging	0.76	0.61	1.25	0.21
Proficiency *Packaging	**0.91**	**0.55**	**1.66**	**0.08**
Language * Proficiency	**1.007**	**0.31**	**3.28**	**0.001**
Language * Packaging	0.61	0.53	1.31	0.26
Language * Event Type	0.02	0.31	0.01	0.99
Proficiency * Event Type * Packaging	0.15	0.34	0.46	0.54
Language * Proficiency * Packaging	**0.72**	**0.27**	**2.58**	**0.01**
Language * Packaging * Event Type	0.05	0.31	0.17	0.86
Language * Packaging * Event Type * Proficiency	0.08	0.14	0.57	0.57

z= distance from the mean in standard deviations, *$p < .05$. **$p < .01$. ***$p < .001$; significant and marginal effects are bolded.

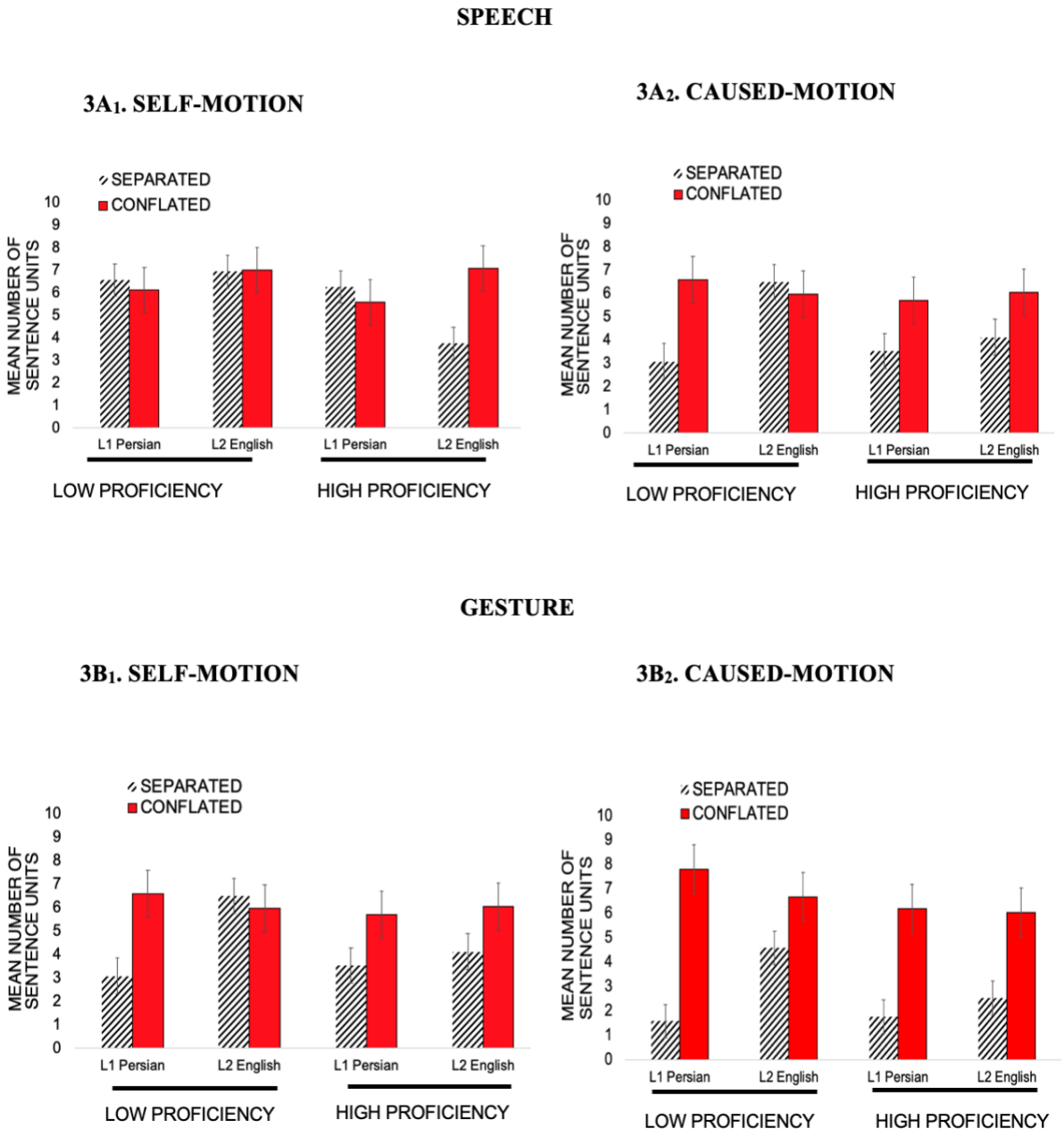


Figure 9. Mean number of numbers of sentence units with separated (striped bars) or conflated (solid red bars) packaging in speech (A panels) and gesture (B panels) produced by bilingual speakers with high vs low proficiency in English when describing self-motion (panels on the left) or caused-motion (panels on the right) events in their first language (L1 Persian) or in their second language (L2 English), error bars represent standard error).

8 DISCUSSION

Gesture largely follows language-specific patterns in speech in first language production. In this study, we asked whether this extends to second language production, and if so, how speaker proficiency, event type, and language type might impact these patterns. Our comparison of the speech and gestures produced by 44 Persian (L1)-English (L2) bilingual speakers—22 with high and 22 with low L2 proficiency—as compared to 22 monolingual English speakers in a motion animation task showed differences. First, comparing *productions only in English* by proficiency and event type, we found that gestures followed the patterns observed in speech across proficiency levels and event types, with a preference for the conflated pattern (i.e., the English monolingual pattern) that synthesizes manner and path components of motion into a single clause or a single gesture. Next, focusing on *productions in both English and Persian* among bilinguals by proficiency level, event type, and language type, we found interactions between the three factors. Overall bilinguals with high proficiency produced more conflated gestures in their L2 English than bilinguals with low proficiency, particularly when describing caused-motion events, mirroring the patterns observed in their speech.

8.1 Talking about motion events in English: effect of proficiency and event type in speech and gesture

Starting with **proficiency**, both groups of bilinguals, particularly those with high proficiency, used mostly conflated patterns in their L2 speech, thus aligning with the speech patterns of native English speakers. This finding is in line with prior research that also showed attainment of target L2 patterns in speech among advanced bilinguals (e.g., Cadierno & Ruiz, 2006; Özçalışkan, 2016; Özyurek, 2002). In earlier work, bilinguals with intermediate or low proficiency were also shown to exhibit difficulties in fully attaining target L2 speech patterns

(e.g., Negueruela et al., 2004). However, this was not completely evident in our study. The bilinguals with low proficiency did not differ reliably from the other groups, but they did tend to rely on the conflated strategy at a lower rate than speakers with high or native English proficiency. One possible explanation for the lack of a reliable proficiency effect in our study might be that all bilingual participants—high or low in proficiency— had passed the English proficiency requirements (TOEFL) and had high scores (B2-C1) in their L2 English and were enrolled in accredited U.S. institutions. Both groups of bilinguals also lived in an English-speaking context (i.e., USA) for an extended amount of time, which, in turn, might have also aided them further in attaining target L2 patterns in speech.

These patterns were largely evident in gesture as well. The gestures produced when speaking English showed a marginal effect of packaging ($p = 0.06$), in line with the patterns observed in speech. Both bilingual groups produced more conflated gestures, mirroring the patterns of native English speakers. Also similar to speech, these patterns tended to be more pronounced among bilinguals with high proficiency than those with low proficiency. These findings align with previous research (Özyürek, 2002), which suggests that only bilinguals with high proficiency and prolonged L2 exposure are likely to adopt target L2 patterns in gesture. It however, contrasts with some other work, which showed that while advanced bilinguals' L2 speech fully reflected target L2 patterns, their gestures still showed characteristics of their L1 (Özçalışkan, 2016; Stam, 2010). The difference in results might be attributable to the length of residency in an English-speaking environment. The participants in Özçalışkan's study had relatively shorter length of residency in the United States (4+ years) compared to our study in which, bilinguals with high proficiency had an average residency of 17-18 years. As such, our

study highlights length of residency as an important factor that contribute to bilinguals' attainment of target gesture patterns in their L2.

Overall, our findings on proficiency show that bilinguals largely followed L2 speech patterns, but only bilinguals with high proficiency fully attained the L2 patterns in speech. The gestures produced by each group mirrored the patterns observed in speech, also with greater attunement to L2 patterns in gesture among bilinguals with high proficiency compared to the ones with low proficiency.

Turning to **event type,** we found no effect of event type on patterns of speech production across English speakers with native, high, or low proficiency. These findings contrast earlier work with monolinguals, which showed greater encoding of manner and consequently more conflated descriptions in caused-motion than self-motion events For example, comparisons between S-framed languages (e.g., English, German) and V-framed languages (e.g., Spanish, Turkish) have shown that all groups, regardless of language typology, rely more on conflated strategies in speech when describing caused-motion than self-motion events (Hendricks & Hickmann, 2015; Özçalışkan, 2000, 2005). The lack of a difference might be attributable to the stimulus we used in our study: both self- and caused-motion events depicted events with unique manner and path information, thus increasing the salience of both components of motion, resulting in more conflated expressions not only in caused- but also in self-motion events

The gestures followed mostly similar patterns, with no effect of event type. However, different from speech, we observed a marginal interaction between packaging and event type. Speakers tended to use more conflated strategies in their gestures when describing caused-motion than self-motion events across groups ($p = 0.07$). This aligns with findings from studies

(e.g., Choi & Lantolf, 2008, Park, 2020, Lewandowski and Özçalışkan. 2021, which showed similar increases in the use of the conflated strategy when describing caused-motion events.

Overall, our findings on event type show that both speech and gestures follow largely similar patterns, with no effect of event type on production. It also shows a marginal difference between the two modalities, with greater use of conflated strategies in caused-motion events in gesture but not in speech—a difference that might be an outcome of the more embodied nature of the gestural modality (i.e., allowing more extended expression of different manners) compared to speech.

Looking **across proficiency and event types**, our findings align with Talmy's (1985, 2000) typology, which classifies English as a satellite-framed language, favoring conflated manner-path constructions (e.g., *run into the house*). It also provides further support for the positioning of the Persian language as closer to a V-language, given the lower expression of conflated strategies in L2 English among bilinguals with low proficiency, particularly compared to the other two groups. The lower use of the conflate strategy might be an outcome of the L1 effect on L2 production within the low or intermediate proficiency group – a pattern that is consistent with earlier work with bilinguals with low proficiency in different languages (e.g., Aktan-Erciyes et al., 2023; Cadierno, 2010; Negueruela et al., 2004; Stam, 2006, 2010).

Persian has mostly been classified as a V-language in earlier work, with a preference for the separated strategy in speech (e.g., Feiz, 2011; Khorvash & Lotfi, 2022).

Our findings showing close alignment between speech and gesture also provide support for theories such as Thinking-for-Speaking (TFS; Slobin, 1996, 2003), which posits that the way a person thinks while speaking is shaped by the grammatical and lexical categories available in their language. Our study showed that bilinguals with high proficiency—but not those with low

proficiency—are able to fully adopt L2 patterns in both their speech and gestures, suggesting that TFS is evident not only in native but also in more advanced non-native speech production contexts. In addition, our findings corroborate with theories of gesture-speech integration (e.g., McNeill, 1992, 2005; Kita & Özyürek 2003), which suggest that gestures and speech form an integrated system originating from the same underlying conceptual structure. Only with advanced proficiency in L2, coupled with immersion in the language environment for an extended amount of time, were bilinguals able to fully adopt the target patterns of their L2.

8.2 Talking about motion events in Persian and English: effect of proficiency, language type and event type in speech and gesture production

Starting with proficiency, we found that bilinguals with lower proficiency produced more speech in their L2 compared to those with high proficiency. This finding aligns with previous research on bilingual speech production in L1 and L2, which highlights how bilinguals tend to be more descriptive when explaining motion events in their weaker language (Nicoladis, Nagpal, Marentette, & Hauer, 2018). One possible explanation for this difference might be the relative weaknesses in vocabulary size between the two groups. Bilinguals with lower proficiency may lack the vocabulary necessary to label different entities (e.g., characters, objects; Bosch, 1983; Garrod, 2001; Lyons, 1977; So, Kita, & Goldin-Meadow, 2009) and compensate for it by using more extended verbal descriptions (Johns, Sheppard, Jones, & Taler, 2016). This effect is particularly pronounced for bilinguals with lower L2 proficiency as shown in a study by So, Kita, and Goldin-Meadow (2013). In this earlier study, researchers examined Mandarin (L1)-English (L2) bilinguals and found that those with low—but not high—proficiency produced a greater amount of speech when speaking their L2 English. Another possible explanation could be tied to the chronicle style of reporting in English language (Nicoladis et al., 2018). English, as a

more “descriptive” language, emphasizes what actually happened, whereas languages such as Greek—another V-language, tend to focus on “why” an event occurred (Ryan, 1993; Tannen, 1980). Persian might follow a similar pattern as Greek (Ryan, 1993), making speech production more concise compared to the descriptive nature of English.

Proficiency had similar effects on gesture among bilinguals. Bilinguals with low proficiency also produced more gestures in their L2—akin to their speech—and also consistent with findings from previous studies (Nicoladis, 2007; Nicoladis, Pika & Marentette, 2009; Nicoladis et al., 2018; So, 2010; Özçalışkan, 2016; Pika et al., 2006; So et al., 2013). These earlier studies showed that bilinguals, especially those with lower proficiency, gestured more frequently in L2 English than monolingual English speakers.

Taken together, our findings highlight the significant role proficiency plays in shaping both speech and gesture production in bilinguals. Bilinguals with lower proficiency tended to compensate for their difficulties in L2 by producing more speech, also accompanying them with more gestures. Moreover, our findings showed that gesture and speech form an integrated system, exhibiting similar patterns in the amount of speech and gesture production in different groups of bilinguals with different proficiency levels. This finding furthers prior work that showed a strong coupling between gesture and speech production across various tasks in first-language production contexts (e.g., Chu & Kita, 2011; Hostetter, Alibali, & Kita, 2007; Theocharopoulou et al., 2015).

Turning to **event type,** we found no effect of event type on speech production. However, both bilingual groups produced more conflated responses in describing caused-motion events regardless of the language they were speaking, as indicated by an event type and packaging interaction. These patterns align with earlier studies (e.g., Lewandowski & Özçalışkan, 2018) on

monolinguals, which showed that while speakers of different languages display cross-linguistic differences in self-motion events, these differences dissipate for caused-motion events, as all speakers tend to converge on the conflated strategy. One possible explanation could be the nature of the caused-motion events, particularly compared to self-motion events. Caused-motion events involve additional information about force dynamics and how the force was applied to move an object or figure, making the inclusion of manner information in the verb a greater necessity, and consequently increasing the likelihood of conflated strategies in caused-motion event descriptions.

The gestures followed the same patterns as speech, showing no effect of event type. However, different from speech, there was no interaction between packaging and event type in gesture. One possible explanation could be the prolonged influence of L1 (Persian) on L2 (English) gesture production. As suggested by earlier work, bilinguals' gestures in both L1 and L2 become more similar as proficiency increases, (Brown & Gullberg, 2008), which might have led to the preference for the conflated pattern in general, regardless of event type in our study.

Next turning to **language type (**Persian vs. English), we found no effect of language. However, there was a significant interaction between language, proficiency, and packaging in speech. Specifically, bilinguals with high—but not low— proficiency produced more conflated packaging in their L2 English compared to their L1 Persian, highlighting the strong influence of proficiency on event packaging in L2 speech. This finding aligns with previous research (Özyürek, 2002; Özçalışkan, 2016; Cadierno & Ruiz, 2006), which demonstrates how proficiency in L2 can shape the linguistic structuring of motion events and how only bilinguals with higher proficiency can fully switch between their two languages. In addition, our findings

are consistent with Aveledo & Athanasopoulos (2023), who showed that high proficiency and increased length of residency——predict both L1 and L2 language patterns.

The gestures followed similar patterns, with no effect of language, but a three-way interaction between language, proficiency, and packaging. Specifically, bilinguals produced more conflated gestures in their L2 English compared to their L1 Persian, suggesting that language proficiency had similar effects on both speech and gesture production.

These findings underscore the impact of language typology and proficiency levels on both speech and gesture, reinforcing the idea that increased proficiency in an L2 with a different typology leads to a shift toward the target language's event packaging patterns (in this case shifting from a V-framed to S-framed). The consistency between speech and gesture further highlights the close relationship between linguistic and non-linguistic encoding of motion events. Future research could examine whether these patterns extend to other language pairs and whether similar effects emerge in bilinguals with varying degrees of immersion in their L2 environment.

In summary, looking across **proficiency, event type and language** in this two-part study we had two sets of findings: Examining productions in English by proficiency and event type showed that bilinguals, especially those with higher proficiency adhered to L2 target patterns in both speech and gesture, suggesting that only with high level of proficiency bilinguals can fully align their multimodal expressions with L2 norms. Examining productions in both English and Persian by proficiency, event type, and language showed an interaction between these three factors: Bilinguals with high but not low proficiency used more conflated strategies when describing caused-motion events both in speech and gesture.

Overall, these results align well with prior research demonstrating that speech and gesture form an integrated system in bilingual motion event encoding (Hostetter et al., 2007; Kita &

Özyürek, 2003). They also support findings that gesture aligns more closely with speech as proficiency and length of residency increase (Özyürek, 2002; Aveledo & Athanasopoulos, 2023). Our findings further contribute to studies such as Brown and Gullberg (2008), and Gullberg (2008), reinforcing the idea that higher proficiency and immersion in L2 promote greater adherence to L2-specific structures rather than a bidirectional influence between L1 and L2.

However, these results also challenge Brown's (2015) "convergence" hypothesis, which suggests that bilinguals develop hybrid gesture patterns distinct from both L1 and L2. At the same time, one major limitation of our study is the lack of a Persian monolingual group, which restricts our ability to determine whether bilinguals' use of conflated patterns in L2 English largely stems from proficiency-driven adaptations or possibly typological similarities between Persian and English. Geopolitical constraints prevented ethics committee approval for data collection in Iran, making it impossible to establish a baseline for Persian monolingual multimodal patterns. Future studies that incorporate Persian monolingual participants can further clarify the influence of L1 structure on bilingual motion event encoding.

The greater amount of gesture production among bilinguals with lower proficiency also showed patterns consistent with existing theories of gesture-speech interaction. For instance, Krauss, Morrel-Samuels, and Colasante (1991) and Butterworth and Hadar (1989) propose that gestures become more frequent in situations where speech difficulties arise. The lexical retrieval theory of gesture production also further suggests that gestures, particularly iconic gestures, facilitate lexical access and enhance speech fluency (Krauss, Chen, & Gottesman, 2000; Rauscher, Krauss, & Chen, 1996). Both theories suggest possible explanations for the greater amount of gesture production among bilinguals with low proficiency in our study.

In conclusion, our findings contribute to the growing body of research on bilingual motion event encoding, highlighting that speech and gesture function as an integrated system across proficiency levels, event types and language types in bilinguals. More specifically, our findings show that the patterns of similarities and differences observed in speech for each of the key factors are also mostly observable in gesture. These findings further highlight the importance of gesture as an additional communicative tool than can shed further light on the communication skills of both monolingual and bilingual speakers.

9 OVERALL CONCLUSION

Research in first language (L1) production contexts has shown a strong connection between speech and gesture use in adults (Kita et al., 2007). Monolinguals use gestures when they speak, with gestures conveying information that is semantically related to speech (Goldin-Meadow, 2007; Hostetter, 2011). Speakers of different languages also systematically differ in their expression of key domains, such as motion in space. They follow language-specific patterns in both speech and co-speech gestures in describing motion, further highlighting the tight link between the two modalities across different languages (e.g., Kita & Özyürek, 2003; Özçalışkan, Lucero, & Goldin-Meadow, 2016a, 2018; Talmy, 2000). However, relatively less is known about the relationship between gesture and speech patterns in L2 production; and the few existing studies present largely inconclusive results.

In this project, we designed two studies, both examining the speech and gestures of 66 adult speakers, including 22 English (L1)-Persian (L2) bilinguals with high proficiency, 22 bilinguals with low proficiency, and 22 English native speakers (monolinguals). In Study 1, we analyzed multimodal patterns of speech and gesture production by proficiency (native, high, low) and discourse context (narratives, explanations), focusing on the amount, diversity, and complexity of production. In Study 2, we examined language-specificity in multimodal production by proficiency (native, high, low), event type (self-motion, caused-motion), and language type (English, Persian). The results of both studies highlight the tight link between gesture and speech in L2 contexts.

In the first study, all speakers, regardless of proficiency level, produced more speech in the narrative context—a pattern that was also reflected in their gestures. However, this positive relationship between speech and gesture was reversed for complexity, with bilinguals with lower

proficiency producing less complex speech, accompanied by more complex gestures that supplemented what was conveyed in speech. This aligns with Interface Hypothesis (Kita & Özyürek, 2003), which posits a close coupling between speech and gesture production. Our findings also align well with McNeill's (1992) Growth Point Hypothesis and de Ruiter's (2000) Sketch Model, both of which consider gestures and speech as interrelated systems, allowing one to compensate for the other when one falls short in communication. Our findings also illustrate how context plays an important role in speech and gesture production and how proficiency interacts with discourse context, with bilinguals with low proficiency using more complex gestures in the cognitively more demanding task of providing explanations.

Turning to the second study, we once again found strong coupling between gesture and speech, with production in English following the same pattern (i.e., more conflated strategies) in speech and gesture. Our findings also suggested—as a trend—that only bilinguals with higher proficiency and prolonged L2 immersion (in our case, 17 years) fully adopted the L2 speech and gesture patterns, resembling those of monolingual English speakers. The finding on event type remained inconsistent: when examining productions only in English we did not find an effect of event type, but when examining productions in both English and Persian, we found a greater preference for conflated expression when describing caused-motion as compared to self-motion events, but only in speech. Our stimuli consisted of events with both salient manner and path components. This in turn, might have increased speakers' attention to both components across trials, resulting in overall greater preference for conflated expressions not only in caused but also in self-motion events. Future studies that examine event types with stimuli involving more subtle variations of manner vs. path of motion might shed further light on this possibility.

Overall, the results of both studies demonstrate that gestures and speech form an integrated system not only in first but also in second language production contexts. This relationship can also be influenced by factors such as proficiency levels, discourse context, event type, and the typology of both the first and second languages.

REFERENCES

Alonso, R. (2020). Boundary-crossing events across languages: A study on English Speakers, Spanish speakers and second language learners. *Review of Cognitive Linguistics*, *18*(2), 316-349.

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*(1), 16-48.

Aktan-Erciyes, A., Akbuğa, E., Kızıldere, E., & Göksun, T. (2023). Motion event representation in L1-Turkish versus L2-English speech and gesture: Relations to eye movements for event components. *International Journal of Bilingualism*. *27*(1), 61-86.

Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition and Neuroscience*, *32*(6), 792-804.

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.

Aveledo, F., & Athanasopoulos, P. (2023). Bidirectional cross-linguistic influence in motion event conceptualization in bilingual speakers of Spanish and English. *International Review of Applied Linguistics in Language Teaching*, *61*(1), 13-36.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Bayazidi, A., Ansarin, A. A., & Mohammadnia, Z. (2019). The relationship between syntactic and lexical complexity in speech monologues of EFL learners. *Applied Research on English Language*, *8*(4), 473-488.

Beattie, Geoffrey & Jane Coughlan. 1998. Do iconic gestures have a functional role in lexical access? An experimental study on the effects of repeating a verbal message on gesture production. Semiotica 119(3/4). 221-249.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British journal of psychology*, *90*(1), 35-56.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127-135.

Bowerman, M., Brown, P., Eisenbeiss, S., Narasimhan, B., & Slobin, D. I. (2002). Putting things in places: Developmental consequences of linguistic typology. In The 31st Stanford Child Language Research Forum (pp. 1-29). Center for the Study of Language & Information.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. Bilingualism: Language and Cognition, *12*(1), 3-11.

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in second language acquisition*, *30*(2), 225-251.

Bosch, P. (1983). *Agreement and anaphora: A study of the roles of pronouns in discourse and syntax*. Academic Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Butterworth, G., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Cognition*, *32*(3), 237–244.

Cadierno, T. (2010). Motion in Danish as a second language: Does the learner's L1 make a Difference. Linguistic relativity in second language acquisition: Thinking for speaking, 1-33.

Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics*, *4*, 183–216.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International journal of language & communication disorders*, *49*(6), 649-671.

Cocks, N., Morgan, G., & Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture*, *11*(1), 24-39.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*(4), 465-475.

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sub lexical levels across languages. *Psychonomic Bulletin & Review*, *13*(6), 972-977.

Chen, L., & Guo, J. (2009). Motion events in Chinese novels: Evidence for an equipollent-framed language. *Journal of pragmatics*, *41*(9), 1749-1766.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83-121.

Choi, S., & Lantolf, J. P. (2008). Representation and embodiment of meaning in L2 communication: Motion events in the speech and gesture of advanced L2 Korean and L2 English speakers. *Studies in Second Language Acquisition*, *30*(2), 191-224.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102-116.

Cruz, A. (2021). A syntactic approach to gender assignment in Spanish–English bilingual speech. *Glossa: A Journal of General Linguistics*, *6*(1).

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and cognition*, *10*(1), 7-21.

Demir, Ö. E., Levine, S., & Goldin-Meadow, S. (2015). A tale of two hands: Children's early gesture use in narrative production. *Journal of Child Language*, *42*(03), 662–681.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge University Press.

Educational Testing Service. (2024). *TOEFL iBT® test prep planner*. Retrieved from https://www.ets.org/toefl.

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

Feiz, P. (2011). Traveling through space in Persian and English: a comparative analysis of motion events in elicited narratives. *Language Sciences*, *33*(3), 401-416.

Garrod, S. (2001). Anaphora resolution. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavior sciences* (pp. 490–494). Amsterdam: Elsevier.

Goldin-Meadow, S. (2007). Gesture with speech and without it. *Duncan Cassell Levy*, 31-50.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Erlbaum.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual review of psychology*, *64*(1), 257-283.

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In C. Dimroth & M. Starren (Eds.), *Information structure, linguistic structure, and dynamics of language acquisition* (pp. 311–328). Amsterdam: Benjamin.

Gullberg, M. (2006). Handling discourse: gestures, reference tracking, and communication strategies in early L2. *Language Learning*, *56*(1), 156–196.

Gullberg, M. (2008). Gestures and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition* (pp. 286-315). Routledge.

Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, *28*(2), 200-236.

Hanna, W., & Barbera, J. (Producers). (1940–present). *Tom and Jerry* [TV series]. Warner Bros.

Han, Z. (2004). Fossilization: five central issues. *International Journal of Applied Linguistics*, *14*(2), 212-242.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. *Nature*, *585*, 357–362 (2020).

Hendriks, H., & Hickmann, M. (2015). Finding one's path into another language: On the expression of boundary crossing by English learners of French. *The Modern Language Journal*, *99*(S1), 14-31.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological bulletin*, *137*(2), 297.

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.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, *9*(3), 90–95.

Ibarretxe-Antuñano, I. (2004). Language typologies in our language use: The case of Basque motion events in adult oral narratives. *Cognitive Linguistics*, *15*(3), 317-349.

Iakovleva, T. (2012). Typological constraints in foreign language acquisition: The expression of voluntary motion by upper intermediate and advanced Russian learners of English. *Language, Interaction and Acquisition*, *3*(2), 231-260.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*, 703.

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.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Khorvash, Z., & Lotfi, A. (2022). Spatial Encoding in English and Persian: Typological Influences on Second Language Acquisition. *International Journal of Foreign Language Teaching and Research*, *10*(41), 49-63.

Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (Vol. 2, pp. 261–283). Cambridge University Press.

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.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*(8), 1212–1236.

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.

Krauss, R. M., Morrel-Samuels, P. & Colasante, C. (1991). Do conversational gestures communicate? *Journal of Personality and Social Psychology*, *61*, 743–754.

Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, *17*(1), 139-152.

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.

Lyons, J. (1977). *Semantics 2*. Cambridge University Press.

Lewandowski, W., & Özçalışkan, Ş. (2021a). How language type influences patterns of motion expression in bilingual speakers. *Second Language Research*, *37*, 27–49.

Lewandowski, W., & Özçalışkan, Ş. (2021b). The specificity of event expression in first language influences expression of object placement events in second language. Studies in Second Language Acquisition, 43(4), 838-869.

Lewandowski, W., & Özçalışkan, Ş. (2018). How event perspective influences speech and co-speech gestures about motion. *Journal of Pragmatics*, *128*, 22-29.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (Ed.). (2000). *Language and gesture* (Vol. 2). Cambridge University Press.

McNeill, D. (2005). *Gesture and thought.* University of Chicago Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (3rd ed.)*. Lawrence Erlbaum Associates.

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.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2017). How what we see and what we know influence iconic gesture production. *Journal of nonverbal behavior*, *41*, 367-394.

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), Proceedings of the 9th Python in Science Conference (pp. 51–56).

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*(4), 1224–1236.

Naigles, L. R., & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science*, *9*(5), 363-369.

Negueruela, E., Lantolf, J. P., Jordan, S. R., & Gelabert, J. (2004). The "private function" of gesture in second language speaking activity: A study of motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics*, *14*(1), 113-147.

Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics*, *28*, 441–454.

Nicoladis, E. (2024). Chapter 17. Bilinguals' lexical choice in storytelling: Testing the weaker-links hypothesis. In E. Babatsouli (Ed.), *Multilingual Acquisition and Learning: An ecosystemic view to diversity* (pp. 452-472). John Benjamins Publishing Company.

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. Developmental psychology, *35*(2), 514-522.

Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked to story-telling style: Evidence from bilinguals. *Language and Cognition*, *10*(4), 641-664.

Nicoladis, E., Pika, S., & Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children? *Journal of Psycholinguistic Research*, *38*, 573–585.

Nicoladis, E., Pika, S., Yin, H., & Marentette, P. (2007). Gesture use in story recall by Chinese–English bilinguals. *Applied Psycholinguistics*, *28*, 721–735.

Oller, D. K., & Eilers, R. E. (Eds.). (2002). *Language and literacy in bilingual children* (*2*). Multilingual Matters.

Olson, D. J. (2024). A systematic review of proficiency assessment methods in bilingualism Research. *International Journal of Bilingualism*, *28*(2), 163-187.

Özçalışkan, S. (2002). Metaphors we move by: A crosslinguistic-developmental analysis of metaphorical motion events in English and Turkish. University of California, Berkeley.

Özçalişkan, Ş. (2005). Metaphor meets typology: Ways of moving metaphorically in English and Turkish. Cognitive Linguistics, 16(1), 207-246.

Özçalışkan, Ş. (2007). Metaphors we move by: Children's developing understanding of Metaphorical motion in typologically distinct languages. *Metaphor and symbol*, *22*(2), 147-168.

Özçalışkan, Ş. (2016). Do gestures follow speech in bilinguals' description of motion? *Bilingualism: Language and Cognition*, *19*(3), 644-653.

Özçalışkan, Ş., & Slobin, D. I. (1999). Learning how to search for the frog: Expression of manner of motion in English, Spanish, and Turkish. In *Proceedings of the 23rd annual Boston University conference on language development* (Vol. 2, pp. 541-552). Somerville, MA: Cascadilla Press.

Özçalışkan, Ş., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in Turkish and English. *Studies in Turkish linguistics*, 259-270.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005a). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101-B113.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005b). Do parents lead their children by the hand? *Journal of Child Language*, *32*(3), 481–505.

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., & Baumann, S. (2017). Early gesture provides a Helping hand to spoken vocabulary development for children with autism, Down Syndrome, and typical development. *Journal of Cognition and Development*, *18*(3), 325–337.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016a). Does language shape silent gesture? *Cognition*, *148*, 10-18.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Is seeing gesture necessary to gesture like a native speaker? *Psychological science*, *27*(5), 737-747.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2018). Blind speakers show language-specific patterns in co-speech gesture but not silent gesture. *Cognitive Science*, *42*(3), 1001-1014.

Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688-704.

Ozturk, S., Pinar, E., Ketrez, F. N., & Özçalişkan, Ş. (2021). Effect of sex and dyad composition on speech and gesture development of singleton and twin children. *Journal of Child Language*, *48*(5), 1048-1066.

Ozturk, S., & Özçalışkan, Ş. (2024). Gesture's Role in the Communication of Adults with Different Types of Aphasia. *American Journal of Speech-Language Pathology*, 1-20.

Park, H. I. (2020). How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism: Language and cognition*, *23*(3), 483-499.

Paul, J., Emerson, S. N., & Özçalışkan, Ş. (2022). Does dialect matter in the expression of motion? Ways of encoding manner and path in Babao and standard Mandarin. *Lingua*, *270*, 103215.

Pawlak, M. (2011). *Speaking and instructed foreign language acquisition.* Multilingual Matters.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., Shultz, E., Bertrand, J., Dufour, N., & Oliphant, T. E. (2011). *Scikit-learn: Machine learning in Python. Journal of Machine Learning Research*, *12*, 2825–2830.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology, 22*, 415–422.

Pika, S., Nicoladis, E., & Marentette, P. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer? *Bilingualism: Language and Cognition*, *9*, 319–327.

Python Software Foundation. (2023). Python (Version 3.x) [Software]. https://www.python.org.

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.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*(1), 27-57.

Ryan, M.-L. (1993). Narrative in real time: chronicle, mimesis, and plot in baseball broadcasts. *Narrative*, *1*, 138–155.

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

Schmidt, A. (Director). (1971). *Die Sendung mit der Maus* [TV series]. WDR.

Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. *SciPy*, *7*(1).

Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 934-944.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In *Relating Events in Narrative, Volume 2* (pp. 219-257). Psychology Press.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–97). Cambridge University Press.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, *33*(*1*), 115-125.

So, W. C. (2010). Cross-cultural transfer in gesture frequency in Chinese–English bilinguals. *Language and Cognitive Processes*, *25*(10), 1335–1353.

So, W. C., Kita, S. & Goldin-Meadow, S. (2013) When Do Speakers Use Gestures to Specify Who Does What to Whom? The Role of Language Proficiency and Type of Gestures in Narratives. *J Psycholinguist Re*s 42, 581–594.

Spreen, O., & Strauss E. (1998). *A compendium of neuropsychological tests (2nd ed).* Oxford University Press.

Stam, G. (2001) Gesture and second language acquisition. Paper presented at *TESOL Convention*, St. Louis, Missouri, March.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture, *International Review of Applied Linguistics in Language Teaching*, *44*(2), 145-171.

Stam, G. (2010). Can an L2 speaker's patterns of thinking for speaking change. *Linguistic relativity in SLA: Thinking for speaking*, 59-83.

Stam, G. (2015). Changes in thinking for speaking: A longitudinal case study. *The Modern Language Journal*, *99*(S1), 83-99.

Stam, G. (2016). Gesture as a window onto conceptualization in multiple tasks: Implications for Second language teaching. *Yearbook of the German Cognitive Linguistics Association*, *4*(1), 289-314.

Stam, G. (2022). How to investigate L2 thinking for speaking. In *Gesture and Multimodality in Second Language Acquisition* (pp. 172-203). Routledge.

Stites, L. J., & Özçalışkan, Ş. (2017). Who did what to whom? Children track story referents first in gesture. *Journal of Psycholinguistic Research*, *46*, 1019-1032.

Stites, L., & Özçalışkan, Ş. (2021). The Time is at Hand: Literacy Predicts Changes in Children's Gestures About Time. *Journal of Psycholinguistic Research*, *50*(5), 967-983.

Tabensky, A. (2008). Expository discourse in a second language classroom: How learners use gesture. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research* (pp. 298–320). Routledge.

Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and semantic description. Vol. III: Grammatical categories and the lexicon* (pp. 36–149). Cambridge University Press.

Tannen, D. (1980). A comparative analysis of oral narrative strategies: Athenian Greek and American English. In W. L. Chafe (Ed.), *The pear stories: cognitive, cultural, and linguistic aspects of narrative production* (pp. 51–87). Ablex Publishing Corporation.

Tannen, D. (1982). The oral/literate continuum in discourse. In D. Tannen (Ed.), *Spoken and written language: exploring orality and literacy* (pp. 1–16). Ablex Publishing Corporation.

Treffers-Daller, J., 2009. Language dominance and lexical diversity: how bilinguals and L2 learners differ in their knowledge and use of French lexical and functional items. In: Richards, B., Daller, M.H., Malvern, D.D., Meara, P., Milton, J., Treffers-Daller, J.

(Eds.), *Vocabulary Studies in First and Second Language Acquisition: The Interface between Theory and Application*. Palgrave MacMillan, New York, pp. 74-90.

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.

Tütüncü, I.S., Paul, J., Emerson, S.N., Şengül, M., Knezevic, M. and Özçalışkan, Ş. (2023), When Gestures Do or Do Not Follow Language-Specific Patterns of Motion Expression in Speech: Evidence from Chinese, English and Turkish. *Cognitive Science*, *47*: e13261.

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open-Source Software*, *6*(60), 3021.

Yoshioka, K. (2008). Gesture and information structure in first and second language. *Gesture*, *8*(2), 236-255.