

**Communication Across Modalities:
An Investigation of Co-Speech Gesture in a Lexical Suppression Task**

Amy L. Rosen

Department of Linguistics, Rutgers, The State University of New Jersey – New Brunswick

Senior Honors Thesis

Advisor: Dr. Kristen Syrett

April 3, 2022

Acknowledgements

This project would not have been possible without the support and encouragement from my thesis advisor, Dr. Kristen Syrett. Thank you, Dr. Syrett, for showing me what it means to be a powerful woman in research. With your encouragement, I challenged myself, thought outside the box, and pursued a project that tackles a fascinating subject. I truly would not be where I am academically and personally without your guidance and advocacy for your students.

I wish to extend another special thank you to the dedicated Lab Coordinator and Research Assistants within the Laboratory for Developmental Language Studies. Thank you, Meg Gotowski, for being an ally in this project. This project is made complete with your research expertise and guidance. To my lab mates, thank you for playing a massive part in my experiment and entertaining me as I constantly talked about my interest in the subject.

To the Linguistics Department, thank you for challenging your students, encouraging them to ask questions, and supporting their academic endeavors. This thesis is a culmination of all I have learned in my linguistics training and made possible with the undergraduate fellowship funding support from the Aresty Research Center.

Last but not least, I would not be where I am without the love and support of my family and friends. To my parents, Ellen and Jeffrey Rosen, and my sister Elizabeth Rosen, thank you for always supporting my dreams and aspirations. Your belief in me allowed me to pursue this project and my other goals confidently. To my friends, fellow thesis writers, seshes (“scribes” per Hieroglyphic Egyptian), Christine Jensen, and Thomas Pflanz, thank you for grounding me and supporting me during this project. This project has been a journey and joy to complete, something I will always cherish.

Abstract

As language users, we have multiple ways of conveying word meaning to our conversational partners, often appealing to lexical items or gestures. Such ‘co-speech gestures’ occur during discourse, contributing varying degrees of information. For bilinguals, there is an additional option to retrieve word meaning from multiple lexicons by code-switching, all the while holding constant the possibility of expressing meaning in the non-spoken modality of gesture. Unbound by articulatory constraints, bimodal bilinguals—proficient in a verbal and signed language—can communicate simultaneously via two modalities much as monolinguals can with spoken language and co-speech gestures. What happens when retrieval of verbal lexical information is suppressed? In this research, we ask whether suppression of word meaning in spoken language gives rise to differential gestures/signing for English monolinguals and bimodal bilinguals, taking this thesis as the starting point to a more general program by focusing on the monolingual population (which are interesting in and of themselves).

We designed a production task where a clue giver interacted with a confederate via Zoom in a word guessing game similar to Taboo involving 36 target words. Clue-givers were told they could not mention the word but could gesture and were given additional constraints regarding prompts. As predicted, the co-speech gestures that surfaced when words are suppressed were dependent on the part of speech and the iconicity of the corresponding manual expression. Participants produced more iconic gestures for more iconic words. Our research provides further insight into the contribution and shape of co-speech gesture, and has implications for the ways in which we think our chosen modes of communication convey meaning to our listeners.

Keywords: co-speech gesture, signing, iconicity, lexical suppression, word meaning

1. Introduction

As language users, there are several ways to convey meaning to our conversational partners. For instance, we can appeal to lexical expressions such as object labels, sentence-level utterances, and gestures like pointing or hand waving. The ways in which we convey meaning are shaped by various aspects of the context, including factors internal and external to the speaker. For native speakers of English or any other spoken language, the spoken word is the predominant method of conveying meaning, while the primary means is different for deaf individuals acquiring a sign language.

When we acquire multiple languages, we have the ability to navigate between them, in a process known as ‘*code switching*.’ In brief, this is the process by which a speaker temporarily ‘ceases’ production of one language and substitutes another in rapid discourse, tapping into the lexicon or morphosyntactic knowledge of both languages in turn. For example, a Spanish-English bilingual conversing in English with an addressee may insert Spanish words within an English dialogue, as in the following retelling of *Borreguita and the Coyote* (Becker, 2001).

- (1) Habia un farmero que tenia una borreguita and he called it a lamb.
 He tied him up in a stick y despues vino un coyote y dijo "Grrrrr!
 Borreguita estas bien rica!"
 "No me comes! No me comes!" (high voice)
 "Por que no?"
 "No estoy gordita, debo de comer todo ya. Esta manana me ves,"
 (high voice).
 "OK. Grrrrr! Manana te veo." Hijole como se va a poner gordita
 esa "Grrrrr!"

This switching often happens naturally, almost instinctually, and at times unbeknownst to the speaker. Code-switching can appear as singular words, short phrases, or entire sentences. During speech production, bilinguals retrieve lexical information from the most cognitively accessible language. The frequency and rate of code-switching may also depend on a speaker's knowledge

about their conversational partner (Lillo-Martin et al., 2016). Bilinguals acquiring two spoken languages are constrained by space and time. The speaker in (1) could not have produced Spanish and English simultaneously. The situation is different for speakers acquiring a spoken language and a signed language, or *bimodal bilinguals*.

Bimodal bilinguals are language users proficient in both a spoken/verbal language and a signed language (Emmorey et al., 2008). While spoken language bilinguals are constrained by the articulatory system, unable to speak both languages simultaneously, bimodal bilinguals are not bound by this construct. Thus, these language users can produce both languages simultaneously in a process known as '*code blending*' (Emmorey et al., 2005; Emmorey et al., 2012, 2013; Herbert & Pires, 2017; Lillo-Martin et al., 2016). Code-blending allows for simultaneous language production via verbal and manual modalities, allowing for one language to either complement, or be additive to the other language.

While the process of code blending as described here is unique to bimodal bilinguals, a somewhat similar process does occur in monolingual speakers of a spoken language, in the pairing of the spoken language with co-speech gesture. These are manual expressions that complement, expand upon, and enrich spoken language. However, there is a key difference code blending between a spoken language and a signed language on the one hand, and the use of co-speech gesture accompanying spoken language on the other. A signed language is a language in its own right, conveying truth-conditional meaning and at-issue content (content that addresses the question under discussion in the discourse), whereas co-speech gesture is claimed to contribute not-at-issue projective content (Ebert & Ebert, 2014; Schlenker, 2018; Tieu et al., 2017, 2018). To illustrate the difference, note that one could directly reject propositions expressed in sign, but not content expressed via a gesture expressing manner of motion,

direction, size, a property of an object, and so on. Moreover (and perhaps more importantly), there are representational and processing differences between a signed language and co-speech gesture internal to the language user: these non-spoken modes of conveying meaning take on a different status internal to the language user, in their cognition and linguistic competence.

Given the qualitatively different nature of sign language and co-speech gesture, a question arises when we compare bimodal bilinguals to spoken language users who exhibit co-speech gestures. What happens when we favor spoken language as the primary means of communication, but temporarily suppress the individual's ability to express lexical content in the spoken language? How do individuals in these two populations differentially address this limitation? This honors thesis is a start¹ at addressing this research question: how does the temporary suppression of the spoken language in a spoken language medium give rise to changes in the production of a signed language (in a bimodal bilingual) or co-speech gesture (in a monolingual speaker of a spoken language)? In addressing this question, this research contributes to our general knowledge of language processing and production, and more specifically to our knowledge of the linguistic status of co-speech gesture, through the lens of lexical suppression.

In section 2, I provide background on code-blending in bimodal bilinguals and co-speech gesture in monolingual speakers, motivating the current study. I then present the current

¹ It is a start, because here, I present the results from one population (the English speakers who gesture) as the first stage of a larger research program. Recruiting efforts from the signing population are ongoing, and pose more significant challenges for a year-long thesis than data collection from spoken monolinguals.

research, an online lexical suppression task in which we manipulate the type of lexical item suppressed, and the level of iconicity of the corresponding sign, to determine if these two factors (part of speech and level of iconicity) modulate the appearance of a non-spoken modality (signing or co-speech gesture). As a reminder, we focus our attention on co-speech gesture in this thesis. While the iconicity value is based on signs, we use this as a rough proxy for iconicity in gesture, in the absence of a comparable rating for co-speech gesture. The results demonstrate that these factors interact in the production of co-speech gesture when words are suppressed. These results add to the broader discussion of how individuals communicate successfully with one another in the face of constraints on their ability to communicate meaning.

2. Background on Code Blending and Co-Speech Gesture

2.1. Code Blending in Bimodal bilinguals

Bimodal bilinguals or “speech-sign bilinguals” are language users proficient in spoken and signed languages (Emmorey et al., 2008). These individuals have access to two languages conveyed by distinct mediums, one of which is manual the other verbal. For purposes of this paper, we will be focusing on ASL-English bimodal bilinguals. As I outlined in the Introduction, unlike spoken-language bilinguals, bimodal bilinguals are not bound by the articulatory constraints of two spoken languages. As such, they are able to engage in a process known as *code blending* (Emmorey et al., 2005; Emmorey et al., 2012, 2013; Herbert & Pires, 2017; Lillo-Martin et al., 2016), which allows for simultaneous language production via verbal and manual modalities. This is in contrast to *code switching*, in which language users must alternate between languages, code-blending among bimodal bilinguals. With code-blending, bilinguals can simultaneously produce linguistic information via the manual and verbal modalities through sign

and speech. Code-blends often occur when singular words and short phrases are spoken instead of entire sentences, and most frequently, the signs align in timing and synchronicity with spoken word (Emmorey et al., 2005; Emmorey et al., 2008).

Code blends may consist of semantically equivalent or semantically non-equivalent code-blends, which signifies whether what is said matches in meaning to what is signed (Emmorey et al., 2005). For code-blends with semantic congruency, the manual production matches in meaning to the parallel speech production. Code-blends may also differ in the grammatical category as well. Finally, Emmorey et al. (2005) found that different word classes participate in code-blends more frequently than others: verbs enter code-blends more often than nouns for ASL-English bilinguals. Apart from actual code blending, ASL-English bilinguals frequently integrate signed and spoken language in conversation, with ASL “intrusions” and even ASL-influenced English (“coda talk”). Characteristics of ASL can appear in speech as either inserted signs or English taking on the syntax of ASL.

It is important to keep in mind that ASL and English differ in word order, and that ASL is not a word for word translation of English; it has its own syntax and topicalization patterns. The signs within ASL can be categorized just as the morphosyntactic and lexical elements of any other language. ASL contains lexical signs for traditional word classes (nouns, verbs, adjectives, etc.), classifier signs (like Mandarin or Japanese), and a series of non-manual markers including facial expressions (e.g., raised/furrowed eyebrows, open/closed mouths, relating to question type).

In addition to signs, a bimodal bilingual can also produce gestures, like other language users can. For these individuals, gesture is heavily influenced by their knowledge of sign language and often takes on characteristics of ASL sign, including role-shifting/perspective-taking, locative

shifts, and information status (Casey & Emmorey, 2008; Ebert & Ebert, 2014). Casey & Emmorey (2008) found that the distinct features of ASL influenced the type, form, and expression of co-speech gestures used by bimodal bilinguals. These individuals produced gestures representing character perspective, handshapes, and unmarked handshape gestures. There may be multiple reasons for a bimodal bilingual to gesture, one of which is the possible failure to suppress ASL or the non-selected language in conversation (Casey & Emmorey, 2008).

As with any other aspect of comprehension and performance natural language, we would want a model that captures the representation, processing, and production planning for any given speaker. Traditionally, researchers have appealed to Levelt's language processing model as the foundation for monolingual language/speech production. (See Figure 1).

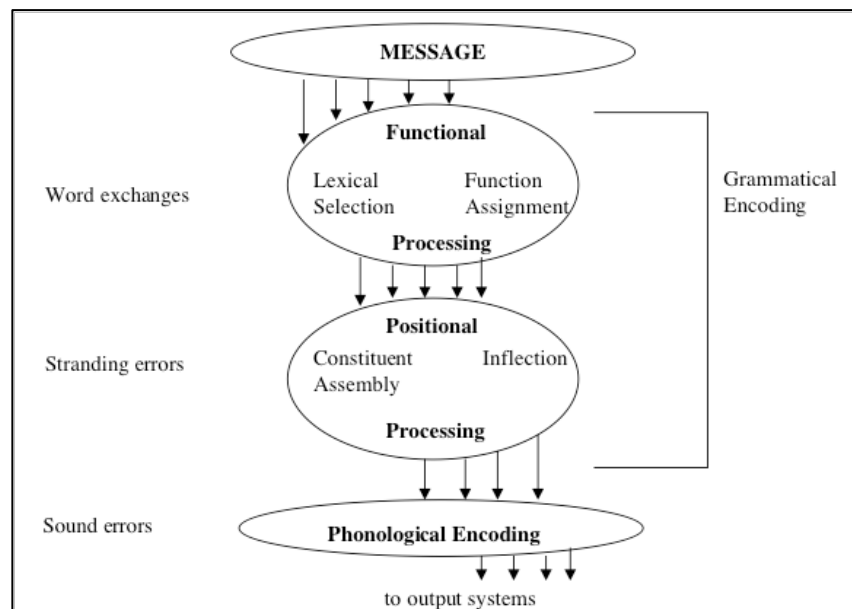


Figure 1. The model of monolingual language processing proposed by Levelt et al. (1989).

This speech production model assumes a unidirectional, linear approach towards language processing, where an individual (1) composes a message through a message generator, (2) proceeds in using various linguistic components, including syntax, morphology, pragmatics, and phonology, (3) conveys said meaning through speech (Levelt, 1989). The physical articulators

are responsible for implementing a thought/message into speech. For monolinguals, this process is straightforward; however, there are additional means of conveying information beyond verbal language, making the language processing system more involved than previously thought. We return to this point shortly.

For bimodal bilinguals, the language production model becomes more complex, as several systems of language processing link together and become integrated. manual linguistic elements of sign and other non-spoken language cues. Emmorey et al. (2008) proposed a language processing model adapted to a bimodal bilingual's speech production process, a process heavily influenced by their knowledge of a signed and spoken language, as shown in Figure 2.

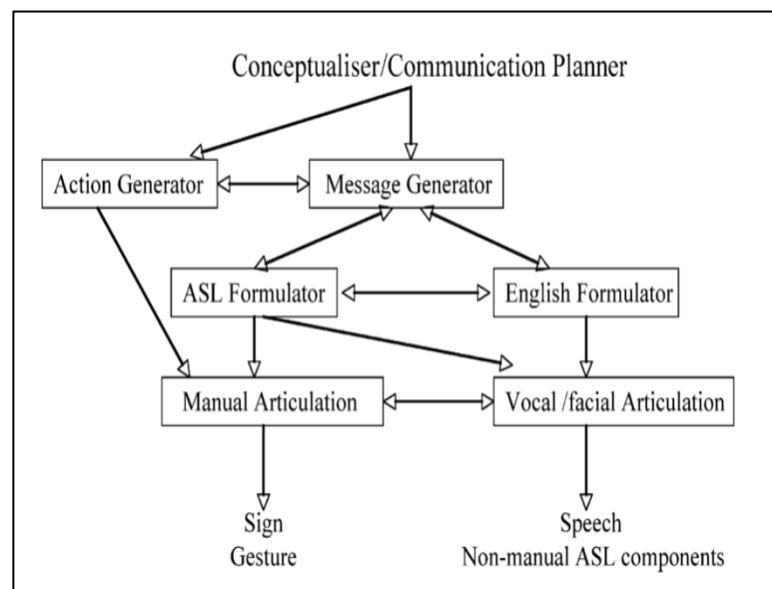


Figure 2. The model of bimodal bilingual (ASL-English) language processing proposed by Emmorey et al., (2008)

The Emmorey et al. (2008) model contains a message generator and action generator which engage simultaneously. Unlike the previous model, this model is multi-directional, meaning that each element is influenced by another. Thus, a bimodal bilingual's language production process includes the option for that message to be conveyed through either a spoken or signed language,

and for them to influence each other in the ways mentioned above. The ASL formulator and Manual Articulation steps in this processing model allow bimodal bilinguals to sign, gesture, and produce verbal speech simultaneously. The manual expressions can be either gestures or signs.

While simultaneously working through complex cognitive processes during language processing, a conversationalist must also consider the most efficient and optimal way to convey meaning to their interlocutor. Studies show that children, and by extension adult language users, show sensitivity to interlocutor language use and alter their language selections accordingly (Lillo Martin et al., 2014). For a bimodal bilingual, this means sifting through multiple lexicons and modalities, and selecting the most practical mode of communication based on their conversational partner's knowledge. Part of this process may actually involve lexical suppression as words are selected in one language instead of the other (Casey & Emmorey, 2008; Lillo-Martin et al., 2014).

2.2. Co-Speech Gesture in Monolinguals

Even for individuals who are monolingual speakers, while they do not know a sign language, there is still a possibility of communicating via the non-spoken modality of gesture. As with code-blending, these gestures occur simultaneously and synchronously with verbal productions. For this reason, we will refer to the gestures of a monolingual speaker as *co-speech gestures* (Casey & Emmorey, 2008; Ebert & Ebert, 2014; Schlenker, 2018; Tieu et al., 2017, 2018). These co-speech gestures may be categorized by type: iconic, deictic, beat, and conventional/symbolic (Casey & Emmorey, 2008). *Iconic* gestures are manual expressions that resemble what they represent (Casey & Emmorey, 2008). For example, extending one's hands away from the body, palms facing outward, to demonstrate the action of pushing is considered iconic. The iconicity of gesture is often measured by whether a word is understood by gesture alone, without a verbal

compliment. Pointing either to a visible or imaginary referent is considered *deictic* gesturing. *Beat* gestures consist of non-iconic manual expressions that match the temporality and synchronicity of speech. These gestures often accompany speech for tonal emphasis or are involuntary. *Conventional* gestures are those largely shared across cultures (e.g., thumbs up, indicating size (large vs. small), directionality, etc.).

Gesture is not just ‘extra hand motion’. It actually serves an important communicative function and helps to represent, consolidate, and convey information. For example, gesture plays a significant role in the linguistic, cognitive, and social development throughout the lifespan, from word learning, through adulthood (Goldin-Meadow, 2004, 2009; McGregor et al., 2009; Hostetter, 2011; Dargue & Sweller, 2018; Vogt & Kauschke, 2018; Aldugom et al., 2020). Moreover, research in cognitive psychology and education has shown that gesture supports learning outside of the language domain, in particular in math instruction (Aldugom et al., 2020). Aldugom et al.’s (2020) found that gesturing helps children consolidate information related to visuospatial working memory in mathematical tasks. Parents use gestures to accompany child-directed speech, and children recruit gestures in word learning and other linguistic tasks (McGregor et al., 2009; Vogt & Kauschke, 2018). For instance, Dargue & Sweller (2018) argued that iconic gestures are beneficial to narrative comprehension (Dargue & Sweller, 2018). Thus, gesturing does not solely co-occur with speech; it plays an essential part in how we communicate and develop cognitively.

Since gesture serves a significant role in discourse coherence, it is worth asking whether producing gestures is speaker- or listener-oriented. There is reason to think that it may have qualities of both. Given the role of gesture in cognition, it may be useful tool for speakers to plan and better understand language. At the same time, it appears that gesture also helps listeners

dissect a speaker's message. In a way, gesture is comparable to dysfluencies such as “um” or “like,” which serve to cue our conversational partner that there may be complex syntax ahead. Gesture may also serve as a form of prosodic prominence, highlighting information structure. How, then, is gesture incorporated into a processing model? Kita and Özyürek (2003) have attempted such a model, integrating spoken language and co-speech gesture. See Figure 3.

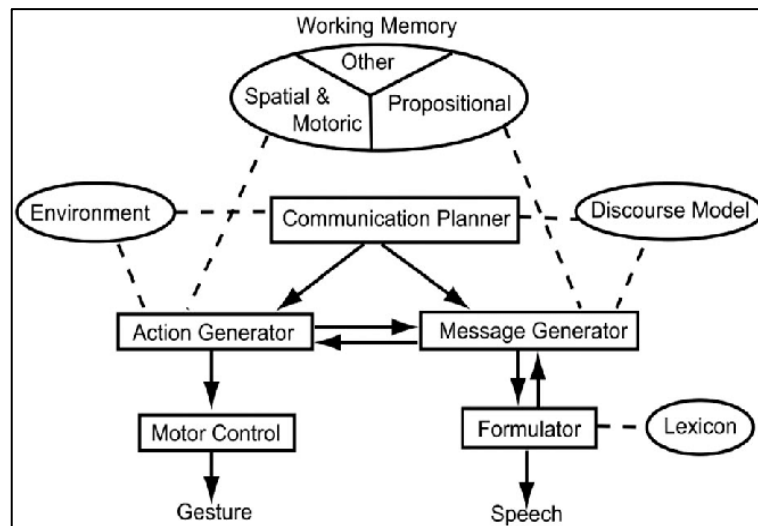


Figure 3. The model of monolingual language processing incorporating a co-speech gesture system by Kita and Özyürek (2003).

Here, there is no longer a unidirectional approach in this model but rather a bidirectional one, meaning that each stage of language processing influences another (as in Figure 2 earlier). In this integrative language processing model, the ability to produce gesture co-exists with producing verbal language, often overlapping. The consequence of the models in Figures 2 and 3 is that whether an individual is a bimodal bilingual or a monolingual speaker, there are multiple, interacting ways to convey meaning through both verbal and manual modalities.

Given the role of co-speech gesture to the communication and comprehension of meaning, and their temporal status relative to spoken language, it is worth asking what the contribution of co-speech gesture is semantically, and how to capture it formally.

Co-speech gesture appears to interact with the truth-conditional content of the expression with which they cooccur. As such, they appear to contribute not-at-issue content (Ebert & Ebert, 2014; Schlenker, 2018; Tieu et al., 2017, 2018). Accordingly, some have argued that co-speech gestures contribute additive information comparable to the meaning conveyed by appositive relative clauses (Ebert & Ebert, 2014). Note that this information is not-at-issue: if one disagrees with proposition expressed by the matrix clause (baseball is very engaging) they can respond, “That’s not true.” But if one disagrees with the information contained in the appositive (baseball is America’s favorite pastime) it is odd to respond to (2) with “That’s not true.”

(2) Baseball, America’s favorite pastime, is very engaging.

In a similar way, it is odd to respond to (3) with “That’s not true” to mean that the mouse is not big. Instead, the response must mean the negation of the statement in (3).

(3) There is a mouse in the kitchen. (gestures with hands to indicate large size)

There are three theories about the status of co-speech gesture: supplemental, supervaluationist, and co-suppositional (Ebert & Ebert, 2014; Schlenker, 2018; Tieu et al., 2017, 2018). The description above represents the supplemental theory of gesture, where co-speech gesture is treated on par appositives and other projective content. Schlenker also proposed a supplemental theory to account for 'post-speech gestures' (gestures that follow speech), which add supporting but not-at-issue information. Schlenker (2018) also argued for a 'cosuppositional' analysis of co-speech gesture, where a co-speech gesture triggers an assertion-dependent presupposition, similar to the presupposition triggered by the change of state verb *stop* in (4). Instead, a gesture like the one in (5) (where ‘UP’ signifies a lifting gesture) is a cosupposition, since it does not indicate information that was backgrounded before the utterance, but accompanies the utterance (Schlenker, 2018).

- (4) It stopped raining. (Presupposition: It rained)
- (5) John helped UP_ his son.

This “assertion-dependent presupposition” indicates that the help John provides his son had to do with lifting. Another example of an assertion-dependent presupposition is shown with the superlative in (6), which presupposes that the other students in the class are weaker than Karlie.

- (6) Karlie is the strongest student in the class.

Schlenker (2018) argues that the content of a co-speech gesture is presupposed to follow from the expression *in which* it is embedded.

Finally, there is the 'supervaluationist' interpretation of co-speech gestures, which argues that the context set, or common ground, should have the same semantic result with or without the presence of a co-speech gesture. Instead of suggesting that co-speech gestures enrich the meaning of a sentence in a cosuppositional fashion, this approach proposes to conjoin the sentence without the co-speech gestures and the sentence with the co-speech gestures. Previous research has explored the impact and influence of co-speech gesture on language comprehension, semantic word meaning, contextual information in conversation, in addition to co-speech gesture interpretation (e.g., Schlenker et al., 2018; Casey & Emmorey, 2008; Ebert & Ebert, 2018; Tieu et al., 2017; Tieu et al., 2018).

2.3. Research Questions

The background above gives us a clear idea of what simultaneous signing and co-speech gestures have in common, and how they differ from each other. In both cases, the non-spoken, manual communication interacts with has the potential to enrich the spoken utterance. What happens, then, when the spoken word is suppressed? When spoken language is made unavailable, does the manual modality intercede to compensate? If so, how? Does the lexical

meaning intended to be conveyed and the type (specifically, the iconicity) of gesture or sign mediate this process? Will we observe a difference in how these two distinct populations—one proficient in a signed language and one not—respond?

One could hypothesize that when bimodal bilinguals are forced to suppress a lexical item, they will simply switch to their other language. However, if their conversational partner is not familiar with sign language, it might be more advantageous to utilize co-speech gestures in their spoken language. As we know from prior research, this co-speech gesture is heavily influenced by signs, as their two languages compete in an integrated language processing system. By contrast, when monolinguals are forced to verbally suppress a lexical item, they do not have another language to fall back on, so they may either search through their space of spoken words or resort to co-speech gestures to act as the support for conveying word meaning.

Given verbal lexical suppression, we generate a prediction that holds across populations, and can be tested within English monolingual speakers, then followed up in bimodal bilinguals beyond this thesis. We predict that when words are suppressed, manual communication will increase. For monolingual speakers, this means increased co-speech gestures. For bimodal bilinguals, this means leaning on their signed language, and possibly also on gestures. For both, the more iconic the sign or gesture corresponding to the suppressed word, the more likely the gesture will occur. And this will happen more for verbs than nouns.

3. Methods

3.1. Participants

34 non-signing native speakers of English were recruited through the Rutgers Department of Linguistics Subject Pool. These individuals received extra credit in their course as compensation

for participating in the study. Participants completed a demographic survey in which they indicated their language background and proficiency.

3.2. Design

The experiment was an elicited production task disguised as a word guessing game. We used the game Taboo as inspiration, adapting it for the purpose of the current research questions. The study was implemented via Zoom. In the task, experiment participants interacted with a confederate (a research assistant), both of whom were given directions by a second research assistant (the author). Each experimental session also included two additional research assistants responsible behind the scenes: the first read the instructions and fed the participant target words one at a time from predetermined lists via private message in the Zoom chatbox, the second maintained a 60-second timer and played auditory prompts.

3.3. Materials

There were 36 target words. See Appendix A. All lexical items were correlated with an ASL sign that is positioned at the torso or above, so the signs or gestures could be observed in the zoom frame. Items were chosen based on iconicity values obtained from the ASL LEX database. Based on raw rating scores (7-point scale: 1 low to 7 high), we sorted lexical items into three iconicity categories: Non-Iconic, Semi-Iconic, and Very Iconic. Lexical items falling in the 1-2.5 range were considered not iconic, 2.5-4.5 semi-iconic, and 4.5-7 very iconic. We based these values on the ASL LEX database ratings by non-signers, deaf signers, and native deaf signers. Nouns were further classified by animacy. Verbs were further classified by frame (transitive v. intransitive/optionally transitive). Words were divided into two lists of 18, assembled in a pseudorandomized order (balancing part of speech and iconicity, and the additional noun and

verb classifications). (See Appendix B.) We also balanced for frequency, based on the COCA database.

3.4.Procedure

Participants were asked to join the Zoom call in a quiet, clear space in preparation for the game. Before the experiment began, both participants and confederates of the study were asked to complete a short consent form that included the study's informed consent and demographic information regarding language background/proficiency. After the participants completed the consent form, they received instruction to adjust their computer camera so that their entire face and torso were visible. They were then ready to begin.

Participants and confederates were given the instructions in Appendix C. Participants were told they were the clue-giver. The second participant (the confederate) was assigned the role of clue guesser. During the game, the participant/clue-giver was fed their list of 18 words one at a time through private message in chat. They were arbitrarily assigned to a given list. Clue givers were told they could not mention the word, but they could mention single words, short phrases, and entire sentences, use words related in meaning to the target word, reference pop culture, literary, and various entertainment sources related to the word prompt, and most significant for our study's investigation, accompany their clues with gestures or hand movements. (See the 'do's and don'ts' in the instructions.)

Once the experimenter revealed the word prompt in the chat, clue-givers attempted to provide the guesser with enough clues relating to the target word to make an accurate guess. Clue guessers had 60 seconds to identify the target word successfully. If the guesser failed to guess the target word in time, clue givers were instructed to move on to the next word. Participants repeated this process until the end of the word list. During the instruction phase,

participants heard the three types of auditory stimuli they would hear during the session: one for a countdown until the word appeared, one for the correct answer, and one for time expired. Each auditory cue either prompted a new word reveal or the end of the round.

After the instruction phase, the participants moved on to a practice phase. At the beginning of this phase, participants were directed to pin their partner's Zoom screen. Doing so allowed the clue giver and guesser to see only themselves side by side on the computer screen. The other two experimenters kept their video cameras off during the task, and were therefore not visible. These measures were taken to help maximize experiment efficiency and maintain participatory attention.

The experimenter began by asking both participants the following questions:

(7) How do you wrap a present?

(8) How do you open a present?

These questions were chosen to get the participants comfortable with interacting with each other, and were chosen specifically to prime participants to gesture during the word guessing game.

The participant and confederate then completed a practice run-through with the word *dinosaur* to ensure they understood gameplay.² After the practice run-through, participants had a chance to ask questions. If there were no questions, gameplay commenced. Participants were recorded via Zoom for later coding and analysis.

4. Results

4.1 Gesture Coding

² For two of the participants, the practice word used was 'dog.'

Following existing literature (Casey & Emmorey, 2008), gestures were classified by type: iconic, deictic, beat, conventional/symbolic, or N/A. Gestures were considered *iconic* if they resembled what they represented, resembled a specific action or object, and was easily inferrable without the need for speech accompaniment. An example of an iconic gesture is in Figure 4, where the participant brings together both hands and opens them to depict a book in the above image.

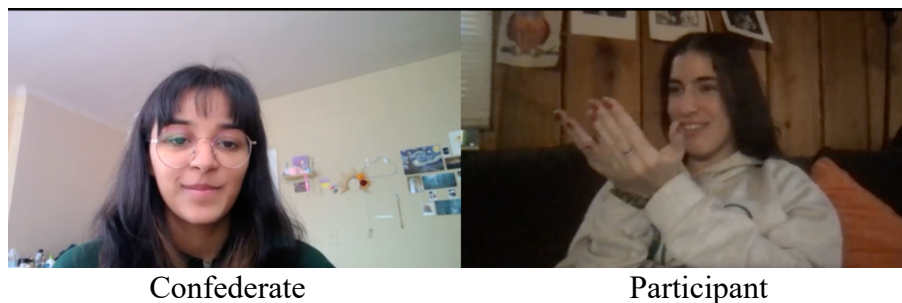


Figure 4. Experimental participant (right) produces an iconic gesture when trying to give the clue corresponding to the target word *book*.

Videos were coded using ELAN, a video editing/annotating tool. Coding consisted of segmenting and annotating what each participant said, what gestures they produced, and documenting specifics of gesture: type, category, and description. See Figure 5, for the prompt in Figure 4.

	00:04:02.000	00:04:03.000	00:04:04.000	00:04:05.000	00:04:06.000	00:04:07.000	00:04:08.000	00:04:09.000	00:04:10.000	00:04:11.000	00:04:12.000	00:04:13.000	00:04:14.000	00:
Target Word [139]	BOOK													
Speech Transcript [137]	Um, before. But you're doing the first word.							I mean not doing. It is the first word that you said.						
Gesture Type [138]	ICONIC							ICONIC						
Gesture Category [138]	OPEN BOOK							OPEN BOOK						
Gesture Descripto [138]	Participant aligns and opens their hands, as if they are holding a book.							Participant aligns and opens their hands, as if they are holding a book.						

Figure 5. ELAN categorization of the spoken language, and gesture information for the prompt *book*, shown in Figure 4.

Figure 6 shows an example of coding pointing as a deictic gesture.

The screenshot shows an ELAN interface with a timeline from 00:04:48.000 to 00:04:56.000. A red vertical line is at 00:04:49.000. The 'Target Word' is 'CAT'. The 'Speech Transcript' is 'It's the thing that's behind you.'. The 'Gesture Type' is 'DEICTIC', the 'Gesture Category' is 'POINT', and the 'Gesture Description' is 'Participant points their left hand towards the camera.'.

Field	Value
Target Word	CAT
Speech Transcript	It's the thing that's behind you.
Gesture Type	DEICTIC
Gesture Category	POINT
Gesture Description	Participant points their left hand towards the camera.

Figure 6. ELAN categorization of the spoken language, and gesture information for the prompt *cat*.

The third gesture type coded were *beat* gestures, manual expressions that align with speech.

Figure 7 shows a participant producing a beat gesture during the target word *window*.

The screenshot shows an ELAN interface with a timeline from 00:05:25.000 to 00:05:36.000. A red vertical line is at 00:05:26.000. The 'Target Word' is 'WINDOW'. The 'Speech Transcript' is 'If i'm looking outside, from a building, [how] what am I looking through?'. The 'Gesture Type' is 'BEAT', the 'Gesture Category' is 'EMPHASIS', and the 'Gesture Description' is 'Participant move'.

Field	Value
Target Word	WINDOW
Speech Transcript	If i'm looking outside, from a building, [how] what am I looking through?
Gesture Type	BEAT
Gesture Category	EMPHASIS
Gesture Description	Participant move

Figure 7. ELAN categorization of the spoken language, and gesture information for the prompt *window*, which elicited pointing.

Finally, gestures were coded as ‘conventional’ (symbolic) gestures if they were understood across cultural contexts. These gestures included size and direction; for instance, if the participant wanted to depict something significant in size, they may move their hands far apart to show the relative size of an object, as shown in Figure 7. The ELAN coding shows that the participant goes on to produce a gesture representing the antonym *small*.

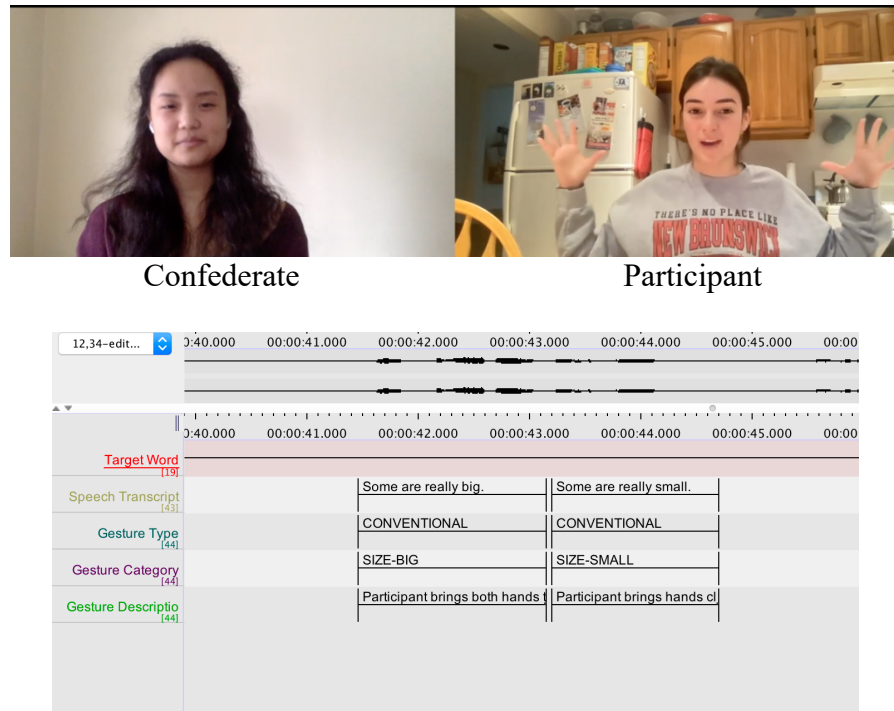


Figure 8. Zoom frames and ELAN categorization of the spoken language, and gesture information for the prompt *dinosaur*, which elicited a conventional gesture indicating size.

4.2 Statistical Analysis

First, we turn to the distribution of gestures types produced. Of the 492 gestures produced, 52.6% (259) were beat gestures, 2.6% (13) were conventional, 7.9% (39) were deictic, and 36.8% (181) were iconic. Beat gestures were produced throughout for emphasis on syllables to mark prominence. The conventional gestures were those that were anticipated, indicating size (e.g., large for *gorilla*, small for *girl*), or position (e.g., low for *kneel*). The deictic gesture of pointing was produced for words such as *see*, *window*, *ask*, with no discernible pattern. The most common iconic gestures were produced for the verbs *paint*, *push*, *read*, *draw*, *create*, *juggle*, *cry*, *kneel*, *see*, *sleep*, and *write*, and the nouns *guitar*, *book*, *picture*, *spider*, and *telephone*.

To determine possible effects of the lexical item category and the iconicity of the corresponding sign on participant's gesturing in the task in the face of lexical suppression, we fit

a mixed-effects logistic regression model with gesture (presence vs. absence) as our dependent variable and part of speech (two levels: verb, noun) and iconicity (three levels: non-iconic, semi-iconic, very iconic) as independent factors, with random intercepts for participant and word (item). There were no main effects. Thus, the factors of part of speech and iconicity did not drive the appearance of gesture.

Level	Coefficient	SE	z value	$p (> z)$
Intercept	-0.167	0.571	-0.292	<.770
Iconicity (semi)	-0.564	0.451	-1.253	<.210
Iconicity (very)	-0.008	0.451	-0.017	<.986
POSverb	0.411	0.369	1.113	0.266

Table 1. Mixed-effects model for Part of Speech and Iconicity

We then turned our attention to those instances in which participants *did* gesture, asking whether or not they are more likely to produce an iconic gesture as a function of these factors. We ran a separate model with gesture type as the dependent variable (iconic vs. non-iconic), with the same independent factors as before (iconicity and part of speech). We ran the model with verb (Table 2) and noun (Table 3) as the baseline. In both instances, there was a significant interaction, with participants more likely to produce iconic gestures when the target word was classified as very iconic.

Level	Coefficient	SE	z value	$p (> z)$
Intercept	-1.737	0.510	-3.404	<.001 ***
Iconicity (semi)	0.631	0.650	0.970	<.332
Iconicity (very)	1.099	0.615	1.787	<.073
POSverb	-0.236	0.620	-0.381	0.703
Iconicity (semi):POSverb	-0.976	0.902	1.082	0.279
Iconicity (very):POSverb	-2.476	0.911	2.720	<.01 **

Table 2. Mixed-effects model for Part of Speech and Iconicity with verb as baseline

Level	Coefficient	SE	z value	$p (> z)$	
Intercept	-1.974	0.524	-3.766	<.001	***
Iconicity (semi)	1.606	0.623	2.579	<.01	**
Iconicity (very)	3.576	0.687	5.207	<.001	***
POSnoun	0.236	0.620	0.381	0.703	
Iconicity (semi):POSnoun	-0.976	0.902	-1.082	0.279	
Iconicity (very):POSnoun	-2.476	0.910	-2.720	<.01	**

Table 3. Mixed-effects model for Part of Speech and Iconicity with noun as baseline

4.3. Discussion

The results above yield two main takeaways: (1) regardless of the part of speech, participants were not more likely to gesture given the iconicity of a target word, and (2) participants who did gesture during the experimental task were influenced by the iconicity of the target word, specifically for iconic verbs. These findings align with our hypotheses, which predict a relationship between the iconicity of a gesture/sign correlated with the suppressed word and the likelihood of gesture production. For many iconic gestures produced, the corresponding verbs represented actions for which individuals use their hands. For instance, when people *write* or *juggle*, they use their hands to utilize a writing utensil or throw balls in the air. In the language processing system that Kita and Özyürek (2003) proposed, both the motor and speech systems work together as individuals gesture and produce speech.

While individuals are influenced by and more likely to produce iconic gestures for iconic words, primarily verbs, our results show that they do not decide to gesture because a word (noun or verb) is iconic. This finding is interesting as it suggests, at least within this experiment, that there is no evident pattern for when individuals produce gestures. We know there is a bidirectional relationship between gesture and speech for monolinguals. However, we do not know what acts as a catalyst for participants to produce a gesture relative to an iconic word.

5. Conclusions

As language users, we can convey meaning using both spoken and non-spoken modalities of communication. For some, this consists of a spoken and signed language (bimodal bilinguals), and for others, this consists of spoken language and co-speech gestures (monolinguals). For monolinguals, co-speech gesture represents an additional method of communication. Co-speech gestures can contribute both non-at-issue and at-issue information to the common ground and seem to be further used when verbal lexical information is temporarily suppressed. In fact, co-speech gestures produced by monolinguals seem to supplement the limited verbal information and thus, increase the likelihood of successful processing of word meaning in conversation. We found that regardless of the part of speech, participants were not more likely to gesture given the iconicity of a target word, but when participants did gesture, they were influenced by the iconicity of the target word, specifically for iconic verbs. These results add to the broader discussion of how individuals with different language backgrounds communicate with one another through varying constraints that may present themselves in real-world conversation.

References

- Aldugom, M., Fenn, K., & Cook, S. W. (2020). Gesture during math instruction specifically benefits learners with high visuospatial working memory capacity. *Cognitive research: principles and implications*, 5(1), 27. <https://doi.org/10.1186/s41235-020-00215-8>
- Alexiadou, A., & Lohndal, T. (2018). Units of language mixing: a cross-linguistic perspective. *Frontiers in Psychology*, 9, 1719. <https://doi.org/10.3389/fpsyg.2018.01719>
- Becker, R. R. (2001). Spanish-English code switching in a bilingual academic context. *Reading Horizons*, 42(2), 99.
- Bishop, M., & Hicks, S. (2005). Orange eyes: Bimodal bilingualism in hearing adults from deaf families. *Sign Language Studies*, 5(2), 188-230,256-257.
<http://dx.doi.org.proxy.libraries.rutgers.edu/10.1353/sls.2005.0001>
- Brysbaert, M. & New, B. (2009) Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*, 41 (4), 977-990.
- Casey, S., & Emmorey, K. (2008). Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes*, 24(2), 290–312. <https://doi.org/10.1080/01690960801916188>
- de Quadros, R. M., Lillo-Martin, D., & Pichler D. C. (2020). Blending languages: Bimodal bilinguals and language synthesis. In Ulrike Zeshan, Jennifer Webster & Anastasia Bradford (Eds.), *Sign Multilingualism*, 201-226. DeGruyter Mouton & Ishara Press.
- de Quadros, R. M., Davidson, K., Lillo-Martin, D., & Emmorey, K. (2020). Code-blending with depicting signs. *Linguistic Approaches to Bilingualism*, 10(2), 290–308.
<https://doi.org/10.1075/lab.17043.qua>

- Dingemanse, M., Perlman, M., & Perniss, P. (2020). Construals of iconicity: Experimental approaches to form–meaning resemblances in language. *Language and Cognition, 12*(1), 1-14. doi:10.1017/langcog.2019.48
- Ebert, C. (2018). A comparison of sign language with speech plus gesture. *Theoretical Linguistics, 44*(3-4), 239-249. <https://doi-org.proxy.libraries.rutgers.edu/10.1515/tl-2018-0016>
- Emmorey, K., Borinstein, H., Thompson, R., & Gollan, T. (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition, 11*(1), 43-61. doi:10.1017/S1366728907003203
- Emmorey, K., Borinstein, H.B., & Thompson, R. (2005). Bimodal Bilingualism: Code-blending between Spoken English and American Sign Language. *ISB4: Proceedings of the 4th International Symposium on Bilingualism, ed. James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan, 663-673*. Somerville, MA: Cascadilla Press.
- Emmorey, K., Petrich, J. A., & Gollan, T. H. (2013). Bimodal bilingualism and the frequency-lag hypothesis. *Journal of deaf studies and deaf education, 18*(1), 1–11. <https://doi.org/10.1093/deafed/ens034>
- Emmorey, K., Petrich, J., Gollan, T. H. (2012). Bilingual processing of ASL-English code-blends: The consequences of accessing two lexical representations simultaneously. *Journal of memory and language, 67*(1), 199–210. <https://doi.org/10.1016/j.jml.2012.04.005>
- Emmorey, K., Li, C., Petrich, J., & Gollan, T. H. (2020). Turning languages on and off: Switching into and out of code-blends reveals the nature of bilingual language

- control. *Journal of experimental psychology. Learning, memory, and cognition*, 46(3), 443–454. <https://doi.org/10.1037/xlm0000734>
- Giimenes, M., & New, B. (2015). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior research methods*, 1-10.
- Goldin-Meadow, S. (2004). Gesture's Role in the Learning Process. *Theory Into Practice*, 43(4), 314–321. <http://www.jstor.org/stable/3701540>
- Goldin-Meadow S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives*, 3(2), 106–111. <https://doi.org/10.1111/j.1750-8606.2009.00088.x>
- Herbert, M., & Pires, A. (2017). Bilingualism and bimodal code-blending among deaf ASL-English bilinguals. *Proceedings of the Linguistic Society of America*, 2, 14:1-15. <https://doi.org/10.3765/plsa.v2i0.4054>
- Hostetter A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137(2), 297–315. <https://doi.org/10.1037/a0022128>
- Lillo-Martin, D., de Quadros, R., & Pichler, D. (2016). The Development of Bimodal Bilingualism: Implications for Linguistic Theory. *Linguistic Approaches to Bilingualism*, 6(6), 719–755. <https://doi.org/10.1075/lab.6.6.01lil>
- Lillo-Martin, D., de Quadros, R. M., Chen Pichler, D., & Fieldsteel, Z. (2014). Language choice in bimodal bilingual development. *Frontiers in Psychology*, 5, 1163. <https://doi.org/10.3389/fpsyg.2014.01163>
- McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: the case of under. *Journal of Child Language*, 36(4), 807–828. <https://doi.org/10.1017/S0305000908009173>

- Petitto, L., Katerelos, M., Levy, B., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, 28(2), 453-496. doi:10.1017/S0305000901004718
- Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics and Philosophy*, 41(3), 295-365. doi:http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/s10988-017-9225-8
- Sehyr, Z. S., & Emmorey, K. (2019). The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: evidence from iconicity and transparency judgments. *Language and Cognition*, 11(2), 208–234. <https://doi.org/10.1017/langcog.2019.18>
- Thompson, b., Perlman, m., Lupyan, g., Sevcikova sehyr, z., & Emmorey, k. (2020). A data-driven approach to the semantics of iconicity in american sign language and english. *Language and cognition*, 12(1), 182-202. Doi:10.1017/langcog.2019.52
- Tieu, L., Pasternak, R., Schlenker, P., & Chemla, E. (2018). Co-speech gesture projection: Evidence from inferential judgements. *Glossa: A Journal of General Linguistics*,3(1), 109. DOI: <http://doi.org/10.5334/gjgl.580>
- Tieu, L., Pasternak, R., Schlenker, P., & Chemla, E. (2017). Co-speech gesture projection: Evidence from truth-value judgement and picture selection tasks. *Glossa: A Journal of General Linguistics*, 2(1), 102. DOI: <http://doi.org/10.5334/gjgl.334>
- Vogt, S., & Kauschke, C. (2017). Observing iconic gestures enhances word learning in typically developing children and children with specific language impairment. *Journal of Child Language*, 44(6), 1458-1484. doi:10.1017/S0305000916000647

Appendix A

Word Information

Target Word	Part of Speech	Iconicity - deaf signer (native)	Iconicity - non signer	Iconicity Category	Animacy	Transitivity	Frequency	Frequency Category
Baby	Noun	6.19	6.833	Very Iconic	Yes	N/A	207,577	Mid
Balloon	Noun	4.79	4.967	Semi Iconic	No	N/A	12,102	Low
Bird	Noun	5.11	4.553	Semi Iconic	yes	N/A	69,662	Mid
Book	Noun	5.33	6.684	Semi Iconic	no	N/A	435,387	High
Camera	Noun	6.41	6.7	Very Iconic	No	N/A	99,855	Mid
Cat	Noun	3.05	4.618	Semi Iconic	Yes	N/A	58,763	Mid
Coffee	Noun	1.95	2.286	Not iconic	no	N/A	70,366	Mid
Father	Noun	1.41	1.259	Not Iconic	Yes	N/A	318,173	High
Fish	Noun	4.05	4.476	Semi Iconic	yes	N/A	78,114	Mid
Girl	Noun	1.57	1.528	Not iconic	yes	N/A	319,120	High
Gorilla	Noun	6.05	6.8	Very iconic	yes	N/A	4,156	Low
Guitar	Noun	6.58	6.967	Very iconic	no	N/A	20,550	Low
Music	Noun	1.74	1.657	Non-iconic	no	N/A	222,767	High
Picture	Noun	2.26	2.167	Non-iconic	No	N/A	165,703	Mid
Pig	Noun	1.32	1.467	Non-iconic	Yes	N/A	22,965	Low
Spider	Noun	5.11	5.792	Very Iconic	yes	N/A	10,153	Low
Telephone	Noun	6.16	7	Very Iconic	No	N/A	29,482	Mid
Window	Noun	4.26	4.19	Semi Iconic	No	N/A	136,043	Mid

Target Word	Part of Speech	Iconicity - deaf signer (native)	Iconicity - non signer	Iconicity Category	Animacy	Transitivity	Frequency	Frequency Category
Ask	Verb	2	2.133	Non-iconic	N/A	Yes	676,596	High
Create	Verb	1.72	2.2	Non-iconic	N/A	Yes	314,631	High
Cry	Verb	6.36	6.567	Very iconic	N/A	No	68,161	Mid
Draw	Verb	4.72	5	Semi Iconic	N/A	Yes	135,022	Mid
Eat	Verb	5.96	6.1	Very Iconic	N/A	No	201,315	Mid
Find	Verb	2.27	1.8	Non-iconic	N/A	Yes	1,051,936	High
Help	Verb	1.61	1.472	Non-iconic	N/A	Yes	606,887	High
Juggle	Verb	6.06	6.371	Very Iconic	N/A	Yes	3,461	Low
Kneel	Verb	5.05	4.111	Semi Iconic	N/A	Yes	9,754	Low
Laugh	Verb	4.71	4.75	Semi Iconic	N/A	No	111,440	Mid
Paint	Verb	5.61	5.189	Very Iconic	N/A	Yes	44,842	Mid
Push	Verb	6.11	7	Very Iconic	N/A	Yes	134,206	Mid
Read	Verb	4.16	4.571	Semi Iconic	N/A	Yes	386,137	High
Run	Verb	2.05	1.667	Non-iconic	N/A	No	465,066	High
See	Verb	4.57	4.63	Semi Iconic	N/A	Yes	1,958,700	High
Sleep	Verb	4.68	4.833	Semi Iconic	N/A	No	95,600	Mid
Work	Verb	2.41	1.733	Non-iconic	N/A	Yes	854,095	High
Write	Verb	5.94	6.571	Very iconic	N/A	Yes	439,865	High

Appendix B**Word Lists**

List 1	List 2
Girl	Pig
Coffee	Father
Music	Picture
Run	Find
Help	Ask
Work	Create
Fish	Cat
Bird	Balloon
Read	Window
Sleep	Laugh
Draw	See
Book	Kneel
Spider	Telephone
Gorilla	Baby
Guitar	Camera
Car	Paint
Write	Eat
Push	Juggle

Appendix C

Instructions

This study is designed to be a game of guessing words!

One of you will be the CLUE GIVER, and one of you will be the GUESSER.

CLUE GIVER, here are your instructions!

- We are going to give you a list of words, one at a time.
- Your task is to try to get the GUESSER to correctly guess each target word using other words or gestures.
- Once you see a target word, you should immediately begin giving clues to the GUESSER.
- Time is up for that target word once the word is guessed or once 60 seconds passes, whichever comes first.

GUESSER, here are your instructions!

- The CLUE GIVER will be shown some words.
- They will try to get you to guess each one.
- Your job is to try your best to guess each word, given the clues you are given.
- You have 60 seconds to guess the word.
- There are no penalties for incorrect guesses.
- There are no limits to how many guesses can be made within the 60 seconds.
- Time is up for that target word once you have correctly guessed it or once 60 seconds passes, whichever comes first.

You have 60 seconds for each target word!

- Before each word appears in the chat area to the CLUE GIVER, you will hear this sound (*play countdown sound*). The word then appears in the chat, and the CLUE GIVER will immediately start to have the GUESSER guess the word.
- If the GUESSER correctly guesses the target word in time, you will hear this sound (*play 'correct answer' sound*).
- After you hear this noise, there will be a countdown (*play countdown sound*), to which you will direct your attention to the chat for the next word. At this time, you should move on to the next target word. Do not wait until 60 seconds has passed.
- If the GUESSER does not guess the word in time, you will hear this sound (*play 'wrong answer' sound*).
- After you hear this noise, there will be a countdown (*play countdown sound*), to which you will direct your attention to the chat for the next word. At this time, you should proceed to the next target word. Do not continue trying to give clues or guess the word.

You should NOT repeat these words outside of the experiment session! If you do, it will RUIN it for others playing the game, and make our data completely unusable.

Do's and Don'ts

There are some RULES the CLUE GIVER must follow!

No-nos!

1. You may not say the target word. (e.g., If the word is 'banana' you cannot say, "banana.")
2. You may not say any part of the target word. (e.g., If the word is "peanut" you cannot say "pea" or "nut." If it is "butterfly" you cannot say "butter" or "fly.")
3. You may not spell the target word in part or in its entirety, or give hints about letters. (e.g., If the word is 'banana' you cannot, for example, say, "It starts with a 'b'." or "There are three 'a's and two 'n's." or "It has 'an' repeated twice.")
4. You may not say the word with a different sound or sound combination substituted in. (e.g., If the word is 'banana' you cannot say, "schmanana" or "bashmana" or "fanana." If it is "shrimp" you cannot say "shrimp" or "shrimp.")
5. You may not say a word that rhymes with your target word. (e.g., If the word is "bridge" you cannot say "fridge" and if it is "hit" you cannot, for example, say "bit," "fit.")
6. You may not make sounds that are associated with the target word or its object. (e.g., If the word is "dog" you cannot making a barking or growling sound or say "bow wow," "ruff.")
7. You may not use antonyms (e.g., if the word is young you cannot say opposite of old)

OK

1. You may say single words, short phrases, or entire sentences.
2. You may say words that are related in meaning in some way.
3. You may accompany your clues with gestures or hand movements.
4. You may make reference to pop culture, literature, poetry, songs, or other sources that are related by association or context.

Clue-giver: Remember, your goal is to have the guesser guess the word. You can use words, phrases, gestures, body movement, etc.