An experimental pragmatic investigation of depictive co-speech gestures

A dissertation presented

by

Christina Zlogar

to

The Department of Linguistics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Linguistics

Harvard University

Cambridge, Massachusetts

May 2019

Dissertation Advisor:
Professor Kathryn Davidson

Author:
Christina Zlogar

An experimental pragmatic investigation of depictive co-speech gestures

Abstract

This dissertation examines the pragmatics and semantics of *co-speech gesture*, gesture produced simultaneously with speech. Using experimental techniques, several patterns are found. First, interlocutors would prefer gestures to not convey communicative content. They instead prefer gestures to be trivial, i.e. truth-conditionally vacuous. Second, linguistic modifiers are preferred over gestural modifiers. This is true even when important information is hypothetically more suited to communication via gesture than via speech. Third, co-speech gesture is not dispreferred as a communication channel simply because it is depictive; at-issue information can be conveyed depictively in a sign language like American Sign Language (ASL). Instead, there seems to be a preference for a single mode of communication. Overall, all studies found pragmatic and semantic asymmetry between gesture and language (spoken or signed).

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I thank my advisor, Kathryn Davidson. Across an ocean and through maternity leave, she met with me regularly on Skype, always patiently pushing me outside of my comfort zone. This dissertation would not have been possible without her.

Thanks to my committee members, teachers, departmental administrators, fellow graduate students, and friends for their mentorship, guidance, and support. I am grateful to Anna Alsop, who assisted with study design and data collection for Chapters 2 and 3, and Kate Henninger, who helped with study design and data collection for Chapter 4. Thanks also to Dorothy Ahn for our ever-useful troll images and Miles Wheeler for our window images. For their valuable feedback, I am indebted to audiences at the Meaning and Modality Linguistics Laboratory at Harvard University, the 2nd Experimental Pragmatics in Italy Conference (XPRAG.it 2018), and the 93rd Annual Meeting of the Linguistic Society of America (LSA 2019).

Finally, I thank my family: my parents, Sue and Al, my brother, Albert, my husband, Miles, and also my sisters-in-law, Cory and Page, who provided much-needed camaraderie during long editing sessions.

To my parents

Chapter 1

Introduction

1.1 Co-speech gesture

Spoken conversation doesn't just involve the mouth, but also the hands. Far from being arbitrary and random, these hand motions are connected to both the words being said and to how they're being spoken. Sometimes, the hands even speak for themselves by iconically depicting an object, action, or idea. Such gestures supplement and complement speech every day, and yet until recently, theoretical linguists haven't paid much attention to them.

A basic feature of gesturing while speaking is that it involves expressing information in two different modes, visual and auditory, simultaneously. This duality raises many interesting questions: What information do we choose to put into gesture instead of (or in addition to) speech? Does communicating something through gesture fundamentally change how that information is packaged and interpreted? Do linguistic intuitions about felicity and meaning even make sense for gestures? This dissertation investigates these questions and more relating to the meaning and pragmatics of gesture with speech.

1.1.1 Motivating examples

The kind of *co-speech gestures* that have been the focus of recent semantic/pragmatic work and that we will be focusing on in this dissertation are gestures made simultaneously with

speech/sign that enrich utterance meaning by iconically depicting an aspect of the situation or event (Kendon 2004; Goldin-Meadow and Brentari 2017); see (2.1–2.6) for examples. In these and future examples, following the notation of Tieu et al. (2017, 2018), we will write co-speech gestures using an English word to approximate their meaning, with capital letters, which will appear immediately after the speech expression it modifies, which is in turn enclosed in square brackets to show the temporal alignment of the gesture with the speech stream.

(1.1) (Kendon 2004:136) (see Figure 1.1 for an illustration)

Context: summarizing the Little Red Riding Hood story:

a. . . . and [took his hatchet]_GRAB-HANDLE

b. and with [a mighty sweep]_LIFT-HANDS-TO-SHOULDER

c. [sliced]_AXE-SWING the wolf's stomach open.

(1.2) (McNeill 1992:79)

. . . and he [bends it way back]_PULL-BACK.

(1.3) (Goldin-Meadow and Brentari 2017:11)

I [ran up the stairs]_SPIRAL-UP.

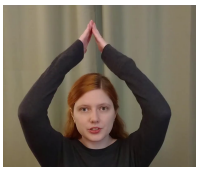
(1.4) (Ebert and Ebert 2016)

One child managed to cut out [a geometrical form]_TRIANGLE.

(1.5) (Schlenker 2018a)

a. John [helped]_UP  his son.

b. John brought [a bottle of beer]_LARGE  .

(1.6) Give me the troll with the wild blue [hair]_WACKY-HAIRSTYLE  .

Consider the series of co-speech gestures in example (2.5) (see Figure 1.1 for an illustration); they serve to *depict* each of the actions described by speech in (a)-(c). However, gestures need not simply mirror the speech they accompany; they can also elaborate on speech by communicating information beyond the concurrent speech. All of the co-speech gestures in (2.1–2.6) convey some information which the speech stream does not. For example, the co-speech gesture PULL-BACK in (2.1) involves the speaker making as if to grip and pull back something flexible which is fixed at the base, such as a sapling. The iconic information communicated by the gesture but not the spoken words is that the object being pulled back is fixed at the base. Similarly, the SPIRAL-UP gesture in (2.2) alone conveys the fact that the staircase in question is a spiral staircase, and the gesture TRIANGLE in (2.3) entails that the geometric shape in question is a triangle. UP and LARGE in (2.4a,b) imply, respectively, that the helping was done by lifting and that the bottle was large. Finally, the gesture WACKY-HAIRSTYLE in (2.6) indicates a particular crazy troll hair style that is not described in the speech stream.

In the next section we define gesture, and then look at gesture classification systems.

1.1.2 What is gesture?

McNeill (1992:11) identifies gestures as spontaneous movements of the arms and hands produced when people are talking that "are closely synchronized with the flow of speech". Kendon (2004:7) defines a gesture as a "visible action when it is used as an utterance part or as a part of an utterance" and, crucially, used with the goal of communicating something. They have "features of manifest deliberate expressiveness", and are readily identified by interlocutors as being deliberate and communicative (Kendon 2004:15). (This is in contrast to unconscious "nervous"-type movements by speakers or functional movements such as picking up a glass of water that are not perceived as intentionally communicative or aligned

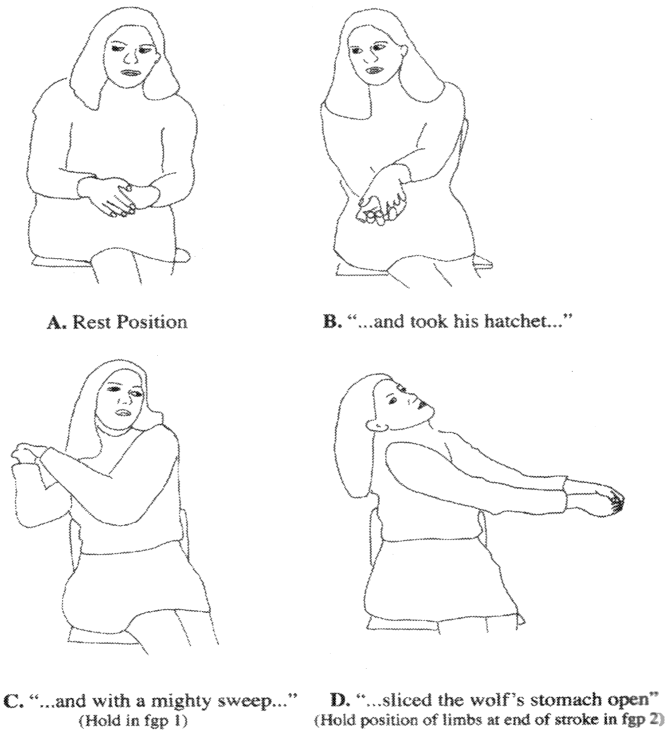


Figure 1.1: *Illustration showing the gestures accompanying example (2.5) (Kendon 2004:136).*

with speech.)

McNeill (1992) urges us to consider gestures and language (speech) as two parts of a single system and the result of a single process of utterance formation. His arguments include that co-speech gestures are synchronous with speech, they are semantically and pragmatically coexpressive with speech, they develop together with speech in children, and they break down together with speech in cases of aphasia. Co-speech gestures can also convey information that the speech stream alone does not, as we saw in the examples in (2.1–2.6). Unlike speech, they are interpreted in an analog gradient fashion, rather than discretely like spoken morphemes; this property is discussed further in Chapter 4.

1.1.3 Gesture classifications

Gestures can be classified according to their communicative function, the kind of meaning they express, their co-occurrence with speech, and their degree of conventionality and arbitrariness, among other factors (Abner et al. 2015; Kendon 2004; McNeill 1992, 2005; among others). Other types of gestures we will not be addressing include *emblems* (culture specific gestures with fixed, arbitrary meanings that can be used with or without accompanying speech, such as the "thumbs-up" gesture), *pantomimes* (sequences of complete gestures that iconically depict scenes and/or events and never co-occur with speech), *beats* (short and simple non-iconic movements that pattern closely with an utterance's prosodic peaks), and *deictics* (pointing gestures commonly used with demonstratives and locative adverbs). *Iconic* co-speech gestures can be distinguished from these other types of gestures by their co-occurrence with speech, their lack of arbitrariness, and their conveyance of meaning through iconicity. However like other gestures, they are naturally occurring linguistic phenomena that are part of the linguistic system (Goldin-Meadow and Brentari 2017; Kendon 1980; McNeill 1985, 1992); they are spontaneously produced by speakers, and are not formally taught to learners.

Gestures can also be classified into foreground and background gestures (Cooperrider 2017). *Foreground gestures* are those gestures that are designed to communicate critical aspects of meaning and are at the foreground of the speaker's and listener's awareness. *Background gestures*, on the other hand, are automatically produced as a byproduct of communication, but are not intentionally communicative. Foreground gestures can be identified by the following four properties:

(1.7) The hallmarks of foreground gesture (Cooperrider 2017)

(i) Concurrent use of a spoken demonstrative

(ii) Absence of speech

(iii) Co-organization of gesture and speaker gaze

(iv) Conspicuous effort in the movement itself

Each of these hallmarks indicates increased intention, awareness, and design on the part of the speaker, and hence foreground gestures can be seen as "listener-oriented", specifically designed to communicate meaning. Background gestures, on the other hand, tend to be "speaker-oriented", since they are spontaneously generated and occur even while speakers are (consciously) not in view of their listeners. While beat gestures are typically background gestures and pantomimes are typically foreground gestures, iconic co-speech gestures can be either foreground or background gestures.

Gestures, and co-speech gestures in particular, closely align with the prosodic structure of speech. A gesture typically consists of a preparatory phase, followed by the gesture *stroke*, defined as the communicative part of the gesture, and finished by a recovery phase, where the hands return to neutral position (Kendon 2004:Ch. 7). Gestures are coordinated with speech so that the gesture strokes align with prosodically prominent units in speech (e.g. nuclear accents).

1.1.4 Dependence of gesture on language

Gesture may be universal, but the *way* we gesture can vary from language to language. Crosslinguistic research with adults, children, and bilinguals have shown that the language spoken influences how speakers represent the same events in iconic gestures (Özyürek 2017).

In a well-known elicitation study, Kita and Özyürek (2003) compare how Japanese, Turkish, and English speakers gesture about an event where the relevant lexical items in each language differ structurally. The event in question is a Sylvester (cat) and Tweety (bird) cartoon, where Sylvester is chasing Tweety. At one point Sylvester has to grab a rope and swing across a gap in order to catch up to Tweety. English, Turkish, and Japanese happen to differ in how they express a "swinging across" motion event linguistically; of the three, only English has an intransitive verb, *to swing*, which encodes the arc-like motion as part of the manner information of the verb. Japanese and Turkish speakers mostly used the equivalent of the manner-less "go across" to describe the scene. Strikingly, in their narration of the

event, most English speakers produced a gesture with an arc trajectory, while most Turkish and Japanese speakers produced a straight gesture without indicating an arc path. The language spoken directly influenced what kind of gesture the speakers made to describe the event.

We will mainly be looking at gestures that accompany English in this dissertation, except in Chapter 4 where we will also be analyzing data from American Sign Language.

1.2 Semantics background

Perhaps the two most well-studied semantic properties of co-speech gestures are their not-at-issueness and their projection behavior. In this section we give a brief introduction to these two topics, and also mention earlier work modeling co-speech gestures within Discourse Representation Theory.

1.2.1 (Not-)at-issueness

Following on very recent work by Ebert and Ebert (2014, 2016), there has been much interest in the formal literature in the propensity of co-speech gestures to be not-at-issue with respect to the discourse topic. This section defines what it means for content to be (not-)at-issue and presents diagnostics for testing at-issueness, and I show, using several diagnostics, that co-speech gestures are generally not-at-issue.

Diagnostics

Tonhauser (2012) proposes three types of diagnostics for determining whether content is at-issue. These diagnostics rely on the following three properties of at-issue content:

(1.8) Properties of at-issue content (Tonhauser 2012)

(i) At-issue content can be directly assented or dissented with.

(ii) At-issue content addresses the question under discussion (QUD).

(iii) At-issue content determines the relevant set of alternatives.

The diagnostics will first be discussed, and then applied to co-speech gestures in Section 1.2.1.

Property (i) says that an addressee can explicitly accept or reject the at-issue content of an utterance. For example, in (1.9) the addressee B can assent/dissent with only the information that Juan lives in Maria's house, not with the fact that Maria has a house; thus the at-issue content is where Juan lives, and the information about Maria having a house is not-at-issue.

(1.9) A: Juan lives in Maria's house.

B: No, that's not true. / Yes, that's true. (Tonhauser 2012)

To assent/dissent with not-at-issue content, the addressee must use a challenging utterance such as "Hey, wait a minute!" or something similar to indicate discourse infelicity (Shanon 1976; von Fintel 2004; Roberts 1996/2012:among others).

Aside from directly asking consultants for intuitions about what is being assented/dissented with, one can also ask consultants to judge the felicity of assents/dissents with positive or adversative continuations. A(n) *positive (adversative) continuation* of an assent is a continuation of an utterance which conveys (the negation of) either the hypothesized at-issue content or the hypothesized not-at-issue content. Consider the example (1.10):

(1.10) (adapted from Tonhauser 2012)

A: That man, my mother's friend, stole your money.

B1: #Yes, true, but he didn't steal my money.

B2: Yes, true, but he's not your mother's friend.

(1.10) is an example of an adversative continuation of an assent. Utterance A contains two pieces of information: (a) that speaker B's money was stolen, and (b) that the thief was A's mother's friend. Response utterance B1 agrees with utterance A, but then qualifies the agreement by dissenting with proposition (a) (i.e. by continuing with the negation of (a)).

Since B1 is infelicitous, we can conclude that the content of (a) is at-issue in utterance A. Utterance B2 similarly assents to utterance A but then dissents with proposition (b); this time, however, the result is felicitous, which, according to property (i), means the content of (b) is not-at-issue.

Property (ii) states that at-issue content has to address the main topic of the discourse; in particular, it needs to be relevant to (i.e. at least partially answer) the Question Under Discussion (QUD). This fact can be used, then, to directly test whether an utterance can be a felicitous response to an explicit QUD. Suppose the content whose status you're trying to determine is in the statement in (1.9), repeated below as (1.11:

(1.11) Juan lives in Maria's house. (Tonhauser 2012)

This statement conveys/entails at least two propositions: (a) Maria has a house, and (b) Juan lives there. Forming an interrogative utterance to be an explicit QUD from each of these propositions allows us to check if (1.11) can be a response to either of these questions.

(1.12) A1: What does Maria have? (Tonhauser 2012)

A2: Where does Juan live?

B: {1:#/2:✓}Juan lives in Maria's house.

Since utterance B can be a response to the explicit QUD A1, the content associated with A1, namely (a), is at-issue. However, B cannot be a response to A2, so the content associated with A2, (b), is not-at-issue.

Property (iii) says that at-issue content in a question determines the relative set of alternatives for that question. A test based on property (iii) is a bit like the reverse of the one for property (ii); whereas the former involved using an answer to pick a QUD containing either the hypothesized at-issue content or not-at-issue content, the latter instead involves using a polar question to see which answer(s) is appropriate for that question. Consider (1.11), rephrased as a polar question in (1.13).

(1.13) A: Does Juan live in Maria's house? (Tonhauser 2012)

B1: Yes, he does/lives in Maria's house.

B2: #Yes, Maria has a house.

The at-issue content of interrogative utterance A dictates the semantic alternatives that count as appropriate responses. In this case, response utterance B1 is felicitous, but response B2 is infelicitous, indicating that the content expressing where Juan lives is at-issue, but the information that Maria has a house is not-at-issue in A.

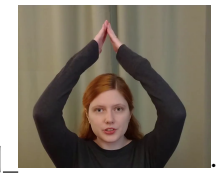
In the next subsection I apply these diagnostics to co-speech gestures to determine their informational status.

Co-speech gesture

Ebert and Ebert (2014, 2016) propose that co-speech-gestures are generally not-at-issue. In this section I apply the above diagnostics for at-issue content to test this hypothesis.

First, at-issue content can be assented/dissented with, and not-at-issue content cannot. We test this using an assent with adversative continuation in (1.14).

(1.14) Test 1: Assent with adversative continuation

A: Sasha found her [toy trolls]_  .

B1: #Yes, true, but she didn't find them.

B2: Yes, true, but they had a different hairstyle.

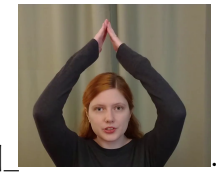
Sentence B1, challenging the proposition that Sasha found trolls, is incompatible with assent; hence, that content is at-issue. Sentence B2, by contrast, felicitously challenges the content of the co-speech gesture (i.e. which hairstyle the trolls had), and so that content is not-at-issue.

Next, only and all at-issue content addresses the current QUD. It is possible therefore to test if the content in question can be directly questioned by an explicit interrogative QUD; this is done in (1.15).

(1.15) Test 2: Pick an appropriate QUD for a given response

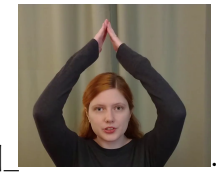
A1: What did Sasha find?

A2: What kind of hairstyle did the toy trolls have?

B: {1:✓/2:#}Sasha found her [toy trolls]_  .

Utterance B can only be the response to interrogative utterance (and explicit QUD) A1, not A2. Assuming that content is at-issue if and only if it addresses the QUD, which is explicitly given here as either question A1 or A2, then this shows that the at-issue content of B is the event of Sasha finding the trolls, and the proposition about the type of hairstyle the trolls had is not-at-issue content.

(1.16) Test 3: Pick an acceptable response with continuation for a given QUD

A: Did Sasha find her [toy trolls]_  .

B1: Yes, she did/found them.

B2: #Yes, the toy trolls/they had hair like that.

Tests 1, 2, and 3 all lead to the same conclusion: co-speech gestural content is typically not-at-issue.

1.2.2 Projection properties of co-speech gesture

In addition to co-speech gestures being not-at-issue, the other observation that has received most of the attention in the semantics literature is that co-speech gestures project out of operators like negation and quantification. Projection is defined below in (1.17):

(1.17) **Definition of Projection** (Simons et al. 2010)

An implication *projects* if and only if it survives as an utterance implication when the expression that triggers the implication occurs under the syntactic scope of an entailment-canceling operator.

Are co-speech gestures projective content? Consider an example utterance (1.18) containing a co-speech gesture SPIRAL, which is made by rotating an extended index finger in an upwards spiral direction and conveys the implication that the relevant motion (here, ascending stairs) is in an upwards, spiral direction (such as up a spiral staircase).

(1.18) Olivia [went up the stairs]_SPIRAL.

Using the "family of sentences" diagnostic (Chierchia and McConnell-Ginet 2000), it can be seen that co-speech gestural content does indeed project. In the variants of (1.18) in (1.19), the implication from the co-speech gesture SPIRAL in (1.18) is still present in polar questions (1.19a) and under the scope of entailment-canceling operators like negation (1.19b), the antecedent of a conditional (1.19c), and epistemic modals (1.19d).

(1.19) a. Did Olivia [go up the stairs]_SPIRAL?

b. Olivia didn't [go up the stairs]_SPIRAL.

c. If Olivia [goes up the stairs]_SPIRAL, she will see the cat at the top.

d. It's possible that Olivia [went up the stairs]_SPIRAL.

In other words, if the speaker is committed to the information conveyed by the SPIRAL co-speech gesture in (1.18), then she must also be committed to it in (1.19a-d).

Since these judgments are sometimes relatively subtle, Tieu et al. (2017, 2018) set out to prove experimentally that co-speech gestures project under negation and quantification. Their work and results are summarized in Section 1.4.

1.2.3 Discourse models

In addition to the work discussed above, Lascarides and Stone (2009) present an integrated logical form (LF) for multimodal discourse including both verbal speech and gesture. They work within the framework of Segmented Discourse Representation Theory (Asher and Lascarides 2003) and utilize rhetorical relations. This integrated LF formalizes their claim that "gesture and speech present complementary, inferentially related information as part of an integrated, overarching speech act with a uniform force" (Lascarides and Stone 2009:404).

1.3 Theoretical work on co-speech gesture projection

In this section we look at several aspects of co-speech gesture projection in more detail: the debate over whether to analyze co-speech gestures as supplemental or cosuppositional; a proposal for a typology of projection properties of iconic enrichments, including co-speech gestures; and an alternative syntactic account.

1.3.1 Compositional semantics: Supplemental or cosuppositional?

There are currently two competing theories for the compositional semantics of co-speech gestures. The first theory, proposed by Ebert and Ebert (2014), says that co-speech gestures are supplements in the sense of Potts (2005). For example, sentence (1.20a) with a co-speech gesture BIG indicating the size of a bottle of water would have a meaning roughly equivalent to (1.20b) with a non-restrictive relative clause expressing the content of the gesture.

(1.20) a. I brought [a bottle of water]_BIG  to the talk.

b. I brought a bottle of water to the talk, which (by the way) was big.

The co-speech gesture is treated like an appositive relative clause.

The second theory, put forward by Schlenker (2018a), is that the semantic contribution of a co-speech gesture is a special kind of presupposition which he calls a conditionalized presupposition, or *cosupposition*. The presupposition is conditionalized on the asserted content of the utterance.

(1.21) Cosuppositions triggered by co-speech gestures (Schlenker 2018a:316–317)

> Let G be a co-speech gesture co-occurring with an expression d, and let g be the content of G. Then G triggers a presupposition $d \Rightarrow g$, where $\Rightarrow$ is generalized entailment (among expressions whose type ends in t).

For example, (1.20a), repeated below in (1.22a), would come with the presupposition (1.22b).

(1.22) a. I brought [a bottle of water]_BIG  to the talk.

b. *Cosupposition*: If I brought a bottle of water to the talk, it would be big.

⇒ The bottle of water was big.

The conditionalized nature of the presupposition is even clearer when the gesture is embedded under negation; consider (1.23).

(1.23) (Schlenker 2018a:297)

a. John didn't [help]_UP  his son.

b. *Cosupposition*: If John had helped his son, he would have done so by lifting him.

Although the asserted content says that John didn't help his son, the co-speech gesture indicates the counterfactual that if John had done so, he would have done so by lifting him up in some way.

Both theories correctly predict the projection of gestural content through negation (see Section 1.2.2). They make different predictions, however, when it comes to duplicating content in the speech stream. Supplements are generally infelicitous when they are trivial; consider the following example.

(1.24) #John brought a big bottle of water, which (by the way) was big, to the talk.

The appositive relative clause in (1.24) cannot duplicate the asserted content that the bottle of water was big. A theory of co-speech gesture as supplement would thus predict that they cannot be trivial, that is, they cannot simply duplicate content already present in the speech stream. However, a theory of co-speech gesture as cosupposition makes no such restriction; presuppositions can be trivial when their content is asserted. These predictions about the ability of co-speech gestures to be trivial will be put to the test in Chapter 2.

1.3.2 An iconic pragmatics

Co-speech gestures can be thought of as iconic additions, or enrichments, to spoken language. Schlenker (2018b) proposes a typology of projection properties for iconic enrichments, including co-speech gestures, in both spoken language and sign language. Two parameters, [±*internal*] and [±*separate_time_slot*], constrain the possible information statuses and types of inferences of iconic content. The *separate time slot* parameter keeps track of whether iconic enrichments occupy their own time slot in the speech/sign stream or whether instead they co-occur with speech/sign. *External* enrichments ([-*internal*]) are defined as modifications to linguistic expressions that can be eliminated/removed without affecting the integrity of the expression they modify; an example is co-speech gesture (4.15), where the gesture is "added" to a spoken expression in a different modality. *Internal* enrichments, on the other hand, involve modulations of words and signs themselves, for example vocal iconic vowel lengthening (4.16).

(1.25) Co-speech gesture: [−*internal*, −*separate_time_slot*]

John brought [a bottle of beer]_LARGE  . (Schlenker 2018b:881)

⇒ the bottle of beer was large

(1.26) Vocal iconic modulation: [+*internal*, −*separate_time_slot*]

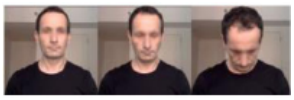
The talk was loooong. (Schlenker 2018b:881)

⇒ the talk was very long

Examples of iconic enrichments that have their own time slot ([+*separate_time_slot*]) include *post-speech/post-sign gestures*, which follow the expressions they modify, and *pro-speech gestures*, which fully replace an expression and thereby fulfill a grammatical function. Some examples are given in (4.17).

(1.27) a. Post-speech gesture (Schlenker 2018b)

John brought [a bottle of beer] — LARGE  .

b. In two minutes, our Chair will DOZE-OFF  .

Schlenker uses these two features to explain and predict the pragmatic statuses and semantic contributions of iconic enrichments in speech and sign. He proposes the following generalization:

(1.28) Iconic enrichment pragmatic typology: Proposed generalization (Schlenker 2018b:887)

a. [±internal]

External enrichments (−*internal*) are not-at-issue: because they are external, it should be possible to disregard them without affecting the main, at-issue content of the clause they appear in. By contrast, internal enrichments (+*internal*) can make any semantic contribution — just like standard words.

b. [±separate time slot]

Enrichments that have a separate time slot (+*separate time slot*) cannot be trivial (=presupposed): because they have their own time slot, they must make a non-trivial contribution to the sentence. By contrast, enrichments that do not have a separate time slot (−*separate time slot*) are not so constrained.

This generalization entails in particular that co-speech/co-sign gestures (−*internal*, +*separate time slot*) have a not-at-issue semantics, while iconic modulations of sign or speech (+*internal*, −*separate time slot*) can be at-issue or not, as the case may be.

1.3.3 Syntactic attachment

Esipova (2018a) seeks to unify the proposals by Ebert and Ebert (2014, 2016) and Schlenker (2018a) for co-speech gestures in quite a different way from Schlenker's (2018b) iconic

pragmatics, under a single syntactic account. She argues that co-speech gestures are a type of adnominal content that can compose as supplements, like appositives, or as modifiers, like adjectives (where a *modifier* is defined to be a piece of content that combines with constituents denoting sets of entities and returns subsets thereof). When they compose like supplements, they attach at the DP-level syntactically, and when they compose like modifiers, they attach at the NP-level. Crucially, the compositional integration of content in this account is modality-blind; gestures integrate compositionally in the same way as spoken content.

This composition-driven analysis, together with the assumption that there is a pragmatic pressure for co-speech gestures to be truth-conditionally vacuous, is designed to explain the projecting behavior of co-speech gestures. Esipova (2018a) presents novel experimental data testing the availability of restrictive (at-issue, non-projective) interpretations of co-nominal gestures. Participants recruited on Amazon Mechanical Turk read context paragraphs, watched videos of sentences uttered in those contexts, and provided acceptability judgments using a slider from 'totally unnatural' to 'totally natural'. Items varied across target interpretation of the co-nominal gesture: projecting non-restricting, restricting, or non-projecting non-restricting. (1.29) is an example with all three target interpretations.

(1.29) *Context: We are going on a group tour. Anna and Maria are responsible for renting a van. Maria just told Anna that Stephanie...*

a. (Projecting non-restricting) *..., who has two pets, a small cat and a large dog, is planning to bring along one of her pets. Anna, who has seen both Stephanie's pets before, says:*

   Do you know which one of Stephanie's pets is coming with us? 'Cause if she's bringing [her cat]_SMALL, we'll be fine, but if she's bringing [her dog]_LARGE, we should get a bigger van.

b. (Restricting) *..., who has two dogs, a small Pug and a large Great Dane, is planning to bring along one of her dogs. Anna, who has seen both Stephanie's dogs before, says:*

Do you know which one of Stephanie's dogs is coming with us? 'Cause if she's bringing [her dog]_SMALL, we'll be fine, but if she's bringing [her dog]_LARGE, we should get a bigger one.

c. (Non-projecting non-restricting) *. . . is planning to bring along her dog. Anna knows that Stephanie only has one dog, but has never seen it. She says:*

Do you know how big Stephanie's dog is? 'Cause if she's bringing [her dog]_SMALL, we'll be fine, but if she's bringing [her dog]_LARGE, we should get a bigger van.

Adjectives and appositives were also tested. The results of the study were that restrictive interpretations were more available for gestures than for appositives, but less available than for adjectives. Compared to adjectives and appositives, judgments for gestures were more gradient and individually variable. Esipova interprets these results as indicating that neither a supplement-only analysis nor a cosupposition-only analysis suffices for co-speech gestures; neither theory on its own sufficiently captures the available gestural interpretations. Assuming that co-speech gestures can compose as either a supplement or a modifier, however, and assuming that gestures are preferentially truth-conditionally vacuous, explains the data patterns.

1.4 Experimental work on co-speech gesture projection

In addition to the theoretical work discussed above, Tieu et al. (2017, 2018) performed a series of experiments to test the availability of projection readings of co-speech gestures in a variety of contexts. They focus on the directional gestures UP and DOWN (Figure 1.2).

The linguistic environments tested include plain affirmative sentences, negative sentences, modal sentences, and quantified sentences (*each*, *none*, and *exactly one*); examples are given in (1.30a-f).

(1.30) (Tieu et al. 2017:9)

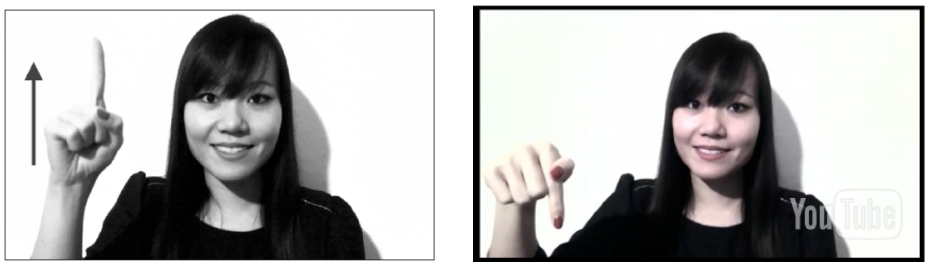


Figure 1.2: *The co-speech gestures UP and DOWN (Tieu et al. 2017, 2018).*

a. The boy will [use the stairs]_UP.

b. The boy will not [use the stairs]_UP.

c. The boy might [use the stairs]_UP.

d. Each of these three boys will [use the stairs]_UP.

e. None of these three boys will [use the stairs]_UP.

f. Exactly one of these three boys will [use the stairs]_UP.

They tested these sentences in the context of images (Figure 1.3) showing cartoon boys and girls either at the top or bottom of a set of stairs, which they could, respectively, potentially go down or up. Some of the stairs were blocked by barriers, in order to facilitate the availability of conditional inferences such as *If the girl were to use the stairs, she would go up the stairs*.

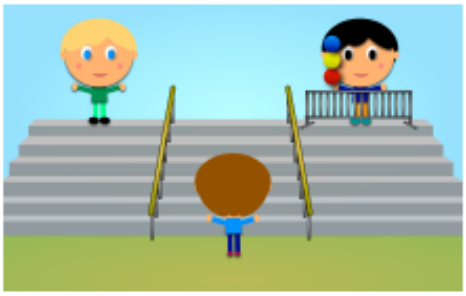


Figure 1.3: *Sample image used in the Tieu et al. (2017, 2018) studies.*

Tieu et al. (2017) report on two experiments and Tieu et al. (2018) on a third. Experiment 1

was a truth value judgment task (TVJT): participants were asked to judge whether the video utterances were true or false of the given pictures. Experiment 2, performed with a different group of participants but using the same videos and images, was a picture selection task: participants were given a video utterance and a pair of pictures and asked to choose which picture fit it best. Experiment 3 tested inferences more directly by asking a third group of participants to rate how strongly the target video utterances would lead them to infer the target inferences.

Taken together, the three experiments yielded evidence of the projection of the gesture's conditional inference (e.g. for DOWN, *If the boy were to use the stairs, he would go down the stairs*) from the scope of negation, evidence of existential projection of the conditional inference under *each*, *none* and *exactly one*, and evidence of universal projection under *none* and *exactly one*. They also found some evidence of local accommodation (i.e. at-issue behavior) under negation and *none*. These results are consistent with the cosuppositional analysis rather than the supplemental analysis of co-speech gestures (see Section 1.3.1).

A limitation of this study is that it only looks at one kind of gesture, the directional UP/DOWN gesture. Since this involves pointing, there is a chance its semantics may interact non-trivially with the semantics of deixis, which might mean it works differently than other iconic co-speech gestures.

1.5 Psycholinguistic work on the pragmatics of co-speech gesture

Lastly, it is worth mentioning the great deal of work on co-speech gesture in the psycholinguistic literature. One interesting line of questioning has to do with the use of iconic size gestures in elicited narratives. Beattie and Shovelton (2006:69) asked the question, "how important are the instances of size encoded gesturally as compared to linguistically within cartoon descriptions?" They filmed one set of participants narrating cartoons that they first saw projected on a wall. Two coders independently, and then collaboratively, identified all instances of iconic gestures in the narrative videos. Another independent set of participants was asked to rate size information from the cartoons based on importance to the storylines

of the cartoons. Using metrics based on the judges' ratings, Beattie and Shovelton found that in the instances where size was explicitly communicated by a participant in at least one modality, that the "high importance" size information was significantly more often represented by gesture only, rather than by speech only or by gesture and speech together, compared to the "low importance" size information. In other words, speakers are more likely to use a gesture to convey size information in a narrative if it is of high consequence to the story than if it is of low consequence.

Following up on this work, Holler and Stevens (2007) look at the role common ground plays in the use of iconic size gestures. Participants were grouped in pairs such that one member was the speaker and the other the listener. The speakers could see an image on the wall behind the listener, who could not see the image. The speaker had to describe the location of a referent so that the listener could afterwards pick out where referent should be in a busy crowded scene from a *Where's Wally?* book. The experimental condition that was manipulated was whether the listener and speaker had a chance to look at and talk about the scene together before the task began. Holler and Stevens found that when speakers talked to "unknowing" listeners who had not seen the image ahead of time, they conveyed size information predominantly in gesture only or in gesture plus speech. However when speakers talked to "knowing" listeners who shared a common ground with including the pictures, they primarily conveyed size information verbally, not gesturally. This is fascinating evidence showing that the common ground influences gestural communication.

1.6 Contributions of this dissertation

1.6.1 The question of gesture felicity

All of the semantic and pragmatic work on co-speech gestures done in the past few years is predicated on the assumption that speakers have intuitions about gestures in the same way they have intuitions about linguistic content. However, this fact has not been established, and should, perhaps, be questioned. It is not clear that people have such clear-cut judgments

about gestures, or that they're accustomed to probing their intuitions about them.

The acceptability rating study described in Chapter 2 of this dissertation is the first of its kind to really take a look at what makes co-speech gestures good (felicitous, natural) in the first place. Chapter 3 also looks at gesture felicity in specific, controlled pragmatic contexts.

1.6.2 New experimental factors studied

The previous section reviewed some of the factors affecting co-speech gesture felicity and meaning that have been experimentally studied in the literature in recent years: felicity under contrastive focus and projection properties. In this dissertation I significantly expand that list of factors to include the following new factors affecting the meaning and use of co-speech gesture:

- context sensitivity
- redundancy/triviality
- nominal vs. verbal modifier
- informativity
- size-and-shape vs. manner modifier
- truth conditions under negation

In Chapter 2 we experimentally test the effects of context on co-speech gestures, in particular on their acceptability ratings. This includes testing the immediate linguistic context for sensitivity to triviality; that is to say, testing whether co-speech gestures like to be truth-conditionally vacuous within an utterance. Using the results of this study we also analyze what, if any, role the type of modifier (nominal or verbal) plays in felicity. In Chapter 3, we experimentally test how pragmatic informativity affects the felicity of co-speech gestures. Finally, in Chapter 4, we systematically examine whether co-speech gestures being size-and-shape versus manner modifiers affects their truth conditions under negation.

1.6.3 Methodological contributions

The studies that comprise this dissertation make use of a new experimental paradigm that can be distributed via crowd-sourcing platforms such as Amazon Mechanical Turk. In order to test various linguistic factors relating to co-speech gestures, we embedded short videos, hosted on a video-hosting site such as YouTube, within an online survey created with survey software (e.g. Qualtrics). This enables participants to see gestures as well as hear speech as part of a complete utterance. The initial use of this paradigm was with an acceptability rating task, but it is easily extended to other semantic and pragmatic tasks, such as truth value judgment tasks, reference identification, and contrastive inference picture selection tasks.

1.7 Overview of findings: The puzzle of co-speech gesture

This section presents an overview of findings from this dissertation.

The first main result from the context study in Chapter 2 is that people would prefer gestures to not convey communicative content. Instead, interlocutors prefer gestures to be trivial, i.e. truth-conditionally vacuous. This is surprising, given the rich depictive ability of co-speech gesture.

The second main finding, from the troll study in Chapter 3, is that interlocutors prefer linguistic modifiers over gestural modifiers. This is true even when important information is hypothetically more suited to communication via gesture than via speech. There is simply a strong preference to communicate pragmatically important information in speech rather than in gesture.

The third main result, from the negation and ASL study in Chapter 4, is that co-speech gesture is not dispreferred as a communication channel simply because it is depictive. At-issue information can be conveyed depictively in a sign language like American Sign Language (ASL). Instead, there seems to be a preference for a *single mode* of communication.

Overall, all studies found pragmatic and semantic asymmetry between gesture and

language (spoken or signed). On the one hand, our result is not surprising given the perspective from psychological work on gesture that speakers often gesture for themselves and that gestures aid language production and are often not intended to be communicative. On the other hand, this is somewhat surprising given recent theoretical linguistic work on gesture that works under the assumption that co-speech gestures have the potential to contribute meaning in a similar way to linguistic content. In general, the lack of previous experimental work in pragmatics of gesture means that this can provide a starting point for future quantitative studies on the way that gestures contribute meaning.

1.8 Existing publications

Chapter 2 was published as joint work with Kathryn Davidson as Zlogar and Davidson (2018). "Effects of linguistic context on the acceptability of co-speech gestures". *Glossa: a journal of general linguistics* 3:73. 1–28.

The data in Chapter 3 was joint work with Anna Alsop and Kathryn Davidson and was presented as a talk, "Description and depiction in a reference game", at XPRAG.it 2018 on June 1, 2018, in Pavia, Italy. The abstract for the talk was published as Zlogar et al. (2018) in *Frontiers in Psychology*.

The data in Chapter 4 was joint work with Kate Henninger and Kathryn Davidson and was presented as a poster, Zlogar et al. (2019). "Negating depictive modifiers in sign and speech", LSA 2019, New York, NY, Jan. 5, 2019.

Chapter 2

Effects of linguistic context on the acceptability of co-speech gestures

2.1 Introduction

One of the most productive developments of the last several decades of formal semantics/pragmatics research has been increasingly sophisticated models of different kinds of meaning that linguistic content can contribute, whether that content is at-issue or backgrounded, entailed or implicated, or updating or proposing updates to a shared common ground between interlocutors, among other distinctions. A classic three-way divide between entailments, implicatures, and presuppositions (see Chierchia and McConnell-Ginet 2000: Chapter 1 for a brief introduction) has been expanded in a number of ways through more complex distinctions. For example, although presuppositional elements like *too*, *again*, etc., often both require particular discourse conditions to be uttered and can "project" through other logical operators so as not to be affected by them, these two properties attributed classically to presuppositions can sometimes be dissociated, which means that attempts to model the contributions of these and other aspects of meaning should be careful to test each property separately.

Recently, this type of formal semantic/pragmatic analysis has been extended beyond

language in a narrow sense to co-speech gesture (Ebert and Ebert 2014, 2016; Tieu et al. 2017; Esipova 2018b; Schlenker 2018a). On the face of it, this is a natural extension, given that co-speech gestures are known to be prosodically integrated with spoken language and have been argued to contribute to a unified semantic content together with speech/sign (McNeill 1992; Kendon 2004; Goldin-Meadow and Brentari 2017). However, there is some reason for caution about assuming that applications of the same linguistic tests for levels of semantic/pragmatic contribution can be applied wholesale to gestural content. One reason might be that speakers of a language like English may be comfortable making metalinguistic judgments about speech, but less so about gesture, which is almost never a target of explicit instruction and so less frequently considered metalinguistically. Another reason is that gestural content may more often be interpreted in an analog way, potentially leading to more gradient grammaticality and/or truth value judgments. Finally, if there is some agreement among formal analyses of co-speech gestures, it is that they are frequently not-at-issue (Ebert and Ebert 2014, 2016; Schlenker 2018a), and not-at-issue content can vary significantly across phenomena and across languages in several respects, including, as we mentioned above, how it projects through various logical operators and whether it imposes any restrictions on the previous discourse. In this paper we focus on this last aspect, by experimentally testing the sensitivity of co-speech gestures to linguistic context, construed broadly as both discourse context and the local context of the simultaneous speech stream. Our hope is that this will contribute an important foundational piece as part of a larger discussion regarding the semantic, pragmatic, and information structural properties of co-speech gestures, and the practical issues involved in creating appropriate contexts in which to test them via fieldwork or experimentation.

2.1.1 Background: Co-speech gestures

The kind of *co-speech gestures* that have been the focus of recent semantic/pragmatic work and that we will be focusing on in this paper are gestures made simultaneously with speech/sign that enrich utterance meaning by iconically depicting an aspect of the situation

or event (Kendon 2004; Goldin-Meadow and Brentari 2017); see (2.1–2.6) for examples. In these and future examples, gestures will be written in all-capital letters, and we will indicate the spoken words that align with the gesture by placing them in square brackets.

(2.1) McNeill (1992:79)

. . . and he bends it way back_PULL-BACK

(2.2) Goldin-Meadow and Brentari (2017:35)

I [ran up the stairs]_SPIRAL-UP.

(2.3) Ebert and Ebert (2016)

One child managed to cut out [a geometrical form]_TRIANGLE.

(2.4) Schlenker (2018a:2)

John [helped]_UP his son.

(2.5) Tieu et al. (2017:3)

I brought [a bottle]_LARGE to the talk.

(2.6) She [scored]_SHOOT the winning point!

All of the co-speech gestures in (2.1–2.6) convey some information which the literal speech stream does not. For example, the co-speech gesture PULL-BACK in (2.1) involves the speaker making as if to grip and pull back something flexible which is fixed at the base, such as a sapling. The iconic information communicated by the gesture but not the spoken words is that the object being pulled back is fixed at the base. Similarly, the SPIRAL-UP gesture in (2.2) alone conveys the fact that the staircase in question is a spiral staircase, and the gesture TRIANGLE in (2.3) entails that the geometric shape in question is a triangle. UP in (2.4) implies that the helping was done by lifting, and LARGE in (2.5) implies that the bottle was large. Finally, the SHOOT gesture in (2.6) indicates that the scoring was done by shooting a basketball.

Kendon (2004:7) defines a gesture as a visible action "used as an utterance or as a part of an utterance" with the goal of communicating something. Gestures can be classified according to their communicative function, the kind of meaning they express, their co-occurrence with speech, and their degree of conventionality and arbitrariness, among other factors (McNeill 1992; Kendon 2004; McNeill 2005; Abner et al. 2015; among others). Other types of gestures we will not be addressing include *emblems* (culture-specific gestures with fixed, arbitrary meanings that can be used with or without accompanying speech, such as the "thumbs-up" gesture), *pantomimes* (sequences of complete gestures that iconically depict scenes and/or events and never co-occur with speech), *beats* (short and simple non-iconic movements that pattern closely with an utterance's prosodic peaks), and *deictics* (pointing gestures commonly used with demonstratives and locative adverbs). Iconic co-speech gestures can be distinguished from these other types of gestures by their co-occurrence with speech, their lack of arbitrariness, and their conveyance of meaning through iconicity. However, like other gestures, they are naturally occurring linguistic phenomena that are part of the linguistic system (Kendon 1980; McNeill 1985, 1992; Goldin-Meadow and Brentari 2017); they are spontaneously produced by speakers, and are not formally taught to learners.

The modest amount of previous formal semantic/pragmatic work on co-speech gestures has generally focused on inferences related to their not-at-issue-ness (Ebert and Ebert 2016; Tieu et al. 2017; Esipova 2018b; Schlenker 2018a), as illustrated by (2.7) where denial of the utterance cannot target the content of the gesture as in (b); instead (2.7a) can be continued by (b′) or (b″).

(2.7) a. John brought a [bottle of beer]_LARGE.

b. . . . #No, it was small.

b′. . . . Yeah, but it was a small one.

b″. . . . Yeah, and it was huge, you're right!

Of not-at-issue types of meaning, Schlenker (2018a) proposes that co-speech gestures are in fact a type of presupposition, specifically an assertion-dependent conditional presupposition he calls a *cosupposition*. Based on informal inference judgments, he demonstrates

that these cosuppositions exhibit standard presuppositional projection behavior in embedded contexts, except that the inference that projects is *conditional*. (In the examples below, reported inferences are indicated with the symbol ⇒.)

(2.8) Projected cosuppositional inferences of co-speech gestures (Schlenker 2018a:3–13)

a. John [helped]_UP his son.

   ⇒ John helped his son by lifting him.

b. John didn't [help]_UP his son.

   ⇒ If John had helped his son, he would have done so by lifting him.

c. If little Johnny takes part in the competition, will his mother [help]_UP him?

   ⇒ If little Johnny takes part in the competition, if his mother helps him, lifting will be involved.

d. None of these ten guys [helped]_UP his son.

   ⇒ For each of these 10 guys, if he had helped his son, this would have involved some lifting.

e. Does Samantha believe that John [helped]_UP his son? p.13

   ⇒? Samantha believes that if John helped his son, lifting was involved.

   ⇒?? If John helped his son, lifting was involved.

In contrast, Ebert and Ebert (2014, 2016) suggest that the way in which co-speech gestures contribute not-at-issue meaning is most analogous to *supplements*, such as expressives or non-restrictive relative clauses (Potts 2005). It is a subtle distinction: consider the case of non-restrictive relative clauses outlined in Chierchia and McConnell-Ginet (2000), which exhibit projection out of several operators in (2.9).

(2.9) Chierchia and McConnell-Ginet (2000:351)

a. Jill, who lost something on the flight from Ithaca to New York, likes to travel by train.

b. Jill, who lost something on the flight from Ithaca to New York, doesn't like to travel by train.

c. Does Jill, who lost something on the flight from Ithaca to New York, like to travel by train?

d. If Jill, who lost something on the flight from Ithaca to New York, likes to travel by train, she probably flies infrequently.

Sentence (2.9a) has the implication that Jill lost something on the flight from Ithaca to New York; sentences (b–d) show that the inference survives under negation, in a question, and in a conditional. However, unlike classic presuppositions, in none of (a–d) is it assumed by the speaker that the hearer already knows that Jill lost something, so these inferences have typically not been classified as presuppositions since they impose no restrictions on the background content.

An example of a supplemental analysis of the meaning of a co-speech gesture is given in (2.10), where Ebert and Ebert (2016) illustrate how the supplemental analysis predicts that the material conveyed by the gesture goes through as an inference in positive contexts (2.10a) but not in negative contexts (2.10b).

(2.10) Supplemental inferences of co-speech gestures (Ebert and Ebert 2016)

a. Some philosopher brought [a bottle of beer]_BIG yesterday.

   ⇒ Some philosopher brought a bottle of beer, which was big.

b. # No philosopher brought [a bottle of beer]_BIG yesterday.

   (Intended inference: No philosopher brought a bottle of beer, which was big.)

Although both the cosuppositional analysis and the supplement analysis of co-speech gestures take them to be not-at-issue, they make different predictions in several respects, one of which is projection behavior under negation, shown in (2.10) above. We will focus on one previously unexplored contrast in this paper: the restrictions that the gestures place on

the discourse, either in previous context or in the same sentence, specifically whether the gesture is *trivial*, duplicating content provided elsewhere. Supplements are awkward when trivial: compare (2.11a), where the supplement is trivial, to (2.11b), a similar structure but where the supplement contains nontrivial information.

(2.11) a. # My friend Jill lost her phone on her flight from Ithaca to New York yesterday. Jill, who lost something on the flight from Ithaca to New York, likes to travel by train.

b. My friend Jill lost her phone on her flight from Ithaca to New York yesterday. Jill, who frequently travels from Ithaca to New York, likes to travel by train.

Taking a supplement analysis of co-speech gesture off the shelf, we would expect degraded acceptability for co-speech gestures that are trivial, by analogy to speech supplements. The cosuppositional analysis makes a less overt prediction about triviality, but by analogy to presupposition we might in fact expect the reverse direction of acceptability: presuppositions are usually expected to be given/trivial, and so the cosuppositional analysis would be consistent with the finding that gestures are more acceptable when trivial.

Past experimental work on the cosupposition/supplement distinction comes from Tieu et al. (2017), who report evidence based on different patterns of projection in favor of the cosuppositional analysis, albeit one with some extra assumptions that need to be made. We note, however, that their work focuses on *inferential* judgments involving gesture, while we take as our starting point the quite different task of *acceptability* judgments of gestures in context. Our motivation for focusing on acceptability is that we begin with the impression that, to the extent that we have acceptability judgments about co-speech gestures, they vary greatly in ways that we cannot predict using current theories of co-speech gesture. This includes examples in existing literature, some of which are entirely natural (e.g., (2.2)) and others which we found to be less so (e.g., (2.8d)), and which cannot in our mind yet be entirely explained by processing constraints like complexity, frequency, familiarity, etc. We are therefore interested in asking what kinds of factors contribute to the acceptability of

co-speech gestures, so that gesture researchers creating linguistic examples can make more informed choices, including for important patterns like inferential judgments.

To summarize, we are interested in focusing on acceptability given the not-at-issue tendency of co-speech gestures, since dependence on previous context (in particular, being trivial) is unexpected for supplements but a hallmark of one type of presuppositional content (so-called "hard triggers"). With an eye toward fieldwork (or, in our case, work on an understudied phenomenon), Tonhauser et al. (2013) propose a typology for projective content. One of the properties that they note forms a class of projective content is the *strong contextual felicity (SCF)* condition, which will be the focus of this paper. By directly applying their suggested methodology for investigating expressions that seem to show projective behavior, we hope to better understand how to classify co-speech gestures in English in comparison to other projective spoken language phenomena, and to gain some more empirical evidence to bear on the theoretical analysis of co-speech gesture.

2.1.2 Background: Testing strong contextual felicity

The goal of our study is to essentially test how easily the content of not-at-issue meaning can be accommodated by an interlocutor. In other words, we are interested in asking whether co-speech gestures are more or less acceptable when they duplicate content or are entailed by content in the preceding discourse, or when they contribute new content. That is, are gestures better when they are informative, either with respect to the past context or the utterance that contains them? We will get at the former by testing whether their content must be entailed by the preceding discourse. Tonhauser et al. (2013) refer to this sensitivity to discourse context as *strong contextual felicity (SCF)*; more specifically, an item is said to be [+SCF] if and only if it is only acceptable in a discourse context that appropriately entails it, where context entailment is defined in (2.12):

(2.12) Tonhauser et al. (2013:75–76)

a. An *m-positive context* is one that entails or implies m.

b. An *m-neutral context* is one that entails or implies neither m nor $\neg m$.

Then the property of strong contextual felicity is defined as follows:

(2.13) *Strong contextual felicity constraint* (Tonhauser et al. 2013:76)

> If utterance of trigger t of projective content m is acceptable only in an m-positive context, then t imposes a *strong contextual felicity constraint with respect to* m.

Informally, if some expression can be uttered in a discourse context that neither supports nor denies an inference m, then that expression does not need to be supported by the context, so it is [−SCF] with respect to m. Formally, the following diagnostic can be used to decide on the strong contextual felicity value of an expression:

(2.14) *Diagnostic for strong contextual felicity* (Tonhauser et al. 2013:76)

> Let S be an atomic sentence that contains trigger t of projective content m.
>
> (i) If uttering S is acceptable in an m-neutral context, then trigger t does not impose a strong contextual felicity constraint with respect to m.
>
> (ii) If uttering S is unacceptable in an m-neutral context and acceptable in a minimally different m-positive context, then trigger t imposes a strong contextual felicity constraint with respect to m.

Part (ii) of the diagnostic captures the intuition that if an expression is not felicitous in a neutral context, but then is felicitous in a nearly identical context in which the only difference is that now m is entailed by the context, then we can safely conclude that the expression is [+SCF]. Using this diagnostic, we devised an experiment using Amazon Mechanical Turk to test if the semantic content of co-speech gestures is [+SCF] or [−SCF] by constructing various scenarios that permit co-speech gestures and manipulating the context to be either m-neutral or m-positive, where m is the proposition conveying the semantic content of the co-speech gesture.

As an example, consider the English additive particle 'too', which is standardly analyzed as having an existence presupposition of a salient parallel alternative proposition.

(2.15) Tonhauser et al. (2013:78–79)

a. [Context: Malena is eating her lunch, a hamburger, on the bus going into town. A woman who she doesn't know sits down next to her and says:]

#Our bus driver is eating empanadas, too.

b. [Context: Same as in (2.15a), but Malena is eating empanadas.]

Our bus driver is eating empanadas, too.

Let *m* be the proposition that somebody besides the bus driver is eating empanadas. The context in (2.15a) is *m*-neutral, because it doesn't specify that anyone else is eating empanadas (Malena is eating a hamburger), or that no one else is eating empanadas. In this context, the sentence with *too* is infelicitous. The minimally different context in (2.15b), by contrast, is *m*-positive, since Malena is now said to be eating empanadas, not a hamburger. In this context, the sentence with *too* is felicitous. Since the same utterance with *too* is infelicitous in an *m*-neutral context but felicitous in a minimally different *m*-positive context, we conclude by the diagnostic in (2.14) that *too* introduces a strong contextual felicity constraint, i.e., *too* is [+SCF] with respect to the implication *m*.

An example of a projection trigger that does not exhibit a strong contextual felicity constraint is the change-of-state construction *stop X*.

(2.16) Adapted from Tonhauser et al. (2013:80)

[Context: Laura, who doesn't live with her parents, visits them and asks them to sit down with her because she wants to tell them something:]

I've stopped eating gluten.

Let *m* be the proposition that Laura used to eat gluten. The context in (2.16) is neutral with respect to *m* since Laura's parents are not asserted or implied to know about Laura's gluten consumption or lack thereof. Because the utterance is felicitous in the *m*-neutral context, we conclude that *stop* does not impose a strong contextual felicity constraint, i.e., is [−SCF] with respect to the implication *m*.

Recall that we are interested in whether iconic co-speech gestures are better when they are *informative*, meaning either with respect to the past context or the utterance that contains

them. We will address the second of these issues by adding an additional dimension to the SCF diagnostic, which is whether the gesture is *trivial* or not, where by trivial we mean the semantic content of the gesture is duplicated in or entailed by the same proposition. This is especially expected to bear on the analysis of co-speech gestures as supplements, which predicts that triviality/duplication should result in lower acceptability of co-speech gestures. Our hypotheses are as follows:

- Hypothesis regarding *strong contextual felicity*: A large category of expressions carrying not-at-issue content require that content to be entailed by the preceding context. We ask whether co-speech gestures have the same requirement.

- Hypothesis regarding matching *speech cue*: One hypothesized analysis of gestures is as supplements (Ebert and Ebert 2014), which are infelicitous when trivial. We ask how matching content in the speech stream affects felicity of co-speech gestures, with the expectation that it should decrease grammaticality under the supplement account.

We describe the methods and procedures of our experiment, including the implementation of both of these factors, in the next section.

2.2 Experiment 1: Co-speech gestures and context sensitivity

2.2.1 Methods

Participants

Participants were 198 adults recruited through Amazon Mechanical Turk, restricted to the United States region. All participants self-identified as native speakers of English. Participants were compensated monetarily for their participation in the questionnaire via Amazon payments of $2. Experimental protocols were approved by the Harvard University Institutional Review Board under approval number IRB16-1331.

Procedure

The questionnaire was created using the Qualtrics Survey Software platform. Amazon Mechanical Turk workers were directed to a Qualtrics link in order to complete the survey. The questionnaire took approximately 10–20 minutes to complete. Participants completed the survey on their own time and on a device of their choosing after receiving a link to the survey; they were instructed at the start to make sure to use a device with a large enough screen to play videos and with working speakers/headphones. The experimental task instructions given at the start of the survey were as follows:

> If you choose to be in the study, you will complete a questionnaire. This questionnaire will help us learn more about co-speech gestures in English. You will be asked to watch short video clips and judge the naturalness of English sentences, and **you may find all of them completely natural, all of them not completely natural, or a combination: some completely natural and some not completely natural**.

The term *co-speech gesture* was not defined and was mentioned nowhere else in the instructions. We also chose not to define the term *natural*, in order to not prejudice the participants toward judging only the spoken words or only the gestures. The experimental design (discussed in detail below) was between-subjects, ensuring that each participant was shown either only videos with co-speech gestures or only videos without gestures, so as to help the gestures seem as natural as possible (since participants who saw gestures saw gestures in every trial). We worried that the participants viewing all no-gesture trials might feel uncomfortable rating every trial as "completely natural", which was the expected rating for non-gesture videos since all of the stimuli were designed to be felicitous English sentences. To counteract this possibility, we included and highlighted the bolded portion of the above instructions so that participants would know that this was a possible outcome and that they did not need to artificially rate some trials lower than others just for the sake of variety in their responses. (We will see in Experiment 2, Section 2.3, that the resulting pattern of judgments remained even after adding infelicitous non-gesture trials.)

For each of the 19 trials of the experiment, the participant was presented with the same

instructions, namely to read the short context paragraph and then click "play" to view the following video. The context paragraphs were one- to two-sentence paragraphs describing a conversation between "Eliza", the speaker in the video, and another named interlocutor. Each video featured the same speaker, a female native English speaker in her early twenties who was identified as "Eliza" from the context paragraph, saying a sentence or two that continued the discourse started in the written context paragraph.

The embedded videos were hosted through YouTube and participants could replay the video if they wished, although no mention of the possibility of replaying videos appeared in the instructions. We did not collect data on the frequency of participants replaying videos.

Beneath the video, participants were prompted to make a binary naturalness judgment in response to the following prompt:

> Please rate how natural you find Eliza's response in the video.

Participants were forced to press either the *completely natural* button or the *not completely natural* button to move on to the next trial. We specifically chose to describe Eliza's utterance (speech and gesture) in the video as a *response* because this speech word seemed to best encompass both her spoken words and her gestures. Other alternatives such as *utterance*, *statement*, *performance*, etc. we deemed to be either too technical or to implicitly bias the participants towards judging only the spoken words or only the gestural production. However, the term *response* implies that Eliza is part of a dialogue; hence we intentionally wrote each pre-video context paragraph as a dialogue, either implicit or explicit, between Eliza and another named interlocutor.

Each participant was presented with three "attention check" trials randomly interspersed between the experimental trials. These attention checks consisted of a context paragraph, written in the same manner as the experimental contexts, and a video in which Eliza informs the participant that this is an attention check and that they should press a particular response and move on to the next trial. One of the attention checks instructed the participant to choose "completely natural"; the other two instructed them to choose "not completely natural". These attention checks were used to filter out responses from participants who

were not watching or not paying sufficient attention to the videos.

Stimuli

For each trial of the survey, participants saw a screen like that shown in Figure 2.1. At the top of the screen were instructions to read the context paragraph and then view the embedded video; below the video was the linguistic naturalness judgment question, along with the two options *completely natural* and *not completely natural*. Participants were forced to choose one or the other in order to proceed to the next trial.

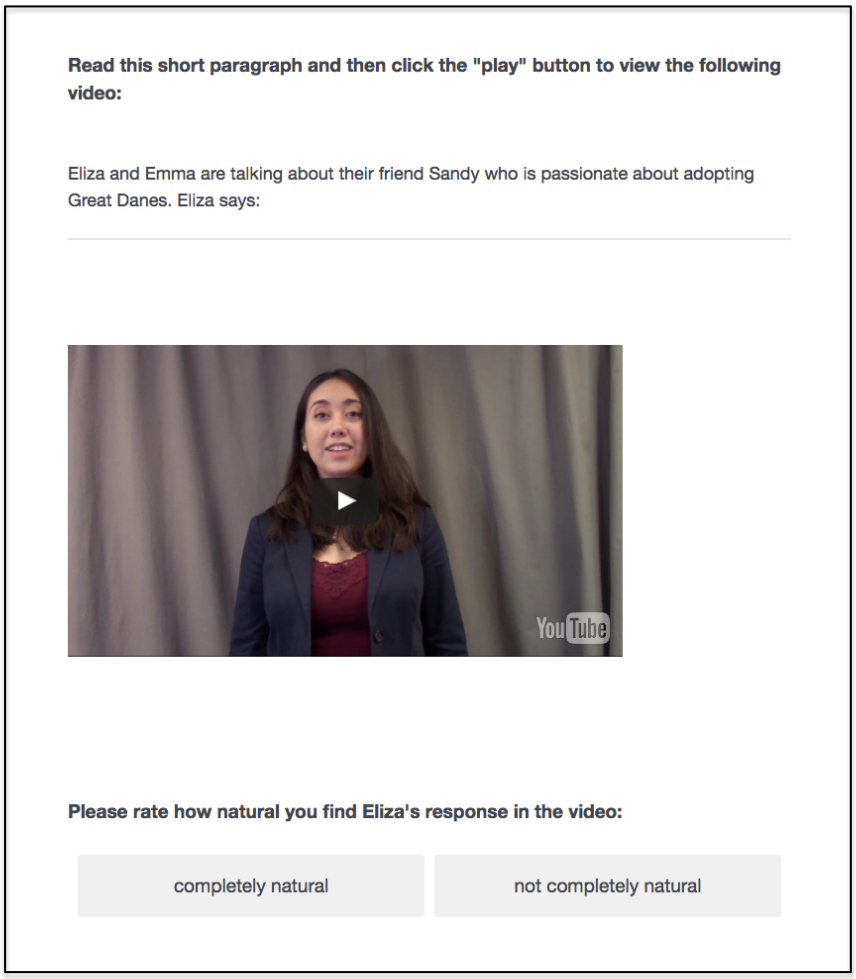


Figure 2.1: *A screenshot of a trial, as seen by a participant.*

Example (2.17) is one of the experimental *scenarios* seen by participants. The topic of

the scenario is that Eliza and a friend are discussing a co-worker's wealth and jewelry. The participant would see either (i) or (ii) as the written context paragraph at the top of the page, and then watch a video with Eliza uttering one of (a)–(d). Utterances (a) and (b) contain co-speech gestures, while (c) and (d) do not. The gesture EARRING conveys the information that the type of diamond jewelry Alicia was wearing was earrings; this information is expressed under *Proposition m* in the example below.

(2.17) Jewelry scenario, Experiment 1 (Scenario 9, Appendix A.1)

(i) *m-neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money. Eliza agrees and says:

(ii) *m-positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money based on her new pair of earrings. Eliza agrees and says:

(iii) *Proposition m for the gesture* EARRING: The type of jewelry was earrings.

a. Alicia was wearing real diamond [jewelry]_EARRING at work this morning.

b. Alicia was wearing real diamond [earrings]_EARRING at work this morning.

c. Alicia was wearing real diamond jewelry at work this morning.

d. Alicia was wearing real diamond earrings at work this morning.

In the appendices, we list the inference licensed by the co-speech gesture in each scenario explicitly as a proposition, so that it is clear what proposition *m* the *m-positive context* is supposed to entail/imply, in the Tonhauser et al. (2013) terminology.

In choosing co-speech gestures for the experiment, we avoided gestures that have conventionalized or codified meanings, such as a "thumbs-up" gesture, as these fall under the category of "emblems" and do not need to be accompanied by speech/sign to convey meaning (Kendon 2004). We also chose not to include pointing gestures, as these presumably intersect non-trivially with the semantics of deictic expressions (e.g., see the treatment in Lascarides and Stone 2009). The choice of gesture target (i.e., the NP or VP semantically

modified by the gesture) turned out to be as important as the choice of gesture in designing the stimuli; in order for a co-speech gesture to contribute non-trivially to the truth conditions of an utterance, the gesture target needs to be semantically underspecified in some way. Modeling our examples after those discussed in Ebert and Ebert (2016) and Schlenker (2018a), we chose gesture targets that are underspecified in manner when the target is a VP, or in adjectival content when the target is an NP; hence the co-speech gestures in our experiment function semantically as either manner adverbials or adjectival modifiers.

Figure 2.2 shows screencaps (captured "mid-action") of six of the co-speech gestures appearing in the videos shown to participants who were assigned to the "gesture-is-present" condition. These six co-speech gestures correspond to Eliza's video utterances like the following:

(2.18) a. Scenario 2 (Appendix A.1)

The basketball match she was in last night was incredible! She [scored]_SHOOT the winning point!

b. Scenario 3 (Appendix A.1)

Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

c. Scenario 6 (Appendix A.1)

The moon was so gorgeous last night — we just sat outside [looking up]_TELESCOPE at it for awhile.

d. Scenario 7 (Appendix A.1)

Alex kept [checking the time]_WRISTWATCH during the date.

e. Scenario 9 (Appendix A.1)

Alicia was wearing real diamond [earrings]_EARRING at work this morning.

f. Scenario 10 (Appendix A.1)

Lisa [performed]_VIOLIN really well at the recital last night!

For a full listing of the written and spoken experimental stimuli for Experiment 1, see

Figure 2.2: *Sample screencaps of co-speech gestures.*

Appendix A.1. Appendix A.2 provides screencaps for all 16 co-speech gestures appearing in the videos.

Design

Each participant saw 19 trials total: 16 experimental trials and 3 attention check trials. Each experimental trial (*scenario*) came in one of 8 types, depending on (i) whether or not there was a co-speech gesture in the video (the GESTURECUE condition); (ii) whether or not there was a *speech cue* in the video, i.e., a linguistic expression verbally expressing the content of the gesture (the SPEECHCUE condition); and (iii) whether the video was presented with a *neutral* context paragraph or a *positive* context paragraph (the CONTEXT condition), according to the Tonhauser et al. (2013) definitions; see Table 2.1 for a list of all eight trial types. We discuss the three experimental factors in more detail below.

At the start of each survey, the participant was randomly assigned to either the [GESTURECUE=yes] group or the [GESTURECUE=no] group. The sixteen scenarios were then presented in a randomized order (with the three attention checks randomly interspersed). For each scenario, a 2x2 Latin square design was used which crossed the two factors CONTEXT and

Table 2.1: *Experiment 1 trial types*

Trial type	GestureCue	speechCue	Context
1	no	no	neutral
2	no	no	positive
3	no	yes	neutral
4	no	yes	positive
5	yes	no	neutral
6	yes	no	positive
7	yes	yes	neutral
8	yes	yes	positive

SpeechCue. Each participant saw a trial for every scenario.

The GestureCue condition. The GestureCue condition encodes the presence ("yes") or absence ("no") of a co-speech gesture in the video utterance. As described above, this condition was manipulated between subjects; each participant saw either only videos with co-speech gestures or only videos without co-speech gestures. This design was chosen so as to help the gestures seem as natural as possible and to elicit more subtle felicity judgments, since there seemed to us to be a very real possibility that participants would rate any video with a gesture as worse than a video without one (as indeed turned out to be the case).

Within a scenario and given identical SpeechCue values, the corresponding videos with and without co-speech gestures were identical; that is, the words spoken by Eliza were exactly the same and were delivered, as nearly as possible, with the same intonation.

In trials with co-speech gestures, one factor that was not systematically controlled for was whether the gesture modified (i.e., whether the gesture *target* was) a verb or a noun. Consider the following two examples:

(2.19) a. Scenario 10 (Appendix A.1)

Lisa [performed]_VIOLIN really well at the recital last night!

b. Scenario 3 (Appendix A.1)

   Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

In (2.19a), the gesture target *performed* is a verb, while in (2.19b) the gesture target is the NP *a dog*. Out of the 16 total experimental scenarios, 13 had gestures modifying VPs, while only 3 had gestures modifying NPs, and one of these three needed to be excluded from analysis due to an accidental name mismatch between the contexts and the videos (Scenario 11). We discuss the results of a follow-up analysis on the data based on this division (noun/verb) in Section 2.2.3.

The SpeechCue condition. The SpeechCue condition encodes the presence ("yes") or absence ("no") of a spoken expression in the video that (approximately[1]) duplicates the same semantic content as the corresponding co-speech gesture for that scenario. We refer to this spoken linguistic expression as a *speech cue*. Consider the following pairs of target utterances (2.20) and (2.21), in which the (a) utterances do not have a speech cue (SpeechCue=no), while the (b) utterances do (SpeechCue=yes); the speech cues in question are bolded in the (b) utterances:

(2.20) Scenario 2 (Appendix A.1)

   a. The match she was in last night was incredible! She [scored]_SHOOT the winning point!

   b. The **basketball** match she was in last night was incredible! She [scored]_SHOOT the winning point!

(2.21) Scenario 3 (Appendix A.1)

   a. Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

   b. Sandy just got [a **big** dog]_BIG yesterday, and I hear it's quite the handful!

[1]The content of the SpeechCue and the co-speech gesture are not strictly equivalent, of course, given the iconic, analog nature of the co-speech gesture.

In (2.20b) the speech cue *basketball* indicates that the sporting event described is a basketball game, and the co-speech gesture SHOOT depicts the act of shooting a basketball; hence both contribute the semantic information that the sport in question is basketball (and so the co-speech gesture is *trivial* in (2.20b)). Similarly, in (2.21b) the speech cue *big* indicates that Sandy's dog is a big dog, and the co-speech gesture BIG manually depicts the same information (and is hence trivial).

Although the speech cues duplicate the semantic content of the designated co-speech gesture for a given scenario, in our stimuli a speech cue can and does appear in trial types for which [GestureCue=no], i.e., in video utterances that do not have a gesture. In these trial types, the speech cue is exactly the same as that which appears in the utterance that has the co-speech gesture for that scenario.

The timing of the SpeechCue with respect to its (roughly) equivalent co-speech gesture varied across scenarios. In some scenarios, the speech cue phrase occurred before the gesture target (2.22); in others, it was contained within or was the entire gesture target phrase (2.23); and in still others it occurred after the gesture target (2.24). (As before, in the following examples the speech cues are bolded for identification purposes.)

(2.22) Scenario 6 (Appendix A.1)

a. The moon was so gorgeous last night — we just sat outside [looking up]_TELESCOPE at it for awhile.

b. The moon was so gorgeous last night — we just sat outside **and took turns with the telescope** [looking up]_TELESCOPE at it for awhile.

(2.23) Scenario 4 (Appendix A.1)

a. Karen [ran]_RUN-DOWN to see what was happening!

b. Karen [ran **down**]_RUN-DOWN to see what was happening!

(2.24) Scenario 7 (Appendix A.1)

a. Alex kept [checking the time]_WRISTWATCH during the date.

b. Alex kept [checking the time]_WRISTWATCH **on his watch** during the date.

In (2.22b), the bolded speech cue *and took turns with the telescope* conveys that the looking up was done with a telescope, just as the gesture TELESCOPE does, and it occurs linearly before the gesture target *looking up*. In (2.23b), on the other hand, the bolded speech cue *down* occurs within the gesture target verb phrase *run down*. Finally, in (2.24b), the bolded speech cue *on his watch* occurs immediately after the gesture target and, like the gesture WRISTWATCH, conveys that Alex checked the time using his wristwatch.

As discussed above, the SpeechCue condition was manipulated within subjects and crossed with Context in a 2x2 Latin square design.

The Context condition. The Context condition encodes whether the written paragraph shown to participants before the video entails or implies the content of the co-speech gesture (a *positive* context), or whether it neither entails/implies the content of the co-speech gesture, nor entails/implies its negation (a *neutral* context). The "positive" and "neutral" terminology follows the naming conventions for *m*-neutral and *m*-positive contexts in Tonhauser et al. (2013), where here *m* is the semantic proposition expressing the information conveyed by the co-speech gesture. The effect is to manipulate the informativity/triviality of gestural content. An example can be seen in (2.25); the minimal differences between the two contexts are bolded for identification purposes.

(2.25) Scenario 2 (Appendix A.1)

a. *Neutral context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **athlete** who Tom hasn't heard of. Eliza says:

b. *Positive context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **basketball player** who Tom hasn't heard of. Eliza says:

c. *Sample target utterance*: The match she was in last night was incredible! She [scored]_SHOOT the winning point!

For this scenario, participants assigned [CONTEXT=neutral] were shown the context paragraph in (2.25a), while participants assigned [CONTEXT=positive] were shown the context paragraph in (2.25b). As seen in the sample target utterance (2.25c), the co-speech gesture for this scenario is SHOOT, which conveys that the scoring event was done by shooting a basketball. The "neutral" context (a) does not specify what type of sporting event is being discussed; hence it entails/implies neither SHOOT nor ¬SHOOT, meeting the criteria of being an *m*-neutral context. By contrast, the "positive" context (b) contains the information that the athlete in question is a basketball player; this entails that were this athlete to score in a game, she would do so by shooting a basketball. This is exactly the content of the gesture SHOOT, so (b) entails the content of the co-speech gesture and meets the criteria of being an *m*-positive context.

Notice that the positive and neutral context paragraphs in (2.25) are extremely similar; they differ only in the replacement of the phrase *athlete* with *basketball player*. In general we kept the neutral/positive context pairs as minimally different as possible, only making those changes from the neutral to the positive that were necessary to entail or imply the proposition expressed by the co-speech gesture. This follows Tonhauser et al.'s (2013) recipe to have an *m*-positive context be minimally different from an *m*-neutral context, in order to tell if *m* does indeed require the support of the context to be used felicitously.

The CONTEXT condition was manipulated within subjects, and, as discussed above, was crossed with SPEECHCUE in a 2x2 Latin square design.

2.2.2 Results

Out of 198 participants, 5 participants' responses were excluded due to those participants failing at least one of the three attention checks. The results discussed below are for responses from the remaining 193 participants.

Recall that the experimental design included eight trial types resulting from all possible combinations of the experimental conditions GESTURECUE, SPEECHCUE, and CONTEXT. Table 2.2 reports the mean acceptance rates, standard deviations, and standard errors across

the eight trial types, and Figure 2.3 shows mean acceptance rates for each trial type with standard error bars. Descriptively, we see an interesting pattern in which the presence of a co-speech gesture decreases acceptability (as indicated by the values of M for trial types 5–8), although the presence of a matching speech cue with the gesture somewhat mitigates this effect (as indicated by the values of M for trial types 7 and 8).

Table 2.2: *Experiment 1 trial types with conditions (*GestureCue, SpeechCue, *and* Context*) and mean acceptance rates*

Trial type	GestCue	SpCue	Context	**M**	**SD**	**N**	**SE**
1	no	no	neut	0.806	0.396	371	0.021
2	no	no	pos	0.775	0.418	360	0.022
3	no	yes	neut	0.771	0.421	363	0.022
4	no	yes	pos	0.734	0.442	361	0.023
5	yes	no	neut	0.607	0.489	364	0.026
6	yes	no	pos	0.604	0.490	359	0.026
7	yes	yes	neut	0.712	0.454	361	0.024
8	yes	yes	pos	0.691	0.463	356	0.025

With respect to the experimental conditions GestureCue, SpeechCue, and Context, trends were as follows. On average, trials with gestures (trial types 5–8) were rejected more often than trials without gestures (trial types 1–4) (with gesture: M=0.65; without gesture: M=0.77). Trials shown with a neutral context (1, 3, 5, and 7) were rejected at approximately the same rate as trials shown with a positive context (2, 4, 6, and 8) (neutral: M=0.72; positive: M=0.70). Finally, trials with a speech cue (3, 4, 7, and 8) were rejected on average at approximately the same rate as trials without a speech cue (1, 2, 5, and 6) (with speech cue: M=0.73; without speech cue: M=0.70).

Table 2.3 gives a breakdown of ratings by scenario. Scenario 11 trials were excluded from analysis because of an accidental name mismatch between the written contexts and the speech in the video. Means for the remaining 15 scenarios ranged from 0.57 (Scenario 1) to 0.82 (Scenario 5), with SDs ranging from 0.38 to 0.50.

Analyses of subjects' judgment responses were conducted using the R programming

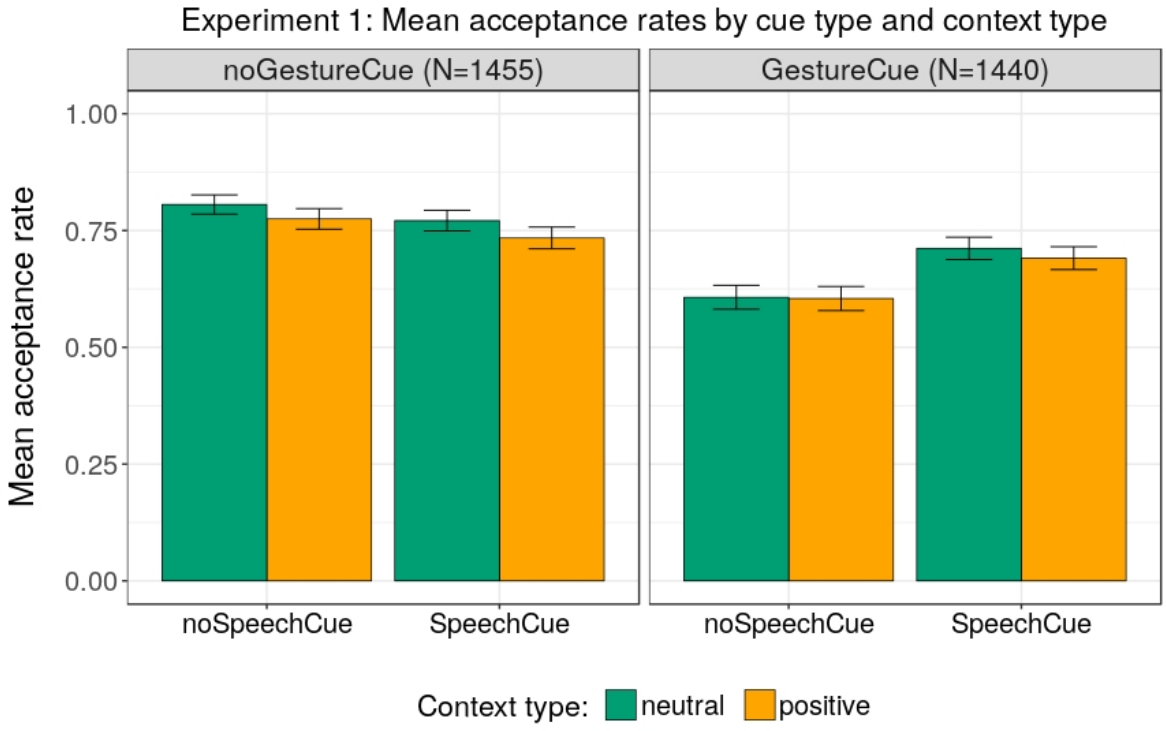


Figure 2.3: *Experiment 1 mean acceptance rates, by experimental conditions* GestureCue, SpeechCue, *and* Context.

language (version 3.2.3) (R Core Team 2016) to build generalized linear mixed effects models (Baayen et al. 2008), using the function *glmer* from the *lme4* package (Bates et al. 2015). In the model with the most data variation coverage (determined by ANOVA testing), the three independent factors were GestureCue (no/yes), SpeechCue (no/yes), and Context (neutral/positive); the binary naturalness judgment (coded as 1="completely natural" and 0="not completely natural") was the dependent variable, and scenario number and participant ID were coded as random effects.

Results of the model indicate that there was a significant main effect of GestureCue (β=−1.148, z=−4.908, p<0.001), general decreased the acceptance rate. but also a significant interaction of the factors GestureCue and SpeechCue (β=0.665, z=2.507, p<0.05). Neither SpeechCue nor Context were significant main effects (p>0.1); these findings agree with

Table 2.3: *Experiment 1 acceptance rates by scenario*

Scen	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**13**	**14**	**15**	**16**
M	.57	.66	.77	.70	.82	.75	.74	.74	.81	.75	.78	.66	.65	.69	.62
SD	.50	.47	.42	.46	.38	.44	.44	.44	.40	.44	.42	.48	.48	.46	.49

other models of the data that we constructed in which the only fixed factor was SpeechCue or Context ($p>0.1$ in each case). There were no other significant interactions between factors.

2.2.3 Discussion

study was how contextual support (within a discourse or the same utterance) influences the felicity of co-speech gestures. The first factor we tested was whether co-speech gestures need to be entailed by their preceding discourse context to be used felicitously. In the terminology of Tonhauser et al. (2013), we wanted to find out if co-speech gestures are [+strong contextual felicity (SCF)] or [−SCF]. Our experimental results indicate that manipulating the discourse context from being *m*-neutral to *m*-positive (where *m* is the proposition representing the semantic content of the co-speech gesture) had no significant effect on participants' acceptance rates of stimuli items. Since participants were equally likely to judge an utterance as completely natural with or without the positive entailment of the content of the co-speech gesture, and assuming that participants were judging the video with the context paragraph in mind, we can tentatively conclude that co-speech gestures are [−SCF]. In other words, co-speech gestures can either be easily accommodated or need no contextual entailment to be used felicitously. This is in clear contrast to other projective content like the English additive particle *too*, a so-called "hard" presupposition trigger (Abrusán 2016); as we discussed in example (2.15) in Section 2.1.2, *too* is [+SCF] with respect to its implication of the existence of a salient parallel alternative proposition in the discourse context. The differing behavior of *too* and co-speech gestures with respect to the content of

the contexts they appear in supports the claims that the former is [+SCF] (Tonhauser et al. 2013) and the latter is [−SCF]. In sum, our experiment has shown that co-speech gestures cannot be classified as "hard" presupposition triggers (assuming participants did indeed read and include the context paragraphs in their linguistic judgments, which concern we will address with Experiment 2 in Section 2.3).

The significant main effect of GestureCue in the data indicates that the presence of a co-speech gesture in a trial in general had a negative effect on the rating of the video. Intriguingly, this negative effect was mitigated when there was a speech cue present along with the gesture in the trial. One way to interpret this is that gestures are used most felicitously when they are semantically duplicating information already present in the speech/sign stream. This redundancy preference clearly pragmatically differentiates co-speech gestures from well-studied non-gestural supplemental material like appositives; appositives are not felicitous when they convey the same information as the main assertion (Potts 2005), as can be seen in (2.26):

(2.26) Adapted from Ebert and Ebert (2016)

> Paul, [the best horse riding instructor in the world], moved to Stuttgart recently (#and is the best horse riding instructor in the world).

The content of the appositive NP in (2.26), shown in brackets, cannot be reiterated as part of the foreground asserted content. However, as we see from the experimental results, partial and even complete redundancy is completely acceptable with co-speech gestures; (2.27) is an example of a trial from the experiment with both a co-speech gesture and a speech cue present (Scenario 3, Appendix A.1).

(2.27) Sandy just got [a **big** dog]_BIG yesterday, and I hear it's quite the handful!

Here the gesture BIG is arguably contributing very similar content to the speech cue *big*.

The significant interaction between GestureCue and SpeechCue indicates that speech cues had a different effect on judgments depending on whether or not there was a gesture; namely, trials with gestures had a higher average acceptance rate with a speech cue than

without, while trials without gestures had a lower average acceptance rate with speech cues than without. Presumably this has something to do with pragmatic calculations about the amount of information present in the utterance and any corresponding implications the hearer (our participants) might have drawn. Clearly, at this point our results raise more questions than answers, especially regarding whether the same pattern would emerge with different kinds of co-speech gestures.

ratings across scenarios, we did several follow-up analyses (which were not planned before the experiment was run) of scenario types to see how properties that varied by scenario affected judgments. We focused on three questions: (i) whether the gesture semantically modified a noun or verb; (ii) whether the speech cue that duplicated the gesture's content came temporally before, during, or after the co-speech gesture; and (iii) whether the belief state of the addressee, given the context paragraph, affected judgments. For the first question, we found that, overall, co-speech gestures that appeared in scenarios where they were modifying nouns were more acceptable than those modifying verbs (**nouns**: *neutral context*: M=0.83, SD=0.38; *positive context*: M=0.77, SD=0.42; **verbs**: *neut*: M=0.634, SD=0.48; *pos*: M=0.628, SD=0.48). However, the number of scenarios is too small to generalize beyond the sample (gestures modified nouns in two scenarios and verbs in thirteen scenarios); see Figure 2.4.

Next, speech cues that occurred after their corresponding gesture were generally less acceptable than those that occurred with or before the gesture (**after the gesture** (N=238): *neutral context*: M=0.63, SD=0.49; *positive context*: M=0.65, SD=0.48; **before the gesture** (N=93): *neut*: M=0.78, SD=0.42; *pos*: M=0.75, SD=0.44; **during the gesture** (N=386): *neut*: M=0.75, SD=0.43; *pos*: M=0.70, SD=0.46), but again, the sample size of each type of scenario is small (speech cues occurred after the gesture in five scenarios, before the gesture in two scenarios, and with the gesture in eight scenarios); see Figure 2.5.

between placement of speech cues (before, during, and after the geseture) may initially appear to be intriguing, but critically there was no difference between these timing conditions for speech cues in trials where the speech cue was present versus those where it was absent.

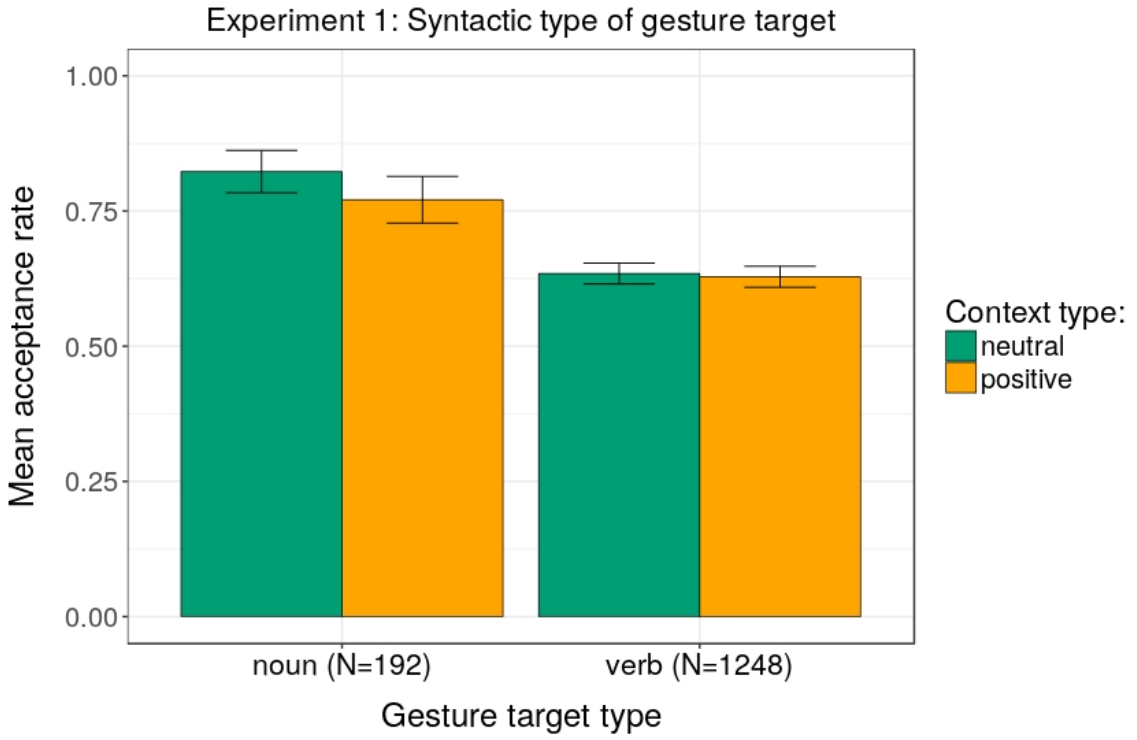


Figure 2.4: *Experiment 1 mean acceptance rates for type of syntactic item modified by gesture (noun vs. verb), by context type.*

Using the *glmer* function in R, we constructed a generalized linear mixed effects model (fit by Maximum Likelihood) on the subset of the data where a gesture was present [GestureCue=yes], with subject ID and scenario number as random effects, the participant's naturalness judgment (Response) as the dependent variable, and SpeechCue (yes/no) and SpeechCueTiming (before/during/after) as the independent variables. Neither the interaction between the presence of a speech cue and the speech cue coming after the gesture, nor the interaction between the presence of a speech cue and the speech cue coming before the gesture, was significant ($p>0.1$ in both cases). This suggests that the difference in ratings between the differing speech cue positions is an artifact of the large variation in trial types and not in the end about the speech cues themselves.

influencing acceptability judgments by participants is the belief state of the addressee in

Figure 2.5: *Experiment 1 mean acceptance rates for speech cue timing relative to gesture timing, by context type.*

the "dialogue" between Eliza and her interlocutor described in the context paragraphs. Does the addressee need to know/believe/be able to infer the content of the gesture proposition in order for the experimental participant to judge Eliza's use of the gesture felicitous?[2] We conducted a follow-up analysis on the data, classifying scenarios (yes/no) according to whether the addressee of the dialogue knows or could infer the content of the proposition in question based on the information provided by the "positive" context. We constructed two separate generalized linear mixed effects models, one on the "yes" subset of the data (where the addressee knows/can infer the semantic content of the gesture based on the positive context), and the other on the remaining "no" subset of the data, with RESPONSE as the dependent variable, SPEECHCUE, GESTURECUE, and CONTEXT as the independent variables,

[2]We thank an anonymous reviewer for raising this interesting question.

and subject ID and scenario number coded as random effects. Neither model showed a significant effect of CONTEXT on RESPONSE ($p>0.1$). From this we conclude that the belief state of the addressee is not a significant factor in acceptability judgments of utterances with co-speech gestures based on prior discourse contexts.

methodological issues which may have had an unintended effect and require follow-up studies. For one, we told participants that they were going to be asked to rate English sentences as *completely natural* or *not completely natural*; however, we did not explain or define the concept of linguistic naturalness. This may in fact have resulted in participants performing different rating tasks from each other due to interpreting "naturalness" in different ways. Some respondents, for example, may have been rating the "acting" performance of the speaker in the videos, rather than rating the felicity of the utterance (speech + gesture) alone. One way to address this might be to provide a simple, non-technical explanation of linguistic naturalness in the survey instructions so as to better communicate to participants what we as researchers are asking them to do. However, at this stage we are still unsure what exactly this would look like, and how one could dissociate performance factors from gesture with a prompt.

A more worrying methodological issue is that from Experiment 1, it is impossible to tell if participants were actually taking the time to read the written context paragraphs and if the contexts factored into their judgments of the video utterances. In fact this is related to our previous concern, because participants might rate the delivery (rather than the linguistic naturalness of the utterance) precisely because they weren't taking context into account. The attention checks built into the experiment could only detect if a participant was not watching or paying proper attention to the videos. Without knowing more about the participants' behavior with respect to the written contexts, we cannot fully conclude whether or not co-speech gestures need to be entailed by the discourse context. We address this concern in a follow-up experiment, Experiment 2, in the next section.

2.3 Experiment 2: Context follow-up

In order to address the methodological concern that participants were not reading the context paragraphs and/or not taking them into consideration in their rating of the videos, we conducted a follow-up study on MTurk in which participants were shown contexts that were pragmatically infelicitous with the video utterance, in addition to the usual neutral and positive contexts. The prediction was that if participants were truly reading the contexts and considering them when rating the videos, then the trials with infelicitous contexts would be accepted at a significantly lower rate. This prediction was indeed borne out.

2.3.1 Methods

Participants

Participants were 90 adults recruited through Amazon Mechanical Turk, restricted to the United States region. All participants self-identified as native speakers of English. Participants were compensated monetarily for their participation in the questionnaire via Amazon payments of $2. Experimental protocols were approved by the Harvard University Institutional Review Board under approval number IRB16-1331.

Procedure

The questionnaire was created using the Qualtrics Survey Software platform. Amazon MTurk workers were directed to a Qualtrics link in order to complete the survey. The questionnaire took approximately 10–20 minutes to complete. Participants completed the survey on their own time and on a device of their choosing after receiving a link to the survey; they were instructed at the start to make sure to use a device with a large enough screen to play videos and with working speakers/headphones. The experimental task instructions given at the start of the survey were the same as for Experiment 1 (see Section 2.2.1).

For each trial of the experiment, the participant was presented with the same instructions,

namely to read the short context paragraph and then click "play" to view the following video. The video stimuli used were a subset of the video stimuli from Experiment 1. The embedded videos were again hosted through YouTube.

One significant difference between Experiment 1 and this follow-up experiment was the judgment task instructions shown to participants. In this study, beneath the video, participants were shown the following prompt:

> Please rate how natural you find Eliza's response in the video, given the context paragraph:

Participants again had to press either the *completely natural* button or the *not completely natural* button to move on to the next trial. The addition of the words *. . . given the context paragraph* to the prompt was intended to signal to participants that they should rate the video utterance in the context of the written paragraph (our concern from Experiment 1).

As in Experiment 1, each participant was presented with three "attention check" trials randomly interspersed between the experimental trials. As before, these attention checks were used to filter out responses from participants who were not paying sufficient attention to the videos.

Stimuli

The stimuli used for this follow-up study were all re-used from Experiment 1, except for the new context paragraphs of the "infelicitous" flavor (see below).

As an example, we return to the "jewelry scenario" in (2.28) (compare to example (2.17) from Experiment 1). In this new experiment, the participants saw one of *three* possible context variants (i, ii, or iii) as a written paragraph at the top of the page. Below this they saw the embedded video, which either contained a co-speech gesture (2.28a) or did not (2.28b). Finally, below the video were the judgment task instructions and the *completely natural* and *not completely natural* buttons. Infelicity was designed not to be egregious but frequently involved, e.g., a switch in names, as in (2.28iii) below. All of the stimuli used in Experiment 2 can be found in Appendix A.3.

(2.28) Jewelry scenario, Experiment 2 (Scenario 9, Appendix A.3)

(i) *Neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money. Eliza agrees and says:

(ii) *Positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money based on her new pair of earrings. Eliza agrees and says:

(iii) *Infelicitous context*: Eliza and Alicia are gossiping about their coworker Nina, and Alicia says that she thinks Nina has a lot of money. Eliza agrees and says:

a. Alicia was wearing real diamond [jewelry]_EARRING at work this morning.

b. Alicia was wearing real diamond earrings at work this morning.

Note that in this study, the experimental condition SpeechCue manipulated in Experiment 1 was eliminated since we only wanted to focus on the relationship between the context paragraphs and the rating of the video utterance. For the [GestureCue=yes] condition, we chose to re-use the "gesture-only" videos from Experiment 1, i.e., those that didn't have a supporting speech cue, as in (2.28a). By contrast, for the [GestureCue=no] condition, we picked the "speech cue-only" videos that did contain a speech cue, as in (2.28b).[3] This ensures that, for a given scenario, each member of the pair of videos (e.g., (2.28a,b)) conveys (roughly) the same overall semantic content, without duplication (again with the caveat that the two are not strictly equivalent, given that the gesture is iconic and interpreted in an analog way).

Design

Exactly as in Experiment 1, each participant saw 19 trials total: 16 experimental trials and 3 attention check trials. There were six distinct trial types, depending on (i) whether or not there was a co-speech gesture in the video (the GestureCue condition); and (ii) whether the

[3]The former condition corresponds to Experiment 1 trial types 5 and 6, and the latter to trial types 3 and 4; see Table 2.1.

video was presented with a "neutral" context paragraph, a "positive" context paragraph, or an "infelicitous" context paragraph (the CONTEXT condition) (Table 2.4). We discuss the two experimental conditions in more detail below.

Table 2.4: *Experiment 2 trial types*

Trial type	GESTURECUE	CONTEXT
1	no	infelicitous
2	no	neutral
3	no	positive
4	yes	infelicitous
5	yes	neutral
6	yes	positive

At the start of each survey, the participant was randomly assigned to either the [GESTURECUE=yes] group or the [GESTURECUE=no] group. The sixteen scenarios were then presented in a randomized order (with the three attention checks randomly interspersed), and for each scenario the participant was randomly assigned one of the three CONTEXT condition values. Each participant saw a trial for every scenario.

The GESTURECUE condition. As in Experiment 1, the GESTURECUE condition encodes the presence ("yes") or absence ("no") of a co-speech gesture in the video utterance. This condition was manipulated between subjects, for the same reasons discussed in Section 2.2.1 above for Experiment 1; thus each participant saw either only videos with co-speech gestures or only videos without co-speech gestures.

The CONTEXT condition. The CONTEXT condition was manipulated within subjects and encodes whether the written paragraph shown to participants before the video either (i) neither entails/implies the content of the co-speech gesture, nor entails/implies its negation (a *neutral* context), (ii) entails/implies the content of the gesture (a *positive* context), or (iii) is written so that the video utterance is a pragmatically infelicitous continuation of the

context paragraph (an *infelicitous* context). Example (2.29) below shows the three context variants seen by participants for Scenario 5; minimal differences between contexts have been boldfaced for ease of comparison.

(2.29) Scenario 5 (Appendix A.3)

a. *Neutral context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there**, comes back to Jamie, and says:

b. *Positive context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there on her computer**, comes back to Jamie, and says:

c. *Infelicitous context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza **doesn't see her there**, comes back to Jamie, and says:

d. *Sample target utterance*: I saw Julia over in the library [writing an essay]_TYPE — it looks like she's a little preoccupied right now.

For this scenario, participants assigned [CONTEXT=neutral] were shown the context paragraph in (2.29a), participants assigned [CONTEXT=positive] were shown (2.29b), and participants assigned [CONTEXT=infelicitous] were shown (2.29c). As seen in the sample target utterance (2.29d), the co-speech gesture for this scenario is TYPE, which conveys that the essay-writing event was done by typing on a computer keyboard (as opposed to, e.g., with a paper and pencil). The "neutral" context (a) does not specify anything about Jamie's actions other than her location in the library; hence it entails/implies neither TYPE nor ¬TYPE, meeting the criteria of being an *m*-neutral context. By contrast, the "positive" context (b) contains the information that Julia is on her computer; this naturally implies that if she is writing an essay, she is doing so on her computer, so TYPE is implied and this is an *m*-positive context. Finally, the "infelicitous" context (c) is logically incompatible with the target utterance because in the context, Eliza doesn't see Julia in the library, but then in the

video utterance she asserts that she saw Julia in the library. We would expect participants shown this infelicitous context to rate the video as *not completely natural*, with or without a co-speech gesture.

to create the infelicitous contexts was to swap the names of Eliza's interlocutor and the person under discussion in the video; this can be seen, for instance, in Scenarios 8, 9, and 10 in Appendix A.3. This was in an effort to keep the infelicitous context paragraphs as minimally different from the neutral contexts as possible, just as we strove to make neutral and positive contexts minimally different.

2.3.2 Results

Out of 90 participants, 2 participants' responses were excluded due to those participants failing at least one of the three attention checks. The results discussed below are for responses from the remaining 88 participants.

Table 2.5 reports the mean acceptance rates, standard deviations, and standard errors across trial types, and Figure 2.6 shows mean acceptance rates for each trial type with standard error bars.

Table 2.5: *Experiment 2 trial types with conditions (*GestureCue *and* Context*) and mean acceptance rates*

Trial type	GestCue	Context	**Mean**	**SD**	**SE**	**N**
1	no	infel	0.403	0.492	0.032	233
2	no	neut	0.782	0.414	0.027	238
3	no	pos	0.815	0.389	0.025	233
4	yes	infel	0.449	0.498	0.033	234
5	yes	neut	0.696	0.461	0.030	230
6	yes	pos	0.725	0.447	0.029	240

Trends across the experimental conditions GestureCue and Context were as follows. On average, trials with gestures (trial types 4–6) were rejected slightly more often than trials without gestures (trial types 1–3) (with gesture: M=0.62; without gesture: M=0.67). Trials shown with an infelicitous context (1 and 4) were rejected at a much higher rate than trials

shown with either a neutral context (2 and 5) or a positive context (3 and 6) (infelicitous: M=0.43; neutral: M=0.74; positive: M=0.77). Note that neutral and positive context trials were rejected at roughly the same rate.

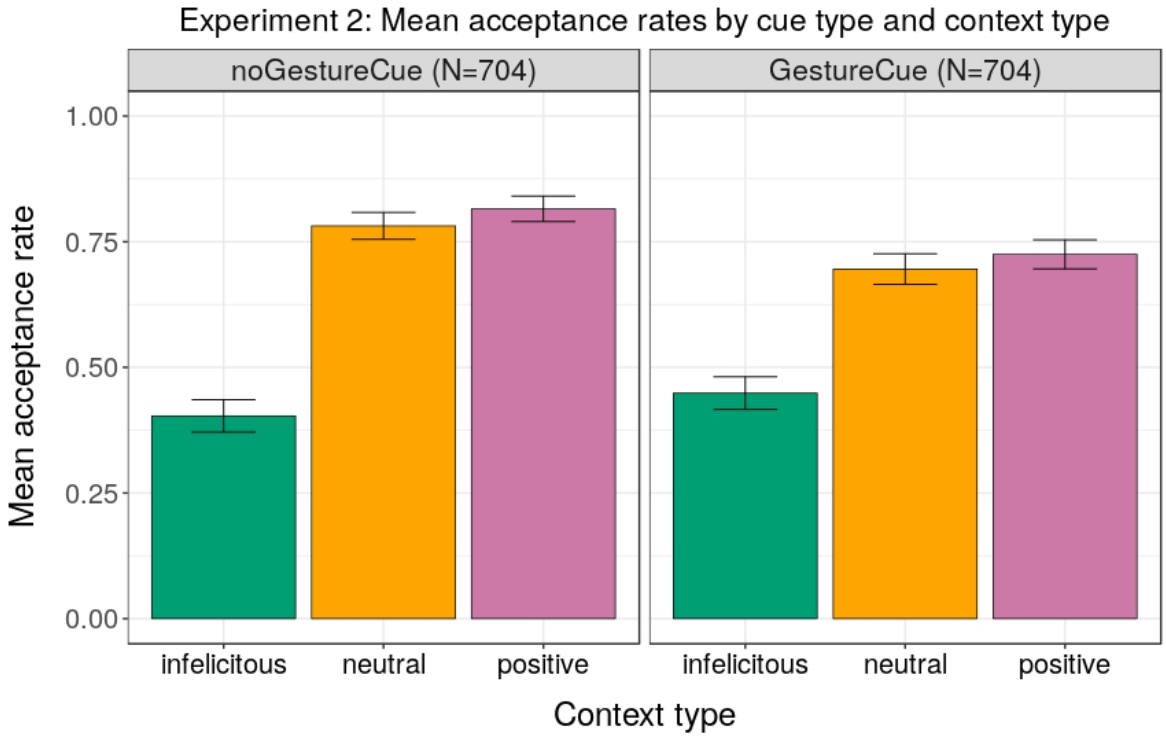


Figure 2.6: *Experiment 2 mean acceptance rates, by experimental conditions* GestureCue *and* Context.

Table 2.6 gives a breakdown of ratings by scenario. Means for scenarios across trials with infelicitous contexts ranged from 0.23 (Scenario 12) to 0.66 (Scenario 2), with SDs ranging from 0.43 to 0.51. Means for scenarios across trials with either neutral or positive contexts, by contrast, ranged from 0.67 (Scenario 16) to 0.93 (Scenario 5), with SDs ranging from 0.25 to 0.49.

Analyses of participants' judgment responses were conducted using the R programming language to build generalized linear mixed effects models using the function *glmer*. In the model with the most data variation coverage (determined by ANOVA testing), the independent factors in the model were GestureCue (no/yes) and Context (neutral/positive/infelicitous); Response (1/0) was the dependent variable, and scenario number

and participant ID were coded as random effects.

Table 2.6: *Experiment 2 acceptance rates by scenario, with infelicitous trials separated out from combined neutral and positive trials*

Scenario	**M** (infel)	**M** (neut/pos)	**SD** (infel)	**SD** (neut/pos)
1	.38	.71	.49	.46
2	.66	.69	.48	.46
3	.53	.79	.51	.41
4	.32	.68	.48	.47
5	.55	.93	.51	.25
6	.41	.80	.50	.41
7	.48	.74	.51	.44
8	.33	.79	.48	.41
9	.62	.80	.49	.41
10	.33	.76	.48	.43
11	.52	.85	.51	.36
12	.23	.79	.43	.41
13	.34	.61	.48	.49
14	.31	.78	.47	.42
15	.38	.68	.50	.47
16	.40	.67	.50	.47

Results of the model indicate that there was a significant main effect of CONTEXT: With "neutral" as the reference value, [CONTEXT=infel] was highly significant (β=$-$1.796, z=$-$7.887, p<0.001), but [CONTEXT=pos] was not significant (p>0.1). In a minimally different model with "infelicitous" as the CONTEXT reference value instead, both [CONTEXT=neut] and [CONTEXT=pos] were highly significant with (β=1.796, z=7.887, p<0.001) and (β=2.034, z=8.590, p<0.001), respectively. In other words, there was no significant difference between the neutral and positive context conditions, but an infelicitous context significantly decreased the acceptance rate compared to either neutral or positive contexts.

Unlike the results from Experiment 1, GESTURECUE was not a significant main effect (p>0.1). We discuss possible reasons for this difference in Section 2.3.3. There were no significant interactions between factors.

We also analyzed the subset of participant responses for which the context presented was either "neutral" or "positive", thereby excluding the data from "infelicitous" CONTEXT trials. Again we used R and the the function *glmer* to build a generalized linear mixed effects model, with independent factors GESTURECUE and CONTEXT, dependent variable RESPONSE, and random effects of scenario number and participant ID. Results of the model indicate no significant main effect of either GESTURECUE or CONTEXT, nor a significant interaction between the two factors ($p>0.1$). This supports the conclusion from the full model reported above that there was no significant difference in acceptance rates between trials with neutral contexts and those with positive contexts.

2.3.3 Discussion

Since the trials with infelicitous contexts were rated significantly lower than other trials, this follow-up study is encouraging, suggesting that participants were reading the context paragraphs and, just as importantly, taking them into account when rating the video utterances. Similarly to Experiment 1, in this follow-up study there was no significant difference in ratings between trials with neutral contexts and trials with positive contexts. Taken together, we can conclude more confidently that co-speech gestures are not sensitive to entailment/implication by the discourse context, and hence are [−SCF] in the Tonhauser et al. (2013) terminology.

The factor GESTURECUE was not a significant main effect in Experiment 2. We suspect that this is likely due to the overwhelming effect of context in Experiment 2 with the introduction of infelicitous trials, minimizing the variability available among felicitous trials. To investigate this further, we compared the means for the same trials across both experiments (the two trial types [SPEECHCUE=yes, GESTURECUE=no] and [SPEECHCUE=no, GESTURECUE=yes]), and found that the range of means for Experiment 1 is 0.60–0.77 (mean SD 0.46), compared to the more compressed range of means in Experiment 2 of 0.70–0.82 (mean SD 0.43). In other words, the same trials in Experiment 2 had higher and more compressed ratings for positive and neutral contexts than those trials did in Experiment 1,

which we attribute to the addition of infelicitous contexts in Experiment 2. We therefore suspect that this compression of contexts of interest (felicitous trials) may have contributed to the lack of effect for GestureCue in Experiment 2 compared to Experiment 1.

2.3.1, we noted that many of the infelicitous contexts for this experiment were created by simply swapping the names of the addressee and the person being talked about in the neutral and positive contexts. One might wonder if participants were paying sufficient attention to these "name-swap"-type infelicitous contexts to detect the cause of infelicity, and if they were not, whether this artificially raised the ratings of those trials. To test this, we coded the scenarios for Experiment 2 according to whether their corresponding infelicitous contexts were made bad by swapping names ("name-swap") or by some other mechanism ("other"), such as introducing a contradiction between the context paragraph and the video utterance (e.g., the infelicitous context of Scenario 5, Appendix A.3). We constructed a generalized linear mixed effects model on the data from just those trials where the participant saw an infelicitous context (condition [Context=infel]), with Response as the dependent variable, type of infelicity (name-swap/other) and GestureCue as the independent variables, and subject ID and scenario as random effects. There was no main effect of type of infelicity on acceptance rates, nor a significant interaction effect between type of infelicity and GestureCue ($p>0.1$ in both cases). We interpret these results as indicating that participants did not behave differently when presented with a "name-swap" infelicitous context than they did with other types of infelicitous contexts, and so the cause of infelicity did not have an overall effect on felicity judgments.

check whether the belief state of the addressee in the "dialogue" between Eliza and her interlocutor described in the context paragraphs influenced participants' judgments. We conducted another follow-up analysis on the Experiment 2 data, using the same classification of scenarios (yes/no) according to whether the addressee of the dialogue knows or could infer the content of the proposition in question based on the information provided by the "positive" context. We constructed two separate generalized linear mixed effect models, one on the "yes" subset of the data (where the addressee knows/can infer the semantic content

of the gesture based on the positive context), and the other on the remaining "no" subset of the data, with RESPONSE as the dependent variable, GESTURECUE and CONTEXT as the independent variables, and subject ID and scenario number coded as random effects. Just as in the Experiment 1 analysis, neither model showed a significant effect of CONTEXT on RESPONSE ($p>0.1$ in both cases). From this we once again conclude that the belief state of the addressee is not a significant factor in acceptability judgments of utterances with co-speech gestures based on prior discourse contexts.

Overall, Experiment 2 lends support to our experimental methodology, providing encouraging evidence that participants judged the overall effect of context plus video when providing their ratings, although the addition of the infelicitous contexts meant that more subtle differences between generally felicitous trials in some cases disappeared. Together, we suggest that Experiments 1 and 2 provide a more clear overall picture of the empirical landscape for co-speech gesture pragmatics.

2.4 Conclusions

2.4.1 Directions for future research

By directly controlling whether the content of co-speech gestures is duplicated in the preceding context and/or the same utterance, the two experiments we report in this paper show how the acceptability of co-speech gestures may be affected by linguistic context. We hope that this can be a first step for future studies, and, in some cases, provide a foundation for interpreting results found in the few already existing studies on the semantics/pragmatics of co-speech gesture.

Through the notion of triviality, we connect to the existing literature on the way that gestures contribute content, which has been the focus of previous discussion of co-speech gestures as either "supplemental" (Ebert and Ebert 2014, 2016) or as "cosuppositional" (Schlenker 2018a). Under a supplemental analysis, it is quite surprising that speech cues aid the acceptability of gestures, given that supplements are typically less acceptable if they are

trivial. On the cosuppositional side, we have shown that co-speech gestures are not "hard" presupposition triggers since they need not be entailed by preceding context; however, many presuppositions are known to be easily accommodated, which could account for their acceptable nature in our study. Under this view, it remains a question why a positive preceding context doesn't also provide the same kind of content support (and our second experiment suggests that it is not due to participants' ignoring written context). Altogether, given the *improvement* of gestures with matching speech cues, our data are much harder to reconcile with the supplemental theory.

It is possible to find examples of parentheticals where they are more trivial. These examples tend to be emphatic or have other specific uses. Consider the following example:

(2.30) Sarah is the valedictorian. Sarah, who did well in her math classes, is going to give my son math lessons.

of the relative clause, that Sarah did well in her math classes, follows from the initial information that Sarah is the valedictorian, so the non-restrictive relative clause is trivial, and grammatical (if a bit marked). It may therefore not be possible to completely rule out the supplemental theory. Either way, this chapter presents an important result about the felicity of co-speech gestures.

One further remaining question raised by these studies has to do with the choice of gesture targets in the development of the stimuli, discussed above in the Methods section (2.2.1) of Experiment 1. Recall that we chose semantically underspecified predicates so that the corresponding co-speech gestures would non-trivially contribute to the truth conditions of the utterances. Some underspecified predicates seem to have a preferred "default" manner or adjectival property that is assumed in the absence of further modification (either by speech cue or gesture cue). Ebert and Ebert (2016) look at some examples of these default interpretations for NPs in German, referring to the phenomena as the "typicality of a gesture for an NP concept". A particularly interesting example they describe involves the NP *Fenster* 'window' and the two different shape properties of either square or circular; they assume the square option to be the more typically expected shape for windows. For an example

from our experimental stimuli, consider the VP *writing an essay*:

(2.31) Scenario 5 (Appendix A.1)

> I saw Julia over in the library [writing an essay]_TYPE – it looks like she's a little preoccupied right now.

In this scenario we chose the gesture TYPE, indicating that the writing was done on a computer as opposed to by hand with pen and paper. In the current decade, it would be a very natural inference to make that *writing an essay* means writing (typing) on a computer; anyone visiting a college campus these days will see ample evidence of this preferred mode of writing essays. Returning to the experiment, in Scenario 5 trials such as (2.31) we chose the gesture option with the "default" or more expected interpretation of TYPE instead of a more uncommon co-speech gesture indicating writing with a pen and paper, call it WRITE-WITH-PEN. An important question is whether this choice of gesture has a significant effect on utterance ratings and whether it interacts in a significant way with the presence of a speech cue in an utterance. The supplement analysis of co-speech gesture, although generally not supported by our findings, would predict that less trivial, less expected gestures (given speech cues) would be improved, and this does match our intuitions about examples like (2.2) from Goldin-Meadow and Brentari (2017), repeated below as (2.32), which we find especially natural/felicitous:

(2.32) I [ran up the stairs]_SPIRAL-UP.

Given that the prototypical staircase is not a spiral staircase, the gesture content is unexpected and hence informative. An experiment could be designed with gestures varying along the dimension of unexpectedness, given context and speech cue.

2.4.2 Summary

The primary goal of these experiments was to diagnose the behavior of co-speech gestures in contexts that do and do not entail/imply their semantic contents, thereby gaining a better understanding of when and how co-speech gestures can be felicitously used in conversation.

We implemented this question formally using the strong contextual felicity diagnostic proposed in Tonhauser et al. (2013), and by varying whether similar information was contained in accompanying speech cues. The results of our experiments show that co-speech gestures do not need to be entailed/implied by their preceding discourse context; hence co-speech gestures are [−SCF] and cannot be considered "hard" (i.e., unaccommodatable) presupposition triggers. We also saw that *speech cues*, or speech expressions that (approximately) duplicate the semantic content of a co-speech gesture, had a significant interaction with the presence or absence of a gesture in the trials. We take this to be an indication that there are other restrictions on the felicitous use of co-speech gestures that we do not yet know about and that involve when and how the gestures can contribute "extra" semantic content to the utterance meaning. We speculate that this may vary depending on the type of content conveyed in the gestures, for example whether they represent size-and-shape or manner information, and whether they modify nouns or verbs, among other potentially relevant dimensions. We hope that the data on co-speech gesture felicity judgments gathered through these experiments will pave the way for future research on co-speech gestures that addresses these larger theoretical questions in interesting and fruitful ways.

Chapter 3

Description and depiction in a reference game

3.1 Introduction

Most studies of natural language pragmatics focus on felicitous use of descriptive modifiers (*red, hungry*), but we know little about the pragmatics of depiction, despite growing acknowledgement of the role of depiction in the compositional semantics of both signed and spoken languages (e.g. Schlenker et al. 2013, Davidson 2015, Henderson 2016, Kuhn and Aristodemo 2017). In this thesis we are investigating the pragmatics of depictive co-speech gestures, and we continue to focus in the current chapter on the role that triviality/informativity plays in felicity judgments, this time in the context of a reference game.

In the previous chapter, we manipulated many different factors to influence the felicity of co-speech gestures, including entailment by the discourse context and presence or absence of a speech cue. The result was highly variable ratings across items, with a strong bias toward gestures accompanied by a speech cue, that is, semantically trivial gestures. A possible confounding factor was that in all of our examples gestures could easily be replaced by their speech cues — there was always an easily accessible linguistic alternative to the information conveyed by the gesture. For instance, in (3.1), the target *work out* plus the

gesture FREE-WEIGHTS could be replaced by the phrase *weight lift*.

(3.1) a. Melissa goes to the gym three times a week to [work out]_FREE-WEIGHTS.

b. Melissa goes to the gym three times a week to **weight lift**.

In this experiment we try to address this issue by restricting our gestures to describing things not easily described linguistically, specifically, wild-looking hairstyles of cartoon trolls. In this way we can separate the descriptive cues (here, color words) from the depictive gesture cues, and manipulate each independently. The new stimuli are of the form:

(3.2) Give me the troll with the wild blue [hair]_HAIRSTYLE.

where the hairstyle is not easy to describe in words and is depicted via gesture.

3.2 Experimental questions

As mentioned above, gestures with close equivalents in speech (e.g. *down* in speech and DOWN in gesture) have been the primary focus of previous studies (e.g. Tieu et al. (2017) on gesture projection), but in this project we dissociated *triviality* (duplication of content elsewhere in the utterance) from *informativity* (here, relevance of the content to the reference resolution task) using modifiers that are difficult in English to match across modes: our verbal modifiers were descriptive color words (e.g. blue), while our gestural modifiers were depictive co-speech gestures of odd shapes (see pictures below). What we varied, through permutations of both distractor items and prompts, was the status of the informativity of the modifier, namely whether with a given utterance and stimuli set the target was identified by the modifier in a way that was:

- **Critically informative**: the stimuli pair (e.g. two trolls) that must be differentiated from each other differed in only the dimension (color vs. shape) matching the modifier
- **Informative**: the stimuli pair differed in both dimensions (color and shape)
- **Uninformative**: the stimuli pair differed only in the dimension that matched the unused modifier

We asked whether participants' accuracy and felicity judgments would vary according to the status of informativity. In addition, we asked whether these dimensions were judged differently by participants in the verbal (descriptive) or gestural (depictive) modes.

3.3 Experiment: Gestures in a reference game

3.3.1 Methods

Participants

Participants were 316 adults recruited through Amazon Mechanical Turk, restricted to the United States region. All participants self-identified as native speakers of English. Participants were compensated monetarily for their participation in the questionnaire via Amazon payments of $1. Experimental protocols were approved by the Harvard University Institutional Review Board under approval number IRB17-0250.

Procedure

The experiment was conducted in two stages with two non-overlapping sets of participants: sub-experiment 1 (N=118) and sub-experiment 2 (N=198). In both stages Amazon Mechanical Turk workers were directed to an online questionnaire designed using the Qualtrics Survey Software Platform. The questionnaire took approximately 5–10 minutes to complete, and participants completed the questionnaire on their own time and on a device of their choosing. Participants were instructed at the start to make sure to use a device with a large enough screen to play videos and with working speakers/headphones. Since some of the survey tasks required color discrimination, participants were also discouraged from taking the survey if they had any reason to suspect they were color-blind.

For both sub-experiments, each participant viewed 16 trials, each consisting of a short video clip and pair of computer-drawn images of "trolls"[1]. During each trial, the participant

[1]Many thanks to Dorothy Ahn for designing and creating the troll images.

was asked to: (i) view a short video description of the "target" troll; (ii) pick the target troll based on the description; and (iii) rate the naturalness of the request, using a continuous slider from 0 ("totally awkward") to 1 ("totally natural") (see Figure 3.1). The only major difference between the two sub-experiments was in the properties of the pairs of trolls used as stimuli; this will be explained in more detail below in Section 3.3.1.

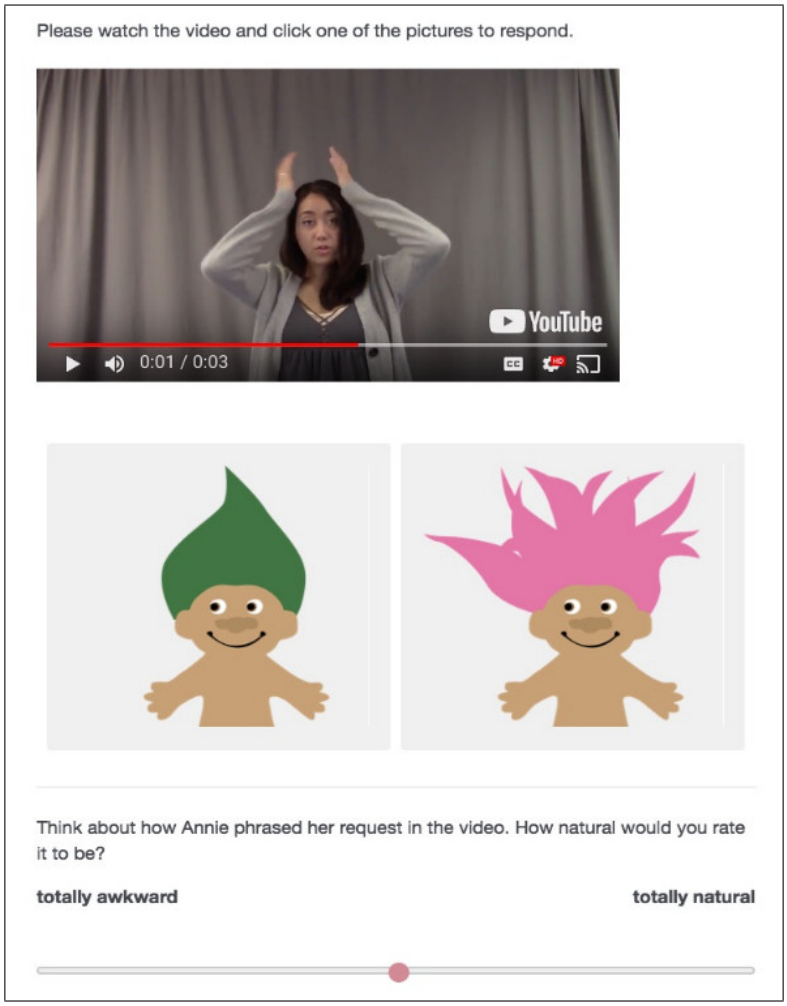


Figure 3.1: *Screenshot of a trial from sub-experiment 1.*

Each video description featured the same speaker, a female native English speaker in her early twenties who was identified as "Annie" in the survey instructions. The embedded videos were hosted through YouTube and participants could replay the video if they wished, although no mention of the possibility of replaying videos appeared in the instructions. We

did not collect data on the frequency of participants replaying videos.

The concept of *linguistic naturalness* was demonstrated to the participants through guided practice trials, rather than explained in words. The survey instructions had participants complete two such practice trials (see Figure B.1 in Appendix B for screenshots). The first practice trial had a written instruction, mimicking what the participant would hear in a video description, that clearly and uniquely identified one of the two images (instruction: *"Give me the red square"*; left image: red square; right image: blue circle), while the second practice trial had a description that did not uniquely specify which of the two images to pick (instruction: *"Give me the triangle"*; left image: yellow triangle; right image: green triangle). For the naturalness ratings, participants were instructed to rate the former practice trial as more natural/less awkward (i.e. to drag the slider to the right) and the latter as less natural/more awkward (i.e. to drag the slider to the left). On the final screen of instructions, participants were reminded that "... there may be many reasons why Annie's request might seem awkward", in an effort to encourage them to consider a wide range of sources of infelicity when judging the felicity of the video utterances.

The guided practice trials were additionally used to screen for participants who were not paying sufficient attention to the instructions or tasks. If participants rated the felicity of the first practice trial, intended to be felicitous, below 0.5, or if they rated the second practice trial, intended to be infelicitous, above 0.5, then their responses were discarded from data analysis.

At the end of the survey there was a short debriefing section, where participants were asked if they had any trouble with the video or audio, trouble telling the troll hair colors apart, or trouble understanding the tasks. They were also given the option to submit feedback or comments on the questionnaire via an open-ended text box.

Design and stimuli

In both sub-experiments, each participant first saw two guided practice trials, as described in the previous section. These were followed by 16 randomized experimental trials: four

trials for each of four types of target description (see Table 3.1). Trial type was defined by the presence of a gestural modifier (±GestureCue), as well as the presence of a linguistic modifier in the speech stream (±LinguisticCue).

Table 3.1: *Trial Types in Experiments 1 and 2*

Trial Type	GestureCue	LinguisticCue
1	no	no
2	yes	no
3	no	yes
4	yes	yes

We include four example target descriptions below; descriptions (i)–(iv) correspond to trial types 1–4. When included, gestural modifiers always described the troll's hair shape (Figure 3.2), while linguistic modifiers provided the troll's hair color. The adjective *wild* was included in all trials to reduce both phonological and pragmatic concerns in cases without a linguistic (verbal) modifier (i, ii below).

(i) Give me the troll with the wild hair. [-GestureCue, -LinguisticCue]

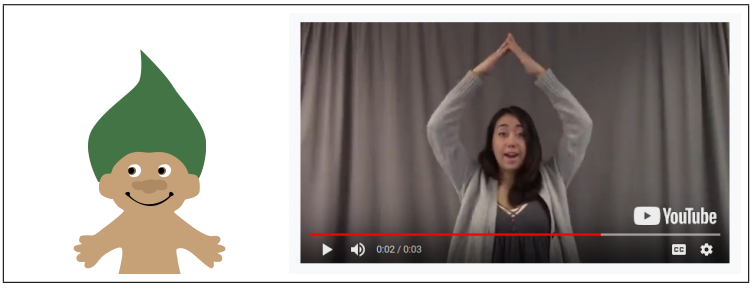
(ii) Give me the troll with the wild [hair]_STYLE-1. [+GestureCue, -LinguisticCue]

(iii) Give me the troll with the wild blue hair. [-GestureCue, +LinguisticCue]

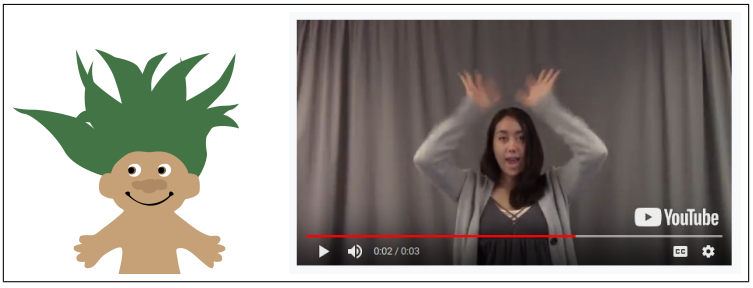
(iv) Give me the troll with the wild blue [hair]_STYLE-1. [+GestureCue, +LinguisticCue]

Descriptions of the target had order randomized by participant, varying color and shape, with the shape gestures aligned with *hair*. Trolls varied by color (4 total) and hairstyle (3 total), which were counterbalanced across four lists across the four cue conditions ((i)-(iv) above). See Appendix B for the complete set of troll images used in the experiment.

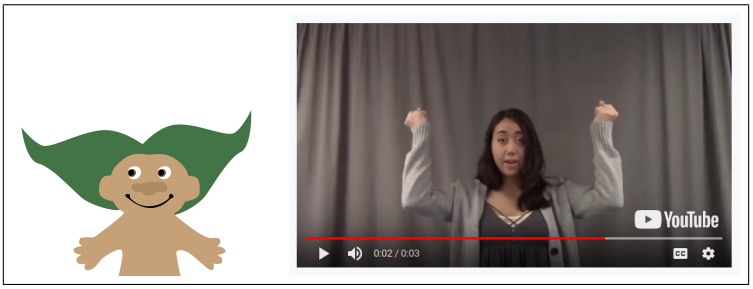
Target trolls were always paired with one (counterbalanced) distractor. In sub-experiment 1, participants (N=118) saw distractor trolls that differed from the target troll on both color and shape, while in sub-experiment 2, participants (N=198) saw distractor trolls that differed either on color, or on shape, but not both. See Figure 3.3. We go through the three target-distractor configurations in more detail below.



(a) *Troll with hairstyle 1 and gesture STYLE-1*

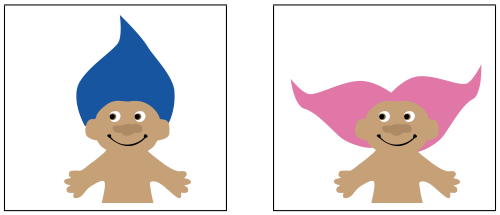


(b) *Troll with hairstyle 2 and gesture STYLE-2*

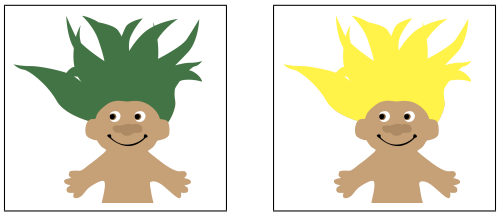


(c) *Troll with hairstyle 3 and gesture STYLE-3*

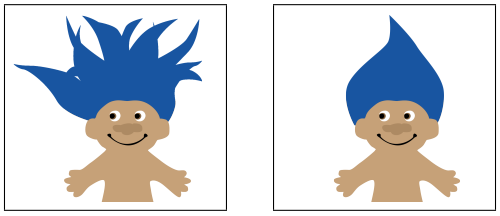
Figure 3.2: *Images and corresponding gesture screencaps for the three troll hairstyles.*



(a) *Sub-experiment 1: different color, different style*



(b) *Sub-experiment 2: different color, same style*



(c) *Sub-experiment 2: same color, different style*

Figure 3.3: *Sub-experiment target/distractor paradigms. (a) In sub-experiment 1, the trolls differ in both color and hairstyle. (b) & (c) In sub-experiment 2, the trolls differ in either color, or hairstyle, but not both.*

Case 1: Trolls differ in color and style. When the two trolls differ in both color and style (Figure 3.3a), both the linguistic cue (color word) and the gesture (hairstyle) are **informative**, but neither is *critically* so. Either just the lingustic cue on its own or just the gesture cue on its own provides enough information for the participant to accurately choose the correct troll.

Case 2: Trolls differ in color only. When the trolls differ in color only (Figure 3.3b), the linguistic cue (color word) is **critically informative**, and the gesture cue is **uninformative**. The participants need the color cue in order to pick the correct troll, and hence we expect them to perform at chance when there is no color cue, in trials (i) and (ii).

Case 3: Trolls differ in style only. In the case where the trolls differ in hairstyle only (Figure 3.3c), the linguistic cue (color word) is **uninformative**. We thus expect participants to perform at chance in choosing the correct troll in trial type (iii). The gesture cue (hairstyle), by contrast, is **critically informative** in this case; it is necessary in order for the participant to discern which troll is being asked for. Because of this, in theory this condition should provide the best chance for gesture accuracy and felicity ratings.

3.3.2 Results

16 participants were excluded from analysis of sub-experiment 1 and 18 from sub-experiment 2, based on the criteria of them failing the guided practice trials at the start of the survey and/or indicating a problem differentiating the troll hair colors in the debrief at the end.

Data were analyzed using linear mixed effects models in R (lmer4 package), with participant ID and trial coded as random effects. Independent variables included GestureCue (presence/absence of a gestural modifier), LinguisticCue (presence/absence of a verbal modifier), and Difference (what distinguishes the target and distractor trolls: color/style/both). Felicity rating and accuracy were analyzed as dependent variables.

The first statistic we'll look at is how accurate participants were at choosing the target troll.

Binary troll choice accuracy results

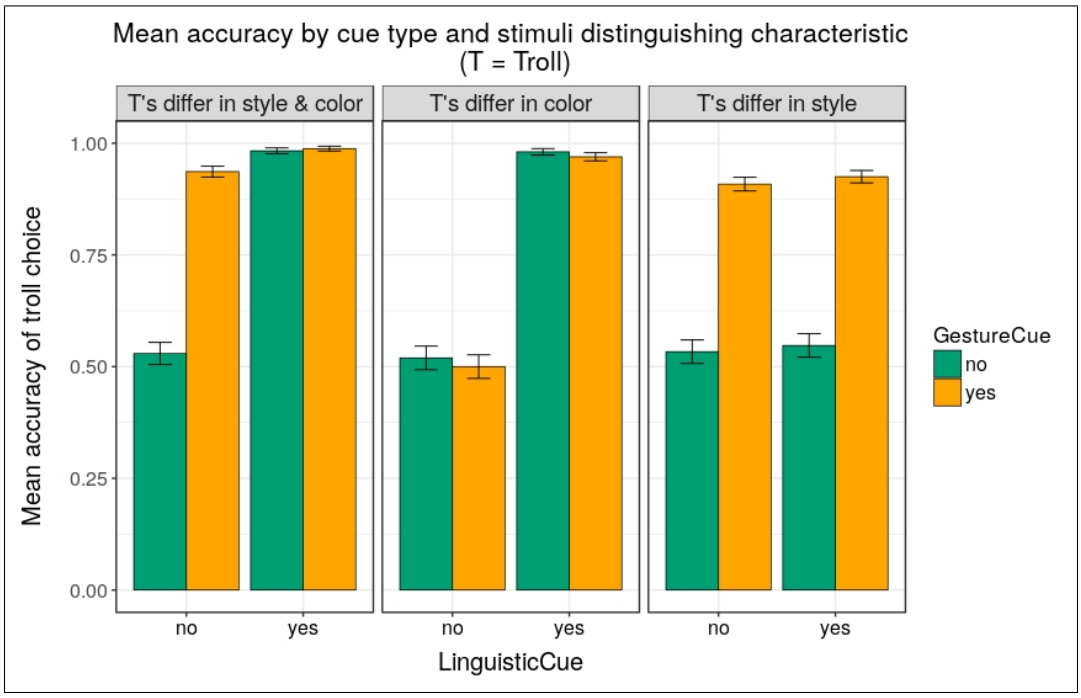


Figure 3.4: *Accuracy results comparing the absence (green bars) and presence (orange bars) of gesture cue, based on target/distractor relationship, and presence of linguistic cue.*

Overall, participants were quite accurate when given sufficient information to choose the correct troll. There were some significant differences, however, between how they utilized the gestures versus the linguistic cues (color words).

When either the linguistic modifier or the gestural modifier were critically informative, participants were significantly more accurate at choosing the target troll ($p < 0.05$) in the condition where the linguistic modifier was critically informative (M = 0.98, SD = 0.14) than in the condition where the gestural modifier was critically informative (M = 0.91, SD = 0.29).

Across both studies, in trials where participants were given sufficient cues to choose the correct troll (i.e. correct not based on what participant takes to be wild, but on shape and/or color), the target troll hairstyle didn't affect acceptability judgments ($p > 0.05$). However, there was a significant effect of target troll hairstyle on participants' accuracy in choosing the target troll ($p < 0.05$), indicating that one of the hairstyles may have been perceived as

more "wild" than the others and participants defaulted to that style when given insufficient cues.

Unsurprisingly, participants performed at chance when they did not have sufficient information to choose the correct troll — trial type 1 when trolls differed in style and color; trial types 1 and 2 when trolls differed in color only; and trial types 1 and 3 when trolls differed in style only.

Felicity Ratings results

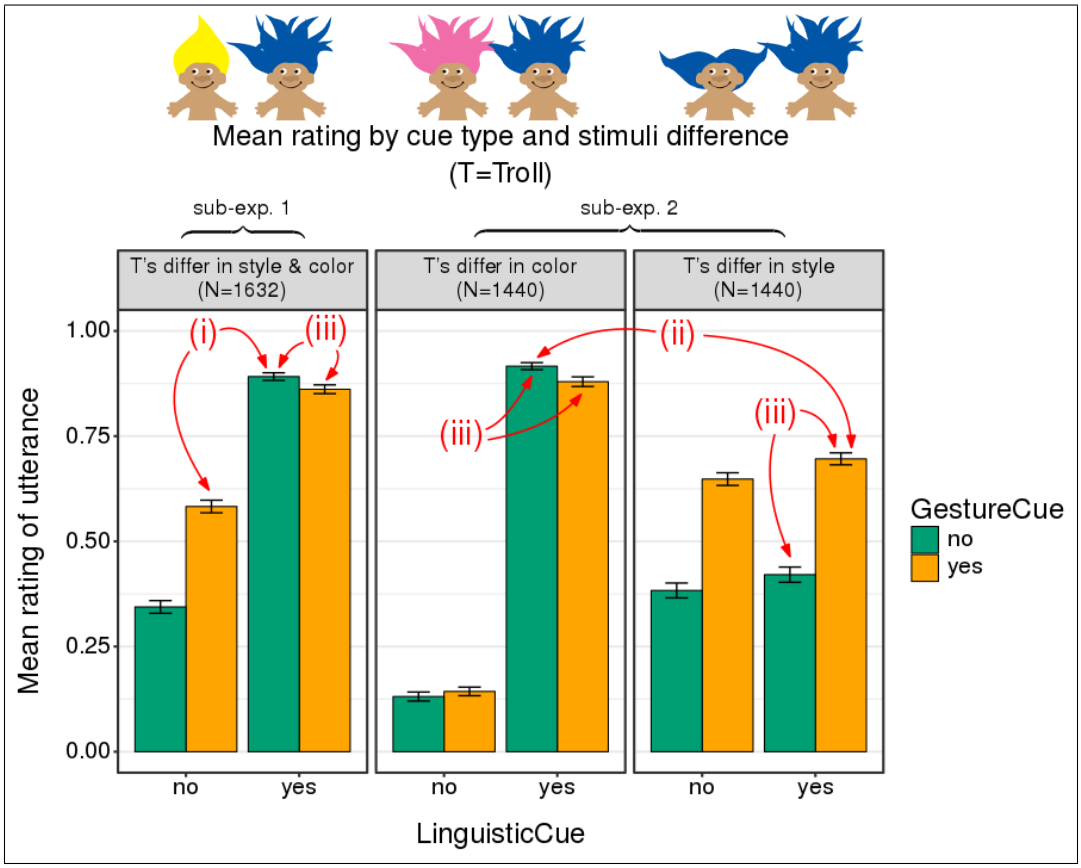


Figure 3.5: *Felicity results comparing the absence (green bars) and presence (orange bars) of gesture cue, based on target/distractor relationship, and presence of linguistic cue. (i) Comparing the gesture-modifier-only condition with the verbal-modifier-only condition when both were informative. (ii) Comparing the the critically informative color cue condition to the critically informative gesture cue condition. (iii) Looking at what happens when adding a gesture to an utterance with a linguistic cue. (See Section 3.3.2 for more information.)*

Results of the felicity ratings are reported in the graph in Figure 3.5. For ease of

understanding we summarize the main results below and annotate the relevant portions of the graph with roman numerals (i) through (iii).

(i) The gesture-modifier-only condition (trial type 2) was rated significantly lower than the linguistic-modifier-only condition (trial type 3) when both were informative for reference disambiguation ($p < 0.05$). In other words, when the trolls differed in both style and color, participants rated the version with only the color word significantly higher than the version with only the hairstyle gesture.

(ii) When the linguistic modifier (color word) was critically informative, the highest mean acceptability rating was 0.92, compared to a much lower mean rating of 0.70 when the gestural modifier (hairstyle) was critically informative. In other words, participants rated the color-word-only trial type when the trolls only differed in color much higher than the gesture trial types when the trolls only differed in hairstyle.

(iii) Overall, adding a gesture to an utterance that contained a linguistic modifier lowered the mean acceptability rating ($p < 0.05$), one exception being in the condition where the gesture was critically informative (i.e. when the troll stimuli differed only in hairstyle).

Finally, when participants were given insufficient information to choose the correct troll, ratings were predictably quite low.

3.4 Discussion and conclusions

The study in the previous chapter examined the felicity of co-speech gestures depending on preceding and simultaneous context. Unfortunately there was a potential confounding factor, in that most of the gestures could be easily replaced by speech near-equivalents. The experimental design of the present study avoids this issue by only using gestures to describe characteristics of the stimuli that do not have a close speech equivalent. Furthermore, we tested the felicity of co-speech gestures in three different pragmatic conditions: when co-speech gestures were uninformative, informative, or critically informative for the purposes

of a reference game. Co-speech gestures were directly compared to color words (linguistic modifiers) so as to compare the behavior and function of depictive and descriptive modifiers in these varying pragmatic conditions.

Our findings suggest that both linguistic and gestural modifiers can be used for reference resolution tasks, but may be prioritized differently since linguistic modifiers were both rated as more natural and used more accurately by participants compared to gestural modifiers. The asymmetries we found between gesture and verbal modifiers lead to further questions about the information structural properties of co-speech gestures. The fact that adding gestures to utterances with linguistic cues decreases the acceptability, except when the gesture was critically informative, indicates that there seems to be some sort of principle at play that says not to gesture unless you need to for your interlocutor's sake. On the other hand, there are also speaker-driven pressures involved with gesture (Goldin-Meadow 2003). More research is needed to tease apart these effects.

A natural concern one might raise about the study is that perhaps participants are rating the *performance* of the gestures, rather than the gestures themselves. However this cannot entirely be the case here, because they are seeing the same performances in the videos across different pragmatic conditions, and yet ratings change. All four trial types were shown in all three pragmatic conditions (trolls differ in style and color; trolls differ in color only; trolls differ in style only) and the results varied based on those pragmatic conditions.

It is an open question as to how these results generalize to all descriptive content; this study was only restricted to color adjectives. However, it may be the case that color adjectives behave pragmatically differently than other descriptive modifiers, in that color seems to be a more intrinsic part of the representation (see, e.g., Sedivy 2003). However, these pragmatic effects have been found in cases where an item has an expected color (e.g., yellow bananas), which is not the case with the trolls.

Chapter 4

Negating depictive modifiers in sign and speech

4.1 Introduction

Thus far this dissertation has compared gesture with speech, looking at the role of context and various pragmatic factors such as informativity and triviality in the felicity of co-speech gestures. In this chapter, though, we shift our focus to comparing gesture in speech with gestural elements of sign languages called classifier predicates. Unlike co-speech gestures, the gestural component of classifier predicates is in the same modality as its co-occurring lexical component. The main thing we are interested in is the compositional properties of co-speech gestures, the extent to which they differ from spoken linguistic content (e.g. in being not-at-issue), and why they might differ.

As discussed in the introduction of this dissertation, depictive gestural modifiers (in particular, content-bearing co-speech gestures) are different from verbal modifiers (e.g. *big*, *yellow*) in that they typically convey backgrounded not-at-issue information (Ebert and Ebert 2014, 2016; Esipova 2019a). One piece of supporting evidence we saw is that the content of the co-speech gesture cannot be directly dissented with, as shown in (4.1).

(4.1) (Ebert and Ebert 2014)

Speaker A: I brought a [bottle of water]_BIG  in the yard.

Speaker B: #That's not true! You actually brought a small bottle.

Further support comes from the behavior of co-speech gestures under negation. We can consider the example of the co-speech gesture LIFT (indicating a lifting event) used without negation in (4.2) versus with negation in (4.3). If LIFT in (4.2a) were at-issue, then its meaning would be roughly equivalent to sentence (4.2b) containing the assertive modifier "like this", where "this" refers to the gesture (Tieu et al. 2017).

(4.2) (Tieu et al. 2017:3)

a. John [helped]_LIFT  his son.

b. John helped his son **like** [**this**]_LIFT  .

Under sentential negation, though, (4.2a) and (b) have different interpretations. The *like this* sentence (4.3b) triggers an implicature that John did help his son (Tieu et al. 2017). In contrast, the co-speech gesture sentence (4.3a) implies that John didn't help his son at all. The sentential negation in (a) seems to semantically target only the verbal content *help* and not the gestural content of LIFT.

(4.3) a. John **didn't** [help]_LIFT  his son.

b. John **didn't** help his son **like** [**this**]_LIFT  .

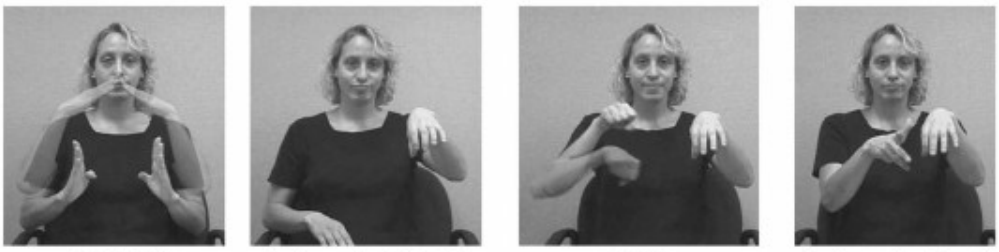
In the case of (4.3a), the co-speech gesture which indicates the manner of helping projects through the negation; the speaker remains committed to some kind of lifting being involved if the helping had happened. Thus the gesture LIFT in (4.2a) cannot be at-issue. Tieu et al. (2017, 2018) go on to show experimentally that the pointing co-speech gestures UP and DOWN similarly project through entailment-canceling operators such as conditionals, negation, modals, and certain quantifiers (including *none*).

This all raises the question: why are co-speech gestures typically not-at-issue, and why can't negation in particular target gestural content? One obvious difference between co-speech gesture and speech is their differing modalities; the gestures we have been discussing are conveyed via a visual-manual modality, while speech is conveyed via an auditory-oral modality. A related fact, discussed in depth in Schlenker (2018b), is that co-speech gestures share a time-slot with auditory content. Perhaps it is this modality mismatch that prevents linguistic negation in spoken languages from targeting the content of co-speech gestures. Another difference between depictive co-speech gestures and verbal modifiers is that the gestures are interpreted via depictive/iconic means. It may be that analog iconic content is not integrated into the grammar in a way that allows it to interact with other operators like negation. To pull apart these factors, we will compare how negation interacts with gradient depictive content in a language where the two *share the same modality*, namely American Sign Language (ASL), and in English, where gestures are in a different modality than speech. In sign languages, the visual-manual modality is used to communicate both linguistic and gestural information.

Recall that the kind of gestures we are interested in are iconic. Iconicity is widespread throughout sign languages (see, e.g., Klima and Bellugi 1979; Taub 2001; Emmorey 2002; Meir 2010; Schlenker 2018d), ranging from signs with iconic forms, like the sign TREE for 'tree' in ASL, which is made with an upright forearm and spread fingers depicting the trunk and branches of a tree (Emmorey 2002), to height specifications on sign language loci (Schlenker 2018d). However, many of the examples of iconicity in sign language are quite different from the iconicity associated with depictive co-speech gestures. For instance,

the sign VOTE meaning 'to vote' in ASL resembles the motion of stuffing a ballot box. This obviously is iconic, representing one method of voting, but the same sign also applies to different methods of voting such as electronic voting. Many examples of iconicity in sign language are like this, where there is an iconic aspect, but not a direct mapping between the signs in signing space and the interpretation. This is in contrast to iconic gestures, whose iconicity is interpreted directly.

In this chapter we will focus instead on a particular iconic construction in sign languages known as a *classifier predicate* (or sometimes "depicting verb"), whose iconicity is interpreted directly, like gestures. Classifier predicates convey iconic, analog information about the relative spatial arrangement and/or movement of referents. They have been found in nearly all of the world's sign languages (Zwitserlood 2012), and their linguistic status has been the subject of much debate. These complex morphosyntactic constructions consist of two parts: (i) a handshape that is semantically specified for a class of objects (the *classifier*), together with (ii) some kind of movement in signing space that iconically depicts the movement and/or location of the referents in real space (the *root*) (Zucchi 2012). Figure 4.1 gives an example of an ASL sentence containing two classifier predicates that iconically depict the (relative) spatial locations of a house and a bicycle.



'The bike is near the house (in this arrangement).'

Figure 4.1: *An ASL sentence containing two* classifier predicate *constructions, one for each referent's location. Images from Emmorey (2002:87).*

The goal of this chapter is to directly compare the negation of co-speech gestures in English with the negation of classifier predicates in ASL, which (like gesture) are analog

and depictive but (unlike gesture) share the same mode as other sentential operators. Using simple scenes to collect informal semantic judgments from English and ASL consultants, we compare and contrast the way negation and depictive content interact in both languages. This chapter is organized as follows. First, Section 4.2 begins with background on classifier predicates in ASL. Next, Section 4.3 reviews two semantic analyses for integrating iconic content into the grammar, and looks at the pragmatic status of such content. Section 4.4 looks at negation in sign language, and also reviews existing examples of negation targeting depictive content in the gesture and sign language literature. Section 4.5 presents and discusses data from a series of truth value judgment tasks in both English and American Sign Language we designed to compare the behavior of depictive content under negation. Finally, Section 4.6 concludes.

4.2 Depiction in sign: Classifier predicates

This section gives an overview of classifier predicates in sign languages: what they are, how they are constructed, how they are categorized, and how they are interpreted.

4.2.1 What are classifier predicates?

Just as spoken words can be analyzed into sublexical units that can be modeled by a phonological feature system (e.g. by place of articulation, nasality, voicing, etc.), signs in sign languages can similarly be broken down into sublexical units organized in a hierarchical feature system (Sandler and Lillo-Martin 2006). Building on the revolutionary work of Stokoe (1960), sign language researchers and linguists have proposed various phonological feature systems for signs, most of which categorize signs according to three parameters: (i) *handshape*, (ii) *location*, and (iii) *movement*. Just as changing the nasality feature on the first segment of the English word /dʌn/ 'done' yields the entirely different word /nʌn/ 'none', minimal pairs exist in ASL which differ in any one of the three sign parameters. Figure 4.2 (images from Sandler and Lillo-Martin 2006:117) gives an example for each type of minimal pair: (a) The signs for APPLE and CANDY have different handshapes but the same

location and movement specifications (side of the chin, rotation of the hand); (b) UGLY and SUMMER differ in location of the sign (nose and forehead, respectively) but have the same handshape and movement (extended index finger, closing of the finger while moving the hand from the contralateral to the ipsilateral side of the face); (c) TRAIN and CHAIR have the same handshape and location, but differ in movement; TRAIN is signed by the fingers of the dominant hand moving forwards and backwards over the fingers of the non-dominant hand, and in CHAIR the dominant hand fingers move down twice touching those of the non-dominant hand.

Classifier predicates are, on the face of it, an exception to the regular phonological system: they are composed of a handshape, similar to the required handshapes of other signs, but the location and movement draw from a broader set of elements and instead of being interpreted in discrete categories, are rather interpreted gesturally to convey spatial information about referents (Emmorey and Herzig 2003). Like classifiers in spoken languages, the handshape represents a specific class of object (e.g., broad and flat; person). The entire construction functions as a predicate. Phonologically, the location and movement parameters are combined into a single iconic depictive unit in classifier predicates. For example, consider a classifier predicate that describes the path of a bicycle (4.4).

(4.4) Classifier predicate for movement of bicycle: (Davidson 2015:491)

$$\underbrace{\text{CL-3}}_{\text{handshape}} \quad \underbrace{(\text{wavy path from point } a \text{ to } b)}_{\text{location+movement}}$$

The classifier construction is made up of a classifier handshape (here, semantically specified for a vehicle) together with an iconic depiction of how and where the bicycle moves, which is the combination of the location (in signing space) and movement (in signing space) morphemes.

Classifier predicate constructions can also be more complex by describing more than one referent's spatial location/movement. Consider the following example that contains three classifier predicates (4.5).

(4.5) WOMAN CL-1(located in point *a*) (Davidson 2015:492)

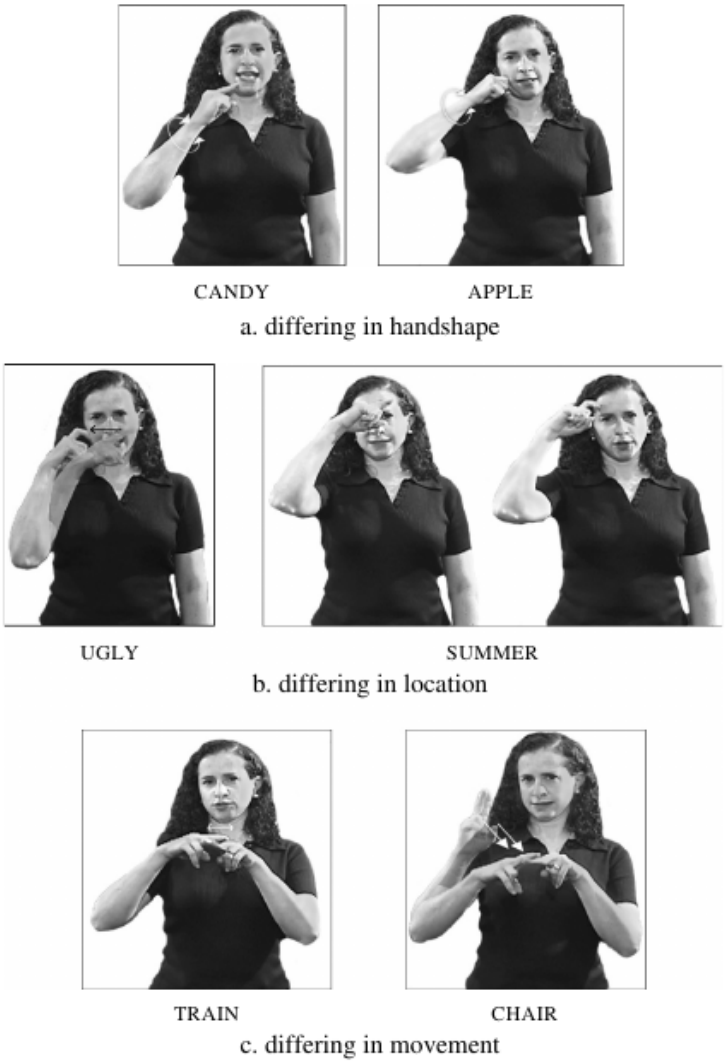


Figure 4.2: *ASL minimal pairs, differing in (a) handshape, (b) location, and (c) movement (Sandler and Lillo-Martin 2006:117).*

MAN CL-1(located in point *b*)

[Right hand] CL-1(straight movement toward center)

[Left hand] CL-1(straight movement toward center).

'The man and woman walked toward each other.'

The first classifier predicates indicates the location of the woman, the second the location

of the man, and the third depicts the two of them walking towards each other using both hands. Thus classifier predicates include aspects outside the regular phonological system, and are interpreted in a depictive way — both the hallmarks of depictive co-speech gestures, and why we consider them here for comparison with gestures under negation.

4.2.2 Classifier handshape and movement categories

Many categorizations of sign language classifier handshapes have been suggested for ASL and other sign languages (Supalla 1982; Liddell and Johnson 1987, among others). One representative example is the classification described in Benedicto and Brentari (2004), based off of Engberg-Pedersen (1993), which proposes four categories of classifier handshapes: (a) *whole entity* handshapes, in which the shape of the hand refers to an entire entity based on that entity's semantic or descriptive properties (e.g. Figure 4.3a); (b) *handling* handshapes, where the handshape refers to the way objects or instruments are held or manipulated (e.g. Figure 4.3b); (c) *extension-and-surface* handshapes, which depict the size/shape, perimeter, width, or surface of an object; and (d) *body part* handshapes, which refer to just a part of the (animate) body, such as a head or limb. Examples of these four types of classifier handshapes are given in (4.6).

(4.6) Examples of types of classifier handshape morphemes (examples from Benedicto and Brentari 2004:750; images from Aronoff et al. 2003:67, Emmorey 2002:319–20)

a. **Whole entity**

i. 'CL-3:vehicle[1]'  , used for vehicles (e.g. cars, trucks, bicycles, trains, boats, submarines, Aronoff et al. 2003:67)

ii. 'CL-B:2D_flat_object'  , used for broad, flat objects, like a bed or

[1]In 4.6, the words after the colons give a short description of the type of object/entity that classifier handshape can be used with, e.g., $\underbrace{\text{CL-1}}_{\text{handshape}}$: $\underbrace{\text{long\_narrow\_object}}_{\text{description of class}}$.

piece of paper

b. **Handling**

i. 'A:grabbable_object'  , used for an object that can be grabbed

c. **Extension-and-surface**

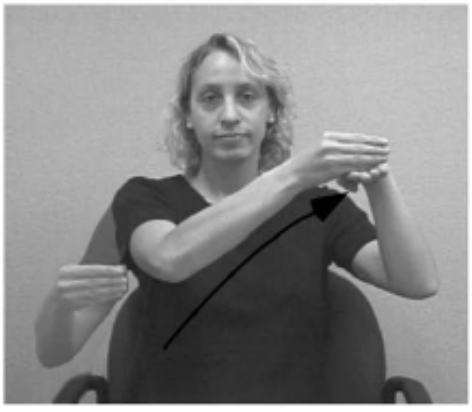
i. 'CL-L-L:square_2D_object'   , used for a square object

d. **Limb/Body part**

i. 'CL-S:head' 

Some researchers try to collapse these into just two main categories: *whole entity classifiers* and *handling classifiers* (Zwitserlood 2012), although see Benedicto and Brentari (2004) for syntactic arguments that body part/limb handshapes should be a separate category.

(a) *'The cat jumps on a shelf.'*



(b) *'Someone puts the book on a shelf.'*

Figure 4.3: *Two illustrations showing different handshape types combined with movement arcs. (a) A* whole entity *classifier handshape for small animals, such as a cat. (b) A* handling *classifier handshape used for handling thick flat objects, such as a book. (Emmorey 2002:89)*

Benedicto and Brentari (2004), again after Engberg-Pedersen (1993), also propose a four-category classification system for the movement+location verbal root of classifier predicates in sign languages. The categories of morpheme are: (a) *motion* movements, where the motion of the hand(s) corresponds to movement of a referent to or from a point; (b) *manner* movements, which describe actions or the path or manner of movement; (c) *position* movements, which are short downward movements of the hand that locates an object at a point in signing space, without indicating the movement of the object; and (d) *extension* movements, which describe the limits or extension of objects or their locations in space, without indicating the movement of the object. Examples of the four types of movement+location morpheme are given in (4.7).

(4.7) Examples of types of classifier movement morphemes: (Benedicto and Brentari 2004:749)

a. **Motion movements**

- MOVE 'to move'
- GO 'to go'

b. **Manner movements**

- BRUSH_TEETH 'brush teeth' (*action*)
- CIRCLE 'to go in circles' (*path*)
- WAVE 'to go/move in an undulating way' (*manner*)

c. **Position movements**

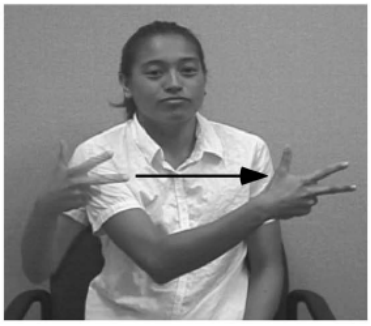
- LOCATE 'to be in, to exist'

d. **Extension movements**

- FLAT_SURFACE 'to be/exist a flat surface object'
- IN_A_ROW 'to be/exist objects in a row'
- SQUARE_BORDER 'to be/exist a square-shaped object in a location'

Figure 4.4 shows two classifier predicates consisting of the classifier handshape representing a vehicle (CL-3), which is simultaneously signed with two different types of movement+location morphemes: either (a) a location "hold" movement, or (b) a "straight-line" movement. We will see in the next section that the movement+location morpheme is in-

(a) CL-3(LOCATE)
'A car is located here.'



(b) CL-3(MOVE_STRAIGHT)
'The car moves in a straight line.'

Figure 4.4: *ASL classifier predicate constructions indicating (a) the location of a vehicle, and (b) the movement path of a vehicle. Images from Emmorey (2002:77).*

terpreted gradiently. One therefore wonders if a long list of morphemes is the right way to categorize this parameter, or if instead it might be better to view this morpheme as gesture-like.

4.2.3 Interpretation: Part lexical, part gestural

How are classifier predicates interpreted? Morphemes in spoken languages are discrete; this includes closed-class morphemes that represent spatial location, such as prepositions and locative affixes. But classifier predicates in sign languages seem to convey a mixture of categorical and gradient iconic information. Emmorey and Herzig (2003) ran a series of quantitative experiments to investigate which aspects of classifier constructions in American Sign Language (ASL) are treated by signers as discrete categorical morphemes and which aspects as gradient, analog information.

The first finding of Emmorey and Herzig (2003) is that the *handshapes* in classifier constructions are represented and interpreted as discrete categorical morphemes. They

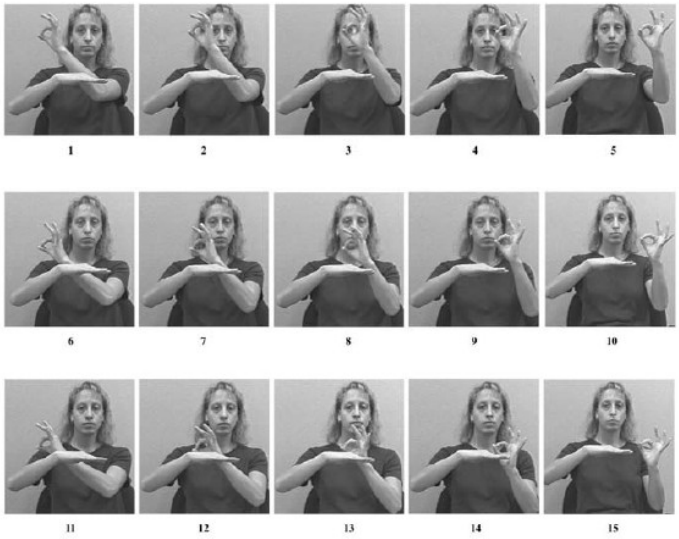
describe a series of interpretation, acceptability, and production studies that involve various sizes of an ASL classifier handshape representing a medallion hung around the neck (Figure 4.5). They found that deaf signers, unlike hearing non-signers, interpret and use the classifier handshapes CL-F and CL-BABY-C categorically rather than gradiently. Furthermore,

Figure 4.5: *A series of images of ASL classifier handshapes designed to test the interpretation of the size of a medallion. Examples 1-5 use an* F*-handshape, while examples 6-10 use a* BABY-C*-handshape. (Emmorey and Herzig 2003)*

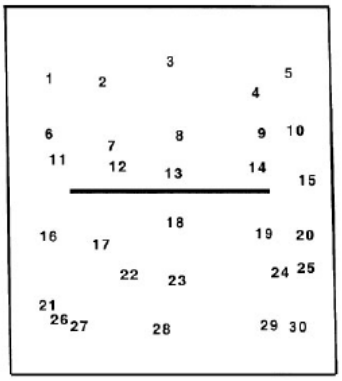
just as English speakers know how to gradiently vary vowel duration in a (discrete) word to iconically convey variation in duration or length (e.g. *long* versus *looooong*), signers know how to manipulate aspects of categorical forms in order to convey gesture-like analog information (Emmorey and Herzig 2003:238). For instance, in the medallion-size production study they found that signers gradiently varied the size of categorical handshapes (representing three relative sizes of a medallion) in order to convey "close" analog deviations from the default

categorical interpretations of these handshapes.

In contrast to classifier handshapes, Emmorey and Herzig (2003) found that ASL signers represent and interpret spatial locations gradiently using an analog system, rather than categorically. Participants were shown a series of images showing a signer signing a classifier predicate that indicated the relative positions of a circle and a bar (Figure 4.6a), and then were asked to mark down on a piece of paper where the circle was with respect to the bar. Despite the closeness of all of the positions, signers did not group multiple stimuli together into linguistic spatial categories, but instead distinguished nearly all of the stimuli into separate positions on the piece of paper (Figure 4.6b). This indicated that the signers were interpreting the classifier predicate locating the circle and bar in a gradient, analog fashion.



(a) *Stimuli (15/30)*



(b) *Responses from deaf participants, by stimuli number*

Figure 4.6: *Testing the interpretation of gradient variation in location: (a) sample stimuli and (b) responses from deaf signers (Emmorey and Herzig 2003)*

Given that the location+movement morpheme of a classifier predicate is interpreted iconically, one might wonder how a gradient, analog item might be integrated into the discrete, compositional semantics of language. The next section explores some approaches to this question.

4.3 The semantics and pragmatics of iconicity

How do the analog iconic and discrete categorical fit together in the grammar? In this section we give an overview of two different proposals for the semantics of iconicity, as well as an overview of the typology of the pragmatic status of iconic content.

4.3.1 Iconic content as demonstrations

Referring to classifier predicates as 'depicting verbs', Liddell writes "What distinguishes depicting verbs from other verbs is that, in addition to their encoded meanings, these verbs also depict certain aspects of their meanings" (Liddell 2003:261). Liddell highlights one fascinating aspect of sign language classifier predicates: they depict or demonstrate the action/event they represent. One semantic analysis of classifier predicates, proposed by Zucchi et al. (2011) and Davidson (2015), takes this demonstration-like component seriously, and formally treats the depictive movement/location morpheme of a classifier predicate as a gestural demonstration that is itself the referent of a covert demonstrative, i.e., a demonstration argument. Intuitively, one would interpret, for example, the classifier predicate (4.8a) expressing the motion of, say, a bicycle, as the paraphrase in (4.8b).

(4.8) a. Classifier predicate for movement of a bicycle along wavy path from point *a* to *b*:
CL-3(*wavy path from* a *to* b)

b. Paraphrase:
the bicycle moved in a way similar to this + (*wavy path from* a *to* b)

The 3 handshape of the classifier predicate is linguistic (with a discrete component of meaning), but the movement/location of the handshape (the "wavy path") is gestural, a demonstration of the way the bicycle moved.

More precisely, Davidson (2015) treats classifier predicates as light verbs requiring a demonstrational argument, akin to the quotative predicate *be+like* in English. She uses a Neo-Davidsonian event-based semantics and a new semantic type *d* for demonstrations, as well as introducing a predicate *demonstration-of* (shorthand: *demonstration*) that takes a

demonstration (type d) and an event (type e) as its arguments. A formal definition of the *demonstration-of* predicate is given in (4.9).

(4.9) (Davidson 2015:487)

> A demonstration d is a *demonstration of* e (i.e. $demonstration(d, e)$ holds) if d reproduces properties of e and those properties are relevant in the utterance context.

Depending on the type of classifier predicate (e.g. whole entity, handling, extension and surface of something, or body part/limb), there are different verbal roots that require a demonstration depicting the relevant location/movement property of the subject. This is formalized in (4.10).

(4.10) (Davidson 2015:494)

> A *classifier predicate* is a verbal root of type $\langle$d,$\langle$e,t$\rangle\rangle$ that is a demonstration of an agent in an event, where the demonstration can show the agent's
>
> (i) *movement* (root is abbreviated: MOVE),
>
> (ii) *location* (LOCATE),
>
> (iii) *manner* (MANIPULATE), or
>
> (iv) *extent* (EXTEND-TO).

Let's see how this works in a few examples. First, consider a classifier predicate that depicts where a woman is located, e.g., CL-1-(LOCATE).

(4.11) WOMAN CL-1-(LOCATE). 'There is a woman here.'

Then the formal semantic derivation of (4.11) would be as follows.

(4.12) WOMAN CL-1-LOCATE(placing the hand at location a).

> **First we give the specific demonstration a name:**
>
> $[\![$placing the hand at location $a]\!] = d_1$
>
> **Next, the denotation of the classifier predicate:**

$[\![$CL-1-LOCATE$]\!] = \lambda d \lambda x \lambda e.\ [theme(e,x) \wedge human(x) \wedge locating(e) \wedge demonstration(d,e)]$

Plugging in d_1 for d we get:

$[\![$CL-1-LOCATE(hand at location a)$]\!]$ =

$\lambda x \lambda e.\ [theme(e,x) \wedge human(x) \wedge locating(e) \wedge demonstration(d_1,e)]$

Plugging in *woman* for entity x we get:

$[\![$WOMAN CL-1-LOCATE(hand at location a)$]\!]$ =

$\lambda e.\ [theme(e,woman) \wedge human(woman) \wedge locating(e) \wedge demonstration(d_1,e)]$

And finally, existential closure yields:

$\exists e.\ [theme(e,woman) \wedge human(woman) \wedge locating(e) \wedge demonstration(d_1,e)]$.

Thus utterance (4.11) is true if and only if there is an event e which is a locating event, the theme of event e is a human woman, and d_1 (showing the location of the woman) is a demonstration of e.

A feature of this proposal is that since the gradient iconic component of classifier predicates are treated as demonstrations, the context can determine which parts of the demonstration are to be interpreted iconically, leaving iconicity interpretations to the pragmatics rather than the semantics. In this sense it is a more conservative approach than that discussed in the next section, which develops a new semantics for iconicity.

4.3.2 Iconic functions in the lexical semantics

A second proposal for the semantic integration of iconic content comes from Schlenker (2018b,d,c, to appear). Unlike Zucchi et al. (2011) and Davidson (2015), Schlenker locates iconic properties directly in the lexical entries of iconic content. For example, consider again the experiment by Emmorey and Herzig (2003) where the relative positions of an F-handshape and a B-handshape are interpreted in a gradient iconic fashion as representing the relative positions of a point and a bar (Figure 4.6). Schlenker (2011) proposes the following semantic rule for this example:

(4.13) (Schlenker 2011:230)

If i is a locus,

$[\![\text{F-classifier}_i\ \text{B-predicate}]\!]^{c,s} = \#$ iff $s(i) = \#$ or $[\![\text{B}]\!]^{c,s} = \#$. Otherwise,

$[\![\text{F-classifier}_i\ \text{B-predicate}]\!]^{c,s} = 1$ iff $\langle i, \text{B}\rangle$ is iconically projective to $\langle s(i), [\![\text{B}]\!]^{c,s}\rangle$ along the 'geometric' dimension.

Here, '$\langle i$,B$\rangle$ is *iconically projectable* to $\langle s(i), [\![\text{B}]\!]^{c,s}\rangle$ along the "geometric" dimension' means that there is a natural geometric projection that maps the relative position of locus i and sign B and the objects they denote, namely $s(i)$ and $[\![\text{B}]\!]^{c,s}$.

Schlenker (2018d) suggests that similar iconic restrictions be built into the lexical entries of iconic content in sign languages. Another example he looks at is the ASL sign GROW, 'to grow, expand'. This sign can be signed neutrally, but it can also be iconically modulated to indicate the rate and degree of expansion, whereby the faster the signer's hands move apart, the faster the growth, and the farther apart the signer's hands end up, the bigger the degree of growth (Schlenker et al. 2013; Schlenker 2018b,d). This is depicted in (4.14).

(4.14) Representation of GROW (Schlenker 2018b:884)

	Narrow endpoints	**Medium endpoints**	**Broad endpoints**
Slow movement	small amount, slowly 	medium amount, slowly 	large amount, slowly 
Fast movement	small amount, quickly	medium amount, quickly	large amount, quickly

He suggests that the stipulation about the relationship between the movement of the hands and the rate of growth be explicitly written into the lexical entry for GROW.

Schlenker's semantics for iconicity is a bigger departure from classical semantics than Zucchi et al. (2011) and Davidson's (2015) proposal, introducing a whole new type of truth conditional requirement. The advantage of this system is its power, its ability to capture a wide variety of iconic phenomena in sign language (Schlenker 2018d,c). This power may

also be a disadvantage, in that one would need to explain why iconic phenomena in spoken languages are not more widespread if such iconic lexical entries are available in the grammar. Another disadvantage is the stipulative nature of the lexical entries: the fact that one's hands being farther apart correlates with more growth in the GROW example, for instance, is not mere coincidence; we have presumably extra-linguistic intuitions about gestural movements and their metaphorical interpretations. It would be redundant to encode these intuitions over and over again in lexical entry after lexical entry.

4.3.3 The pragmatic status of iconic content

As for the pragmatic status of classifier predicates, Schlenker (2018b) proposes a typology of projection properties for iconic enrichments in both spoken language and sign language. Two parameters, [±*internal*] and [±*separate_time_slot*], constrain the possible information statuses and type of inferences of iconic content. The *separate time slot* parameter keeps track of whether iconic enrichments occupy their own time slot in the speech/sign stream or whether instead they co-occur with speech/sign. *External* enrichments ([-*internal*]) are defined as modifications to linguistic expressions that can be eliminated/removed without affecting the integrity of the expression they modify; an example is co-speech gesture (4.15), where the gesture is "added" to a spoken expression in a different modality. *Internal* enrichments, on the other hand, involve modulations of words and signs themselves, for example vocal iconic vowel lengthening (4.16).

(4.15) Co-speech gesture: [−*internal*, −*separate_time_slot*]

John brought [a bottle of beer]_LARGE  . (Schlenker 2018b:881)

⇒ the bottle of beer was large

(4.16) Vocal iconic modulation: [+*internal*, −*separate_time_slot*]

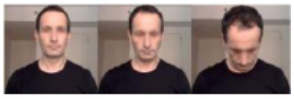
The talk was loooong. (Schlenker 2018b:881)

⇒ the talk was very long

Examples of iconic enrichments that have their own time slot ([+*separate_time_slot*]) include *post-speech/post-sign gestures*, which follow the expressions they modify, and *pro-speech gestures*, which fully replace an expression and thereby fulfill a grammatical function. Some examples are given in (4.17).

(4.17) a. Post-speech gesture (Schlenker 2018b)

John brought [a bottle of beer] — LARGE  .

b. In two minutes, our Chair will DOZE-OFF  .

Schlenker uses these two features to explain and predict the pragmatic statuses and semantic contributions of iconic enrichments in speech and sign. He proposes the following generalization:

(4.18) Iconic enrichment pragmatic typology: Proposed generalization (Schlenker 2018b:887)

a. [±internal]

External enrichments (−*internal*) are not-at-issue: because they are external, it should be possible to disregard them without affecting the main, at-issue content of the clause they appear in. By contrast, internal enrichments (+*internal*) can make any semantic contribution — just like standard words.

b. [±separate time slot]

Enrichments that have a separate time slot (+*separate time slot*) cannot be trivial (=presupposed): because they have their own time slot, they must make a non-trivial contribution to the sentence. By contrast, enrichments that do not have a separate time slot (−*separate time slot*) are not so constrained.

This generalization entails in particular that co-speech/co-sign gestures (*−internal*, *+separate time slot*) have a not-at-issue semantics, while iconic modulations of sign or speech (*+internal*, *−separate time slot*) can be at-issue or not, as the case may be.

What does this mean for classifiers? Schlenker does not directly address any examples with classifier predicates, but we can try to infer how they fit into the picture. One possibility is that sign language classifier predicates have the same properties as co-sign gestures, since they consist of a lexical and a gestural component signed simultaneously (*−separate time slot*). The other possibility is that classifier predicates are more similar to verbs in sign languages that have been iconically modified, like the sign GROW 'to grow, expand' discussed above in example (4.14), which can be at-issue. In fact Schlenker provides an example where the verb GROW is under the scope of negation, precisely the kind of situation we're interested in looking at for classifier predicates. This example is given in (4.19).

(4.19) (Schlenker 2018b:12)

Context: We are discussing the recent history of the speaker's research group.

LAST-YEAR POSS-1 GROUP NOT ...

a. $\text{GROW}_{\text{neutral}}$.

b. $\text{GROW}_{\text{large}}$.

c. $\text{GROW}_{\text{small}}$.

'Last year, my group didn't a. (really) grow / b. grow a lot / c. grow (even) a little.'

The interpretations given for (4.19) seem to suggest that negation can target the iconic part of GROW. Although GROW is iconically enriched rather than being a true classifier predicate, we will see in the results of our study below that classifier predicates in ASL appear to behave in the same way, unlike co-speech gestures.

Next we take a closer look at negation in sign languages and present a series of examples from the literature that do consist of classifier predicates under the scope of negative operator.

4.4 Negation

4.4.1 Negation in sign languages

Just like spoken languages, sign languages have many types of negation, including clause negators/negative particles (in ASL, NOT), negative pronouns (NONE), negative responses (NO), and negative adverbs (NEVER), in addition to other irregular negative forms (e.g. negative modals, CANNOT) (Zeshan 2004; Quer 2012). In addition to *manual* markers of negation for the above categories, sign languages also heavily utilize *non-manual* negation marking, which is negative marking that is conveyed by head movements and/or facial expressions. Non-manual markings in sign are often seen as analogous to intonation patterns in speech.

With respect to negation strategies, sign languages broadly fall into two categories: non-manual dominant languages and manual dominant languages (Zeshan 2004, 2006). The former are languages where non-manual negation marking is obligatory for clausal negation, and in fact can be used as the sole marking of clausal negation. Manual dominant languages, by contrast, require manual markers of negation, and accompanying non-manual negation marking is optional. ASL is a non-manual dominant sign language; headshake is obligatory with, e.g., the manual sign for sentential negation (Sandler and Lillo-Martin 2006). Some sign researchers indicate the non-manual headshake in their glosses, while others don't; we follow the various authors' glossing conventions in the examples that follow, and in our own examples in Section 4.5 we do not indicate non-manual negation markers because it is always present.

Like many sign languages, ASL has a manual sign for sentential negation that negates the truth of the proposition, NOT (Figure 4.7a). It takes scope over the predicate and is quite similar to the English *not* in its meaning and its function (Wood 1999). It can appear in preverbal position or in sentence-final position (4.20).

(4.20) (Wood 1999:21–22)

a. MARY NOT BREAK FAN. (preverbal)

'Mary did not break the fan.'

(a) **(b)**

Figure 4.7: *ASL manual signs for (a) the sentential negation particle* NOT*, and (b) the negative determiner* NONE/NO^{o} .

b. JOHN BREAK FAN NOT. (sentence-final)

'John did not break the fan.'

In addition to a manual sign for clausal negation, ASL also has two manual negative determiners that are often used for general negation. Wood (1999) writes these as NOTHING and NO^{o} [2], with the superscript on the latter indicating the coarticulated non-manual facial expression of rounded lips (Figure 4.7b)[3]. An example of the ASL negative determiners is given in (4.21).

(4.21) JOHN BREAK FAN NOTHING/NO^{o} . (Wood 1999:40)

'John did not break any (part of the) fan.'

Syntactically, NOTHING takes only an internal argument as its complement, while NO^{o} can take either an internal or external argument as its complement. Semantically, the two

[2]Wood (1999) notes that NO^{o} is more commonly written as NONE, and in our stimuli later we use this notation.

[3]The configuration of the mouth distinguishes NO^{o} from yet another negator NO^{-}, which is a propositional negator and is produced with the same manual movements but with the mouth instead flattened and drawn downward (Wood 1999:36).

negative determiners receive a partitive interpretation; for example, sentence (4.21) means not only that John didn't break the fan, but also that he didn't break any part of the fan.

In the stimuli we created to systematically test the behavior of ASL classifier constructions and English depictive gestures under negation, we chose to use both sentential negation examples and examples with a negative determiner. Sentential negation was the obvious (uncomplicated) choice, and the negative determiner was chosen because it appeared in all of the naturally occurring depictive examples in the literature that we found, which we review in Section 4.4.2 below.

4.4.2 Negation/depiction examples from the literature

There are few existing examples in the literature of negation targeting either depictive gestures or depictive classifier predicates. For ASL, those that do exist tend to express negation with the negative quantifiers/determiners (NOTHING, NONE) rather than with sentential negation. Below we review examples that we did find in the literature for speech and sign.

Speech and gesture

(Schlenker 2018a) provides several examples of co-speech gestures under negation, including sentential negation (4.22) and with the quantifier *none* (4.23).

(4.22) (Schlenker 2018a)

John didn't [punish]_SLAP  his son.

(4.23) (Schlenker 2018a)

None of these 10 philosophers found [a bottle he liked]_LARGE  .

Tieu et al. (2017) test the deictic gesture UP under sentential (4.24a) and quantificational (4.24b) negation.

(4.24) (Tieu et al. 2017)

a. The boy will not [use the stairs]_UP  .

b. None of these three girls will [use the stairs]_UP  .

And finally, we previously saw example (4.3a) from Tieu et al. (2017), repeated below as (4.25).

(4.25) John **didn't** [help]_LIFT  his son. (=4.3a)

Sign and classifier predicates

In order to construct natural stimuli involving negation and classifier predicates, we wanted to start with existing examples and modify them, but it turned out to be surprisingly difficult to find such examples. Below are the examples we did find; note that they are all elicited, rather than from corpora.

Schlenker (to appear) constructs a complex example involving a classifier predicate showing the movement of a large helicopter and the negative quantifier *none*, shown in (4.26).

(4.26) (Schlenker to appear :50)

[*Context: our company has one helicopter and one airplane.*]

WITHIN 1-HOUR OUR COMPANY BIG HELICOPTER BOSTON$_a$ NEW-YORK$_b$ NONE CL-HELICOPTER-LARGE(MOVE from *a* to *b*, with a smooth detour mid-way).

'Within an hour none of our company's big helicopters will fly (with the assumption that this would involve a smooth detour) from Boston to New York.'

This example was experimentally tested; the projective inference that if any of the helicopters had gone from Boston to New York, they would have done so with a smooth detour was rated 6.5/7 on an inferential strength scale.

Benedicto and Brentari (2004) provide several nice examples of classifiers under the scope of a negative determiner which they gloss as NOTHING. (4.27) consists of a whole entity classifier, depicting a person bowing, under a negative determiner:

(4.27) (Benedicto and Brentari 2004:760)

[*Context: A curtain call after a theater performance*]

ACTOR CL-1(bowing motion) NOTHING.

'None of the actors bowed.'

Another example of a whole entity classifier under a negative determiner is (4.28a), where the B-handshape represents a flat object (a book).

(4.28) (Benedicto and Brentari 2004:771)

BOOK CL-B(MOVE) NOTHING.

'None of the books fell down (on its side).'

A similar example that uses a handling classifier instead of a whole entity classifier to show the movement of a book (by an unnamed agent) is (4.29).

(4.29) BOOK CL-C(MOVE) NOTHING.

'S/he didn't put any book down (on its side).'

Finally, Wood (1999) gives an example of a handling/instrumental classifier under the scope of the negative determiner NOTHING; the classifier depicts someone opening a window.

(4.30) (Wood 1999:78)

PRO-1 WINDOW CL-OPEN-WINDOW NOTHING.

'I did not open any of the window(s).'

This last example was the inspiration for the "window scene" examples we used as part of our stimuli in Section 4.5 below.

It is interesting that all of these examples, with the exception of (4.26), have negation at the very end. This is despite the fact that ASL is an SVO language, like English. Gonzalez et al. (to appear) observe that negation can be put sentence-finally in ASL to get widest scope readings; perhaps that explains the negative placement in these examples. As previously mentioned, we included a negative determiner, NONE, in the design of our survey, along with basic sentential negation; we put NONE at the end of the survey sentences in order to stay faithful to the literature examples (see Section 4.5.2).

4.5 Testing depictive modifiers under negation

One aim of the following experiment is to confirm previous observations that suggest that co-speech gestures behave differently under negation than spoken linguistic content. To test this, we designed an experiment that would directly compare the behavior of co-speech gestures and color words under negation. If our experiment confirms previous observations and shows that co-speech gestures, unlike color words, cannot be targeted by negation in English, we are then interested in investigating why this is true. One possible reason could be that a modality mismatch prevents the negation from targeting the co-speech gesture; in other words, negation in the auditory-oral modality cannot target gesture, which is in the visual-manual modality. A different possibility is that co-speech gestures are not integrated into the grammar in a way that allows them to interact with other operators like negation. In the following experiment, we attempt to determine whether the first or

second of these hypotheses is more likely. In order to isolate the first variable of modality, we compared the behavior of co-speech gestures under negation in English (where negation and depictive modifier are in different modalities) to that of classifier predicates under negation in ASL (where negation and depictive modifier are in the same modality). We established in previous sections that classifier predicates in ASL are very similar in some ways to co-speech gestures plus speech in English, having both linguistic and depictive components. Thus, if our research shows that depictive modifiers in ASL can be targeted by negation more effectively than co-speech gestures, this would indicate that a modality mismatch (rather than some kind of compositional clash) is the more important factor in preventing negation from targeting co-speech gestures.

4.5.1 Methods

We created several simple experimental scenes and asked both English and ASL consultants to provide informal truth value judgments for sentences containing various types of negation and depictive modifiers. These results were contrasted with truth value judgments for minimal pair sentences containing non-depictive modifiers.

7 non-signing English speakers and 4 deaf ASL consultants (3 native/early signers) were shown slides each containing a written context paragraph, a simple scene, and a short video that they were asked to watch. Each video presented a sentence under negation. The contexts and the scenes were the same for the English and ASL consultants, while the videos differed according to language. For each scene-video pairing, the consultants were asked to perform a Truth Value Judgment Task (TVJT) and answer the question: *Is the sentence in the video true in the context of the written paragraph and scene?*

The videos were hosted on YouTube. See Figure 4.8 for a screenshot of one of the troll scene slides. There were 12 scene-video pairings per language, and each participant saw all 12. All 12 scene-sentence pairings can be found in Appendix C. English judgments were informally collected over Skype; ASL judgments were collected by a deaf collaborator.

Figure 4.8: *Screenshot of a troll scene slide shown to English and ASL consultants to elicit truth-value judgments.*

4.5.2 Design factors

The dependent variable for this experiment is the truth value judgments described above. The independent variables include:

- type of modifier (depictive or non-depictive) in sentences under negation
- type of depictive modifier (size-and-shape or manner-and-path)
- type of negation in the sentence (sentential or quantificational)
- whether the depictive modifier is clausemates with negation

Our research questions are:

(i) Do any of the above factors affect the truth value judgments of English and ASL consultants?

(ii) Does a modality (mis)match between the negation and the target of negation affect the truth value judgments of consultants? In other words, is there a difference between the

truth value judgments of the English and ASL consultants?

We take a closer look at each of these design factors in turn below.

Size-and-shape vs. manner-and-path modifiers

In designing our sentences, we tried to cover at least two categories of gesture: size-and-shape, and manner-and-path. *Size-and-shape* gestures provide information about the physical qualities of entities, while *manner-and-path* gestures demonstrate the way(s) in which an action/event occurs. Examples of each type are given in (4.31).

(4.31) a. Size-and-shape depictive modifier: (Ebert and Ebert 2014, 2016)

I brought [a bottle of water]_BIG  to the talk.

⇒ entails that the water bottle was big in size

b. Manner-and-path depictive modifier: (Tieu et al. 2017)

The boy will [use the stairs]_UP  .

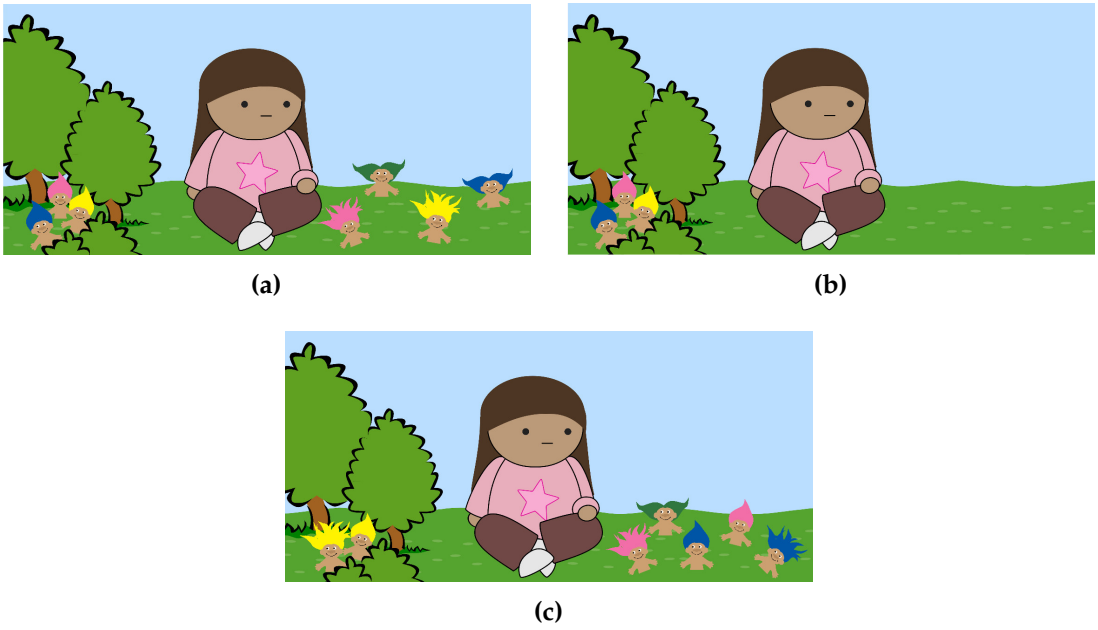
⇒ entails that the path taken on the stairs will be upwards

For our stimuli, we found classifier predicates that corresponded most closely with each of these types of gestures, and thus we came up with a canonical gesture/classifier predicate for each of the size-and-shape and manner-and-path categories. The size-and-shape modifier was instantiated by the (hard-to-describe) hairstyle of a toy troll, while the manner-and-path modifier was instantiated by the way a particular type of window opens. For each of these modifier types, we designed a series of scenes to contextually support them, against which the stimuli sentences would be evaluated.

The first series of scenes, designed for the size-and-shape modifier, features a girl named Sasha sitting in her backyard, playing with her toy trolls. (We refer to these scenes hereafter as the "troll scenes".) In the troll scenes Sasha is surrounded by her trolls, but she can't

find some of them because they are "hidden" in the bushes. The depictive modifiers used in these videos pick out one particular hairstyle shape worn by the trolls, a "single-point" style (Figure 4.12a), in contrast to trolls with other hairstyles.

There were three variants of the scene: In the first (Figure 4.9a), the hidden trolls are the ones with the single-point hairstyle; in the second (Figure 4.9b), none of the trolls are found and all are hidden in the bushes; and in the third (Figure 4.9c), the trolls with yellow hair are hidden in the bushes.



(a) **(b)** **(c)**

Figure 4.9: *Troll scene types. (a) Some trolls were found, but not those with the single-point hairstyle. (b) None of the trolls were found. (c) Some trolls were found, but not those with yellow hair.*

To test the manner-and-path depictive modifier, we created a second series of scenes showing a wall of windows of various types — some open horizontally (Figure 4.10a), and some open vertically (Figure 4.10b). (We refer to these scenes hereafter as the "window scenes".) In addition to varying whether the windows are open or shut, some versions of the window scenes have one of the windows broken, with a baseball and shards of glass lying in front it. Like in the troll scenes, some of the window scenes were designed to test the interpretation of negation with a depictive gestural modifier, and others are designed to

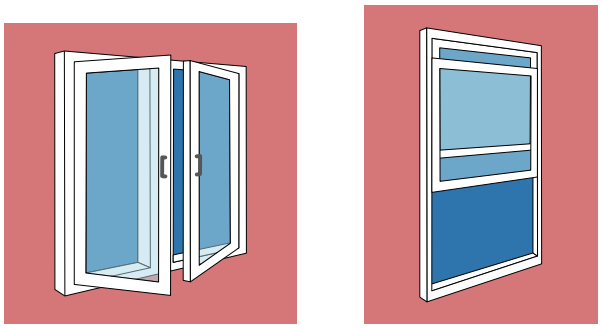


Figure 4.10: *Windows that open (a) horizontally and (b) vertically.*

test negation with a non-depictive color word. The depictive modifiers used in the stimuli with these scenes convey the manner in which a window would be opened, while the non-depictive modifiers refer to the color of the windows.

There were four variants of the window scene. In the first scene only a vertically-opened window is broken (Figure 4.11a); in the second, a green, horizontally-opened window is broken (Figure 4.11b); in the third, the two vertically-opening windows are open (Figure 4.11c); and in the fourth, none of the windows are open or broken (Figure 4.11d).

Depictive and non-depictive modifiers

The most important factor we manipulated was whether the modifier in the sentence was depictive or non-depictive. The depictive modifiers were co-speech gestures in the English sentences and classifier predicates in the ASL sentences. The non-depictive modifiers were lexical color words such as 'yellow'. This comparison enabled us to test in a controlled setting the interpretation of these depictive modifiers under negation and see how they differ from non-depictive lexical items.

In the sentences that go with the troll scenes, the depictive modifier picks out one particular troll hairstyle, the "single-point" hairstyle (Figure 4.12a). For the English sentences, the speaker in the videos uses the co-speech gesture SINGLE-POINT shown in Figure 4.12b,

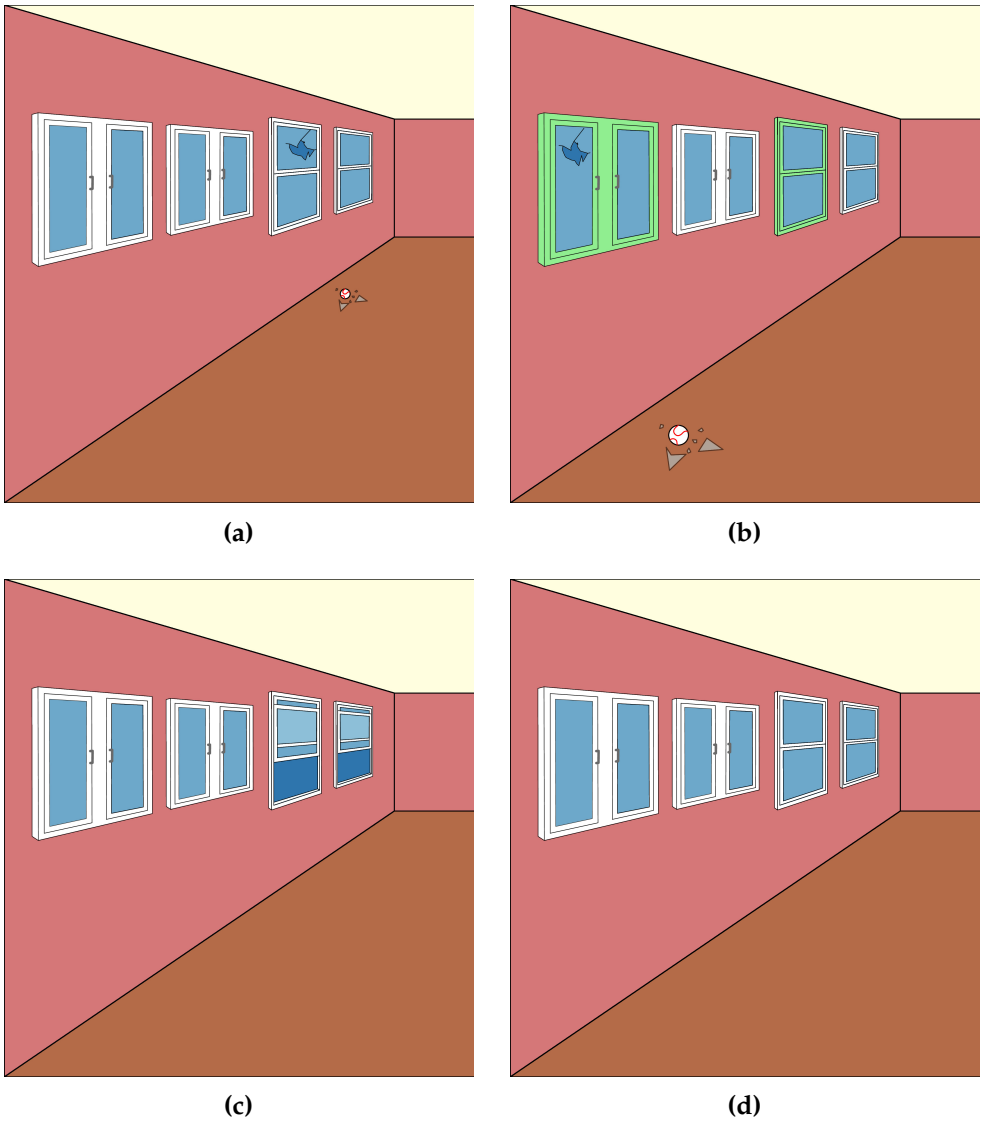


Figure 4.11: *Window scene types. (a) A vertically-opening window is broken. (b) A green window is broken. (c) Only the vertical windows are open. (d) None of the windows are open or broken.*

while the signer in the ASL videos uses a classifier predicate CL:FLAT O (Figure 4.12c). This is an "extension-and-surface"-type classifier in the schema we discussed by Benedicto and Brentari (2004); see (4.6) in Section 4.2.2

The English and ASL sentences in (4.32) are an example with the "single-point" depictive modifier, and those in (4.33) are the other half of the minimal pair with the non-depictive

(a) *A troll with a "single-point" hairstyle.*

(b) *SINGLE-POINT (English co-speech gesture)*

(c) CL:FLAT O *(ASL classifier predicate)*

Figure 4.12: *Depictive motions illustrating the "single-point" troll hairstyle of picture (a), in (b) English, as a co-speech gesture, and (c) ASL, as a classifier predicate.*

color word "yellow".

(4.32) Depictive modifier example for the "single-point" hairstyle:

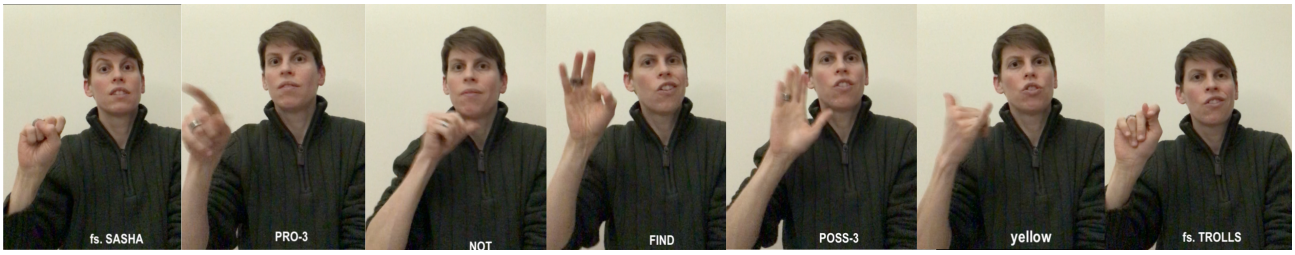
a. (English) Sasha didn't find her [trolls]_**SINGLE-POINT**.

b. (ASL) fs.SASHA PRO-3 NOT FIND fs.TROLLS **CL:FLAT O**.

(4.33) Non-depictive modifier example with 'yellow':

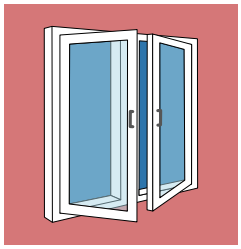
a. (English) Sasha didn't find her **yellow** trolls.

b. (ASL) fs.SASHA PRO-3 NOT FIND POSS-3 **YELLOW** fs.TROLLS.

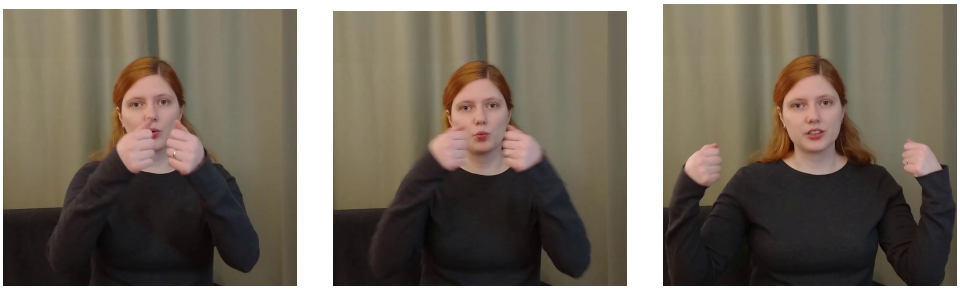


The depictive modifier used with the window scenes is a manner-and-path modifier that depicts opening a window whose doors open horizontally, like that in Figure 4.13a.

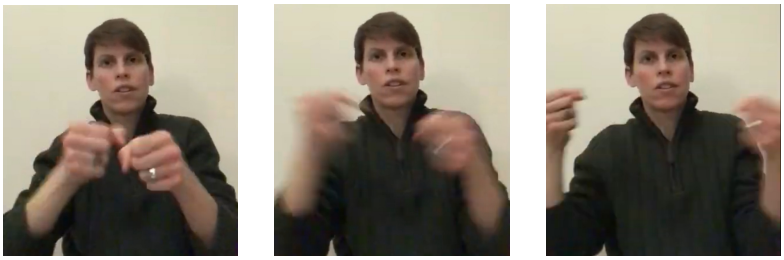
The English sentences contain the co-speech gesture HORIZONTAL-OPEN (Figure 4.13b), while the ASL sentences use the classifier predicate CL:S OPEN-WINDOW (Figure 4.13c). This classifier has a handling handshape; see (4.6) in Section 4.2.2. For a minimal pair of examples



(a) *A window that opens horizontally*



(b) *HORIZONTAL-OPEN (English co-speech gesture)*



(c) CL:S OPEN-WINDOW *(ASL classifier predicate)*

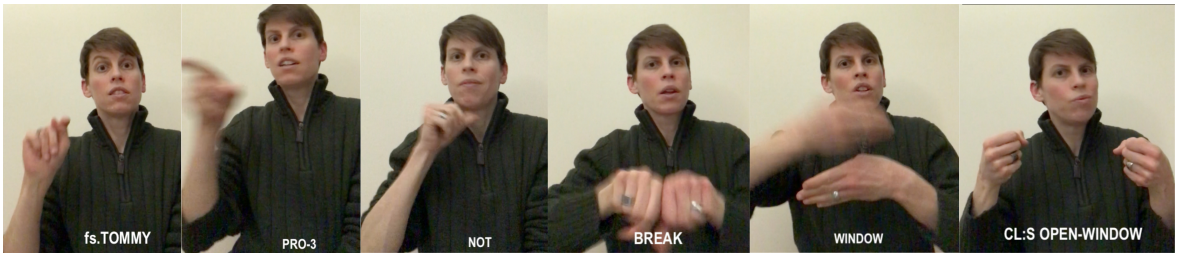
Figure 4.13: *Depictive motions illustrating the horizontal opening of a window like (a), in (b) English, as a co-speech gesture, and (c) ASL, as a classifier predicate.*

for manner-and-path depictive and non-depictive (color word) modifiers see (4.34) and (4.35).

(4.34) Depictive modifier showing windows being opened horizontally:

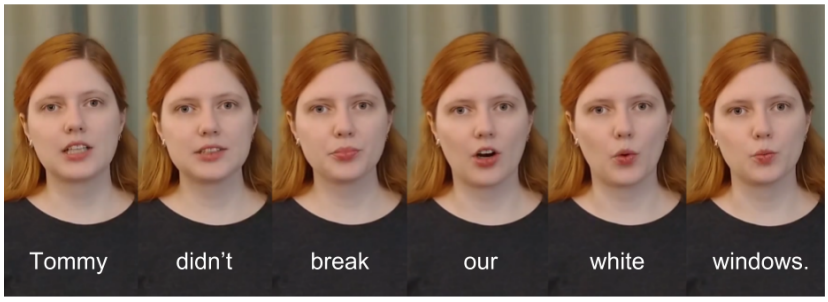
a. (English) Tommy didn't break our [windows]_**HORIZONTAL-OPEN**.

b. (ASL) fs.TOMMY PRO-3 NOT BREAK WINDOW **CL:S OPEN-WINDOW**.



(4.35) Non-depictive modifier 'white':

a. (English) Tommy didn't break our **white** windows.



b. (ASL) fs.TOMMY PRO-3 NOT BREAK WINDOW **WHITE**.

Type of negation

All of the test sentences contained negation, but not all the same type of negation. We constructed parallel examples with either sentential negation, realized as NOT in ASL, or

with the negative quantifier *none*/NONE. This reflected the two types of negation we found in examples of negation targeting depictive predicates in the literature (sign and gesture).

(4.36) Depictive size-and-shape modifier with 'none' (cf. 4.32):

a. (English) Sasha found **none** of her [trolls]_SINGLE-POINT.

b. (ASL) fs.SASHA PRO-3 FIND fs.TROLLS CL:FLAT O **NONE**.

Example (4.36) is an example with quantificational negation *none*/NONE. It can be compared to example (4.32) above which has sentential negation *not*/NOT.

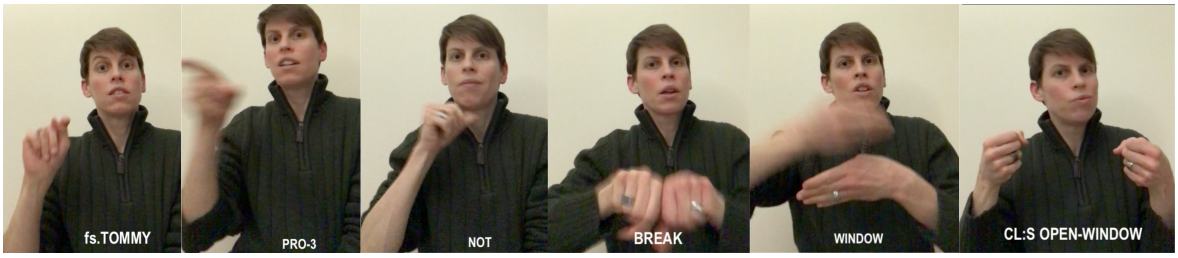
Clausemates with negation

Looking back at (4.34), repeated here as (4.37), one concern due to the understudied nature of ASL is that in the ASL example (b), the classifier predicate was not clausemates with negation, being embedded in something like a relative clause.

(4.37) (= 4.34)

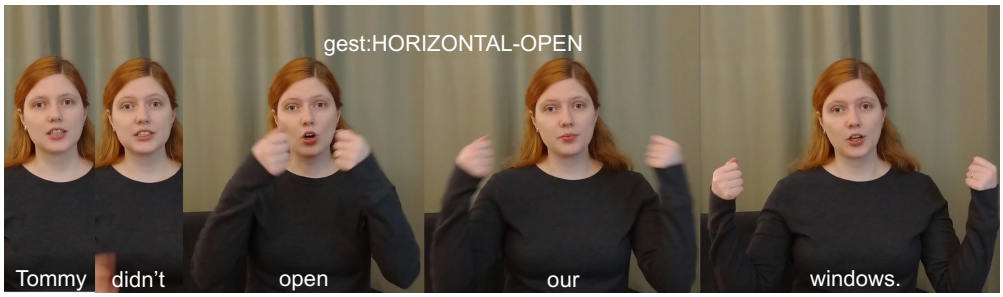
a. (English) Tommy didn't break our [windows]_**HORIZONTAL-OPEN**.

b. (ASL) fs.TOMMY PRO-3 NOT BREAK WINDOW **CL:S OPEN-WINDOW**.



To address this concern, we included a variant (4.38) where the depictive modifier targets the main verb, ensuring that it is within the same clause as the sentential negation.

(4.38) a. (English) Tommy didn't [open]_HORIZ-OPEN our windows.



b. (ASL) fs.TOMMY PRO-3 NOT CL:S OPEN-WINDOW OUR WINDOW.

Controls

We included one control sentence (4.39) in the survey which contained no modifiers, depictive or otherwise. This sentence was expected to be judged true for the scene it was

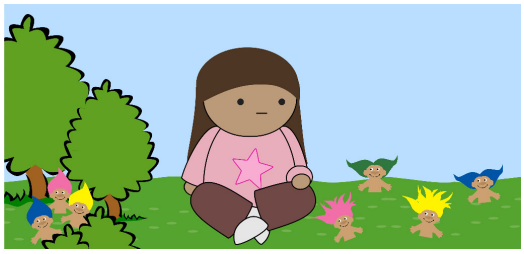
paired with (see (C.11) in Appendix C), as in that scene none of the windows were broken.

(4.39) Control item — no modifiers

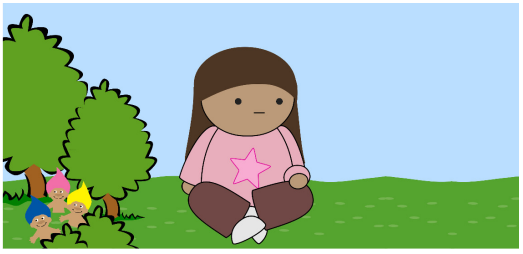
a. (English) Tommy didn't break our windows.

b. (ASL) fs.TOMMY PRO-3 NOT BREAK OUR WINDOW.

Finally, in two instances we manipulated the scenes themselves, changing the number of objects "found" (size-and-shape troll scenes) or "broken" (manner-and-path window scenes), from the default "some but not all" condition (Figure 4.14a) to the "none" condition (Figure 4.14b). All of the test sentences were true in the "none" scenes regardless of the interpretation of the depictive modifier, so these served as additional control items. See Appendix C for the two control examples of this type, (C.2) and (C.7).



(a) *Some but not all trolls found by the girl.*



(b) *No trolls found by the girl.*

Figure 4.14: *Two variants of a scene designed to go with size-and-shape modifier test items. Trolls hidden in the bushes to the left are not found by the girl.*

4.5.3 Results

In the survey, English consultants judged the **depictive** examples with negation false for both size-and-shape and manner-path modifiers (6/7 consultants), while ASL consultants judged the very same depictive examples to be true (4/4 consultants). The one English consultant who diverged from the others behaved exactly like the ASL consultants and judged all critical trials to be true. For the **non-depictive** negation examples, both English and ASL consultants judged them to be true for both types of modifiers (E:7/7, ASL:4/4). **Other factors** (type of negation, size-and-shape vs. manner-path, etc.) did not influence truth value judgments, but were sometimes reported to influence naturalness during debriefing sessions. See Table 4.1 for a breakdown of responses by trial type.

Table 4.1: *Truth value judgment results by trial, from 7 English consultants and 4 ASL consultants. T stands for "true" and F for "false". Example numbers refer to stimuli listed in Appendix C.*

Ex	Trial description	**T**(Eng)	**F**(Eng)	**T**(ASL)	**F**(ASL)
(C.1)	Depictive modifier, size-shape, *not*	1	6	4	0
(C.3)	Non-depictive modifier, size-shape, *not*	7	0	4	0
(C.4)	Depictive modifier, size-shape, *none*	1	6	4	0
(C.5)	Non-depictive modifier, size-shape, *none*	7	0	4	0
(C.9)	Depictive modifier, manner-path, *not*	1	6	4	0
(C.10)	Non-depictive modifier, manner-path, *not*	7	0	4	0
(C.6)	Depictive modifier, manner-path, *none*	2	5	4	0
(C.8)	Non-depictive modifier, manner-path, *none*	7	0	4	0
(C.12)	Depictive modifier, manner-path, on main verb	1	6	4	0
	Controls:				
(C.2)	Depictive modifier, size-shape, no trolls found	7	0	4	0
(C.7)	Depict. modifier, manner-path, no win. broken	7	0	4	0
(C.11)	No modifiers, manner-path	7	0	4	0

These results are rather striking. The very same utterances in the same contexts are judged to be false by English speakers, but true by ASL signers. That is to say, sign language depictive modifiers have *different truth conditions under negation* than depictive co-speech gestures do. We conclude that there does seem to be a difference in the ability of ASL classifier predicates to be negated vs. English co-speech gestures, suggesting that the inability of negation to target depictive content in English is due to the interaction of the

two different modalities rather than a compositional clash in the semantics.

4.6 Conclusion

This chapter compares depiction in speech with depiction in sign. Sign languages have complex verbal constructions called classifier predicates, which combine a discrete categorical handshape morpheme with an analog gestural morpheme that represents the relative position and/or movement of referents. This gestural component is interpreted gradiently and iconically, such that small changes in the position and/or movement of the hands in signing space are interpreted as meaningful and representing small changes in the relative positions/movements of referents. The depictive element of these constructions has been likened to depictive gesture accompanying speech. Here we test that hypothesis and directly compare the behavior of classifier constructions under negation in ASL with depictive co-speech gestures under negation in English.

The results show that in controlled contexts, co-speech gestures exhibit different truth conditions under negation than parallel ASL classifier constructions. Negation can target the content of classifier predicates in ASL, where both the negation and the classifier predicates are in the same mode, but cannot target gestural content in English, where they are in different modes. These different patterns lead us to conclude that analog/depictive content can be the target of operators like negation if it is in the same mode.

A natural prediction, then, would be that depictive content in the auditory modality can be targeted by spoken negation. Indeed, this seems to be true; consider (4.40).

(4.40) John did not give a *loooong* talk, he (just) gave a long talk.

The sentential negation *not* can target the extra-long interpretation of the iconically modified *loooong* and contrast it with normal *long*.

These results are especially interesting given some recent data on code blending in bimodal bilinguals — adults who speak both a sign and a spoken language natively. Code blending is when aspects of speech and sign are produced simultaneously (Emmorey et al.

2016), such as saying and signing the lexical item *cat*/CAT at the same time. Quadros et al. (2019) looked at occurrences of classifier predicates (which they refer to as "depicting signs") in spontaneous code-blended speech from hearing bimodal bilinguals, and found that that speech does sometimes co-occur with these constructions. Classifier predicates were blended with either a verb or with a vocal gesture. It is tempting to think of the case of blending with a verb as akin to using a depictive gesture with speech. However, this would mean that language treats gestures and classifier predicates the same way, which we have seen is not the case. It would be interesting to see how both types of code-blended classifier constructions would be interpreted under negation or other operators by bimodal bilinguals.

This study should serve as a warning for those of us studying speech, gesture, and sign. Goldin-Meadow and Brentari (2017) suggest that sign languages would perhaps be better compared to speech plus gesture rather than simply speech alone. That may be the case, but this study shows that a note of caution is in order. Depictive content in sign languages that appear to have a gestural component, like classifier predicate constructions, may not in fact be semantically equivalent to parallel speech plus gesture constructions. More study is needed on the effects of modality on the semantics of gesture and sign.

Chapter 5

Conclusion

5.1 Summary

Chapter 2 looked at the effect of context on the felicity of co-speech gestures. Participants recruited via Amazon Mechanical Turk filled out a Qualtrics survey in which they had to read a context paragraph, watch a short video, and then rate the felicity of the video utterance. Some of the participants saw videos with co-speech gestures. The context paragraphs were manipulated to either entail the content of a gesture or to not entail either it or its negation, utilizing the strong contextual felicity diagnostic proposed in Tonhauser et al. (2013). A second experimental factor manipulated whether the video utterance contained a speech cue, a linguistic expression that (approximately) duplicated the content of the co-speech gesture. We asked whether the positive context paragraphs that entailed the content of a gesture increased the acceptability rating of that gesture, and also whether the presence of a speech cue had any effect on the acceptability.

The results indicated that manipulating the context paragraphs had no effect on the felicity of the co-speech gesture. From this we concluded that co-speech gestures are not hard presupposition triggers requiring the common ground to entail them. If they are presuppositional, either co-speech gestures are easily accommodated or are some other kind of presupposition, such as a conditional presupposition ("cosupposition") as proposed

by Schlenker (2018a). (A followup survey confirmed that participants were indeed taking the context paragraphs into account in their acceptability ratings.) A second finding of the original study was that the presence of a speech cue significantly increased the acceptability of a co-speech gesture. We interpreted this to mean that co-speech gestures are preferentially truth-conditionally vacuous; they prefer to simply duplicate content already present in the speech stream. This was a point against the supplement analysis of co-speech gestures (Ebert and Ebert 2014, 2016), as supplements are typically infelicitous when they duplicate asserted content.

In Chapter 3, the experiment was motivated by a desire to separate semantic triviality of co-speech gestures from pragmatic informativity and to manipulate the latter. The study from Chapter 2 indicated that gestures are rated higher when they duplicate content in the speech stream. To disentangle this preference for triviality from informativity, we designed an experiment where the gestures had no close equivalents in speech, depicting hairstyles for toy trolls. Participants were asked to play a reference game in which they had to watch a short video and then choose which troll the video was asking for. Additionally they were asked to rate the naturalness of the video utterance. The video stimuli asked for trolls using hairstyle gestures, color words, or both. Depending on the differences between the target and distractor trolls, the gestures were either uninformative, informative, or critically informative for the reference identification task.

When both the color word and the gesture cue were informative, the gesture-only condition was rated significantly lower than the linguistic-cue only condition. Adding a gesture to a linguistic cue lowered the rating, except in the case when the gesture was critically informative. When either the linguistic cue or the gesture cue was critically informative, the highest mean rating was much higher in the linguistic cue case than in the gesture cue case. Additionally, participants were slightly but significantly more accurate in choosing the correct troll using a linguistic cue (color word) than a hairstyle gesture. Altogether these results show a bias toward expressing informative and critically informative information in speech rather than in gesture, when possible.

Chapter 4 compared co-speech gestures to classifier constructions in sign languages, specifically looking at the behavior of both under negation. Classifier predicates are complex morphosyntactic constructions that combine a discrete, categorical handshape with a gesture component that iconically depicts the relative locations and movements of referents. The depictive element of these constructions has been likened to depictive gesture accompanying speech. In an informal survey, several English speakers and American Sign Language (ASL) signers were recruited to look at simple scenes, watch short videos, and provide informal true or false judgments for the videos in the context of the scenes. There were two sets of scenes: one set tested size-and-shape-type gestures/classifier predicates, while the other tested manner-path-type gestures/classifier predicates. The sentences in the videos always embedded the gesture or classifier predicate under negation, either sentential or quantificational.

The results of the survey were surprising: co-speech gestures did not behave the same way under negation as classifier predicates did. The negation could target the content of the classifier predicate in the ASL examples, where both the negation and the classifier predicate are in the same mode, but could not target the co-speech gesture in the English sentences, where they are in different modes. We take this to mean that analog/depictive content can be the target of operators like negation if it is in the same mode. Once again, this is evidence that gesture does not contribute its meaning in the same way as language proper, whether it be speech or sign.

5.2 Discussion

In this thesis, we have seen that co-speech gestures prefer to duplicate information already present in speech, rather than contribute new information. They tend to be packaged as background information that can't directly be targeted by operators such as negation; we conclude from this that gesture does not fully integrate with linguistic material. Grammaticality judgments seem to be a lot fuzzier than for language proper, and, in a laboratory setting, examples with gestures tend to be rated lower than examples without, except when

the gesture is critical for the pragmatic task at hand.

Our findings indicate that gestures are not treated on a pragmatic par with speech and sign. In three very different behavioral studies, we saw the same thing again and again — interlocutors prefer important information to be conveyed via speech, not gesture. This is true even if gesture is the only good way to communicate the content, such as with the troll hairstyles in the experiment in Chapter 3. We saw this both in an acceptability ratings task and in a reference identification game. Moreover, this asymmetry between gesture and speech is seemingly not due to the fact that co-speech gestures are depictive; otherwise, we would not expect classifier predicates in ASL (also depictive constructions) to have different truth condition patterns under negation compared to depictive gestures in English.

Recent work in formal semantics and pragmatics (Schlenker 2018a,b; Esipova 2019b) treats the composition of co-speech gesture no differently than semantic composition in spoken language, but these three studies show that this isn't necessarily the case. In fact, one might go so far as to say that gestures aren't any more integrated with language than other contextual information, as there is little evidence of compositional integration in these three studies. Gestures can be clearly communicative, but that doesn't mean that they share the same kind of compositional properties as spoken language. The domain of dance, for instance, has been recently studied using formal linguistic tools (Charnavel 2019; Patel-Grosz and Grosz 2019), but that's quite different from claiming that dance composes with linguistic meaning in the usual way.

This work also raises the question, what should we be comparing co-speech gestures to? Gestures are often compared to close counterparts in speech, but in Chapter 2 we saw that gestures were rated significantly worse than their corresponding speech cues. Alternatively, gesture plus speech is sometimes compared to sign language (Goldin-Meadow and Brentari 2017); however, the study in Chapter 4 reveals that even close depictive counterparts in sign behave differently than combinations of speech and gesture. Chapter 3 demonstrated that gestures specifically designed to have no close speech equivalents do not behave like their speech equivalents in critically informative pragmatic contexts. In sum, gestures don't

work like linguistic modifiers, or like sign modifiers either. These results should serve as a cautionary note for linguistic researchers working on gesture and sign: whether because it is in a different mode than speech or for other as of yet unknown reasons, gesture seems to behave qualitatively differently than language.

5.3 Limitations

A limitation of the current work is that it does not include a production study. The studies in this dissertation had participants watching videos of utterances that contained gestures, and this methodology allowed us tight control over the experimental factors we wanted to test. Instead of guessing which factors affect gesture felicity and designing experiments around those, though, a production study could be a complimentary way to answer the question of what information speakers choose to put into gesture and would provide more naturally-occurring examples of co-speech gesture as a basis for future research.

A second potential limitation of the video methodology is that it is unclear if people know how to react to gestures in a laboratory setting. The lower and highly varied ratings for trials with gestures in the first study in Chapter 2 may indicate that pre-recorded videos with gestures strike participants as inherently unnatural in some way. Perhaps gestures are only seen as natural when they are perceived real-time and in person by interlocutors. More research on the naturalness of gesture is needed to answer this question.

5.4 Future directions

5.4.1 Co-speech gesture under contrastive focus

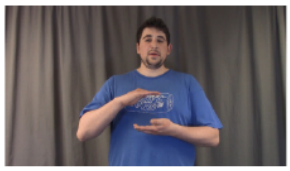
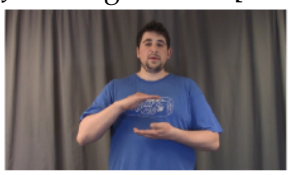
Future work could experimentally test more factors that might affect the felicity of co-speech gestures. One such factor, contrastive focus, has been recently tested by Esipova (2019a) using the methodology described in this dissertation.

Esipova (2019a) presents a study on the effects of contrastive focus on the acceptability of co-speech gestures. Examples like (5.1) were constructed with contrastive focus markers

on two complex word-gesture expressions, where the speech was the same in each but they differed in their gestures.

(5.1) John might order [a beer]_**SMALL**  or [a beer]_**LARGE**  .

In order for the contrastive focus to be felicitous in these examples, the co-speech gestures must be interpreted as at-issue, so that (5.1), for instance, approximately means that John might order a small beer or a large beer. Esipova (2019a) compares such at-issue co-speech gesture constructions with controls where the locus of variation is in the speech that accompanies the gesture rather than in the gesture, as in (5.2).

(5.2) John might order [**a beer**]_SMALL  or [**a cocktail**]_SMALL  .

In an online survey, participants were asked to watch short videos and rate the acceptability of utterances like (5.1) and (5.2). In addition to "scalar" gestures like SMALL and LARGE, shape tracing gestures were also tested. An additional factor of "emphasis" was tested, whereby sometimes the gestures were gesticulated more emphatically.

The main result of the study was that examples like (5.1) were rated significantly less acceptable than the control examples like (5.2). Esipova (2019a) takes this to mean that co-speech gestures by default make not-at-issue contributions, and it is costly to force at-issue readings. Patterns of variation at the individual level suggest further that co-speech

gestures cannot be uniformly treated as supplements, as proposed by Ebert and Ebert (2014). The results are, however, compatible with a cosupposition analysis of co-speech gesture (Schlenker 2018a). Neither manipulating the type of gesture (scalar or shape) nor the emphasis on the gesture made any significant difference in acceptability judgments.

5.4.2 Co-speech gestures and contrastive inference

As another way of further studying the pragmatics of gesture, a second future direction would be to look at whether co-speech gestures trigger contrastive inferences. Building on work by Alsop et al. (2018); Alsop (2018), we designed and ran a pilot study to take a preliminary look at this question.

This experiment was designed to test how native English speakers compute contrastive inferences using information expressed by depictive co-speech gestures compared to verbal non-depictive linguistic items. Our main experimental question was whether co-speech gestures can trigger contrastive inferences. In particular, we sought to understand how the inferences compare between co-speech gestures and other linguistic content known to trigger contrastive inferences, like color words. Since the content of co-speech gestures is not-at-issue, we would expect to have trouble contextually focusing it, and hence that participants would disprefer the gestural content compared to verbal content as a contrastive clue for reference identification.

In particular we sought to understand how interlocutors interpret gestures when the speaker's description also contains a non-depictive verbal color word. Do they ever interpret the gesture contrastively and not the color word? In which pragmatic contexts? Unfortunately, the results of the pilot study were inconclusive; there were no significant patterns in the data. Future work on this topic would be needed to determine why the pilot failed and to answer the original research questions.

5.5 Conclusion

Overall, this dissertation has found significant asymmetry between depictive gesture and language (spoken or signed). There is a pragmatic preference for gesture to be semantically trivial and to not convey pragmatically important information. From a cognitive point of view this is perhaps not too surprising, given that gestures have been shown to often aid the speaker in language production, as indicated by the fact that speakers gesture when their interlocutors are not visible (Iverson and Goldin-Meadow 1998; Bavelas et al. 2008). This is worth keeping in mind when looking at recent theoretical work (Ebert and Ebert 2014, 2016; Schlenker 2018a; Esipova 2019b) on co-speech gesture, which assumes that co-speech gestures can potentially contribute meaning in a similar way to linguistic content. The asymmetry between language and gesture casts doubt as to whether all of our normal linguistic tools work for analyzing gesture, including, for instance, grammaticality judgments. We should similarly be wary of assuming the standard Gricean principles are at play, as this might be a domain that is primarily speaker-driven.

References

Abner, Natasha, Kensy Cooperrider, and Susan Goldin-Meadow. 2015. Gesture for linguists: A handy primer. *Language and Linguistics Compass* 9:437–449. DOI: https://doi.org/10.1111/lnc3.12168.

Abrusán, Márta. 2016. Presupposition cancellation: Explaining the 'soft-hard' trigger distinction. *Natural Language Semantics* 24:165–202. DOI: https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/s11050-016-9122-7.

Alsop, Anna. 2018. An experimental investigation of modification and pragmatic contrast in English and Japanese. Honors Thesis, Harvard University, Cambridge, MA.

Alsop, Anna, Elaine Stranahan, and Kathryn Davidson. 2018. Testing contrastive inferences from suprasegmental features using offline measures. *Proceedings of the Linguistic Society of America* 3:71–115. DOI: 10.3765/plsa.v3i1.4386.

Aronoff, Mark, Irit Meir, Carol Padden, and Wendy Sandler. 2003. Classifier constructions and morphology in two sign languages. In *Perspectives on classifier constructions in sign languages*, ed. Karen Emmorey, 53–84. Mahwah, NJ: Lawrence Erlbaum Associates.

Asher, N., and A. Lascarides. 2003. *Logics of conversation*. Cambridge: Cambridge University Press.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48. DOI: https://dx.doi.org/10.18637/jss.v067.i01.

Bavelas, Janet, Jennifer Gerwing, Chantelle Sutton, and Danielle Prevost. 2008. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language* 58:495–520.

Beattie, G., and Heather Shovelton. 2006. When size really matters: How a single semantic feature is represented in the speech and gesture modalities. *Gesture* 6:63–84.

Benedicto, Elena, and Diane Brentari. 2004. Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory* 22:743–810.

Charnavel, Isabelle. 2019. Steps toward a universal grammar of dance: Local grouping structure in basic human movement perception. *Frontiers in Psychology* 10:1364. URL https://www.frontiersin.org/article/10.3389/fpsyg.2019.01364.

Chierchia, Gennaro, and Sally McConnell-Ginet. 2000. *Meaning and grammar: An introduction to semantics*. Cambridge, MA: MIT press.

Cooperrider, Kensy. 2017. Foreground gesture, background gesture. *Gesture* 16:176–202.

Davidson, Kathryn. 2015. Quotation, demonstration, and iconicity. *Linguistics and Philosophy* 38:477–520.

Ebert, Cornelia, and Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Slides from a talk given at *Semantics and Philosophy in Europe (SPE 7)*, ZAS, Berlin, June 2014. Available at https://semanticsarchive.net/Archive/GJjYzkwN.

Ebert, Cornelia, and Christian Ebert. 2016. The semantic behavior of co-speech gestures and their role in demonstrative reference. Slides from an LSCP LANGUAGE seminar talk at the Institut Jean-Nicod, Département d'Études Cognitives, École Normale Supérieure, Paris, Dec. 2016. Available at http://www.cow-electric.com/neli/talks/CE-Paris-2016.pdf.

Emmorey, Karen. 2002. *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Emmorey, Karen, Marcel R. Giezen, and Tamar H. Gollan. 2016. Psycholinguistic, cognitive, and neural implications of bimodal bilingualism. *Bilingualism: Language and Cognition* 19:223–242.

Emmorey, Karen, and Melissa Herzig. 2003. Categorical versus gradient properties of classifier constructions in ASL. In *Perspectives on classifier constructions in signed languages*, ed. Karen Emmorey, 222–246. Mahwah, NJ: Lawrence Erlbaum Associates.

Engberg-Pedersen, Elisabeth. 1993. *Space in Danish Sign Language*. Hamburg: Signum Verlag.

Esipova, Maria. 2018a. Composition and projection of adnominal content across modalities. Ms. Available at https://ling.auf.net/lingbuzz/004377.

Esipova, Maria. 2018b. Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive focus. In *Proceedings of sinn und bedeutung (sub) 22*, ed. Uli Sauerland and Stephanie Solt. Berlin: Leibniz-Centre General Linguistics.

Esipova, Maria. 2019a. Acceptability of at-issue co-speech gestures under contrastive focus. *Glossa: a journal of general linguistics.* 4:11. DOI: 10.5334/gjgl.635.

Esipova, Maria. 2019b. Composition and projection in speech and gesture. Doctoral Dissertation, New York University.

von Fintel, Kai. 2004. Would you believe it? The king of France is back! Presuppositions and truth value intuitions. In *Descriptions and beyond*, ed. A. Bezuidenhout and M. Reimer, 315–341. Oxford: Oxford University Press.

Goldin-Meadow, Susan. 2003. *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap.

Goldin-Meadow, Susan, and Diane Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* 40:e46. DOI: https://doi.org/10.1017/S0140525X15001247.

Gonzalez, Aurore, Kate Henninger, and Kathryn Davidson. to appear. Answering negative questions in American Sign Language. NELS Proceedings.

Henderson, R. 2016. A demonstration-based account of (pluractional) ideophones. In *Proceedings of Semantics and Linguistic Theory (SALT) 26*, ed. M. Maroney, C.-R. Little, J. Collard, and D. Burdorf, 664–683. LSA.

Holler, Judith, and R. Stevens. 2007. The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology* 26:4–27.

Iverson, Jana M, and Susan Goldin-Meadow. 1998. Why people gesture when they speak. *Nature* 396:228.

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In *The relationship of verbal and nonverbal communication*, ed. Mary Ritchie Key, 207–227. The Hague: Mouton.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, Sotaro, and Aslı Özyürek. 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* 48:16–32. DOI: https://doi-org.ezp-prod1.hul.harvard.edu/10.1016/S0749-596X(02)00505-3.

Klima, Edward S, and Ursula Bellugi. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.

Kuhn, Jeremy, and Valentina Aristodemo. 2017. Pluractionality, iconicity, and scope in French Sign Language. *Semantics and Pragmatics* 10. DOI: http://dx.doi.org/10.3765/sp.10.6.

Lascarides, Alex, and Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics* 26:393–449. DOI: https://doi.org/10.1093/jos/ffp004.

Liddell, Scott, and Robert Johnson. 1987. An analysis of spatial-locative predicates in American Sign Language. Paper presented at 4th International Symposium on Sign Language Research, Lappeenranta, Finland.

Liddell, Scott K. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.

McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review* 92:350–371. DOI: http://dx.doi.org.ezp-prod1.hul.harvard.edu/10.1037/0033-295X.92.3.350.

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

McNeill, David. 2005. *Gesture and thought*. Chicago: University of Chicago Press.

Meir, Irit. 2010. Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language* 86:865–896.

Özyürek, Aslı. 2017. Function and processing of gesture in the context of language. In *Why gesture? How the hands function in speaking, thinking and communicating*, ed. R. Breckinridge Church, Martha W. Alibali, and Spencer D. Kelly, 39–58. Amsterdam: John Benjamins Publishing Company.

Patel-Grosz, Pritty, and Patrick Grosz. 2019. Steps towards a formal semantics of dance. Ms.

Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.

Quadros, Ronice, Kathryn Davidson, Diane Lillo-Martin, and Karen Emmorey. 2019. Code-blending with Depicting Signs. To appear in *Linguistic Approaches to Bilingualism*.

Quer, Josep. 2012. Negation. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 316–339. Berlin: De Gruyter Mouton.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roberts, Craige. 1996/2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5:1–69. DOI: http://dx.doi.org/10.3765/sp.5.6.

Sandler, Wendy, and Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.

Schlenker, Philippe. 2011. Iconic agreement. *Theoretical Linguistics* 37:223–234.

Schlenker, Philippe. 2018a. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41:295–365. DOI: https://doi.org/10.1007/s10988-017-9225-8.

Schlenker, Philippe. 2018b. Iconic pragmatics. *Natural Language and Linguistic Theory* 36:877–936.

Schlenker, Philippe. 2018c. Sign language semantics: Problems and prospects. *Theoretical Linguistics* 44:295–353.

Schlenker, Philippe. 2018d. Visible meaning: Sign language and the foundations of semantics. *Theoretical Linguistics* 44:123–208.

Schlenker, Philippe. to appear . Iconic presuppositions. Ms. Available at `https://ling.auf.net/lingbuzz/004116`.

Schlenker, Philippe, Jonathan Lamberton, and Mirko Santoro. 2013. Iconic variables. *Linguistics and Philosophy* 36:91–149. DOI: `https://doi.org/10.1007/s10988-013-9129-1`.

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32:3–23.

Shanon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language* 14:247–249.

Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. What projects and why. In *Proceedings of Semantics and Linguistic Theory (SALT) 20*, ed. N. Li and D. Lutz, 309–327. Ithaca, NY: CLC Publications.

Stokoe, William C. 1960. Sign language structure: An outline of the visual communication systems of the American deaf. In *Studies in linguistics: Occasional papers*, volume 8. Buffalo, NY: University of Buffalo.

Supalla, Ted. 1982. Structure and acquisition of verbs of motion and location in american sign language. Doctoral Dissertation, University of California, San Diego.

Taub, Sarah F. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Tieu, Lyn, Robert Pasternak, Philippe Schlenker, and Emmanuel Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: a journal of general linguistics* 2:1–27. DOI: `https://doi.org/10.5334/gjgl.334`.

Tieu, Lyn, Robert Pasternak, Philippe Schlenker, and Emmanuel Chemla. 2018. Co-speech gesture projection: Evidence from inferential judgments. *Glossa: a journal of general linguistics* 3:1–21.

Tonhauser, Judith. 2012. Diagnosing (not-)at-issue content. In *Proceedings of Semantics of Under-represented Languages in the Americas (SULA) 6*, 239–254.

Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89:66–109. DOI: `https://doi.org/10.1353/lan.2013.0001`.

Wood, Sandra K. 1999. Semantic and syntactic aspects of negation in ASL. Master's thesis, Purdue University.

Zeshan, Ulrike. 2004. Hand, head, and face: Negative constructions in sign languages. *Linguistic Typology* 8:1–58.

Zeshan, Ulrike. 2006. *Interrogative and negative constructions in sign languages*. Nijmegen: Ishara Press.

Zlogar, C., A. Alsop, and K. Davidson. 2018. Description and depiction in a reference game. *Frontiers in Psychology* Conference abstract: XPRAG.it 2018 in Pavia, Italy. doi: 10.3389/conf.fpsyg.2018.73.00015.

Zlogar, Christina, and Kathryn Davidson. 2018. Effects of linguistic context on the acceptability of co-speech gestures. *Glossa: a journal of general linguistics* 3:73. 1–28. DOI: https://doi.org/10.5334/gjgl.438.

Zlogar, Christina, Kate Henninger, and Kathryn Davidson. 2019. Negating depictive modifiers in sign and speech. Poster presented at LSA 2019, New York, NY, Jan. 5,2019.

Zucchi, S., C. Cecchetto, and C. Geraci. 2011. Event descriptions and classifier predicates in sign languages. In *Presentation at the meeting of the Venice FEAST Conference, Venice, Italy, 21 June 2011*.

Zucchi, Sandro. 2012. Formal semantics of sign languages. *Language and Linguistics Compass* 6:719–734.

Zwitserlood, Inge. 2012. Classifiers. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 158–185. Berlin: De Gruyter Mouton.

Appendix A

Appendix for Chapter 2

A.1 Stimuli for Experiment 1

This appendix contains all the stimuli (video transcript and written text) used for Experiment 1.

We use the following conventions: gestures will be written in all-capital letters, and we will indicate the spoken words that align with the gesture by placing them in square brackets. In the positive contexts, material which has been changed from or added to the corresponding neutral contexts has been bolded. By *proposition* we mean the inference licensed by the positive context that conveys the semantic content contributed by the co-speech gesture, and also by its corresponding speech cue (noting that the two are not strictly equivalent given the iconic, analog nature of the co-speech gesture).

(1) **Scenario**: Boy in treehouse

- *Neutral context*: Billy is sad because he's not with his friends. He starts to cry and his dad goes over to him. Their neighbor Eliza watches and then tells her friend Sarah what happened. Eliza says:
- *Positive context*: Billy is sad because he's not **up in the treehouse** with his friends. He starts to cry and his dad goes over to him. Their neighbor Eliza watches and then tells her friend Sarah what happened. Eliza says:

- UP *proposition*: The helping was done by lifting the son up.[1]

a. John [helped]_UP his son and now he's not crying anymore.

b. John [helped]_UP his son get up into the treehouse and now he's not crying anymore.

c. John helped his son and now he's not crying anymore.

d. John helped his son get up into the treehouse and now he's not crying anymore.

(2) **Scenario**: Favorite athlete

- *Neutral context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new athlete who Tom hasn't heard of. Eliza says:
- *Positive context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **basketball player** who Tom hasn't heard of. Eliza says:
- SHOOT *proposition*: The scoring was done by shooting a basketball.

a. The match she was in last night was incredible! She [scored]_SHOOT the winning point!

b. The basketball match she was in last night was incredible! She [scored]_SHOOT the winning point!

c. The match she was in last night was incredible! She scored the winning point!

d. The basketball match she was in last night was incredible! She scored the winning point!

(3) **Scenario**: Pet dog

---

[1]Example (1a) is modeled on an example in Schlenker (2018a:3) (rewritten in (i) using our notation):

(i) John [helped]_UP his son.
⇒ John helped his son by lifting him. (*cosupposition*)

Note that Schlenker writes the cosupposition (inference) triggered by (i) differently than we did above by including *John*, the subject of the VP, in the inference. To us, it seemed more instinctively natural to not include the subject of a VP modified by a co-speech gesture in its propositional content. However, it is not clear to us what the ramifications of this choice point are, and so we leave it to future research.

- *Neutral context*: Eliza and Emma are talking about their friend Sandy who is passionate about adopting animals. Eliza says:
- *Positive context*: Eliza and Emma are talking about their friend Sandy who is passionate about adopting **Great Danes**. Eliza says:
- BIG *proposition*: The dog was big.

a. Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

b. Sandy just got [a big dog]_BIG yesterday, and I hear it's quite the handful!

c. Sandy just got a dog yesterday, and I hear it's quite the handful!

d. Sandy just got a big dog yesterday, and I hear it's quite the handful!

(4) **Scenario**: Library disturbance

- *Neutral context*: Eliza is talking to Martin about her friend Karen, who was studying in the library when she suddenly heard a commotion erupt in the distance. Martin asks what happened next, and Eliza says:
- *Positive context*: Eliza is talking to Martin about her friend Karen, who was studying in the library when she suddenly heard a commotion erupt **from a floor below**. Martin asks what happened next, and Eliza says:
- RUN-DOWN *proposition*: The running was in a downward direction.

a. Karen [ran]_RUN-DOWN to see what was happening!

b. Karen [ran down]_RUN-DOWN to see what was happening!

c. Karen ran to see what was happening!

d. Karen ran down to see what was happening!

(5) **Scenario**: Writing

- *Neutral context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her there, comes back to Jamie, and says:

- *Positive context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her there **on her computer**, comes back to Jamie, and says:
- TYPE *proposition*: The writing was done by typing on a computer.

a. I saw Julia over in the library [writing an essay]_TYPE — it looks like she's a little preoccupied right now.

b. I saw Julia over in the library [typing out an essay]_TYPE — it looks like she's a little preoccupied right now.

c. I saw Julia over in the library writing an essay — it looks like she's a little preoccupied right now.

d. I saw Julia over in the library typing out an essay — it looks like she's a little preoccupied right now.

(6) **Scenario**: The moon

- *Neutral context*: Bryan and Eliza are discussing how they spent their Sunday nights with their families. Eliza says:
- *Positive context*: Bryan and Eliza are discussing how they spent their Sunday nights with their families. **Bryan found out that Eliza just got a new telescope.** Eliza says:
- TELESCOPE *proposition*: The looking was done using a telescope.

a. The moon was so gorgeous last night — we just sat outside [looking up]_TELESCOPE at it for awhile.

b. The moon was so gorgeous last night — we just sat outside and took turns with the telescope [looking up]_TELESCOPE at it for awhile.

c. The moon was so gorgeous last night — we just sat outside looking up at it for awhile.

d. The moon was so gorgeous last night — we just sat outside and took turns with the telescope looking up at it for awhile.

(7) **Scenario**: Check the time

- *Neutral context*: Eliza and Sophie are talking about Eliza's recent date with Alex. Eliza complains:
- *Positive context*: Eliza and Sophie are talking about Eliza's recent date with Alex, **who wore an expensive wristwatch**. Eliza complains:
- WRISTWATCH *proposition*: The time-checking was done by looking at a wristwatch.

a. Alex kept [checking the time]_WRISTWATCH during the date.

b. Alex kept [checking the time]_WRISTWATCH on his watch during the date.

c. Alex kept checking the time during the date.

d. Alex kept checking the time on his watch during the date.

(8) **Scenario**: Cleaning

- *Neutral context*: Eliza and Mary are talking about their roommate Tom, who never seems to help with the housework. Mary says that Tom doesn't know how to clean anything, but then Eliza reminds her:
- *Positive context*: Eliza and Mary are talking about their roommate Tom, who never seems to help with the housework. Mary says that Tom doesn't know how to **vacuum**, but then Eliza reminds her:
- VACUUM *proposition*: The floor-cleaning was done by vacuuming.

a. Last Mother's Day, Tom surprised his mom by cleaning the floorVACUUM without complaining.

b. Last Mother's Day, Tom surprised his mom by vacuuming the floorVACUUM without complaining.

c. Last Mother's Day, Tom surprised his mom by cleaning the floor without complaining.

d. Last Mother's Day, Tom surprised his mom by vacuuming the floor without complaining.

(9) **Scenario**: Jewelry

- *Neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money. Eliza agrees and says:
- *Positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money **based on her new pair of earrings**. Eliza agrees and says:
- EARRING *proposition*: The type of jewelry was earrings.

a. Alicia was wearing real diamond [jewelry]_EARRING at work this morning.

b. Alicia was wearing real diamond [earrings]_EARRING at work this morning.

c. Alicia was wearing real diamond jewelry at work this morning.

d. Alicia was wearing real diamond earrings at work this morning.

(10) **Scenario**: Violin concert

- *Neutral context*: Eliza and Jacob are talking about their niece's recital from the night before. Eliza says:
- *Positive context*: Eliza and Jacob are talking about their niece's **violin** recital from the night before. Eliza says:
- VIOLIN *proposition*: The performing was done on a violin.

a. Lisa [performed]_VIOLIN really well at the recital last night!

b. Lisa [performed]_VIOLIN really well at the violin recital last night!

c. Lisa performed really well at the recital last night!

d. Lisa performed really well at the violin recital last night!

(11) **Scenario**: Backpack (Note: excluded from Exp. 1 data analysis due to name mismatch: *Alicia* versus *Stephanie*)

- *Neutral context*: Eliza and James are looking for their friend Stephanie at the library. James says that he needs some help carrying all the books he checked out. Eliza replies:
- *Positive context*: Eliza and James are looking for their friend Stephanie at the library. **They recognize her by her new backpack.** James says that he needs some help carrying all the books he checked out. Eliza replies:
- BACKPACK *proposition*: The bag is (more specifically) a backpack.

a. I don't think Alicia can help — it looks like her bagBACKPACK is super heavy.

b. I don't think Alicia can help — it looks like [her backpack]_BACKPACK is super heavy.

c. I don't think Alicia can help — it looks like her bag is super heavy.

d. I don't think Alicia can help — it looks like her backpack is super heavy.

(12) **Scenario**: Hat tip

- *Neutral context*: Eliza just moved to a new town, and is telling her friend Sue about her friendly neighbor, an older gentleman. She says:
- *Positive context*: Eliza just moved to a new town, and is telling her friend Sue about her friendly neighbor, an older gentleman **who tips his hat to people walking by**. She says:
- HAT-TIP *proposition*: The greeting was done by tipping a hat.

a. He [greeted]_HAT-TIP us as he went to get the paper this morning.

b. He [tipped his hat]_HAT-TIP to us as he went to get the paper this morning.

c. He greeted us as he went to get the paper this morning.

d. He tipped his hat to us as he went to get the paper this morning.

(13) **Scenario**: Gambling

- *Neutral context*: Eliza and Louis are talking about how much fun they had on a recent trip to the casino with their friend Valerie. They all started with $200. Louis asks Eliza how Valerie did overall, and Eliza replies:
- *Positive context*: Eliza and Louis are talking about how much fun they had on a recent trip to the casino with their friend Valerie. They all started with $200. **Valerie preferred the tables with dice games.** Louis asks Eliza how Valerie did overall, and Eliza replies:
- DICE *proposition*: The gambling was done using dice/by playing a dice game.

a. Valerie [gambled]_DICE all her money away at the casino last weekend!

b. Valerie [gambled]_DICE all her money away at the dice game last weekend!

c. Valerie gambled all her money away at the casino last weekend!

d. Valerie gambled all her money away at the dice game last weekend!

(14) **Scenario**: Working out at the gym

- *Neutral context*: Eliza and Paul are discussing their friend Melissa's recent efforts to get in shape. Eliza says:
- *Positive context*: Eliza and Paul are discussing their friend Melissa's recent efforts to **build up her muscle tone**. Eliza says:
- FREE-WEIGHTS *proposition*: The working out was done by lifting weights.

a. Melissa goes to the gym three times a week to work outFREE-WEIGHTS.

b. Melissa goes to the gym three times a week to weight liftFREE-WEIGHTS.

c. Melissa goes to the gym three times a week to work out.

d. Melissa goes to the gym three times a week to weight lift.

(15) **Scenario**: Making soup

- *Neutral context*: Eliza is telling Hannah about their friend Derek's recent unsatisfying culinary experience. Derek hates to be interrupted while cooking, especially when he's at the stove. Eliza says:
- *Positive context*: Eliza is telling Hannah about their friend Derek's recent unsatisfying culinary experience. Derek hates to be interrupted while cooking, especially when he's **stirring a pot** at the stove. Eliza says:
- STIRRING *proposition*: The soup-making was done by stirring.

a. Derek was [making soup]_STIRRING when the doorbell rang.

b. Derek was [stirring the soup]_STIRRING when the doorbell rang.

c. Derek was making soup when the doorbell rang.

d. Derek was stirring the soup when the doorbell rang.

(16) **Scenario**: Cold weather

- *Neutral context*: Eliza and Ethan are talking about the recent snowstorm and how even their friend Josh, who usually never gets cold, had to put on his winter clothing. Eliza says:
- *Positive context*: Eliza and Ethan are talking about the recent snowstorm and how even their friend Josh, who usually never gets cold, had to put on his winter **hat and scarf**. Eliza says:
- WRAP-SCARF *proposition*: The bundling-up was done by wrapping a scarf around his neck.

a. Josh [bundled up]_WRAP-SCARF before going out to shovel the driveway.

b. Josh [bundled up]_WRAP-SCARF with a scarf before going out to shovel the driveway.

c. Josh bundled up before going out to shovel the driveway.

d. Josh bundled up with a scarf before going out to shovel the driveway.

(17) **Attention checks**

(A) Hi, this is an attention check — click “completely natural” and go on to the next question!

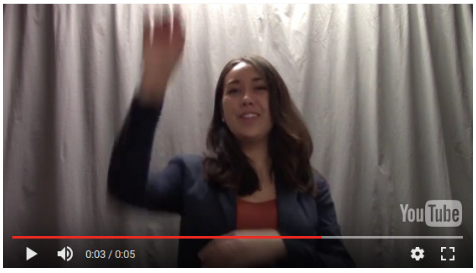
(B) Hi, this is an attention check — click “not completely natural” and go on to the next question!

(C) Hi, this is an attention check — click “not completely natural” and go on to the next question!

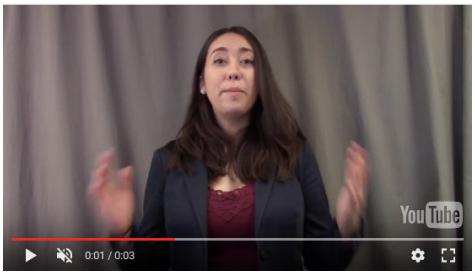
A.2 Gesture screencaps

This appendix shows screencaps for all 16 gestures used in the videos for Experiments 1 and 2.

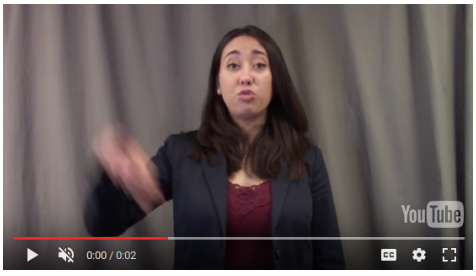
Scenario 1: UP



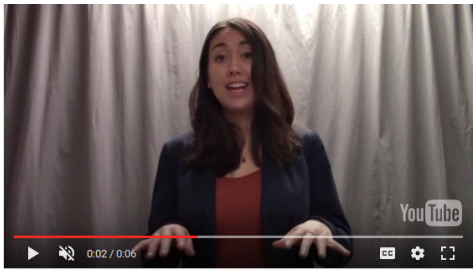
Scenario 2: SHOOT



Scenario 3: BIG

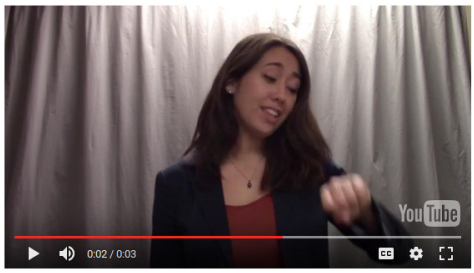


Scenario 4: RUN-DOWN

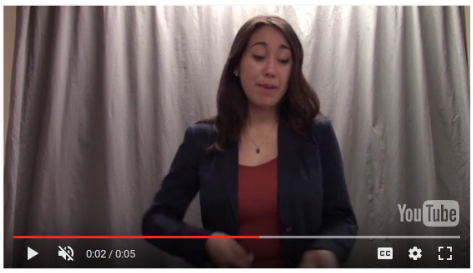


Scenario 5: TYPE

Scenario 6: TELESCOPE



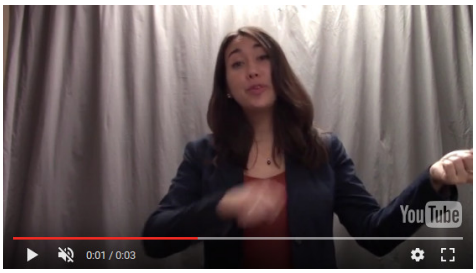
Scenario 7: WRISTWATCH



Scenario 8: VACUUM

Figure A.1: *Screencaps of co-speech gestures: Scenarios 1–8.*

Scenario 9: EARRING



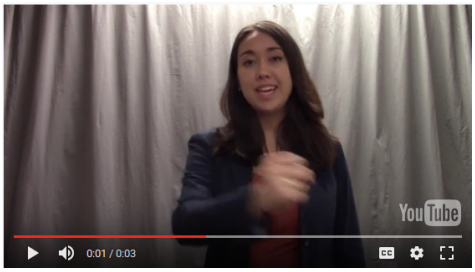
Scenario 10: VIOLIN

Scenario 11: BACKPACK

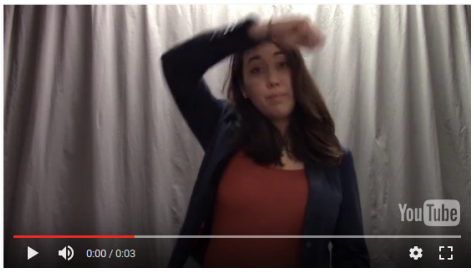
Scenario 12: HAT-TIP

Scenario 13: DICE

Scenario 14: FREE-WEIGHTS



Scenario 15: STIRRING



Scenario 16: WRAP-SCARF

Figure A.2: *Screencaps of co-speech gestures: Scenarios 9–16.*

A.3 Stimuli for Experiment 2

This appendix contains all the stimuli (video transcript and written text) used for Experiment 2.

We use the following conventions: gestures will be written in all-capital letters, and we will indicate the spoken words that align with the gesture by placing them in square brackets. Minimal differences between context types are bolded. By *proposition* we mean the inference licensed by the positive context that conveys the semantic content contributed by the co-speech gesture, and also by its corresponding speech cue (noting that the two are not strictly equivalent given the iconic, analog nature of the co-speech gesture).

(1) **Scenario**: Boy in treehouse

- *Neutral context*: Billy is sad because he's not **with his friends**. He starts to cry and his dad John goes over to him. Their neighbor Eliza watches and then tells her friend Sarah what happened. Eliza says:
- *Positive context*: Billy is sad because he's not **up in the treehouse with his friends**. He starts to cry and his dad John goes over to him. Their neighbor Eliza watches and then tells her friend Sarah what happened. Eliza says:
- *Infelicitous context*: Billy is sad because he's not with his friends. He starts to cry and his dad John goes over to him. Their neighbor Eliza watches and then **says to Billy's dad**:
- UP *proposition*: The helping was done by lifting the son up.

a. John [helped]_UP his son and now he's not crying anymore.

b. John helped his son get up into the treehouse and now he's not crying anymore.

(2) **Scenario**: Favorite athlete

- *Neutral context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **athlete** who Tom hasn't heard of. Eliza says:

- *Positive context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **basketball player** who Tom hasn't heard of. Eliza says:
- *Infelicitous context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her new favorite **team she's rooting for**. Eliza says:
- SHOOT *proposition*: The scoring was done by shooting a basketball.

a. The match she was in last night was incredible! She [scored]_SHOOT the winning point!

b. The basketball match she was in last night was incredible! She scored the winning point!

(3) **Scenario**: Pet dog

- *Neutral context*: Eliza and Emma are talking about their friend Sandy who is passionate about adopting **animals**. Eliza says:
- *Positive context*: Eliza and Emma are talking about their friend Sandy who is passionate about adopting **Great Danes**. Eliza says:
- *Infelicitous context*: Eliza and Emma are talking **to** their friend Sandy who is passionate about adopting animals. Eliza says:
- BIG *proposition*: The dog was big.

a. Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

b. Sandy just got a big dog yesterday, and I hear it's quite the handful!

(4) **Scenario**: Library disturbance

- *Neutral context*: Eliza is talking to Martin about her friend Karen, who was studying in the library when she suddenly heard a commotion erupt **in the distance**. Martin asks what happened next, and Eliza says:

- *Positive context*: Eliza is talking to Martin about her friend Karen, who was studying in the library when she suddenly heard a commotion erupt **from a floor below**. Martin asks what happened next, and Eliza says:
- *Infelicitous context*: Eliza is talking to **Karen** about her friend **Martin**, who was studying in the library when **he** suddenly heard a commotion erupt in the distance. **Karen** asks what happened next, and Eliza says:
- RUN-DOWN *proposition*: The running was in a downward direction.

a. Karen [ran]_RUN-DOWN to see what was happening!

b. Karen ran down to see what was happening!

(5) **Scenario**: Writing

- *Neutral context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there**, comes back to Jamie, and says:
- *Positive context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there on her computer**, comes back to Jamie, and says:
- *Infelicitous context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza **doesn't see her there**, comes back to Jamie, and says:
- TYPE *proposition*: The writing was done by typing on a computer.

a. I saw Julia over in the library [writing an essay]_TYPE — it looks like she's a little preoccupied right now.

b. I saw Julia over in the library typing out an essay — it looks like she's a little preoccupied right now.

(6) **Scenario**: The moon

- *Neutral context*: Bryan and Eliza are discussing how they spent their Sunday nights with their families. Eliza says:
- *Positive context*: Bryan and Eliza are discussing how they spent their Sunday nights with their families. **Bryan found out that Eliza just got a new telescope.** Eliza says:
- *Infelicitous context*: Bryan and Eliza are discussing how they spent their Sunday **mornings** with their families. Eliza says:
- TELESCOPE *proposition*: The looking was done using a telescope.

a. The moon was so gorgeous last night — we just sat outside [looking up]_TELESCOPE at it for awhile.

b. The moon was so gorgeous last night — we just sat outside and took turns with the telescope looking up at it for awhile.

(7) **Scenario**: Check the time

- *Neutral context*: Eliza and Sophie are talking about Eliza's recent date **with Alex**. Eliza complains:
- *Positive context*: Eliza and Sophie are talking about Eliza's recent date **with Alex, who wore an expensive wristwatch**. Eliza complains:
- *Infelicitous context*: Eliza and **Alex** are talking about Eliza's recent date with **Sophie**. Eliza complains:
- WRISTWATCH *proposition*: The time-checking was done by looking at a wristwatch.

a. Alex kept [checking the time]_WRISTWATCH during the date.

b. Alex kept checking the time on his watch during the date.

(8) **Scenario**: Cleaning

- *Neutral context*: Eliza and Mary are talking about their roommate Tom, who never seems to help with the housework. Mary says that Tom doesn't know how **to clean anything**, but then Eliza reminds her:
- *Positive context*: Eliza and Mary are talking about their roommate Tom, who never seems to help with the housework. Mary says that Tom doesn't know how **to vacuum**, but then Eliza reminds her:
- *Infelicitous context*: Eliza and **Tom** are talking about their roommate **Mary**, who never seems to help with the housework. **Tom** says that **Mary** doesn't know how to clean anything, but then Eliza reminds **him**:
- VACUUM *proposition*: The floor-cleaning was done by vacuuming.

a. Last Mother's Day, Tom surprised his mom by cleaning the floorVACUUM without complaining.

b. Last Mother's Day, Tom surprised his mom by vacuuming the floor without complaining.

(9) **Scenario**: Jewelry

- *Neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has **a lot of money**. Eliza agrees and says:
- *Positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has **a lot of money based on her new pair of earrings**. Eliza agrees and says:
- *Infelicitous context*: Eliza and **Alicia** are gossiping about their coworker **Nina**, and **Alicia** says that she thinks **Nina** has a lot of money. Eliza agrees and says:
- EARRING *proposition*: The type of jewelry was earrings.

a. Alicia was wearing real diamond [jewelry]_EARRING at work this morning.

b. Alicia was wearing real diamond earrings at work this morning.

(10) **Scenario**: Violin concert

- *Neutral context*: Eliza and Jacob are talking about their niece Lisa's **recital** from the night before. Eliza says:
- *Positive context*: Eliza and Jacob are talking about their niece Lisa's **violin recital** from the night before. Eliza says:
- *Infelicitous context*: Eliza and **Lisa** are talking about their **nephew Jacob**'s recital from the night before. Eliza says:
- VIOLIN *proposition*: The performing was done on a violin.

a. Lisa [performed]_VIOLIN really well at the recital last night!

b. Lisa performed really well at the violin recital last night!

(11) **Scenario**: Backpack

- *Neutral context*: Eliza and James are looking for their friend Alicia at the library. James says that he needs some help carrying all the books he checked out. Eliza replies:
- *Positive context*: Eliza and James are looking for their friend Alicia at the library. **They recognize her by her new backpack.** James says that he needs some help carrying all the books he checked out. Eliza replies:
- *Infelicitous context*: Eliza and **Alicia** are looking for their friend **James** at the library. **Alicia** says that **she** needs some help carrying all the books **she** checked out. Eliza replies:
- BACKPACK *proposition*: The bag is (more specifically) a backpack.

a. I don't think Alicia can help — it looks like her bagBACKPACK is super heavy.

b. I don't think Alicia can help — it looks like her backpack is super heavy.

(12) **Scenario**: Hat tip

- *Neutral context*: Eliza just moved to a new town, and is telling her friend Sue about her friendly neighbor, **an older gentleman**. She says:
- *Positive context*: Eliza just moved to a new town, and is telling her friend Sue about her friendly neighbor, **an older gentleman who tips his hat to people walking by**. She says:
- *Infelicitous context*: Eliza just moved to a new town, and is telling **her new neighbor, an older gentleman**, about her **best friend Sue**. She says:
- HAT-TIP *proposition*: The greeting was done by tipping a hat.

a. He [greeted]_HAT-TIP us as he went to get the paper this morning.

b. He tipped his hat to us as he went to get the paper this morning.

(13) **Scenario**: Gambling

- *Neutral context*: Eliza and Louis are talking about how much fun they had on a recent trip to the casino with their friend Valerie. They all started with $200. Louis asks Eliza how Valerie did overall, and Eliza replies:
- *Positive context*: Eliza and Louis are talking about how much fun they had on a recent trip to the casino with their friend Valerie. They all started with $200. **Valerie preferred the tables with dice games.** Louis asks Eliza how Valerie did overall, and Eliza replies:
- *Infelicitous context*: Eliza and **Valerie** are talking about how much fun they had on a recent trip to the casino with their friend **Louis**. They all started with $200. **Valerie** asks Eliza how **Louis** did overall, and Eliza replies:
- DICE *proposition*: The gambling was done using dice/by playing a dice game.

a. Valerie [gambled]_DICE all her money away at the casino last weekend!

b. Valerie gambled all her money away at the dice game last weekend!

(14) **Scenario**: Working out at the gym

- *Neutral context*: Eliza and Paul are discussing their friend Melissa's recent efforts **to get in shape**. Eliza says:
- *Positive context*: Eliza and Paul are discussing their friend Melissa's recent efforts **to build up her muscle tone**. Eliza says:
- *Infelicitous context*: Eliza and **Melissa** are discussing their friend **Paul**'s recent efforts to get in shape. Eliza says:
- FREE-WEIGHTS *proposition*: The working out was done by lifting weights.

a. Melissa goes to the gym three times a week to work outFREE-WEIGHTS.

b. Melissa goes to the gym three times a week to weight lift.

(15) **Scenario**: Making soup

- *Neutral context*: Eliza is telling Hannah about their friend Derek's recent unsatisfying culinary experience. Derek hates to be interrupted while cooking, especially **when he's at the stove**. Eliza says:
- *Positive context*: Eliza is telling Hannah about their friend Derek's recent unsatisfying culinary experience. Derek hates to be interrupted while cooking, especially **when he's stirring a pot at the stove**. Eliza says:
- *Infelicitous context*: Eliza is telling **Derek** about their friend **Hannah**'s recent unsatisfying culinary experience. **Hannah** hates to be interrupted while cooking, especially when **she**'s at the stove. Eliza says:
- STIRRING *proposition*: The soup-making was done by stirring.

a. Derek was [making soup]_STIRRING when the doorbell rang.

b. Derek was stirring the soup when the doorbell rang.

(16) **Scenario**: Cold weather

- *Neutral context*: Eliza and Ethan are talking about the recent snowstorm and how even their friend Josh, who usually never gets cold, had to put on **his winter clothing**. Eliza says:
- *Positive context*: Eliza and Ethan are talking about the recent snowstorm and how even their friend Josh, who usually never gets cold, had to put on **his winter hat and scarf**. Eliza says:
- *Infelicitous context*: Eliza and **Josh** are talking about the recent snowstorm and how even their friend **Ethan**, who usually never gets cold, had to put on his winter clothing. Eliza says:
- WRAP-SCARF *proposition*: The bundling-up was done by wrapping a scarf around his neck.

a. Josh [bundled up]_WRAP-SCARF before going out to shovel the driveway.

b. Josh bundled up with a scarf before going out to shovel the driveway.

(17) **Attention checks**

(A) Hi, this is an attention check — click "completely natural" and go on to the next question!

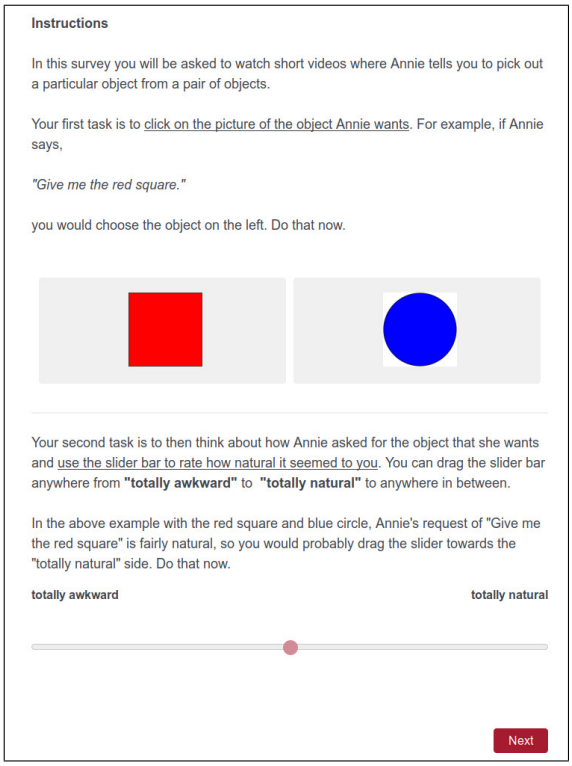
(B) Hi, this is an attention check — click "not completely natural" and go on to the next question!

(C) Hi, this is an attention check — click "not completely natural" and go on to the next question!

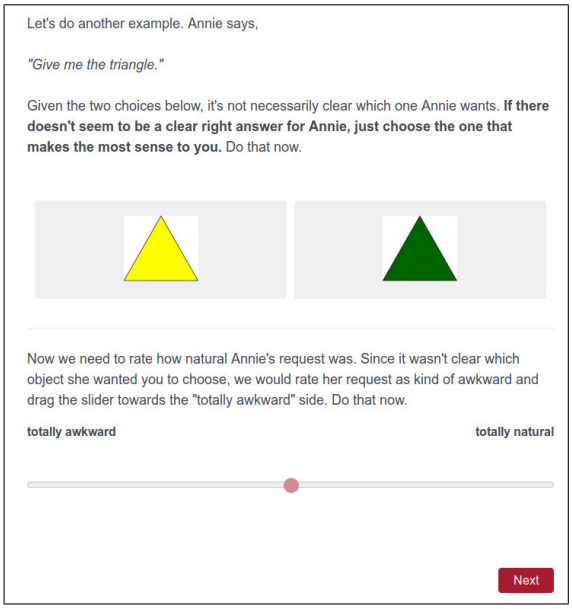
Appendix B

Appendix for Chapter 3

B.1 Experimental materials



(a) *Practice trial 1*



(b) *Practice trial 2*

Figure B.1: *Survey instructions for sub-experiments 1 and 2.*

(a) BLUE1 **(b)** BLUE2 **(c)** BLUE3

(d) GREEN1 **(e)** GREEN2 **(f)** GREEN3

(g) PINK1 **(h)** PINK2 **(i)** PINK3

(j) YELLOW1 **(k)** YELLOW2 **(l)** YELLOW3

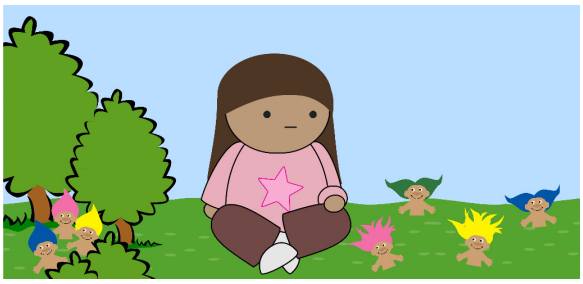
Figure B.2: *Troll images used in stimuli target/distractor pairs.*

Appendix C

Appendix for Chapter 4

C.1 Survey stimuli

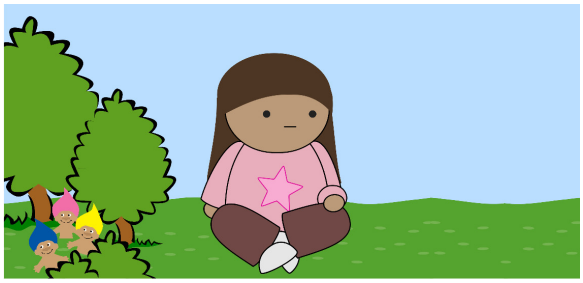
(C.1) Stimulus 1

Scene: 

a. (English) Sasha didn't find her [trolls]_**SINGLE-POINT**.

b. (ASL) fs.SASHA PRO-3 NOT FIND fs.TROLLS **CL:FLAT O**.

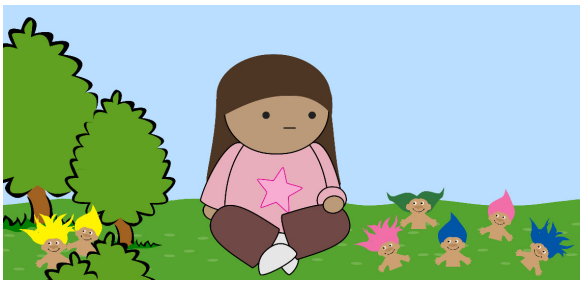
(C.2) Stimulus 2

Scene: 

a. (English) Sasha didn't find her [trolls]_**SINGLE-POINT**.

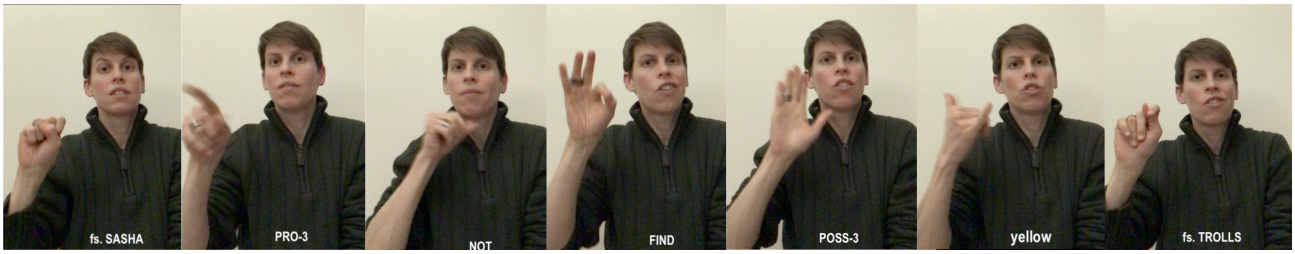
b. (ASL) fs.SASHA PRO-3 NOT FIND fs.TROLLS **CL:FLAT O**.

(C.3) Stimulus 3

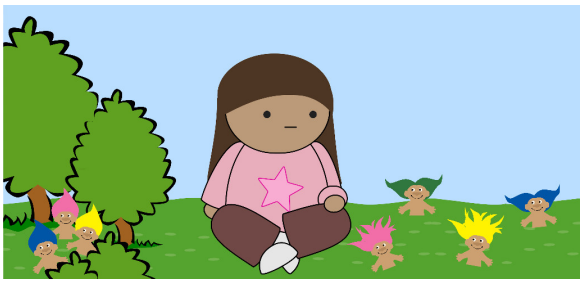
Scene: 

a. (English) Sasha didn't find her **yellow** trolls.

b. (ASL) fs.SASHA PRO-3 NOT FIND POSS-3 **YELLOW** fs.TROLLS.



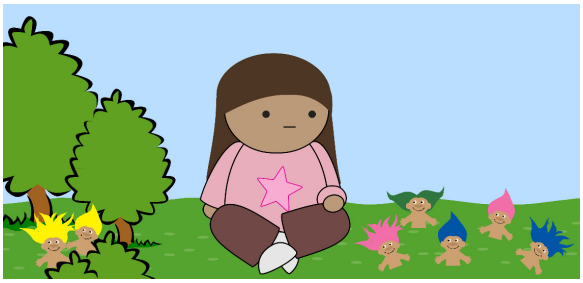
(C.4) Stimulus 4

Scene: 

a. (English) Sasha found **none** of her [trolls]_SINGLE-POINT.

b. (ASL) fs.SASHA PRO-3 FIND fs.TROLLS CL:FLAT O **NONE**.

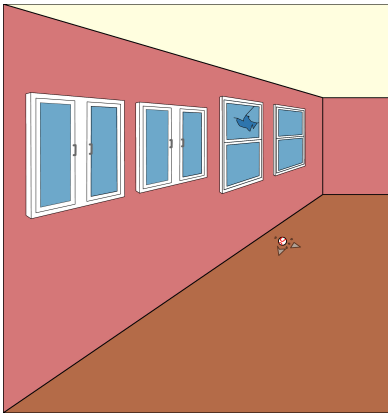
(C.5) Stimulus 5

Scene: 

a. (English) Sasha found **none** of her yellow trolls.

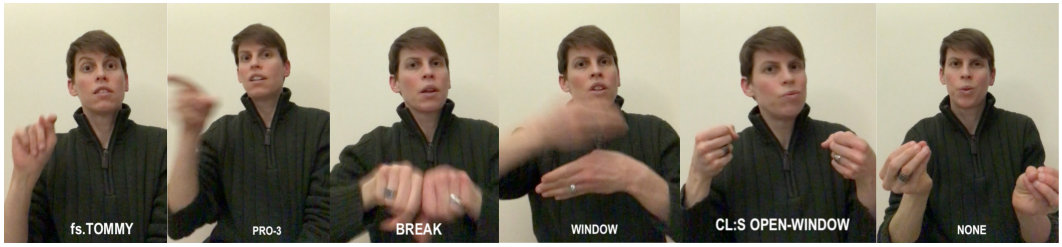
b. (ASL) fs.SASHA PRO-3 FIND YELLOW fs.TROLLS **NONE**.

(C.6) Stimulus 6

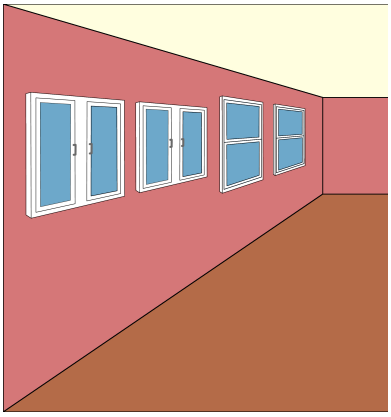
Scene: 

a. (English) Tommy broke **none** of our [windows]_HORIZONTAL-OPEN.

b. (ASL) fs.TOMMY PRO-3 BREAK WINDOW CL:S OPEN-WINDOW **NONE**.

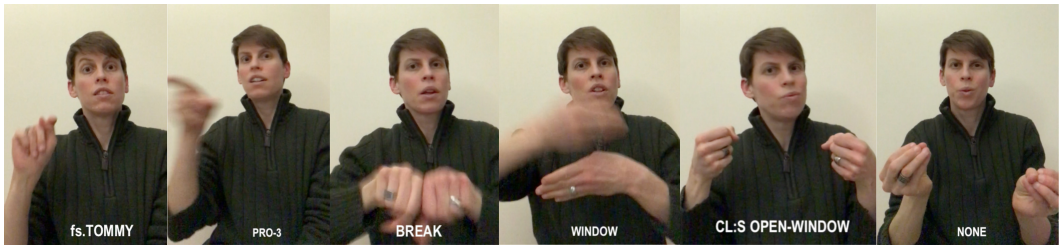


(C.7) Stimulus 7

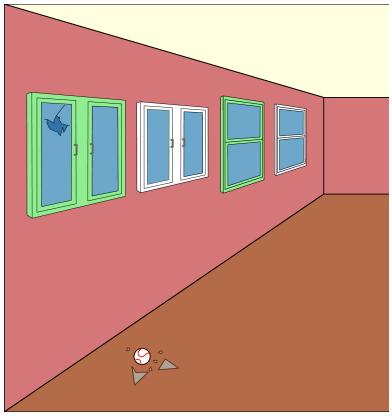
Scene: 

a. (English) Tommy broke **none** of our [windows]_HORIZONTAL-OPEN.

b. (ASL) fs.TOMMY PRO-3 BREAK WINDOW CL:S OPEN-WINDOW **NONE**.



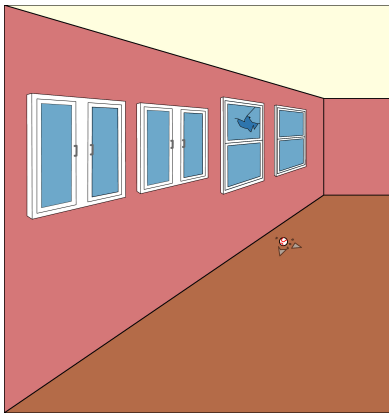
(C.8) Stimulus 8

Scene: 

a. (English) Tommy broke **none** of our white windows.

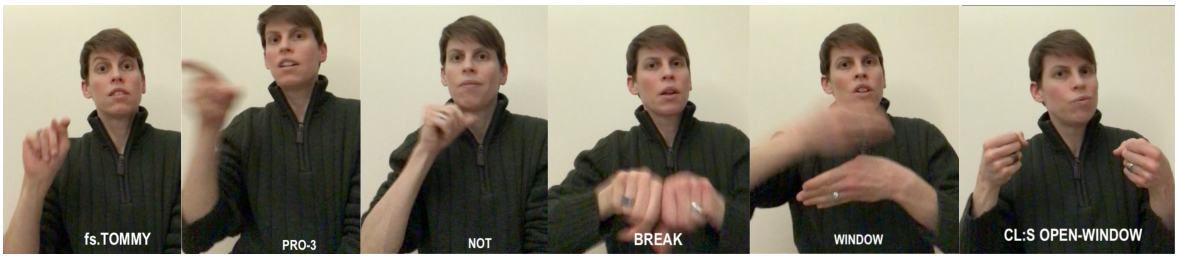
b. (ASL) fs.TOMMY PRO-3 BREAK WINDOW WHITE **NONE**.

(C.9) Stimulus 9

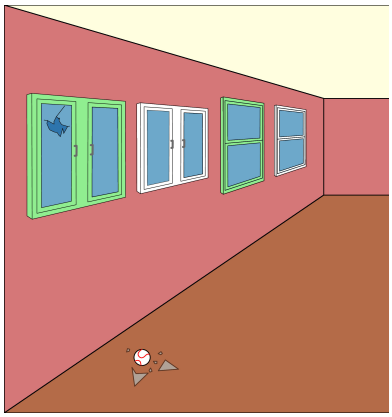
Scene: 

a. (English) Tommy didn't break our [windows]_**HORIZONTAL-OPEN**.

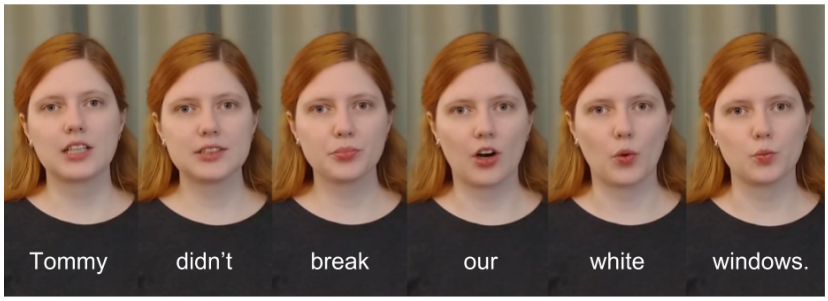
b. (ASL) fs.TOMMY PRO-3 NOT BREAK WINDOW **CL:S OPEN-WINDOW**.



(C.10) Stimulus 10

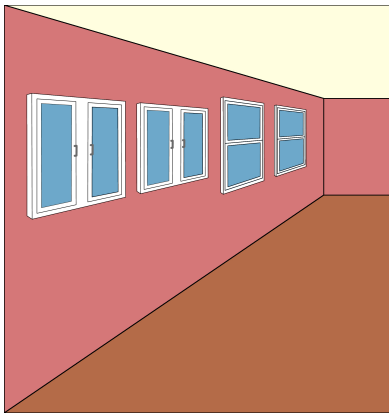
Scene: 

a. (English) Tommy didn't break our **white** windows.

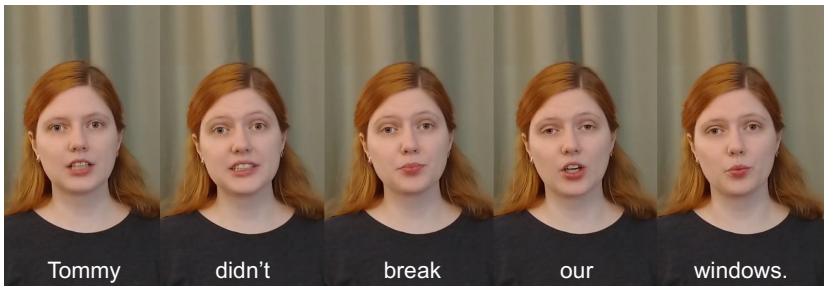


b. (ASL) fs.TOMMY PRO-3 NOT BREAK WINDOW **WHITE**.

(C.11) Stimulus 11

Scene: 

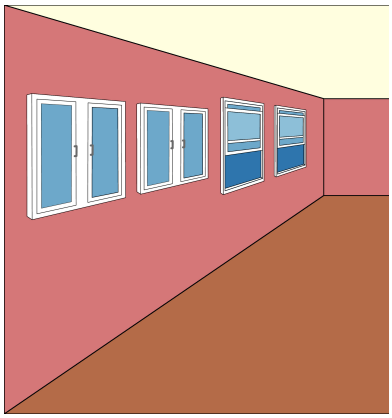
a. (English) Tommy didn't break our windows.



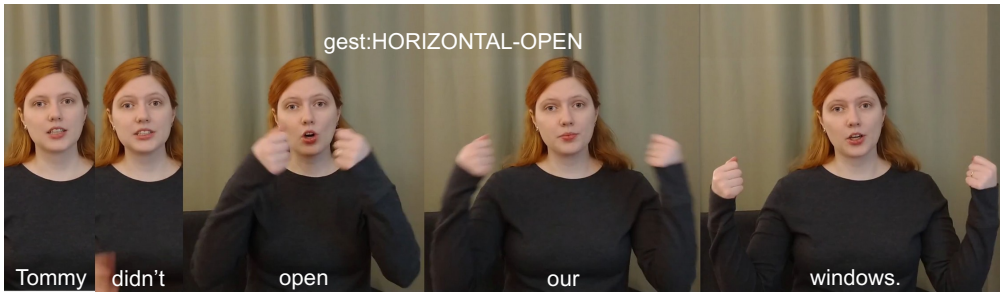
b. (ASL) fs.TOMMY PRO-3 NOT BREAK OUR WINDOW.

(C.12) Stimulus 12

Scene:



a. (English) Tommy didn't [open]_HORIZ-OPEN our windows.



b. (ASL) fs.TOMMY PRO-3 NOT CL:S OPEN-WINDOW OUR WINDOW.


fs.TOMMY
PRO-3
NOT
CL:S OPEN-WINDOW
OUR
WINDOW