Composition and projection of co-speech gestures[*]

Maria Esipova
New York University

Abstract In this paper I argue against the common assumption in recent literature that content-bearing gestures co-occurring with speech project in a single uniform way determined by their co-speech status. Instead, I propose a composition-driven, modality-neutral approach whereby which projection mechanisms are available or enforced for a given piece of compositionally integrated content, spoken or gestural, is determined by how it composes in the syntax/semantics. I adduce experimental data supporting this view from an acceptability judgement task comparing co-nominal gestures to adnominal adjectives and appositives. More broadly, this paper establishes the need to treat gestures as bona fide linguistic objects at all levels of representation in order to understand how they contribute to the meaning of utterances as well as which aspects of grammar are modality-(in)dependent.

Keywords: composition, projection, co-speech gestures, non-restricting modifiers, supplements, presuppositions, acceptability judgements

1 Introduction

There has been a recent upsurge in studying how content-bearing gestures contribute to the meaning of otherwise spoken utterances from the perspective of formal semantics and pragmatics (Ebert & Ebert 2014; Ebert 2017; Tieu, Pasternak, Schlenker & Chemla 2017, 2018; Hunter 2018; Schlenker 2018a,b; Zlogar & Davidson 2018; Esipova 2019, a.o.). Much of this research has been tackling the problem of *gesture projection*, i.e., if and how gestures come to be interpreted outside the semantic scope of various operators (negation, conditional and question operators, modals, quantifiers, etc.) in whose syntactic scope they seem to appear. One type of gestures that this literature has focused on are *co-speech gestures*, i.e., content-bearing gestures co-occurring and associating with spoken expressions, as in (1).[1]

1 Approximate word equivalents of gestures are written in ALL CAPS. Co-speech gestures are written as SUBSCRIPTS, with underlining indicating their approximate temporal alignment. Gestures are sometimes illustrated by pictures placed before the approximate onset of the gesture.

(1) Stephanie might bring her dog_{LARGE}.
→ Stephanie's dog is large.

There is a general consensus in this literature that such gestures contribute projecting inferences by default (Ebert & Ebert 2014; Ebert 2017; Tieu et al. 2017, 2018; Schlenker 2018a,b), but there is no consensus on what mechanisms assure said projection. All of this literature, however, relies on an implicit or explicit assumption that the projection profile of co-speech gestures is uniquely determined by their co-speech status. In particular, the issue has been presented as a choice between the view whereby co-speech gestures uniformly project as supplements (Ebert & Ebert 2014) and the view whereby co-speech gestures uniformly project as *cosuppositions*, i.e., assertion-dependent presuppositions (Schlenker 2018a,b).

In this paper I argue that this assumption that there is a single unique way in which co-speech gestures project is incorrect, and the debate above is insubstantial. I propose instead a uniform, modality-neutral view of projection whereby how a given piece of compositionally integrated content, spoken or gestural, projects is determined by how it composes in the syntax/semantics. Gestures like in (1) can compose either as *modifiers* (akin to attributive adjectives), in which case they can be *restricting* (truth-conditionally non-vacuous and non-projecting) or *non-restricting* (truth-conditionally vacuous and projecting as pragmatically triggered presuppositions), or as supplements, in which case they have to project. I adduce the results of an acceptability judgement experiment comparing co-nominal gestures to adnominal adjectives and appositives that support this composition-driven view.

As a more general point, I conclude that in order to build a rigorous theory of gesture projection—or projection of any secondary modality content, such as facial expressions or iconic voice modulations—we need to ask theoretically meaningful questions about how this content integrates into utterances at various levels of representation and how this integration is constrained at the interfaces between these levels. To address these questions, we need to approach gestures and other secondary modality content as bona fide linguistic objects that abide by linguistic rules at all levels of representation. The goal of an interested researcher is thus to determine what these rules are, when they are the same as the rules for ordinary linguistic expressions, and when and why they are different.

The rest of the paper is organized as follows. Section 2 outlines the composition-driven, cross-modal theory of projection. It establishes the projection patterns of modifiers and supplements and explains how gestures fit into the picture. Section 3 reports on the acceptability judgement experiment comparing co-nominal gestures to spoken adnominal adjectives and appositives. It shows that the view that co-speech gestures are supplements across the board undergenerates restricting

readings of gestures, which are shown to be marginally available, and the view that co-speech gestures trigger cosuppositions across the board overgenerates unattested readings for DP-adjoining gestures. Section 4 concludes.

2 A composition-driven, cross-modal theory of projection

2.1 Modifiers

2.1.1 Composition of modifiers

Following the basic working definition in Morzycki 2015, I will define a *modifier* as an expression of type $\langle \tau, \tau \rangle$, and I will also assume that there are further morphosyntax-based rules for determining which of the two sisters (if either) in a given configuration is the modifier and which one is the expression being modified. For example, in cases of intersective modification, the expression being modified is the one that projects its syntactic label or features to the parent node.

In this paper I will only focus on subsective modifiers, such as in (2), because (i) the projection behavior of non-subsective modifiers, such as in (3), is poorly understood, and (ii) non-conventionalized gestures, which are the main focus of this paper, are typically subsective when they are construed of as modifiers.

(2)   Zoe is a {blond / skillful} stuntwoman.
         → Zoe is a stuntwoman.

(3)   Daisy is an alleged murderer.
         ↛ Daisy is a murderer.

In a subsective modification configuration, the result of modification entails the expression being modified via generalized entailment. A sub-propositional expression α whose type ends in t entails another expression of the same type β via *generalized entailment* ($\alpha \Rightarrow_\forall \beta$) iff α entails β via *generalized material implication* ($\alpha \Rightarrow \beta$) for any values of the arguments α and β take. Generalized material implication is a lifted version of material implication that can apply to sub-propositional expressions, just like generalized conjunction (Partee & Rooth 1983) is a lifted version of logical conjunction (see also Schlenker 2018a: fn. 25).

Subsective modifiers come with this *subsective entailment* as their characteristic property. I maintain that all subsective modifiers follow the schema in (4), whose left conjunct assures that the subsective entailment goes through.

(4)   ***Subsective modification* (composition schema)**
         In a tree $\alpha_{\langle \tau_1 ... \tau_n, t \rangle}$ whose daughters are γ and $\beta_{\langle \tau_1 ... \tau_n, t \rangle}$, γ is a *subsective modifier* iff γ is a modifier and:
         $[\![\alpha]\!] = \lambda X_{\tau_1}^1 ... \lambda X_{\tau_N}^n . [\![\beta]\!](X^1)...(X^n) \wedge (...)$

The right conjunct in (4) can be as simple as $[\![\gamma]\!](X_1)...(X_n)$ for intersective modifiers, such as *blond* in (5) (assuming those compose via predicate modification). Non-subsective modifiers, such as *skillful* in (6), are usually assumed to take the expression they modify as an argument. I maintain that *skillful* and its kin still follow the schema in (4), with the left conjunct responsible for the subsective entailment instead of the latter being an idiosyncratic property of the skillful relation.

(5) $[\![\text{blond stuntwoman}]\!] = \lambda x \lambda w.\text{stntwmn}(x)(w) \wedge \text{blond}(x)(w)$

(6) $[\![\text{skillful stuntwoman}]\!] = \lambda x \lambda w.\text{stntwmn}(x)(w) \wedge \text{skillful}(\text{stntwmn})(x)(w)$

2.1.2 Projection of modifiers

When and how do modifiers project? All subsective modifiers are *restrictive* in that they have the compositional potential to restrict the expressions they modify, i.e., roughly, to yield logically stronger expressions; this potential is the result of having the right conjunct in the schema in (4). However, not all specific instances of modifiers realize this potential. Leffel (2014) defines *non-restricting modifiers*[2] (NRMs) as modifiers that don't affect the truth conditions of the utterance in which they appear. He illustrates this observation via examples like (7). Under the world-knowledge based assumption that not all chemicals are harmful, but all toxins are, (7a) doesn't entail a version of itself without the adjective, but (7b) does.

(7)   a. I will eliminate every harmful chemical.   *restricting modifier*
           $\nrightarrow$ I will eliminate every chemical.

       b. I will eliminate every harmful toxin.   *non-restricting modifier*
           $\rightarrow$ I will eliminate every toxin.

Leffel's (partial) definition of NRMs is given in (8); I propose a refined version in (9).

(8)   ***Non-restricting modifier* in Leffel 2014: (3.58a) (partial)**
       An occurrence α_n of a modifier α in $[_{\text{DP}}...\alpha_n...\text{N}...]$ or $[_{\text{DP}}...\text{N}...\alpha_n...]$ is *non-restricting* with respect to index i iff the speaker believes that $[\![\text{DP}]\!]^i = [\![\text{DP}[\alpha_n/\varepsilon]]\!]^i$, where $X[a/b]$ is the expression just like X except with a token of a replaced by one of b and ε is the empty string.

(9)   ***Non-restricting modifier* (adopted here)**
       If an utterance u contains a subtree α whose daughters are γ and β and γ is an occurrence of a subsective modifier, γ is *non-restricting* at an index i iff the speaker of i believes that $[\![u]\!]^i = [\![u[\gamma/\varepsilon]]\!]^i$, where $X[a/b]$ is just like X but with a token of a replaced by one of b and ε is the empty string.

---

2 He uses the terms "non-restrictive" and "non-restricting" differently from me.

In contrast to Leffel's, my definition is not restricted to adnominal modifiers and is explicitly restricted to subsective modifiers (Leffel makes this restriction later). More importantly, however, I switch to global, utterance-level equivalence from local, constituent-level one. This is necessary, because Leffel's definition is at odds with the projection behavior of inferences contributed by NRMs, and we need to separate the definitional utterance-level truth-conditional vacuity of NRMs, which holds globally, from the inference about the local equivalence of the expression being modified and the result of modification, which is sensitive to local contexts.

While NRMs are truth-conditionally vacuous, they still make a meaningful contribution in the form of *non-restricting modifier inferences* (NRM inferences) that the expression being modified entails the result of modification ($\beta \Rightarrow_\forall \alpha$).[3] For example, in (7b), the NRM inference is that being a toxin entails being a harmful toxin, or, simply put, that all toxins are harmful.

However, unlike what is suggested by Leffel's definition in (8), this entailment doesn't always hold in all of the speaker's belief worlds. Rather, what the speaker has to believe is that this entailment holds relative to the *local context* (Karttunen 1974; Stalnaker 1974; Heim 1983; Schlenker 2009, a.o.) of the modification configuration. For example, in both (10) and (11), the speaker intends the modifier as truth-conditionally vacuous. However, in (10), the (propositional) local context of the modification configuration is equivalent to the Stalnakerian global context set C (Stalnaker 1974), thus, we get a global inference that (the speaker believes that) being a sausage entails being a deadly sausage, i.e., that all sausages are deadly. In (11), however, we get a conditionalized inference that (the speaker believes) that if processed meat causes cancer, all sausages are deadly.

(10) *Context: The speaker believes that processed meat causes cancer.*
      I shouldn't eat so many deadly sausages.
      → All sausages are deadly.

(11) *Context: The speaker just read an article saying that processed meat might be causing cancer, but they are not ready to embrace it as a fact just yet.*
      {Maybe processed meat causes cancer and / If processed meat causes cancer,} I shouldn't eat so many deadly sausages.
      → If processed meat causes cancer, all sausages are deadly.

For comparison, local context sensitivity of lexical presuppositions is illustrated for *know* below.

(12) Zoe knows that she can't continue the race.
      → Zoe can't continue the race.

---

3 Or, rather, that the two entail each other ($\alpha \Leftrightarrow_\forall \beta$), however the entailment from the result of modification to the expression being modified ($\alpha \Rightarrow_\forall \beta$) is assured by the modifier being subsective.

(13) {Maybe Zoe's car has been damaged and / If Zoe's car has been damaged,}
      she knows that she can't continue the race.
      → If Zoe's car has been damaged, she can't continue the race.

While NRM inferences project similarly to lexical presuppositions, we don't have to maintain that they are triggered in the semantics, be it lexically or as part of a composition rule (as is proposed in Leffel 2014). I propose that NRM inferences are triggered pragmatically whenever the addressee concludes that the speaker intends for a given instance of a modifier to be non-restricting, i.e., truth-conditionally vacuous. Thus, I define *non-restricting modifier inferences* as follows:

(14) ***Non-restricting modifier inference***
      If an elementary clause p contains a subtree α whose daughters are γ and β and the addressee infers that the speaker intends γ as a non-restricting modifier (as defined in (9)), the addressee will generate the *non-restricting modifier inference*: $c'' \Rightarrow_\forall (\beta \Rightarrow \alpha)$, where c'' is the sub-propositional local context of α in the propositional local context c' of p.

I assume Schlenker's (2009; 2010) pragmatic approach to local contexts, which defines the local context c' of any expression d whose type ends in t as the strongest possible restriction one could make before interpreting d in a Stalnakerian context set C. The propositional local context c' for the elementary clause p containing the modifier is computed, for example, by updating the global context C with the left conjunct under *maybe* or the antecedent of a conditional, as in (11). The sub-propositional local context c'' is computed within p for the result of modification α and is always relativized to c'. In simple cases this means that the world argument of c'' is restricted to the worlds of c'.

Thus, if the addressee recognizes that the speaker intends *deadly* as truth-conditionally vacuous in (10), they will generate the inference that (the speaker believes that) in all the worlds of the global context C being a sausage entails being a deadly sausage, i.e., all sausages are deadly. But in (11), the NRM inference would be that all sausages are deadly in all the worlds in which processed meat causes cancer.[4]

One crucial difference between lexical presuppositions and NRM inferences is that while the former can sometimes be interpreted locally under semantic operators via *local accommodation* (e.g., Heim 1983; Schlenker 2009), this is not a meaningful option for NRM inferences. Local accommodation makes the affected piece of content truth-conditionally non-vacuous, however, if a modifier is meant to be truth-conditionally non-vacuous, it can't be non-restricting, so no NRM inference will be generated in the first place. Empirically, the contrast is illustrated in (15).[5]

---

4 For simplicity I am assuming that the sub-propositional local context c'' is maximally broad in both cases, i.e., that it denotes $\lambda x \lambda w. [\![c']\!](w) \wedge D_e(x)$, where D_e is the domain of individuals in the model.
5 **Bold** indicates prosodic contrastive focus marking.

(15)   *Context: Zoe, Lucy, and Pam are going on a camping trip. Pam is coming from Boston to New York to join the rest of the group. Zoe and Lucy are discussing how to get to the camping site from New York. ...*

a. *...They have agreed that they need a car, no matter how big. Lucy:*

? I don't know if Pam **has** a car, but if she's coming in her **car**, we can use it to get to the camping site.
'...if (she has a car and she's coming in her car)...'

b. *...They have agreed that they need a large car to fit all their supplies. Lucy knows that Pam has exactly one car. Lucy:*

I don't know if Pam's car is **large**, but if she's coming in {her **large** car / her large **car**}, we can use it to get to the camping site.
Intended: '...if ({she has a large car / her car is large} and she's coming in her large car)...'

In (15a), the existence inference of the definite description *her car* cannot be satisfied globally, but it can be marginally interpreted locally under *if*, resulting in an imperfect but relatively acceptable utterance. However, locally accommodating the existence inference of the definite description is not enough in (15b); the modifier has to be non-restricting as Pam only has one car,[6] but its NRM inference can't project in the context, which results in complete infelicitousness.

2.1.3   NRM inferences and cosuppositions

The definition of NRM inferences in (14) is reminiscent of *cosuppositions*, i.e., assertion-dependent inferences that project like presuppositions, originally proposed in Schlenker 2018a as inferences triggered by co-speech gestures across the board and extended to other types of content in Schlenker 2018b,c. In (16), I give Schlenker's (partial) generalized definition of cosuppositions from Schlenker 2018c.[7]

(16)   **Cosuppositions in Schlenker 2018c: (67) (partial, w/notational changes)**

a. A cosupposition is triggered when an elementary expression pp' has an entailment p' which is presented as being unimportant, and for this reason the global Context Set C should guarantee that, relative to its local context c', pp' should be equivalent to p, i.e.: (i) $c' \Rightarrow_\forall (pp' \Leftrightarrow p)$

---

6 I assume that each modifier is either restricting or non-restricting, and, furthermore, restricting modifiers have to pick out a (locally) non-empty part of the denotation of the expression they modify.

7 While (16) only makes reference to the global context set C and a single local context c', Philippe Schlenker (p.c.) has confirmed to me that for a non-propositional pp' we need to compute two local contexts, a propositional one and a sub-propositional one, just like in (14).

b. (i) is equivalent to the standard definition of cosuppositions: (ii) $c' \Rightarrow_\forall$ $(p \Rightarrow p')$[8]

Schlenker's algorithm in (16) is, on the one hand, unconstrained in that it isn't linked to any specific compositional configuration (it can even apply if p and p' don't combine compositionally at all and are instead two pieces of a single lexical entry pp'). This leads to overgeneration when the purported cosupposition trigger cannot be a modifier (as defined in this paper), as we will see in the next section.

On the other hand, Schlenker assumes that p and p' are conjuncts within pp' (whether or not they combine compositionally) and that p' is, thus, an entailment of pp', which is too restrictive, as this prevents cosuppositions from arising in cases where p and p' are of different semantic types (pp' can't entail p' if they are of different types, nor can p and p' conjoin in this case). The consequence of that is that while Schlenker's cosupposition algorithm generates the same inferences for intersective NRMs as the algorithm in (14), it can't apply to non-intersective modifiers, which are not of the same type as the expression they modify or the result of modification.[9] The inability to handle non-intersective modifiers is also a problem for Leffel's (2014) analysis of NRM inferences as semantic presuppositions hard-coded into the composition rule for NRMs (more generally, it's unclear how Leffel's analysis of NRMs would generalize beyond adnominal intersective modification).

Note that simply switching to the 'daughter entails parent' formulation from 'sister entails sister' to allow cosuppositions for non-intersective modifiers doesn't fill the need for constraining cosuppositions, which is the main issue with Schlenker's view. Rethinking cosuppositions as NRM inferences both constrains them in the right way (as we'll see in section 3) and allows generating them for non-intersective modifiers. In other words, fixing the technical bug doesn't address the bigger issue, but addressing the bigger issue fixes the technical bug along the way.

2.2 Composition and projection of supplements

2.2.1 Composition of supplements

Supplements, most obviously exemplified by appositives, compose with anchor expressions and contribute propositions of a special kind about those anchors. For adnominal appositives, whose anchors are individuals, this means that they compose with DPs (Determiner Phrases, type e or $\langle\langle e, st\rangle, t\rangle$)—as opposed to attributive

---

8 Here Schlenker refers to the formulation of cosuppositions in Schlenker 2018a.

9 One of the cases mentioned in Schlenker 2018c is the degree modifier 'completely' in Italian Sign Language and spoken Italian discussed by Aristodemo (2017), who argues that both cases give rise to a projecting inference (e.g., 'full' entails 'completely full'), which she analyzes as a cosupposition. But degree modifiers are non-intersective, so the algorithm in (16) can't yield this result.

adjectives, which compose with NPs (Noun Phrases, type $\langle e, st \rangle$). For example, in (17), the appositive *(who is) a stuntwoman*[10] composes with the DP *Zoe* denoting the individual Zoe and contributes the proposition that Zoe is a stuntwoman.

(17)   I invited Zoe, (who is) a stuntwoman.
        → Zoe is a stuntwoman.

Whichever specific account of appositives one assumes, the general outcome is the same: appositives are *non-restrictive*, because the way they compose with their anchors doesn't allow them to restrict anything. For example, Potts (2005) assumes bidimensional semantics whereby the anchor fills the argument slot of the appositive yielding a proposition of a special, conventional implicature type. Koev (2013) and AnderBois, Brasoveanu & Henderson (2013) assume a dynamic unidimensional setup instead, whereby appositives introduce propositional discourse referents of a special kind; the link between appositives and their anchors is anaphoric.

The consequence of this composition that appositives can never be restricting is borne out empirically. For example, in (18a), the two contrasted nominals are prosodically integrated,[11] adjoin at the NP-level,[12] are, thus, modifiers and can, thus, be restricting. However, (18b) is infelicitous, precisely because appositives, which are packaged into their own prosodic phrases and adjoin at the DP-level, can't ever be restricting.

(18)   a. ($_{\text{PrP}}$ I invited Zoe the **stuntwoman**), ($_{\text{PrP}}$ not Zoe the **politician**).

        b. # ($_{\text{PrP}}$ I invited Zoe), ($_{\text{PrP}}$ the **stuntwoman**), ($_{\text{PrP}}$ not Zoe), ($_{\text{PrP}}$ the **politician**).

2.2.2 Projection of supplements

Supplements project, and typically they project very strongly, i.e., they typically cannot be interpreted locally under semantic operators, not even under pressure:

(19)   a. If you invite Zoe, a stuntwoman, you should show her your muscle car.
           → Zoe is a stuntwoman.

        b. # I don't know if Zoe is a stuntwoman, but if you invite Zoe, a stuntwoman, you should show her your muscle car.
           Intended: '...if (Zoe is a stuntwoman and you invite her)...'

---

10 Here I assume that nominal appositives contain silent syntactic structure making them similar to full-blown appositive relative clauses (cf. *Zoe, (who is) often called the most talented stuntwoman*). This assumption isn't crucial, however, and whether there any syntactic differences between the two structures is immaterial for the purposes of this paper.

11 $_{\text{PrP}}$ stands for "prosodic phrase", whether it is an intermediate or an intonational phrase.

12 I follow Matushansky 2008 in that proper nouns have an NP layer denoting the property of being named in a certain way.

How exactly one accounts for supplement projection is immaterial for the purposes of this paper. For example, in both Potts 2005 and AnderBois et al. 2013, projection of supplements is assured by the propositions they contribute being special in a way that makes them impervious to semantic operators. In Potts 2005, these propositions have a special semantic type, and in AnderBois et al. 2013 they are a special type of discourse referents. Either option works for me.

Schlenker (2013) discusses apparent exceptions to the projection requirement on appositives, such as (20).

(20)   If tomorrow I call the Chair, who in turn calls the Dean, we will be in deep trouble.
        ↛ If tomorrow I call the Chair, they will call the Dean.
        ≈ 'If (tomorrow I call the Chair and they call the Dean)...'

Such interpretations are very constrained, however (see Jasinskaja & Poschmann 2018 for generalizations on when they are available). The crucial difference between modifiers and supplements with respect to projection is, thus, that for a modifier not to project means to be restricting, and this interpretation is in principle available to any modifier by default. Whether a given instance of a modifier is interpreted as restricting (and, therefore, non-projecting) or non-restricting (and, therefore, projecting) is determined by a variety of utterance-external factors. In supplements, however, projection is systematically triggered by a certain compositional configuration (e.g., that of an appositive or a sentence-level adverb), and local interpretations are only available under very limited circumstances.

This difference leads me to conclude that we do need two separate projection mechanisms for NRMs and supplements. A reductionist story like the one in Morzycki 2008, whereby NRMs contribute Pottsian (2005) conventional implicatures, fails to capture this contrast in availability of local interpretations. Even if we use the same machinery to assure lack of interaction with semantic operators for the two types of projecting inferences (i.e., projection proper), we would still need two separate triggering mechanisms. In subsection 2.1.2, I proposed that triggering of NRM inferences is pragmatic in nature. I am not proposing any new mechanism for supplements, but it is clear that triggering of projecting inferences in supplements is much more systematic, much more conventional, and, thus, much more semantic, which is how it is implemented in most existing analyses of supplements.

2.3 Fitting gestures into the picture

I propose that the setup above applies straight-forwardly to compositionally integrated secondary modality content, including gestures. I maintain a "no gesture-specific composition" principle whereby if a gesture integrates compositionally into

an utterance, it can do it in all and only ways available to spoken content. Thus, gestures can in principle compose as modifiers or supplements, as long as their iconic content is compatible with both construals, and there are no further constraints on their composition. In particular, non-conventionalized size gestures such as in (1) can be easily construed of as representing a property (akin to *large*), in which case they adjoin to NPs and compose as modifiers, or as representing an object (akin to *a large object*), in which case they adjoin to DPs and compose as supplements.

However, I agree with the general view in the existing literature that co-occurrence with spoken expressions does make gestures more likely to be perceived as truth-conditionally vacuous. For supplement gestures, this doesn't have any additional consequences, since those are typically truth-conditionally vacuous and project as is. For modifier gestures, this has the consequence that they are more likely to be interpreted as non-restricting and, thus, contributing projecting inferences as well. Yet, this preference can be overridden, and gestures whose iconic content is in principle compatible with a modifier construal can also be restricting, in which case they are truth-conditionally non-vacuous and don't project.

Of course, compositional configurations available to a given instance of a gesture will be constrained by various factors, such as its iconic content, its conventionalized morphosyntactic and lexical properties (if any), and various interface considerations. For example, in Esipova 2018a,b, I discuss how modifier construals are unlikely for prosodically independent (i.e., non-co-speech) gestures in English because of the English linearization rules as well as interaction of universal articulatory considerations with the English rules of syntax-sensitive prosodic grouping. However, under the modality-neutral, composition-driven approach to projection that I advocate for here, the co-speech status of a gesture does not directly and uniformly determine how this gesture will project.

The summary of the view presented in this section is given in Table 1.

3 Experiment: adjectives, appositives, gestures

3.1 Goals of the study

The composition-driven theory of gesture projection presented in the previous section as well as the alternative theories in Ebert & Ebert 2014 (the supplemental theory, under which all co-speech gestures are supplements) and Schlenker 2018a (the cosuppositional theory, under which all co-speech gestures trigger cosuppositions) make specific predictions regarding how co-speech gestures such as in (1) can and cannot be interpreted. In particular, while all the three theories predict that co-speech gestures prefer to project, they differ on whether they predict the possibility of local (i.e., non-projecting) interpretations, and if yes, what kind of local

(Subsective) modifiers	Supplements
Compose with β, yielding α such that α entails β	Compose with β, yielding a proposition of a special kind about β
Can be restricting or non-restricting	Can never be restricting
Trigger projecting inferences when non-restricting; triggering is pragmatic	Almost always trigger projecting inferences; triggering is conventional
Examples (target content bolded): • attributive adjectives (*her **large** dog*) • restrictive relative clauses (*her dog **that's large***) • degree modifiers (***completely** full*) • NP-level gestures (*her [[$_{NP}$ dog]$_{LARGE}$]*)	Examples (target content bolded): • appositives (*her dog, **(who is) a large animal***) • sentence-level adverbs (***Surprisingly**, Zoe won the race*) • DP-level gestures ([[$_{DP}$ *her dog*]$_{LARGE}$])

Table 1   Modifiers vs. supplements

interpretations they predict to be possible. There are at least two conceivable local interpretations for co-speech gestures. One is that of a restricting modifier (which I will label 'Restricting'), and the other is a hypothetical interpretation corresponding to a locally interpreted NRM inference or a locally interpreted supplement (which I will label 'Non-projecting non-restricting'), as illustrated in (21).

(21)   If Stephanie's bringing her dog$_{LARGE}$, we should get a bigger van.
Projecting non-restricting interpretation: $\rightarrow$ Stephanie's only dog is large.
Restricting interpretation: $\approx$ 'If Stephanie's bringing her large dog, but not her small one...'
Non-projecting non-restricting interpretation: $\approx$ 'If (Stephanie's only dog is large and she's bringing her large dog)...'

Previous experimental studies on gesture projection (Tieu et al. 2017, 2018; Esipova 2019) concluded that local interpretations of co-speech gestures are in principle possible, but none of them was designed to distinguish between the two types of local interpretations above. The goal of this study is to do so and to test the relevant predictions of the three theories via an acceptability judgement task.

3.2   Methods

Participants were recruited on Amazon Mechnical Turk and paid $1 each. Those who failed the attention check or reported being non-native speakers of English

were excluded. The final number of participants was 122 (33 female, 89 male).

Participants read context paragraphs, watched videos of sentences uttered in those contexts, and assessed these sentences by dragging a slider on a pseudo-continuous scale from 'Totally unnatural' to 'Totally natural' (mapped to 0–100).

The items differed along two fully-crossed factors: Content Type (Adjective, Appositive, Gesture) and Interpretation (Projecting non-restricting, Restricting, Non-projecting non-restricting), yielding 9 conditions. The target interpretation was enforced within each video by setting up the question under discussion (QUD) and explicitly contrasting two alternatives; the written contexts were meant to maximally support this interpretation. 4 complete test paradigms (for 4 different scenario types) were constructed, yielding 36 test items in total. All the videos used in the experiment can be found at https://osf.io/fr5xt/.

A complete sample paradigm is given in (22). All the written contexts (in italics) in this paradigm started in the same way, setting up the van-renting scenario. They then continued in three different ways, setting up the subscenario supporting one of the three target interpretations. Three videos were then recorded for each of the three written contexts: with adjectives, appositives, and gestures.

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

     a. *Projecting non-restricting: ...Stephanie, 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 (i) her small **cat**, (ii) her **cat**, a **small** animal, (iii) her **cat**$_{SMALL}$, we'll be fine, but if she's bringing (i) her large **dog**, (ii) her **dog**, a **large** animal, (iii) her **dog**$_{LARGE}$, we should get a bigger van.

     b. *Restricting: ...Stephanie, 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 (i) her **small** dog, (ii) her dog, a **small** animal, (iii) her **dog**$_{SMALL}$, we'll be fine, but if she's bringing (i) her **large** dog, (ii) her dog, a **large** animal, (iii) her **dog**$_{LARGE}$, we should get a bigger van.

     c. *Non-projecting non-restricting: ...Stephanie 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 (i) her **small** dog, (ii) her dog, a **small** animal, (iii) her **dog**<u></u>$_{\text{SMALL}}$, we'll be fine, but if she's bringing (i) her **large** dog, (ii) her dog, a **large** animal, (iii) her **dog**$_{\text{LARGE}}$, we should get a bigger van.

In (22a), the expressions of interest can't be restricting, as Stephanie only has one cat and one dog,[13] hence the 'non-restricting' part of the interpretation label. Furthermore, the QUD is about which of her two pets Stephanie is bringing, not their size, so the contributions of the expressions of interest are allowed to project, hence the 'projecting' part. In (22b), the expressions of interest are restricting the expressions they combine with, as Stephanie has two dogs and these expressions are meant to pick out one of them based on size. In (22c), the expressions of interest aren't restricting the expressions they combine with, as Stephanie only has one dog, hence the 'non-restricting' part of the interpretation label. Furthermore, the QUD is about the size of Stephanie's dog, and it's clear from the question Anna asks that she doesn't know what size it is, so the contributions of the expressions of interest have to be interpreted locally under *if*, hence the 'non-projecting' part.

The design was within subjects. Each participant saw one randomly selected item per condition (thus, 9 test items) and 2 additional items (thus, 11 items in total), presented in random order. One of the additional items was an attention check where the participant was instructed to drag the slider all the way to the left or to the right. The other one had a mismatch between the written context and the video sentence; the context corresponded to the 'Non-projecting non-restricting' interpretation, but the sentence in the video had restricting adjectives. This item was added to assess whether participants were paying attention to the written contexts and as an informal baseline for degradedness.

3.3 Predictions of the different theories of co-speech gestures

Let us now review the predictions made by the three theories of co-speech gestures (supplemental, cosuppositional, composition-driven) regarding the thee potential interpretations in (21). All the three theories predict 'Projecting non-restricting' interpretations to be highly acceptable for co-speech gestures; the differences arise for the other two interpretation types.

The supplemental analysis predicts that co-speech gestures should pattern with appositives regarding the relative acceptability of 'Restricting' and 'Non-projecting

---

13 The assumption here is that the extensions of *cat* and *dog* are restricted to Stephanie's cats and dogs only, either via domain restriction or via indices on *cat* and *dog*. Leffel (2014) uses the latter option for possessive DPs, but concedes the need for pragmatic domain restriction for some other cases.

non-restricting' interpretations; the expectation is that the acceptability ratings should be relatively low for both interpretation types for both content types.

The predictions of the cosuppositional analysis are a bit harder to assess. In general, it allows for local interpretations of co-speech gestures, either by locally accommodating the cosupposition or by failing to generate it in the first place (both are assumed to be costly). In both cases, the result is equivalent to simply conjoining the gesture to the spoken expression it adjoins to. For co-nominal gestures, if a gesture adjoining to an NP (type $\langle e, st \rangle$) is interpreted locally, the cosuppositional analysis predicts the (possibility of) the 'Restricting' interpretation for it. However, if a co-nominal gesture adjoins to a DP (type $\langle \langle e, st \rangle, t \rangle$[14]), the local interpretation predicted by the cosuppositional theory depends on the lexical semantics one posits for DP-level gestures, which Schlenker himself doesn't discuss.[15]

Under a natural assumption that an $\langle \langle e, st \rangle, t \rangle$-type *LARGE* is an indefinite akin to 'a large object', conjoining it to the spoken DP *her dog* would yield the interpretation 'Stephanie's bringing her dog and a large object' for *Stephanie's bringing [[_{DP} her dog]_{LARGE}]*, which is very clearly unattested. Giving indefinite-like semantics to DP-level gestures also yields unattested results if the cosupposition is generated and projects when the spoken DP is a quantifier like *no dog*. To introduce the missing link between the spoken DP and the gesture adjoining to it and to block gestures from adjoining to DPs that don't introduce discourse referents, one could say that DP-level co-speech gestures are anaphoric to the DPs they adjoin to, as in (23). This essentially amounts to giving DP-level gestures appositive-like semantics while trying to preserve cosupposition-like composition and projection.

(23) $[\![LARGE_i]\!]^g = \lambda P \lambda w. P(g(i))(w) \wedge \mathsf{large}(g(i))(w)$
       $\approx$ 'that object, and that object is large'

Assuming such semantics for DP-level gestures would yield decent results for projection of the purported cosupposition (if it is further assumed that its local context is also anaphorically linked to the spoken DP). However, locally interpreting the gesture in *Stephanie's bringing [[_{DP} her dog]_{LARGE}]* would yield 'Stephanie's bringing [her dog]$_i$ and it$_i$ is large', which corresponds to the 'Non-projecting non-

---

14 I assume that the DP will have to be interpreted as a generalized quantifier, as it is unclear what it would mean for two individuals to entail one another.

15 One can't tell if a co-nominal gesture adjoins at the NP- or DP-level based on its temporal alignment, at least not in the items used in this experiment. In particular, the preparatory phase of the gesture often starts before the spoken DP (as can be seen in the videos for this experiment and the ones in Esipova 2019; in both cases the gesture models were instructed to produce the gestures in the way that was most natural to them). I, thus, assume that both syntactic construals are in principle possible for all the gestures in the experiment. Timing the gesture with the NP only is likely to be unnatural. It's furthermore unclear if doing so would in fact force the NP-level construal, and if it does, what it would give us, since it's the DP-level construal that is more constrained in how it can be interpreted.

restricting' interpretation in (22c). Since under the cosuppositional analysis all local interpretations of co-speech gestures result from the same mechanism (local accommodation or non-generation of the cosupposition), they should be equally available regardless of where the gesture adjoins. Thus, if one assumes the semantics like in (23) for DP-level gestures, 'Restricting' and 'Non-projecting non-restricting' interpretations of co-speech gestures should be equally available.

The composition-driven view predicts that 'Restricting' interpretations should be available, because modifier gestures can in principle have them, but degraded as compared to 'Projecting non-restricting' ones, because co-speech gestures do prefer to be truth-conditionally vacuous. 'Non-projecting non-restricting' interpretations are expected to be as unavailable for co-speech gestures as for adjectives and supplements. Modifier gestures can't have them because only truth-conditionally vacuous modifiers can be non-restricting, and the modifier would not be truth-conditionally vacuous under the 'Non-projecting non-restricting' interpretation. Supplement gestures can't have them, because supplements have to project.

The predictions of the three theories are summarized in Table 2.

supplemental	cosuppositional	composition-driven
gestures should pattern like appositives	PNR > (R = NPNR) (w/further assumptions)	PNR > R > NPNR

Table 2   Predictions for the three interpretations in the Gesture condition

3.4   Results and discussion

Spreadsheets with the raw and cleaned up data are published along with this paper. All statistical tests and plots were done in R (2019); the code is published along with this paper. The data were subsetted based on Content Type, and a mixed effects linear regression model was run for each subset with Interpretation as a fixed effect, and Participant and Scenario Type as random effects. Once the significant effect of Interpretation was established, pairwise comparisons between different interpretations were performed via similar models. The results are visualized in Figure 1, and the statistics are given in Table 3. The mean % acceptability for mismatch items was 39, which is not significantly different from that of 'Non-projecting non-restricting' items with adjectives (Beta $= -.041$, $t = -.858$, $p = .393$).

As expected, 'Projecting non-restricting' and 'Restricting' interpretations of adjectives are equally and highly acceptable, but 'Non-projecting non-restricting' ones are much less acceptable. Also as expected, only 'Projecting non-restricting' interpretations are highly acceptable for appositives; 'Restricting' and 'Non-project-
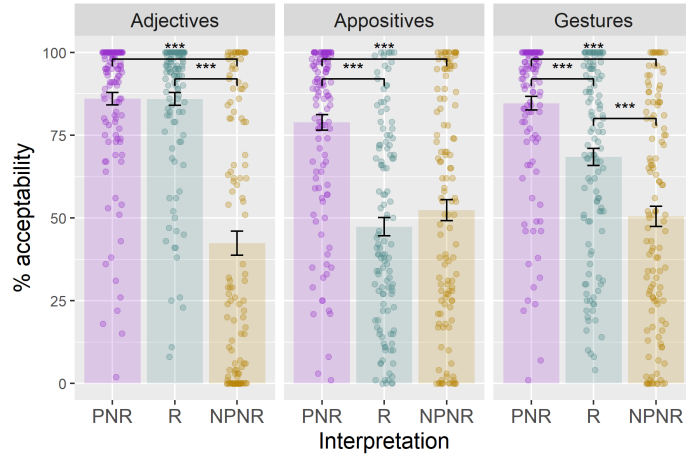
Figure 1 Bar chart showing % acceptability of different interpretations per content type. Error bars show standard error. Dots represent individual responses (jitter was added for better visualization).

ing non-restricting' interpretations are much less acceptable, with no significant difference between the two. 'Projecting non-restricting' interpretations of gestures are highly acceptable, just like for adjectives and appositives. 'Non-projecting non-restricting' interpretations of co-nominal gestures are much less acceptable, again, just like for adjectives and appositives. Crucially, however, 'Restricting' interpretations of gestures are marginal; they are significantly more acceptable than 'Non-projecting non-restricting' ones (in contrast to appositives), but significantly less acceptable than 'Projecting non-restricting' ones (in contrast to adjectives).

The data at hand match the predictions of the composition-driven theory only. Contra the supplemental theory, co-speech gestures do not pattern with appositives regarding the acceptability of 'Restricting' interpretations. Contra the cosuppositional theory (under additional assumptions), 'Restricting' and 'Non-projecting non-restricting' interpretations of co-speech gestures are not equally acceptable.

4 Conclusion

In this paper I have argued that contra the pervasive assumption in the literature, there is no single way in which co-speech gestures project. While the observation that co-speech gestures prefer to be truth-conditionally vacuous made in the previous literature (Ebert & Ebert 2014; Schlenker 2018a; Tieu et al. 2017, 2018) is correct, how exactly this truth-conditional vacuity is assured depends on how a given instance of a gesture composes in the syntax/semantics. In this respect ges-

Content	Mean % acceptability				Comparisons	
	PNR	R	NPNR	PNR/R	PNR/NPNR	R/NPNR
Adjectives	86.0	85.9	42.4	NS	***	***
				Beta = .001	Beta = .565	Beta = .568
				$t = .03$	$t = 10.71$	$t = 10.842$
				$p = .976$	$p < 2\mathrm{e}{-16}$	$p < 2\mathrm{e}{-16}$
Appositives	78.9	47.4	52.4	***	***	NS
				Beta = .49	Beta = .398	Beta = $-.077$
				$t = 10.35$	$t = 8.057$	$t = -1.628$
				$p < 2\mathrm{e}{-16}$	$p = 6.28\mathrm{e}{-13}$	$p = .106$
Gestures	84.6	68.4	50.5	***	***	***
				Beta = .301	Beta = .513	Beta = .277
				$t = 6.298$	$t = 10.95$	$t = 5.459$
				$p = 5.43\mathrm{e}{-09}$	$p < 2\mathrm{e}{-16}$	$p = 2.59\mathrm{e}{-07}$

Table 3   Statistics for acceptability of different interpretations per content type

tures are no different from spoken expressions. In particular, gestures construed of as modifiers can be restricting or non-restricting and only project when they are non-restricting; restricting interpretations of co-speech gestures are dispreferred, because restricting modifiers are truth-conditionally non-vacuous, but available. Gestures construed of as supplements can't be interpreted locally at all, because supplements in general can't. This view fits the empirical picture from section 3, which cannot be captured by theories that rely on the assumption that there is a single way in which all co-speech gestures project.

Going back to the point raised in the Introduction, treating gestures as bona fide linguistic objects at all levels of representation is what allowed us to arrive at this more nuanced and empirically adequate view of gesture projection. Furthermore, the results of the present study suggest that narrow syntax and compositional semantics are modality-blind, i.e., how a given piece of content integrates compositionally into an utterance doesn't depend on whether it's spelled out as a spoken item from the lexicon or as a gesture that is not necessarily stored in the lexicon. Modality-specific effects, however, manifest in the interaction of linearization and pragmatics. In particular, co-occurrence with material in a more primary modality (which is, of course, an impossible configuration for spoken lexical items) has been confirmed to result in a preference for truth-conditional vacuity for gestures. Grounding future work on gesture and other secondary modality content in a systematic interface-based approach can, thus, both yield a more fine-grained understanding of such content and provide insights into which aspects of linguistic behavior are modality-independent and which are not.

References

AnderBois, Scott, Adrian Brasoveanu & Robert Henderson. 2013. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32(1). 93–138. doi:10.1093/jos/fft014.

Aristodemo, Vita Maria Valentina. 2017. *Gradable constructions in Italian Sign Language*: Institut Jean-Nicod PhD dissertation. http://www.theses.fr/2017EHES0147/abes/Aristodemo-Vita_Maria_Valentina_these_2017.pdf.

Ebert, Cornelia. 2017. Co-speech vs. post-speech gestures. Talk given at *Language and cognition workshop in memory of Peter Bosch*, Osnabrück, February.

Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Talk given at *Semantics and Philosophy in Europe (SPE 7)*, Berlin, June.

Esipova, Maria. 2018a. Focus on what's not at issue: Gestures, presuppositions, supplements under contrastive focus. In Uli Sauerland & Stephanie Solt (eds.), *Sinn und Bedeutung 22*, 385–402. Berlin: Leibniz-Centre General Linguistics.

Esipova, Maria. 2018b. Gestures at the interfaces. Talk given at the workshop *Foundational Topics in Semantics*, New York University. https://osf.io/9dp36/.

Esipova, Maria. 2019. Acceptability of at-issue co-speech gestures under contrastive focus. *Glossa: A Journal of General Linguistics* 4(1). 11. doi:10.5334/gjgl.635.

Heim, Irene. 1983. On the projection problem for presuppositions. In Michael Barlow, Daniel Flickinger & Michael Wescoat (eds.), *West Coast Conference on Formal Linguistics (WCCFL) 2*, 114–125.

Hunter, Julie. 2018. Relating gesture to speech: reflections on the role of conditional presuppositions. *Linguistics and Philosophy* 1–16. doi:10.1007/s10988-018-9244-0.

Jasinskaja, Katja & Claudia Poschmann. 2018. Attachment in syntax and discourse: towards an explanation for the variable scope of non-restrictive relative clauses. In Sireemas Maspong, Brynhildur Stefánsdóttir, Katherine Blake & Forrest Davis (eds.), *Semantics and Linguistic Theory (SALT) 28*, 433–453. doi:10.3765/salt.v28i0.4438.

Karttunen, Lauri. 1974. Presuppositions and linguistic context. *Theoretical Linguistics* 1. 181–194. doi:10.1515/thli.1974.1.1-3.181.

Koev, Todor. 2013. *Apposition and the structure of discourse*: Rutgers University PhD dissertation. https://ling.rutgers.edu/images/dissertations/Koev-2013-Dissertation-Final-Version.pdf.

Leffel, Timothy. 2014. *The semantics of modification: Adjectives, nouns, and order*: New York University PhD dissertation. https://ling.auf.net/lingbuzz/002212.

Matushansky, Ora. 2008. On the linguistic complexity of proper names. *Linguistics*

and Philosophy 31(5). 573–627. doi:10.1007/s10988-008-9050-1.

Morzycki, Marcin. 2008. Nonrestrictive modifiers in nonparenthetical positions. In Christopher Kennedy & Louise McNally (eds.), *Adjectives and adverbs: Syntax, semantics, and discourse* Studies in Theoretical Linguistics, 101–122. Oxford: Oxford University Press.

Morzycki, Marcin. 2015. *Modification*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511842184.

Partee, Barbara & Mats Rooth. 1983. Generalized conjunction and type ambiguity. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use and interpretation of language*, 361–383. Berlin: de Gruyter. doi:10.1515/9783110852820.361.

Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199273829.001.0001.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org.

Schlenker, Philippe. 2009. Local contexts. *Semantics and Pragmatics* 2(3). 178. doi:10.3765/sp.2.3.

Schlenker, Philippe. 2010. Presuppositions and local contexts. *Mind* 119(474). 377–391. doi:10.1093/mind/fzq032.

Schlenker, Philippe. 2013. Supplements without bidimensionalism. Ms., Institut Jean-Nicod and New York University. http://www.semanticsarchive.net/Archive/jgwMjNmM/.

Schlenker, Philippe. 2018a. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295–365. doi:10.1007/s10988-017-9225-8.

Schlenker, Philippe. 2018b. Iconic pragmatics. *Natural Language & Linguistic Theory* 36(3). 877–936. doi:10.1007/s11049-017-9392-x.

Schlenker, Philippe. 2018c. Iconic presuppositions. Ms., Institut Jean-Nicod and New York University. https://ling.auf.net/lingbuzz/004116.

Stalnaker, Robert. 1974. Pragmatic presuppositions. In Milton Kunitz & Peter Unger (eds.), *Semantics and philosophy*, 197–213. New York University Press.

Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: A Journal of General Linguistics* 2(1). 102. doi:10.5334/gjgl.334.

Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2018. Co-speech gesture projection: Evidence from inferential judgments. *Glossa: A Journal of General Linguistics* 3(1). 109. doi:10.5334/gjgl.580.

Zlogar, Christina & Kathryn Davidson. 2018. Effects of linguistic context on the acceptability of co-speech gestures. *Glossa: A Journal of General Linguistics*

3(1). 73. doi:10.5334/gjgl.438/.

Maria Esipova
Department of Linguistics
10 Washington Place
New York, NY 10003, USA
masha.esipova@nyu.edu