

Composition and projection of co-speech gestures

Overview The current consensus in the literature is that the content contributed by gestures co-occurring with speech, *co-speech gestures*, as in (1), tends to project from under semantic operators (e.g., Ebert & Ebert 2014; Tieu et al. 2017, 2018; Schlenker 2018). But there is no consensus on what mechanism assures said projection. Ebert & Ebert (2014) claim that co-speech gestures are akin to appositives and project as such. Schlenker (2018) argues that co-speech gestures trigger assertion-dependent presuppositions, *cosuppositions*. I show that neither analysis can on its own account for all available interpretations of nominal gestures without yielding any unavailable ones. I propose instead that how a gesture projects depends on how it composes. When it adjoins to an NP (type *et*), it composes as a *modifier*, akin to an adjective. However, there is a pragmatic preference for co-speech gestural modifiers to be non-restricting, which gives rise to projecting inferences akin to those of non-restricting adjectives. I re-purpose Schlenker’s cosuppositions as a modality-neutral mechanism for generating such inferences. When a gesture adjoins to a DP (type $\langle et, t \rangle$), it composes and projects as a *supplement*, akin to an appositive. This proposal is compliant with a view whereby there are no gesture-specific composition or projection strategies.

(Un)available interpretations of adnominal adjuncts: experimental data Design MTurk participants were asked to read context paragraphs and watch videos of sentences uttered in those contexts. They were then asked to assess these sentences on a pseudo-continuous scale from ‘Totally unnatural’ to ‘Totally natural’ (mapped to 0–100). The items differed across two factors: content type (adjective, appositive, gesture) and interpretation (projecting non-restricting (PNR), restricting (R), non-projecting non-restricting (NPNR)), resulting in 9 conditions. The target interpretations were enforced within the videos by setting up the QUD and contrasting two alternatives; the written contexts were meant to maximally support this interpretation. 4 complete test paradigms were constructed; a sample paradigm is given in (2) (names are abbreviated for anonymity). Each participant saw 1 randomly selected item per condition. The final number of participants was 122 (33 female, 89 male).

Results The results are visualized in Figure 1. They suggest that no content type allows for non-projecting non-restricting interpretations. For adjectives, projecting non-restricting and restricting interpretations are both available. For appositives, only projecting non-restricting, but not restricting interpretations are available. For gestures, projecting non-restricting interpretations are easily available, and restricting interpretations are marginally available.

Discussion The contrast between appositives and gestures regarding availability of restricting interpretations suggests that the appositive-like strategy cannot be the only interpretation strategy available to co-speech gestures, contra Ebert & Ebert 2014.

Cosuppositions of NP- and DP-level gestures Schlenker’s (2018) cosupposition strategy alone isn’t sufficient to account for the data at hand either: it yields good predictions for co-nominal gestures adjoining to NPs, but not for those adjoining to DPs.

Gestural cosuppositions A gestural cosupposition of the configuration $[[S]^G]$ has the form $S \Rightarrow G$, where S is the denotation of the spoken expression the gesture adjoins to, G is the denotation of the gesture, and \Rightarrow is generalized entailment. When this cosupposition projects, a pragmatically computed local context (Schlenker 2009) c' of $[[S]^G]$ has to entail it: $c' \Rightarrow (S \Rightarrow G)$. Schlenker also allows for local accommodation under pressure (e.g., Heim 1983); an accommodated cosupposition is conjoined to S : $S \& (S \Rightarrow G)$, which is equivalent to $S \& G$ ($\&$ is generalized conjunction). For this mechanism to apply, the denotations of S ,

G , and c' all have to be of the same semantic type, otherwise generalized entailment and conjunction are impossible. Thus, the type of the constituent the gesture adjoins to matters.

Cosuppositions of NP-level (et) gestures Given the right set of assumptions about the denotations of S , G , and c' , the mechanism above yields good results for NP-level gestures. The results for *Stephanie brings her* $[[dog]^{LARGE}]$ are given in (3). When the cosupposition projects, the resulting inference is roughly equivalent to ‘If Stephanie brings a dog of hers, that dog is large’. Given that size is an individual-level property, it is natural to conclude that Stephanie’s dog is large tout court. When the cosupposition is locally accommodated, the result is equivalent to treating the gesture as a restrictive modifier, yielding ‘large dog’.

Cosuppositions of DP-level ($\langle et, t \rangle$) gestures Regardless of the specific assumptions, allowing local accommodation for DP-level gestures yields unattested interpretations. For example, if we assume that in *Stephanie brings* $[[her\ dog]^{LARGE}]$ the gesture denotes something like ‘a large object’, conjoining it to the DP *her dog* will result in a completely unattested interpretation ‘Stephanie brings her dog and a large object’. Introducing an anaphoric link between *her dog* and the gesture yields the non-projecting non-restricting interpretation (see (4)), which was shown to be unavailable by the experiment above. Thus, if we wanted to maintain that the cosupposition strategy is the only one available to co-speech gestures, we would either have to say that local accommodation can apply to NP-level, but not DP-level gestures, or that co-nominal gestures can’t adjoin to DPs. Neither option seems motivated.

Proposal: composition determines projection I propose that how gestures project is determined by how they compose. But how do gestures compose?

No gesture-specific compositionality I propose that syntax and semantics proper are modality-blind, i.e., when gestures compositionally integrate into utterances, they do so just like spoken content. Any modality-specific effects arise in pragmatics or phonology and its interfaces. This is a conceptually appealing view, compatible with late lexical insertion.

Two composition strategies exist in the nominal domain; both are available to gestures.



Modifiers, exemplified by adjectives, adjoin to NPs in the syntax; semantically, they compose with sets of individuals and return subsets thereof, i.e., modifiers are necessarily *restrictive*. However, the subsets they return don’t have to be proper subsets of the input, i.e., modifiers don’t have to be *restricting*. (The *restrictive* vs. *restricting* distinction was made explicit in Schlenker To appear.) Non-restricting modifiers are truth-conditionally vacuous, but they give rise to strongly projecting inferences about all the elements of the input set. I re-purpose Schlenker’s cosuppositions as a modality-neutral mechanism yielding such inferences. To allow for $\langle et, et \rangle$ -type modifiers, I propose that cosuppositions require that the NP being modified entail the result of the modification (instead of the modifier).

Supplements, exemplified by appositives, adjoin to DPs and contribute obligatorily projecting propositional content about them; they aren’t restrictive and thus can’t be restricting. Any existing implementation will do for my purposes (e.g., Potts 2005; Koev 2013).

Preference for non-restricting gestures I follow Schlenker’s original intuition that there is a pragmatic pressure for co-speech gestures to be truth-conditionally vacuous. This means that when co-nominal gestures are modifiers, they prefer to be non-restricting. The pragmatic nature of this constraint allows for gradience and variability in acceptability of restricting co-speech gestures, which is in line with the data above and those in Esipova To appear.

Future work My next step is to show that the modifier vs. supplement distinction exists for spoken and gestural adjuncts in the verbal domain and can be analyzed in the same way.

Notation: In ‘word^{GESTURE}’ GESTURE co-occurs with *word*; underlining doesn’t show syntactic attachment. Each gesture is illustrated once. **Bold** indicates prosodic contrastive focus.

- (1) If Stephanie’s bringing her dog^{LARGE} , we should get a bigger van.
 → Stephanie’s dog is large.
- (2) *Context: We are going on a group tour. A. and M. are responsible for renting a van. ...*
 a. *PNR: ... M. just told A. that Stephanie, who has two pets, a small cat and a large dog, is planning to bring along one of her pets. A., 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. *R: ... M. just told A. that Stephanie, who has two dogs, a small Pug and a large Great Dane, is planning to bring along one of her dogs. A., 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. *NPNR: ... M. just told A. that Stephanie is planning to bring along her dog. A. 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^{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.
 (Intended: ‘...if (Stephanie’s dog is large and she is bringing her (large) dog)...’)
- (3) Stephanie brings her $[[_{NP} \text{dog}]^{\text{LARGE}}]$.
 a. $[[S]] = \lambda x. \text{dog}(x)$ b. $[[G]] = \lambda x. \text{large}(x)$ c. $[[c']] = \lambda x. \text{bring}(s, x) \wedge \text{poss}(s, x)$
 d. projection ($c' \Rightarrow (S \Rightarrow G)$): $\forall x. (\text{bring}(s, x) \wedge \text{poss}(s, x)) \rightarrow (\text{dog}(x) \rightarrow \text{large}(x))$
 e. local accommodation ($S \& (S \Rightarrow G)$): $\lambda x. \text{dog}(x) \wedge \text{large}(x)$
- (4) Stephanie brings $[[_{DP} \text{her dog}]_i^{\text{LARGE}_i}]$.
 a. $[[S_i]]^g = \lambda P. P(\iota x. \text{dog}(x) \wedge \text{poss}(s, x))$ b. $[[G_i]]^g = \lambda P. P(g(i)) \wedge \text{large}(g(i))$
 c. local accommodation: $\lambda P. P(\iota x. \text{dog}(x) \wedge \text{poss}(s, x)) \wedge \text{large}(\iota x. \text{dog}(x) \wedge \text{poss}(s, x))$

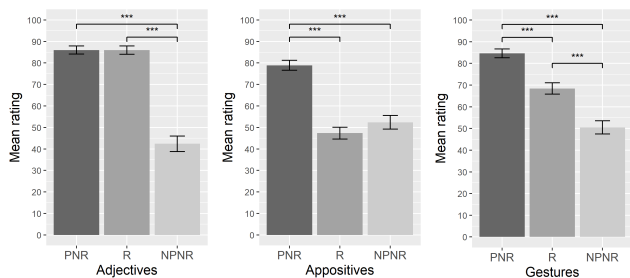


Fig. 1: Acceptability ratings of different interpretations for different content types.

References Ebert & Ebert. 2014. *SPE* 7. Esipova. To appear. *Glossa*. Heim. 1983. *WCCFL* 2. Koev. 2013. PhD thesis. Potts. 2005. OUP. Schlenker. 2009. *S&P* 2(3). Schlenker. 2018. *L&P* 41(3). Schlenker. To appear. *Blackwell Companion to Semantics*. Tieu et al. 2017. *Glossa* 2(1). Tieu et al. 2018. *Glossa* 3(1).