Towards a uniform cross-modal typology of composition and projection

Background Formal semantics/pragmatics is seeing an upsurge in studying *secondary modality content* like gestures and facial expressions (Ebert & Ebert 2014; Schlenker 2018a,b; Tieu et al. 2017, 2018, a.o.). Much of this work focuses on how such content *projects*, e.g., how the inferences arise in examples with *co-speech gestures* like (1).[1]

(1) a. If Lea brings her dog$_{\text{LARGE}}$, we'll need a bigger van.   b. If Zoe shoots at the target$_{\text{LONGBOW}}$, I'll give you \$10.
→Lea's dog is large.   →If Zoe shoots, she'll be shooting a longbow.

Most of this literature assumes that projection of secondary modality content is governed by independent rules. This approach is most prominent in Schlenker 2018b, which aims to predict if/how a piece of secondary modality content projects, based on (i) whether it co-occurs with something in a more primary modality, and (ii) whether it's "syntactically eliminable". Thus, co-speech gestures co-occur with speech and are eliminable; Schlenker claims that such content uniformly triggers *cosuppositions* (assertion-dependent presuppositions).

Proposal I argue that projection behavior of compositionally integrated content is guided by the same linguistic principles in all modalities. Specifically, how a piece of content projects is determined by how it composes in the syntax/semantics. A typology like Schlenker's is thus both untenable and unnecessary.

Modifiers vs. supplements I focus on two composition strategies and associated projection behavior patterns.

(Subsective) modifiers, e.g., spoken adjectives, combine with an expression β yielding α that entails β via generalized entailment. Modifiers can but don't have to be restricting; *non-restricting modifiers* (NRMs) are subsective modifiers intended by the speaker as truth-conditionally vacuous (Leffel 2014). I argue that NRMs pragmatically trigger inferences that β entails α that project like presuppositions, i.e., relative to local contexts:

(2) a. *(Context: the speaker believes that processed meat causes cancer.)* I shouldn't eat so many deadly sausages.
→I shouldn't eat so many sausages.   (truth-conditional vacuity)
→Being a sausage entails being a deadly sausage (i.e., all sausages are deadly).   (NRM inference)
b. Maybe processed meat causes cancer and I shouldn't eat so many deadly sausages.
→Maybe processed meat causes cancer and I shouldn't eat so many sausages.   (truth-conditional vacuity)
→If processed meat causes cancer, then all sausages are deadly.   (NRM inference)

Supplements, e.g., spoken appositives, combine with an expression yielding a proposition of a special kind about it (e.g., Potts 2005; Koev 2013). They don't have the compositional potential to be restricting, and they typically must project:

(3) #If Lea brings her dog, a **small** animal, that's OK, but if she brings her dog, a **large** animal, we'll need a bigger van.
✗ ≈If Lea brings her ⟨**small**, **large**⟩ dog...   ✗ ≈If (Lea brings her dog and it is ⟨**small**, **large**⟩)...

Case studies Secondary modality content can compose as modifiers or supplements, too, with similar contrasts in projection.

Co-speech gestures I follow Esipova 2019 in that gestures like in (1) can be construed as modifiers (cf. *large*) or supplements (cf. *a large object*). As modifiers, they can be restricting, but marginally so, as they prefer to be truth-conditionally vacuous:

(4) (*Context: Lea has two dogs.*)   (see Esipova 2019 for experimental data)
?If Lea brings her **dog**$_{\text{SMALL}}$, that's OK, but if she brings her **dog**$_{\text{LARGE}}$, we'll need a bigger van.

The gestures in (1) aren't conventional, though, and their potential construals are constrained by their iconic content only. I observe that some conventional gestures can be modifiers, too, with similar consequences for projection, e.g.:

(5) If you bring a semanticist$_{\text{CRAZY}}$ to my talk, I'll likely fight with them.
✓ →All semanticists are crazy.   ?✓ ≈If you bring a **crazy** semanticist to my talk...

But some conventional gestures can only be supplements (cf. sentence-level adverbs like *hopefully*), e.g.:

(6) If a friend of mine wins the race$_{\text{FINGERS-CROSSED}}$, I'll buy them a drink.
✓ →I want a friend of mine to win the race.   ✗ ≈If (a friend of mine wins the race and I wanted them to win)...

Schlenker can't account for the different projection patterns in (5) vs. (6), but excluding conventional gestures from consideration will miss the analogy between (1) and (5). My view, whereby composition determines projection across modalities (and various considerations, iconic or conventional, can constrain composition), applies uniformly to all gestures.

Co-speech/gesture facial expressions I observe that mirative facial expressions (eyes wide open, O_O) and spoken adverbs can both compose as sentence-level supplements, which must project, or as degree modifiers, which don't project by default:

(7) a. When, surprisingly, a friend of mine gets drunk, I sometimes comment on that.
b. When a friend of mine gets drunk$_{\text{O\_O}}$, I sometimes comment on that.   (under the sentence-level construal)
→When a friend of mine gets drunk, this fact is {surprising, impressive}.

(8) a. When a friend of mine gets surprisingly drunk, I sometimes comment on that.
b. When a friend of mine gets drunk$_{\text{O\_O}}$, I sometimes comment on that.   (under the degree modifier construal)
↛When a friend of mine gets drunk, they do it to a surprising extent.

The same holds if we replace the word *drunk* with a conventional Russian gesture $DR\{I,U\}NK$ (flicking a finger on the neck) to eliminate the effect of voice modulations on *drunk* in (8b). Schlenker can't account for the two uses of O_O. The uniform composition-driven view captures both easily.

Selected references Ebert & Ebert. 2014. *SPE 7*. Esipova. 2019. *SALT 29*. Koev. 2013. PhD thesis. Leffel. 2014. PhD thesis. Potts. 2005. *OUP*. Schlenker. 2018. *L&P*. Schlenker. 2018. *NLLT*. Tieu et al. 2017. *Glossa*. Tieu et al. 2018. *Glossa*.

[1] In word$_{\text{GESTURE}}$ the gesture co-occurs with the word; **bold** indicates prosodic contrastive focus marking.