Contrastive inferences across languages and across modalities

Introduction. Based on increasing recent interest in how linguistic meaning may interact and/or compose with meaning beyond language (Schlenker, 2018; Tieu et al., 2019; Pasternak & Tieu, 2022; Storment, 2024; among many others), we focus here on depictive gestures accompanying language. While several recent semantic analyses build on the observation that co-*speech gestures* are frequently non-restrictive (Ebert & Ebert, 2014; Schlenker, 2018; Esipova, 2019), their pragmatic contribution remains under-explored, and entirely unexplored in a sign language, where co-linguistic depictions share a modality with the language. Our cross-linguistic study asks whether depictive gestures in English, Mandarin, and ASL support *contrastive inference*, a phenomenon in which language users attribute the use of modification (e.g., *tall glass*) to the need for referential informativity (e.g., the presence of more than one glass) (Sedivy et al., 1999; Alsop et al., 2018; Kreiss & Degen, 2020). If depictive co-speech/sign gestures in general are non-restrictive, we expect a lack of contrastive inference across the board; if such inference does not arise in speech but does arise in sign, we may posit that gestures' inability to restrict in speech is due to a modality divide between depiction and verbal content and/or their inability to bear focus in speech. Finally, if contrastive inference is nonetheless observed in more gradient ways, perhaps varying by language and/or modality, it suggests the need to consider more complex possibilities to encode depiction in nominal semantic structures.

Methods. For all three languages, we employed an offline forced-choice paradigm adapted from Alsop et al. (2018). We selected this approach over standard visual-world eye-tracking for contrastive inference (e.g., Sedivy et al., 1999) for two methodological reasons driven by modality: First, all of our stimuli are visual (videos containing gestures), so they demand the participant's visual attention. In an eye-tracking paradigm, attending to the video stimulus competes with scanning the target objects, a confound not present in audio-only studies. Second, while eye-tracking captures rapid, incremental processing, offline measures are better suited to capturing the final interpretive output of potentially complex pragmatic calculus. If the pragmatic enrichment triggered by gestures requires more reasoning than adjectival modification, an offline task provides the necessary window to observe the final interpretation.

Procedure. Participants completed a Qualtrics questionnaire. They were familiarized with a scenario involving a speaker and a listener where a visual presentation of objects was hidden from view. In each trial, participants viewed a video of a request (e.g., "Pass me the [modifier] noun") and were asked to infer the likely pair of objects by choosing between two arrays: a non-contrastive set (target and different-category distractor) and a contrastive set (target and same-category competitor). Crucially, participants were informed that trials might not have a single "correct" answer and were instructed to select the option that "made the most sense."

Figure 1: Choices for the item "bowl"
Left: non-contrastive set; Right: contrastive set

Design. We manipulated the modification of noun phrases in the request as the independent variable, using a within-subject design. This factor had three levels in English and Mandarin (unmodified; adjective; gesture), and two in ASL (unmodified; modified: SASS classifiers), where ASL size and shape classifiers (SASS) are compared to both spoken adjectives and gestures since these modifiers have a dedicated syntactic and prosodic position (like adjectives) but depict specific sizes and shapes (like gestures). The dependent variable was array choice.

Stimuli. We created 15 critical items for English and Mandarin and 16 for ASL. We used scalar or dimensional adjectives (e.g., *small*, *square*) in the adjective condition, and corresponding depictive gestures (normed for acceptability) in the gesture condition. Following a Latin Square design, participants viewed each item once. Participants completed 19 (English: 15 critical, 4 catch), 18 (Mandarin: 15 critical, 3 catch), or 20 (ASL: 16 critical, 4 catch) fully randomized trials.

Table 1: Example stimuli for the item "bowl" in each language, per condition

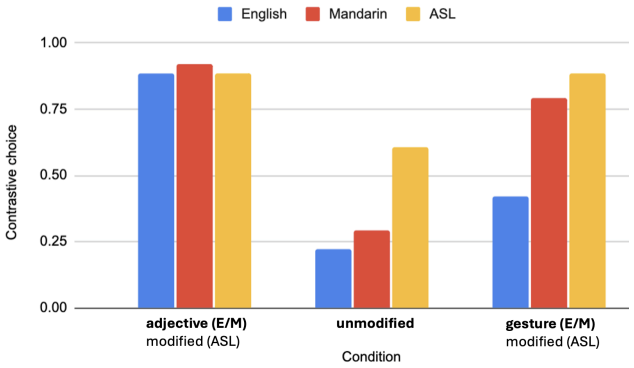
	unmodified	adjective	gesture
English	"Pass me the bowl"	"Pass me the small bowl"	"Pass me the $[SMALL]_g$[1] bowl"
Mandarin	"Digei wo zhe ge wan" (Pass me this Cl. bowl)	"Digei wo zhe[2] ge xiaode wan" (Pass me this Cl. small bowl)	"Digei wo zhe ge $[XIAODE]_g$ wan" (Pass me this Cl. $[SMALL]_g$ bowl)
ASL	 BOWL 2-GIVEo-1	**modified**  Nmm: oo BOWL SMALL 2-GIVEo-1	

Participants. We analyzed data from 74 native English speakers, 65 native Mandarin speakers (both recruited via Prolific), and 20 native ASL signers (snowball sampling), after exclusion.

Results. Data were analyzed using Generalized Linear Mixed-Effects Models with a binomial family, including random intercepts for participants and items. Treatment coding was adopted, with the unmodified condition as the reference level. We observed a robust main effect of modification in all three languages, for both gesture and adjective. (***: $p < 0.001$)

	English	Mandarin	ASL	
gesture	β=1.37 ***	β=2.89 ***	**modified**	β=1.92 ***
adjective	β=4.42 ***	β=4.13 ***		

Figure 2: Cross-Linguistic Contrastive Inference Comparison



Discussion. Robust contrastive inference was observed in the adjective and gesture conditions for English and Mandarin, and in the modified condition for ASL, suggesting depictive co-speech/sign gestures do elicit contrastive inference. Rather surprisingly, there is significantly higher contrastive inference in the gesture condition in Mandarin than in English, raising intriguing hypotheses on differences in prosody and/or determiner semantics for future experimentation.

Conclusion. Our cross-linguistic findings are in tension with the view that depictive co-speech/sign gestures are universally non-restrictive; rather, we find variability by language and modality. Under pragmatic accounts where informativity of a definite noun phrase is measured by its extension, non-restrictive modifiers should not elicit contrastive inference. That they do so, albeit with varying strength across languages, invites us to reassess the semantic status of depictive co-linguistic gestures or refine our pragmatic measures of informativity.

Selected References. ❖ **Alsop, A., Stranahan, E., & Davidson, K. (2018)**. Testing Contrastive Inferences from Suprasegmental Features Using Offline Measures. *Proc. LSA, 3*(1), 71:1–15 ❖ **Ebert, C., & Ebert, C. (2014)**. Gestures, demonstratives, and the attributive/referential distinction. *Semantics and Philosophy in Europe*. ❖ **Esipova, M. (2019)**. *Composition and projection in speech and gesture*. NYU Dissertation. ❖ **Kreiss, E., & Degen, J. (2020)**. Production Expectations Modulate Contrastive Inference. *Proc. CogSci, 42*, 259-265. ❖ **Pasternak, R., & Tieu, L. (2022)**. Co-linguistic content inferences: From gestures to sound effects and emoji. *QJEP, 75*(10), 1828-1843. ❖ **Schlenker, P. (2018)**. Gesture projection and cosuppositions. *L&P, 41*(3), 295–365. ❖ **Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999)**. Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147. ❖ **Storment, J. (2024)**. Going ✈️ lexicon? The linguistic status of pro-text emojis. *Glossa, 9*(1), 1-43. ❖ **Tieu, L., Schlenker, P., & Chemla, E. (2019)**. Linguistic inferences without words. *PNAS, 116*(20), 9796-9801.

[1] Gestures co-occurring with the immediately following spoken word are represented by this bracket notation: $[\,]_g$.

[2] We used a demonstrative determiner in the Mandarin stimuli because Mandarin doesn't have a definite determiner, and that it would be very inappropriate to use bare nouns due to their compatibility with indefinite interpretations.