Picturing Syntax: Cross-Linguistic Variation in the Interpretation of Emoji Strings

Katherine Gerke and Dennis Ryan Storoshenko - University of Calgary

Issue Cohn (2016) proposes that Jackendoff's model for the interaction of Phonology, Syntax, and Conceptual Structure can be extended to include graphical modes of communication. We use emoji strings to test whether syntactic structure can be overlaid onto entirely graphical modalities. We conclude: (i) emoji string interpretation is language-dependent, not merely image-based, (ii) speakers of different languages use their syntax to differing degrees in this task, and (iii) reading direction is also important in the processing of emoji strings.

Study Design Our stimuli are three-character emoji strings containing two animal faces (the nouns), and one object denoting a verb. These were presented in text message conversation screenshots where one interlocutor guesses the meaning of an emoji string sent by the other. Using a forced-choice task, participants judge whether the guess is acceptable or not. The verb's position was manipulated to be first, second, or third in the string. Guesses always include a transitive verb, and taking the agent to be either animal, we test all six S-V-O permutations. There is an additional verb-direction variable: four verbs are neutrally-directed (e.g. telephone "call"), four are left-directed (e.g. gun "shoot"), and four are right-directed (e.g. knife "stab"). Each emoji string was seen once by each participant, with each verb used in each of the three positions, yielding 36 strings (and 72 guesses). Participants saw each string once, with counterbalanced lists of guesses, plus 46 fillers. Our participants are 12 L1-English speakers, 12 L1-Japanese speakers and 6 L1-Arabic speakers (with 6 more planned). Stimuli were prepared in Apple iMessage, using the regional settings for each language to make the stimuli as natural as possible.

Results Our statistical analyses cover rating and reaction time data across all conditions within each language; for space, here we only present the acceptance rates for the verb-medial conditions.

	The mouse V-ed the frog			The frog V-ed the mouse		
	🐭🔫🐸	🐭☎️🐸	🐭🔪🐸	🐸🔫🐭	🐸☎️🐭	🐸🔪🐭
Arabic	33.33%	58.33%	41.67%	83.33%	91.67%	100.00%
English	62.50%	100.00%	100.00%	62.50%	29.17%	29.17%
Japanese	41.67%	87.50%	66.67%	66.67%	25.00%	0.00%

In the Generalized Linear Model analysis for the English rating data, we find significant preferences for verb-medial strings, and for left-positioned agents. There is no significant effect of verb direction, nor any interaction of this with other variables. (though direction seems to play a role in the gun strings). For Japanese, there is again a significant preference for left-positioned agents. We also find interactions between verb position and direction, between verb and agent position, and a significant three-way interaction. While Arabic data collection is not complete, we preliminarily report a significant effect of agent position, now strongly preferred on the right.

Discussion and Conclusion The cross-linguistic effects are clear (claim i). Japanese is however most clearly sensitive to the visual composition of the string over syntax. This is seen in the outright rejection of *The frog stabbed the mouse* where the knife points at the frog; however, most telling is that among all conditions where the verb emoji was rightmost, which should mimic the verb-final syntax of Japanese, the highest acceptance rate was 37.5% for an SOV reading of the neutrally-directed verbs. Much as Cohn et al. (2012) find differences between American and Japanese comic books, we find differences in the reading of emoji (claim ii). Looking at Arabic results where the agent is positioned rightmost, we believe we are observing an effect of reading direction; these would be SVO, which is a possible word order in Arabic. Note that the highest preference among these goes opposite to the verb direction, suggesting this is not as salient a feature for these

speakers (claim iii). Beyond showing that Cohn's extension of Jackendoff's model is supported, these findings have ramifications for industry, as emoji direction is not consistent across phone models, and does not reverse with the reading direction of the language settings.

References

Cohn, Neil. 2016. A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition* 146:304–323.

Cohn, Neil, Amaro Taylor-Weiner, and Suzanne Grossman. 2012. Framing attention in Japanese and American comics: Cross-cultural differences in attentional structure. *Frontiers in Psychology* 3:1–2.