

(2022). In R. Page, C. Lee, M. Zappavigna, & J. Unger (Eds.), *Researching the language of social media: A student guide*. Taylor & Francis.

Researching the Syntax of Emoji Sequences on Sina Weibo

Susan C. Herring and Jing Ge-Stadnyk

The linguistic study of emoji is a new area where established research methods are lacking. Our research analyzes the syntactic properties of sequences of non-identical emoji, treating the sequences like sentences. There is debate about whether or not emoji are a “language” (e.g., Gawne & McCulloch, 2019). Syntax is an essential component of language; thus we applied syntactic categories and constructs from verbal language to emoji sequences to see how well they fit and what patterns (if any) they would reveal. Our data were emoji sequences in posts by celebrities and their followers on the Chinese microblogging platform Sina Weibo. We discovered that although the emoji sequences lack certain grammatical features, indicating that they are not yet a full language, there is suggestive evidence of “syntacticizing” patterns, similar to those reported for Pidgin and Creole languages (Givón, 1979). Aware that this claim could be controversial, we needed our syntactic evidence to be robust.

The most challenging aspect of the research was translating the sequences into English. It was important to do this well, because the validity of our syntactic analyses depended on the accuracy of our translations. However, exact translation of emoji is challenging, due in part to their wholistic, context-dependent, and subjective nature (e.g., the single emoji 🤡 in our data could mean “a clown” or “wearing colorful makeup”). Moreover, the meaning of an emoji sequence as a whole may differ from the sum of the meanings of its individual parts, as defined by Weibo or Emoji Unicode. Decoding these sequences requires understanding of popular culture, including TV shows and online slang.

It helped that one author is a native Chinese speaker who could translate the Chinese text accompanying some emoji sequences, and who was familiar with China’s popular culture. She initially translated all of the sequences, taking their surrounding discourse into consideration, while consulting with the other author on about half of them. However, we were unsure of some translations, so additional steps were taken. 1) Via WitMart (a Chinese crowdsourcing platform similar to Amazon’s Mechanical Turk), we recruited crowdsourced emoji interpreters who were active Weibo users and who followed some of the sample celebrities’ Weibo accounts. These culturally-savvy interpreters resolved most of the ambiguous cases. 2) We also consulted with three proficient emoji users recommended by a Chinese graphicon design firm, who provided translations of the remaining cases.

Finally, after analyzing the syntactic phenomena in the sequences (parts of speech, transitivity, word order, etc.), we reviewed each sequence to check for consistency in our interpretations. For sequences with more than one possible syntactic interpretation (e.g., 🍷🍷🍷👍👍👍 “These shows are great!” or “(We) really like the shows!”), we chose the

translation most consistent with other examples. Translations were also adjusted in a few cases based on consultation with Chinese students who were identified as proficient emoji users by a professor in mainland China.

It may not be necessary to follow all these steps to ensure robust emoji translations. But some of the strategies we employed could prove useful in emoji research that requires translation.

#advice: Leverage collective intelligence to interpret uses of graphical icons in social media.

References

Givón, T. (1979). From discourse to syntax: Grammar as a processing strategy. In T. Givón (Ed.). *Discourse and syntax* (pp. 81-112). New York: Academic Press.

Gawne, L., & McCulloch, G. (2019). Emoji as digital gestures. *Language@Internet*, 17, article 2.