

Predictive power of surprisal for multimodal language cues

Beata J. Grzyb¹, Stefan L. Frank², Gabriella Vigliocco¹

¹ Experimental Psychology Department, University College London, London, United Kingdom

² Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

The ecology of language use is face-to-face: children learn language in interaction with caregivers, adults use language mostly in face-to-face contexts. In these contexts, speakers dynamically modify their communicative behaviour in ways that support listener's comprehension. When addressing a child, caregivers use slower speaking rate, longer word duration, and higher pitch (Fernald & Simon, 1984). These modulations are considered to be characteristic of Child Directed Language (CDL). Importantly, these modifications are not only tied to speech. Caregivers use iconic gestures (i.e., gestures that represent referents by depicting aspects of their shape or manipulation) especially in communicative conditions that are more difficult for the child (Vigliocco et al., 2019). Prosodic exaggerations and gestures have been argued to support language comprehension and learning (e.g., Vosoughi et al., 2010; Goldin-Meadow & Wagner Alibali, 2013). But what mechanisms drive these modifications in speakers?

When addressing an adult (Adult Directed Language, ADL), speakers tend to utter unpredictable words, which are more informative, more slowly and they utter more quickly predictable words, which provide less information to the listener (see Jaeger & Buz, 2017, for review). Iconic gestures usually start before their lexical affiliates (Ter Bekke, Drijvers, & Holler, 2020), and have been shown to facilitate listeners' prediction of upcoming words (e.g., Zhang et al., 2021). In addition, there is initial evidence that lexical affiliates of iconic gestures are often less predictable from context (at least in spontaneous narratives) (Beattie & Shovelton, 2000).

The informativeness of a word in context can be measured in terms of surprisal, an information-theoretic measure of how unexpected the word is given the prior linguistic context (Levy, 2008; Shannon, 1948). Existing studies have shown a link between surprisal and word durations in adult-to-adult conversational speech, showing that low-surprisal words (i.e., more predictable) have shorter duration than high-surprisal words (i.e., less predictable) (Demberg et al., 2012; Seyfarth, 2014). We do not know whether a similar link exists in CDL. We also do not know whether surprisal predicts if speakers (either addressing a child or an adult) will produce an iconic gesture.

Here, we examine whether surprisal predicts word durations and gestures in both CDL and ADL. We quantified prediction difficulty in terms of surprisal obtained from recurrent neural network language models trained on either CDL (i.e., 42 data sets of child-directed speech extracted from the CHILDES corpus) or on ADL (Spoken British National Corpus 2014). We then assessed how well the surprisal measures predict modulations of word duration and gesture presence in the ECOLANG corpus of semi-naturalistic interactions between a caregiver and a child (aged 3-4), or between two adults.

We found that speakers use both prosodic modulation and iconic gestures more often in conjunction with high surprisal words (i.e., words harder for the comprehender to predict from the previous context), above and beyond lexical factors (e.g., frequency and length) known to increase production difficulty. This suggests that speakers adapt their multimodal communication on-the-fly to prediction difficulty from the perspective of a child and of an adult, thus suggesting that the production system is optimized for communicative success. These effects were found in child- and adult-directed language suggesting that adaptation to listeners' on-line predictions is a general mechanism in language use.

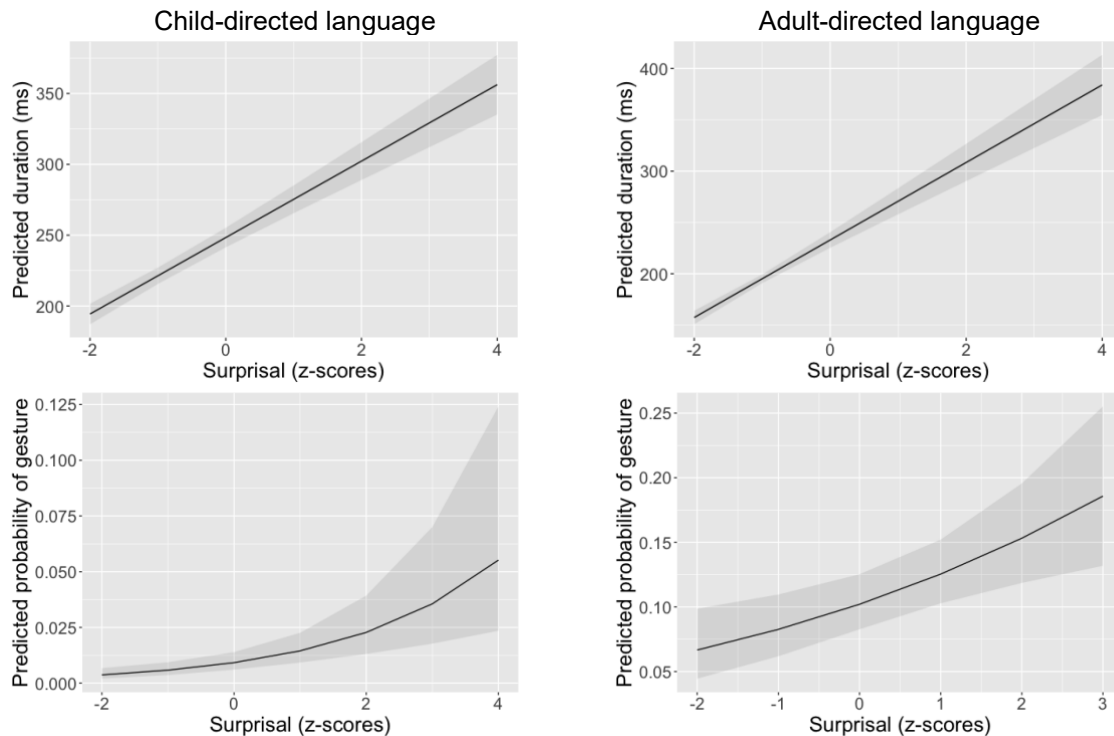


Figure 1. Predicted word durations and probability of iconic gestures in Child-directed and Adult-directed speech.

References

- Beattie, G., & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. *British Journal of Psychology*, *91*, 473 – 491.
- Demberg, V., Sayeed, A., Gorinski, P., & Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *EMNLP 2012* (pp. 356–367). Jeju Island, Korea.
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113.
- Holle H & Gunter TC. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175-92.
- Jaeger, T. F., & Buz, E. (2017). Signal Reduction and Linguistic Encoding. In *The Handbook of Psycholinguistics* (pp. 38–81). John Wiley & Sons, Ltd.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.
- Goldin-Meadow, S., & Alibali, M.W. (2013). Gesture's role in speaking, learning, and creating language. *Annual review of psychology*, *64*, 257-283.
- Seyfarth, S. (2014). Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140-155.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423.
- Ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech, <https://doi.org/10.31234/osf.io/b5zq7>
- Vosoughi, S., Roy, B. C., Frank, M. C., & Roy, D. (2010). Effects of caregiver prosody on child language acquisition. In *Proceedings of the 5th International Conference on Speech Prosody*. Chicago.
- Vigliocco, G., Murgiano, M., Motamedi, Y., Wonnacott, E., Marshall, C., Milan-Maillo, I., & Perniss, P. (2019). Onomatopoeias, gestures, actions and words: How do caregivers use multimodal cues in their communication to children? *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1171-1177).