

MODALITY-SPECIFIC EFFECTS ON LANGUAGE EMERGENCE - AND WHY LOOKING AT INTERACTION ALSO MATTERS

VINICIUS MACUCH SILVA^{1*}, JUDITH HOLLER^{2,3}, ASLI ÖZYÜREK^{4,5}, AND
SEÁN G. ROBERTS⁶

*Corresponding Author: vini.macuch@gmail.com

1. Department of Linguistics, University of Tübingen, Tübingen, Germany
2. Language and Cognition Department, Max Planck Institute for Psycholinguistics, The Netherlands
3. Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands;
4. Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, 6500 AH, 5. Nijmegen, The Netherlands
5. Multimodal Language and Cognition Lab, Radboud University Nijmegen, 6525 HP, Nijmegen, The Netherlands
6. Department of Anthropology and Archaeology, University of Bristol, Bristol, UK

As with other highly specialized scientific fields, attention within the field of language evolution tends to orbit around specific themes, sometimes with limited integration and cross-talk between different subdomains of interest. Take the following case as an example: modality has lately become a prominent topic in linguistic evolutionary research, particularly in the context of experimental studies of early language emergence (e.g., Fay et al. 2013, 2014; Perlman & Cain, 2014; Perlman et al., 2015) and intergenerational language transmission/evolution (e.g., Motamedi et al., 2017a, 2017b; Verhoef et al., 2014). Similarly, interaction has recently started receiving more attention among language evolution researchers (e.g., Macuch-Silva & Roberts, 2016; Micklos, 2014; Roberts & Levinson, 2017), following a wave of large-scale cross-cultural and cross-linguistic studies in interactional linguistics (e.g., Dingemanse et al., 2015; Floyd et al., 2014; Stivers et al., 2009). However, despite gaining momentum within the wider domain of language evolution, topics such as modality and interaction seldom inform one another at a deeper level. Here we present an

experimental study of language emergence which has sought to *combine* both modality and communicative interaction. It draws on quantitative analyses to allow us to directly test the influence of one dimension on the other and ultimately on the early bootstrapping of communication systems.

The experiment

In order to investigate how modality might affect the creation of new communicative symbols, we invited 15 pairs of participants to the lab to play a communication game in which they had to describe items to one another without using words or conventionalized gestures. Following the structure of similar non-verbal referential communication tasks (Fay et al., 2013, 2014), pairs of participants were allocated to separate experimental conditions (n=5 dyads per condition), namely one condition in which players could use only non-linguistic vocalizations (vocal-only condition), one in which they could use only non-conventionalized manual gestures (gesture-only condition), and one in which they could use both vocalizations and gestures (multimodal condition).

The task was to describe novel stimuli which were either auditory or visual in nature and which did not refer to entities, actions, or qualities with conventionalized signals. More concretely, auditory stimuli consisted of 8 sounds resembling both generic natural sounds (e.g., wings flapping) and human-made/ artificial sounds (e.g., door creaking), whereas visual stimuli consisted of 8 images of circles filled with different patterns and shapes (e.g., lines). On each trial, one participant had to communicate an item to their partner, who in turn had to select the correct target item out of a 3-option array. Participants reversed roles after each trial and the game advanced until all items had been communicated by both members of the dyad. Participants' performance was quantified in terms of (i) accuracy (how well dyads did at correctly guessing items), and (ii) efficiency (how long it took them in communicating and guessing those same items). Crucially, we also measured the degree to which participants interacted in the experiment, which we operationalized as the number of turns a dyad needed to complete communication about each item.

Results

The results show differences between gestural and vocal communication, as well as between strictly unimodal communication and a combined use of modalities. For auditory items, participants in the multimodal condition were more efficient than the other conditions, as would be predicted by theories which recognize the

power of multimodal communication. Unexpectedly, for visual items, participants in the vocal-only condition were more efficient than participants in other conditions, though they were less accurate. Additional analyses show that participants in the multimodal condition deployed the vocal and gestural modality to different extents when describing auditory and visual stimuli. Multimodal signals were produced more for visual stimuli in comparison to auditory.

In addition to the above analyses, which show the relative power of each modality both in isolation and in conjunction with one another, we looked at the interplay between participants' overall task performance and their interactive patterns of communication. Our analyses show that accuracy and efficiency are modulated by the amount of interaction participants engage in, as measured in terms of the number of communicative turns taken by members of a dyad in any given trial. Specifically, we found that if participants engage in more trials with extended interaction, i.e. trials in which there is at least one matcher turn in addition to the initial director turn, their accuracy and efficiency improves in subsequent trials. Interestingly, while participants in the gestural and multimodal conditions engaged in extended trials in 5-10% of all trials, participants in the vocal condition engaged in practically no extended trial (only one such trial was found in the entire data set). In other words, participants in the vocal condition interacted considerably less than in other conditions, which might explain their reduced accuracy in describing visual items.

Conclusion

We present results of an experimental study of language emergence which focuses both on communication modality and interaction. Based on a mix of confirmatory and exploratory analyses, we show that modality affects how fast and accurately participants communicate to one another, but also how much they interact with one another, which in turn impacts efficiency and accuracy in the long run. We interpret the results of our analyses both in terms of different representational affordances provided by the vocal and the gestural modalities, and in terms of different constraints imposed by these modalities on spontaneous interaction and task-related negotiation. Crucially, we show that the modality in which participants communicate affects not only their immediate referential power, as measured in terms of trial-by-trial communicative accuracy and efficiency, but ultimately too their overall communicative performance, as evidenced by interaction-mediated boosts to the abovementioned measures. As such, our work highlights the importance of employing complementary analyses

aimed at different dimensions of interest, which might ultimately reveal more fine-grained pictures of one's object of study. In the case at hand, we investigated the role of modality on language emergence taking into account the mediating role of communicative interaction. We demonstrate that both modality and interaction shape how communication is achieved in the absence of conventionalized communicative symbols.

References

- Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., ... & Rossi, G. (2015). Universal principles in the repair of communication problems. *PloS one*, *10*(9), e0136100.
- Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356-1367.
- Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in psychology*, *5*.
- Floyd, S., Rossi, G., Enfield, N. J., Baranova, J., Blythe, J., Dingemanse, M., ... & Zinken, J. (2014). Recruitments across languages: A systematic comparison. In *the 4th International Conference on Conversation Analysis [ICCA 2014]*.
- Macuch Silva, V., & Roberts, S. G. (2016). Language adapts to signal disruption in interaction. In *11th International Conference on the Evolution of Language (EvoLang XI)*.
- Micklos, A. (2014). The nature of language in interaction. In *10th International Conference on the Evolution of Language (EVOLANG X)*.
- Motamedi, Y., Schouwstra, M., & Kirby, S. (2017a). An evolutionary approach to sign language emergence: From state to process. *Behavioral and Brain Sciences*, *40*.
- Motamedi, Y., Schouwstra, M., Culbertson, J., Smith, K., & Kirby, S. (2017b, July). The cultural evolution of complex linguistic constructions in artificial sign languages. In *Proceedings of the 39th annual meeting of the cognitive science society* (Vol. 39). CogSci.
- Perlman, M., & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*(3), 320-350.
- Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, *2*(8), 150152.
- Roberts, S. G., & Levinson, S. C. (2017). Conversation, cognition and cultural evolution. *Interaction Studies*, *18*(3), 402-429.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in

conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587-10592.

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57-68.