Modality Matters

MODALITY MATTERS: AN INTRODUCTION

HANNAH LITTLE[*1] and ASHLEY MICKLOS[2]

[*]Corresponding Author: hannah.little@uwe.ac.uk
[1]University of the West of England
[2]University of California, San Diego

1. Introduction

Modality (the mode in which language is expressed) is a fundamental topic within language evolution. Most notably, modality is at the centre of the debate of whether language emerged originally as gesture-first, speech-first, or multimodal from the start. Further, the affordances provided to users of existing communication systems are modality-dependent. Modality can affect how language is grounded, transmitted and used in interaction and, as a result, feeds into the language evolution debate at every level. Despite this, much work in evolutionary linguistics, especially in the domain of models and artificial language experiments, tends to extrapolate results from only one modality to language generally. However, in order to justify doing this, we need to first better understand the role of modality in linguistic emergence.

Language evolution, and perhaps linguistics more broadly, places a huge amount of importance in finding linguistic universals. This relies on identifying linguistic features and evolutionary processes that are independent of linguistic modality. Modality-based constraints on structures and strategies may allow one system to develop on a particular trajectory, while another cannot. In this workshop, we will celebrate processes where modality matters, as these examples give us the opportunity to identify exactly what physical properties of a modality are affecting linguistic emergence, which in turn allows us to isolate what might be universal at a more cognitive level.

2. Modality Matters

In this workshop, we focus specifically on the role of modality in the emergence of language and linguistic features. Broadly, talks fall into 2 categories: 2.1) how modality affects the emergence of linguistic forms and, 2.2) how modality affects the emergence of mechanisms in interaction (e.g. repair, feedback, turn-timing, etc.). These are each expanded on below. We encourage discussion of linguistic emergence as a result of these mechanisms.

2.1. *Emergence of Linguistic Form*

Modality can affect forms in many ways, some more obvious than others. Modality can affect whether a form is auditory or visual. Modality can affect whether the articulators are paired to each other, larger or smaller, hidden, or coupled to respiration. It can also affect whether signals are fast or slow fading, or how iconic they are. All of these variables (and many more) constrain the forms the signals can take. These constraints may further have knock on effects to other aspects of form. For example, Galantucci, Kroos, and Rhodes (2010) found that modalities having faster rapidity of fading would cause signals to emerge with more combinatorial structure. Other aspects of modality can also influence combinatorial structure in signals. For example, Little et al. (2017) argued that the manual modality has more signal space dimensions than the spoken modality, causing emerging sign languages to go through a phase where they do not use combinatorial structure. With more possibilities for making distinctions between signals, then more signals can exist before combination of those signals becomes necessary. Others have argued that combinatorial structure emerges later in emerging sign languages because of a prevalence for iconicity available in the signed modality (Sandler et al., 2011). Whether the manual modality really is more iconic than the spoken modality is a question asked in this workshop by Ortega and Ozyurek (2018) and Perlman, Little, and Thompson (2018).

There are also questions of whether modality may affect structure at compositional levels. For example, we know that the manual modality allows for simultaneous articulation of signs, where the spoken modality does not. It may be more "natural" to produce (and perceive) signs simultaneously in some contexts. Some have argued that the manual modality used in silent gesture experiments is a good tool to understand "natural" word orders. However, it is important that we understand whether biases for certain word orders orders is modality-specific. In this workshop, there is work exploring this question of whether word order is affected by modality in different contexts (Mudd, Kirby, & Schouwstra, 2018; Towns & Schouwstra, 2018).

2.2. *Emergence of mechanisms in interaction*

It has been argued that pragmatic universals of interaction underlie not only communication, but also the emergence of communicative modes (Evans & Levinson, 2009; Levinson, 2016). That is, universal features of human communicative interaction, such as the mechanisms for turn-taking, feedback, and repair, are pre-requisites for language. Nonetheless, while pragmatic universals hold across modalities, they may be realised differentially (as discussed by Scott-Phillips, Granito, Kendal, and Tehrani, 2018, in this workshop). Pointedly, Sicoli (2016) asks: "What are the affordances carried by the medium and how may that affect the organization of our linguistic practice?" (p. 428).

For instance, in the documentation of other-initiated repair across modalities, it has been found that while repair itself is a ubiquitous feature of interaction (Dingemanse & Enfield, 2015; Dingemanse et al., 2015; Sicoli, 2016), it is carried out with respect to the affordances of a given modality. Restricted requests, for example, are best suited to modalities that allow for parsable units to be efficiently replicated by the addressee such that it can be recognised as initiating repair. Whistled and signed languages do not allow for this affordance as readily as spoken languages, therefore we do not see restricted requests as frequently in these modes. As Sicoli (2016) suggests, "[w]e should consider the possibility that modality constraints might be involved in differences of repair practiced through different communicative modalities" (p. 426). Though repair strategies themselves differ across modalities, they do so in ways predicted by communicative mediums.

Turn-timing analyses, on the other hand, has revealed similarities across modalities. The timing gap between speakers' turns in spoken language is 200 ms (Stivers et al., 2009); de Vos and colleagues (2015) have found that sign language users have similar turn-timing, including isolate signed languages like Kata Kolok (de Vos, 2018). This suggests that generalizations can be made regarding certain interactional mechanisms across modality.

Each communicative mode provides certain affordances for doing interaction. This will be further explored by Silva, Holler, Özyürek, and Roberts (2018), with particular regard to communicative turns. In this workshop we highlight the differences in interaction as governed by modality.

3. Workshop Objective

While modality is a topic that is addressed in many domains within the evolutionary linguistics literature, researchers between fields may not interact enough. We envision that this workshop will be a unique opportunity to assemble people working on topics of modality in many different disciplines. We expect this cross-disciplinary discussion will generate new perspectives, hypotheses, debates and collaborations.

References

de Vos, C. (2018). Does the time-pressure of turn-taking shape the prosody of emerging sign languages? In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 12th international conference (evolangxii).* NCU Press.

de Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology*, *6*, 268.

Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages:

towards a typology of conversational structures. *Open Linguistics*, *1*(1), 96–118.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., Rossi, G., & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLoS ONE*, *10*(9), 1–15.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *The Behavioral and brain sciences*, *32*(5), 429–448; discussion 448–494.

Galantucci, B., Kroos, C., & Rhodes, T. (2010). The effects of rapidity of fading on communication systems. *Interaction Studies*, *11*(1), 100–111.

Levinson, S. C. (2016). Turn-taking in Human Communication - Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Little, H., Eryılmaz, K., & de Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1–15.

Mudd, K., Kirby, S., & Schouwstra, M. (2018). Improvised word order biases are not modality specific: evidence from non-linguistic vocalizations. In H. Little & A. Micklos (Eds.), *The proceedings of the evolang xii modality matters workshop.* Online at urlhttp://hlittle.com/MM/allpaperpdfs/EvoLangMM$_paper6.pdf$.

Ortega, G., & Ozyurek, A. (2018). Systematic allignment between different types of iconicity and semantic domains in silent gesture: raw materials of sign language emergence. In H. Little & A. Micklos (Eds.), *The proceedings of the evolang xii modality matters workshop.* Online at urlhttp://hlittle.com/MM/allpaperpdfs/EvoLangMM$_paper7.pdf$.

Perlman, M., Little, H., & Thompson, R. L. (2018). Iconicity in signed and spoken lexicons: implications for language evolution. In H. Little & A. Micklos (Eds.), *The proceedings of the evolang xii modality matters workshop.* Online at urlhttp://hlittle.com/MM/allpaperpdfs/EvoLangMM$_paper2.pdf$.

Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*, *29*(2), 503-543.

Scott-Phillips, T., Granito, C., Kendal, J., & Tehrani, J. (2018). Intercultural contact and representational form. In H. Little & A. Micklos (Eds.), *The proceedings of the evolang xii modality matters workshop.* Online at url{http://hlittle.com/MM/allpaperpdfs/EvoLangMM_paper4.pdf}.

Sicoli, M. A. (2016). Repair organization in Chinantec whistled speech. *Language*, *92*(2), 411–432.

Silva, V. M., Holler, J., Özyürek, A., & Roberts, S. G. (2018). Modality-specific effects on language emergence - and why looking at interaction also matters. In H. Little & A. Micklos (Eds.), *The pro-*

ceedings of the evolang xii modality matters workshop. Online at url{http://hlittle.com/MM/allpaperpdfs/EvoLangMM_paper3.pdf}.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Peter, J., Ruiter, D., Yoon, K.-e., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *PNAS*, *106*(26), 10587–10592.

Towns, R., & Schouwstra, M. (2018). Interpreting patient animacy in silent gesture. In H. Little & A. Micklos (Eds.), *The proceedings of the evolang xii modality matters workshop.* Online at urlhttp://hlittle.com/MM/allpaperpdfs/EvoLangMM$_paper5.pdf$.

ICONICITY IN SIGNED AND SPOKEN LEXICONS: IMPLICATIONS FOR LANGUAGE EVOLUTION

MARCUS PERLMAN[1], HANNAH LITTLE[2] and ROBIN L. THOMPSON[1]

[*]Corresponding Author: m.perlman@bham.ac.uk

[1]The University of Birmingham, UK

[2]The University of the West of England, Bristol, UK

1. Introduction

Many current theories of language origins posit that *iconicity*, or resemblance between form and meaning, played a critical role in grounding the creation of the first symbol systems (e.g. Fay, Arbib, & Garrod, 2013; Kendon, 2008; Perlman, Dale, & Lupyan, 2015). Some popular gesture-first accounts of language origins hinge on the premise that visible gestures afford dramatically more iconicity than audible vocalizations (e.g. Armstrong & Wilcox, 2007; Tomasello, 2008; Goldin-Meadow, 2016). Therefore, the rise of spoken words must have depended on the already established use of iconic signs, which served originally to ground the connection between spoken form and meaning. Part and parcel of this hypothesis is the assumption that many signs of modern signed languages clearly derive from iconic (and indexical) gestures (e.g. Klima & Bellugi, 1979), whereas, in general, the words of spoken languages are arbitrary to their core, stemming back to their original coinage (e.g. Pinker & Bloom, 1989).

However, considerable evidence now shows that the lexicons of all languages exhibit a significant amount of iconicity (Perniss, Thompson, & Vigliocco, 2010; Dingemanse et al., 2015). Moreover, this research suggests that there may be some important differences in how iconicity is spread across signs in comparison to words. Specifically, some kinds of meanings, but not others, may lend themselves to highly iconic signs, while different kinds of meanings may afford more or less iconicity in words (Dingemanse et al., 2015; Perlman & Cain, 2014). For example, Dingemanse et al. (2015) propose that qualities like spatial relations and visual shape might be easier to represent with iconic signs, but harder with iconic words. Conversely, qualities related to sound and loudness might easily be iconically represent with words, but not signs. This research

challenges the basic assumption of gesture-origin theories that signed languages are categorically more iconic than spoken languages. Furthermore, it raises the possibility that the first languages were iconic and multimodal (Perlman, 2017). The first symbol systems were built utilizing the differential suitability of each modality for the iconic representation of different kinds of meanings. Here, we investigate what such an original system might have looked like, based on a more detailed understanding of how iconicity is manifested in the lexicons of signed and spoken languages. We use iconicity ratings to compare iconicity in the vocabularies of two signed languages - American Sign Language (ASL) and British Sign Language (BLS) - with two spoken languages - English and Spanish.

2. Methods

Our study used publically available iconicity ratings of 993 ASL signs (Caselli et al. 2017), 604 BSL signs (Vinson et al., 2008), 3001 English words (Winter et al., 2017), and 637 Spanish ratings (Perry et al. 2015). These ratings were collected by different procedures. In both languages, signs were rated on a scale from 1 (arbitrary) to 7 (iconic). ASL signs were rated by non-signers, as were some BSL signs, but some were rated by native BSL signers. Spoken words were all rated by native speakers on a scale from -5 (sounds like the opposite of what it means) to 5 (sounds like what it means), with arbitrary signs being 0.

In our analysis, we examined 1) the correlation in iconicity ratings between the languages; 2) the relationship between iconicity and an array of semantic variables; 3) how iconicity varies between broad lexical classes; and 4) between more specific semantic categories (e.g. manual verbs, clothes, colors).

Our analyses made use of ratings for several variables related to the semantics of the signs and words, including concreteness (Brysbaert et al., 2014), imageability (Cortese & Fugett, 2004), sensory experience (Juhasz & Yap, 2013), and perceptual strength for vision, audition, touch, smell and taste ratings (Lynott & Connell, 2009, 2013; Winter, 2016). Notably, all of these ratings were collected only for English glosses. In addition, for the 220 meanings for which we had iconicity ratings in all four languages, we categorized the lexical class of each sign and word based on English gloss assignments in the SUBTLEX-US corpus (Brysbaert & Keuleers, 2012). More particular semantic categories for these meanings were determined post hoc by the authors.

3. Results

The results show several notable patterns that characterize how iconicity is spread across the vocabularies in the four languages, including similarities and differences between signed and spoken languages. Overall, we found substantial correlation in the iconicity ratings between the languages, including English with ASL ($r = 0.16$, $p < 0.001$), BSL ($r = 0.22$, $p = 0.001$), and Spanish ($r = 0.16$, $p =$

0.0003). The highest correlation was between ASL and BSL (r = 0.68, p < 0.001), suggesting iconicity may be more robust in signs than words.

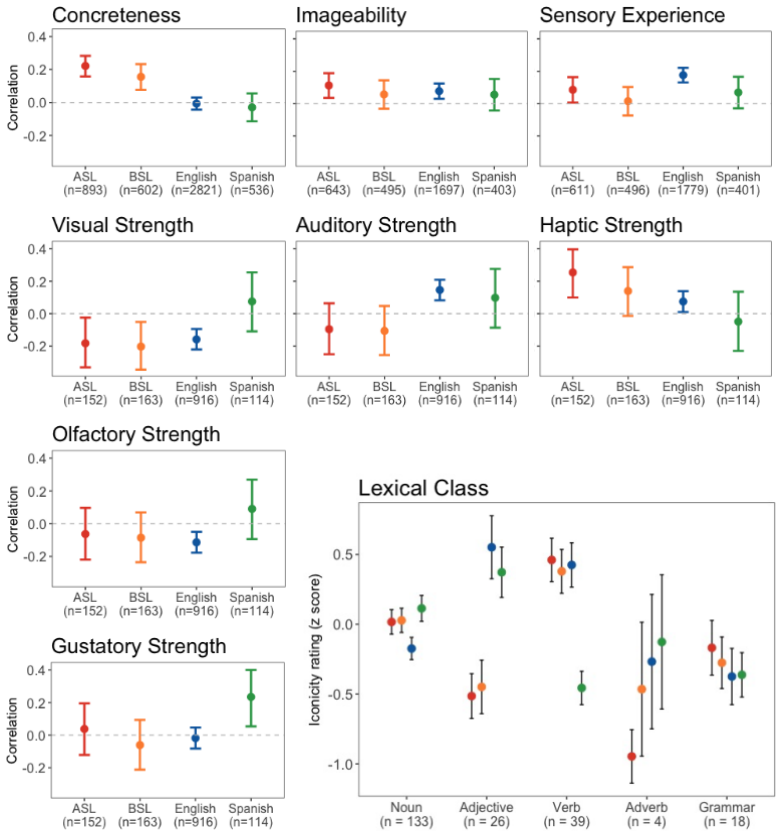

Figure 1. Top left plots show Pearson's correlations (r) between iconicity ratings and ratings of semantic properties. Error bars show 95% confidence interval. Below each language, n indicates the number of signs or words for which we had ratings. Bottom right plot shows normalized mean iconicity ratings and standard errors by (English) lexical class.

Shown in Figure 1, iconicity in each language was distributed according to the eight semantic variables in ways that reflect the semiotic affordances of the language modality. Across languages, signs and words for more sensorial meanings tended to be more iconic. More concrete meanings were more iconic in signs, but not words. Haptic strength was strongly correlated with iconicity in signs, while auditory strength was strongly correlated with iconicity in words.

Figure 1 also shows how iconicity is distributed across lexical classes in each language. For example, in the signed languages and in English, verbs were especially high in iconicity. Adjectives were relatively high in the two spoken

languages, but low in the signed languages. Counter to our prediction, nouns were about average in iconicity in ASL and BSL.

There were also distinct patterns of iconicity between signed and spoken languages in the more specific semantic categories. For example, signs for manual actions were particularly iconic in ASL and BSL, while signs for colors were particularly low. In comparison, in English and Spanish, words for perceptual properties were especially iconic, which was not the case in the signed languages.

4. Discussion and Conclusion

Modern languages, both spoken and signed, exhibit considerable iconicity across their vocabularies (Perniss et al., 2010; Dingemanse et al, 2015). This suggests a possibility that the first symbol systems were built from iconic *vocalizations*, as well as from iconic gestures, optimizing iconicity for different kinds of meanings. To examine how iconicity may have been balanced between modalities, we compared iconicity in the lexicons of modern signed and spoken languages. Our study utilized previously collected iconicity ratings, which, as a result, were collected under different protocols. Nevertheless, our findings suggest some characteristic ways that the iconicity of signs and words appears to be influenced by the language modality depending on their meaning.

Notably, our findings also have implications for understanding how modality may affect the emergence of structure in the evolution of communication systems. For example, some hypotheses on the evolution of phonological patterning in emerging sign languages make implicit assumptions about the categorically higher potential for iconicity in gestures versus vocalizations. Goldin-Meadow and McNeill (1999) and Sandler (1996) have suggested that signed languages may acquire phonological patterning a lot later than spoken languages because their high level of iconicity inhibits the reanalysis of signals as recombinable phonological units. This hypothesis is supported by a growing body of experimental work (Little et al., 2017; Roberts et al., 2015; Verhoef et al., 2015). However, these studies all use artificial signalling modalities, and the implications of this work for the real world rely on a more detailed understanding of iconicity in different kinds of natural languages. The present findings suggest that, depending on the language modality, iconicity may operate on the emergence of phonological structure differently in different semantic domains of vocabulary. This may explain, for example, why onomatopoeia in spoken languages may tend to have less standardized or "wild" phonology (Rhodes, 1994).

Altogether, our findings provide a preliminary, empirically-grounded, and detailed account of how iconicity is spread across the lexicons of signed languages in comparison to spoken languages. They demonstrate the prevalence of iconicity across human languages, no matter the modality, and suggest that theories of language evolution must consider the potential for iconicity in gesture and vocalization alike.

References

Armstrong, D. F., & Wilcox, S. E. (2007). *The Gestural Origin of Language*. Oxford ; New York: Oxford University Press.

Brysbaert and Keuleers (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44, 991–997.

Brysbaert, M., Warriner, A. M., & Kuperman, V. (2014). Concreteness ratings for 40 thousands generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.

Caselli, N., Sevcikova, Z., Cohen-Goldberg, A., Emmorey, K. (2017). ASL-Lex: A Lexical Database for ASL. *Behavior Research Methods*. doi:10.3758/s13428-016-0742-0.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384-387.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603-615.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356-1367.

Goldin-Meadow, S. (2017). What the hands can tell us about language emergence. *Psychonomic Bulletin & Review*, *24*, 213-218.

Goldin-Meadow, S., & McNeill, D. (1999). *The role of gesture and mimetic representation in making language the province of speech* (pp. 155-172).

Kendon, A. (2008). Some reflections on the relationship between 'gesture'and 'sign'. *Gesture*, *8*(3), 348-366.

Klima, E., & Bellugi, U. (1979). *The Signs of Language*. Cambridge, Mass.: Harvard University Press.

Little, H., Eryılmaz, K., & de Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1-15.

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*(2), 558-564.

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior research methods*, *45*(2), 516-526.

Perlman, M. (2017). Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies*, *18*, 376-401.

Perlman, M. & Cain, A.A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*, 320-350.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, *2*(8), 150152.

Perniss, P., Thompson, R. L. & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00227

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*(9), e0137147.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Rhodes, R. (1994). Aural images, In L. Hinton, J. Nichols, & J. Ohala (Eds.). *Sound Symbolism* (pp. 276-292). Cambridge: Cambridge University Press.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, 141, 52–66.

Sandler, W. (1996). Phonological features and feature classes: The case of movements in sign language. *Lingua*, 98(1), 197–220.

Tomasello, M. (2008). *Origins of human communication.* The MIT Press.

Verhoef, T., Kirby, S., & Boer, B. (2015). Iconicity and the emergence of combinatorial structure in language. *Cognitive science.*

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods, 40*, 1079-1087.

Winter, B. (2016). The sensory structure of the English lexicon. PhD thesis, University of California Merced.

Winter, B., Perlman, M., Perry, L.K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies, 18*, 430-451.

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. ROBERTS[6]

*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.

INTERCULTURAL CONTACT AND REPRESENTATIONAL FORM

THOM SCOTT-PHILLIPS 1[*1,2], CARMEN GRANITO[2],JEREMY KENDAL[2] & JAMIE TEHRANI[2]

[*]Corresponding Author: scott-phillipst@ceu.edu
[1] Department of Cognitive Science, Central European University, Budapest, Hungary
[2] Department of Anthropology, Durham University, Durham, UK

1. An intriguing parallel

Pictorial representations are, like words, highly versatile. They can visualise simple physical objects as well as very complex and abstract concepts and situations. They are also ubiquitous in almost all human societies. Cultures around the world have made images to convey information about living kinds, objects and ideas for at least 75,000 years, in forms as diverse as cave paintings, religious icons and emojis.

There are several points of difference between languages and pictorial representations. Modality is one (pictorial representations are by definition limited to one specific modality). Structure is another (languages are by definition highly structured). A third, on which we shall focus here, is style and transparency. While the dogma that linguistic form is wholly arbitrary should be questioned – onomatopoeia is the obvious counter-example, and it is increasingly recognised that many of the component parts of natural languages sometimes do possess a degree of iconicity – it is clearly the case that pictorial representations show variation in their degree of figurativeness more readily and more obviously than the component parts of languages typically do. (By 'figurative', we mean the extent to which an image is inter-subjectively recognisable as a depiction of objects, people, animals, scenes, and so on (see also Healey et al., 2007). This is not the same as 'iconicity', because unlike icons figurative images do not necessarily have a perceptual resemblance to their intended referent. They can, in particular, be inter-subjectively recognisable as one thing, but refer to another.

This is the case, for example, in the way that many emojis are used. The aubergine emoji is often used to refer to sexual genitalia.)

There is however an intriguing parallel between these two domains (languages and pictorial representations) that seems to somewhat transcend their various points of difference. There is in the history and anthropology of art a pattern that instances of intercultural contact often lead to changes in artistic style, and in particular in the degree of figurativeness used (e.g. Morphy & Layton, 1981; Verstegen, 2012; Shatzmiller, 2013; Versluys, 2017; see also Figure 1, below). And there is in sociolinguistics and language evolution the hypothesis, increasingly supported by data, that the degree of contact that a language community has with outsiders can be a factor in shaping linguistic form (Wray & Grace, 2007; Lupyan & Dale, 2010; McWorter, 2011; Trudgill, 2011). In short, these two fields have independently hit on the idea that intercultural contact can causally affect form and structure.

Figure 1: ***Aboriginal art as a real world example of how intercultural contact can affect forms of pictorial representation***. Left: Example of Yolngu art, which increased figurativeness after extensive exchange with Europeans. Narritjim Maymuru, Bamabama, 1976. Right: Two examples of Papunya art from Central Australia, which targeted an isolated audience of initiates and developed an increasingly abstract style of representation. Charlie Eagle Tjapaltjari, Wallaby Dreaming in the Sandhills, 1977 (top); Tim Leutra Tjapaltjarri, Possum Dreaming at Kurningka, 1977 (bottom).

2. Objectives

Building on the observations above, this presentation has two goals.

First we will present an experimental study that shows, clearly and unambiguously, that pictorial styles can be causally shaped by intercultural contact. More specifically, we use experimental methods borrowed from language

evolution (and cultural evolution more generally), to show that drawings produced by connected groups tend to retain a degree of figurativeness that ensures that they are – at least somewhat – transparent to outsiders, whereas in isolated groups drawings tend to become abstract and highly opaque. (See below for an abridged description of the methods.) This pattern is, as we said, also observed in the corpus data of natural languages, but its existence and causality is far clearer to see in the case of pictorial representation, because of features particular to that domain.

We will then, second, discuss how the various features of different communicative modalities, such as spoken language and pictorial representation, constrain and enable different means of human communication. We will draw in particular on recent developments and insights in cognitive pragmatics, which emphasise how the various differences of relevance here – between, for instance, linguistic and non-linguistic communication, between the iconic and the symbolic, and between meaning and showing – are mostly matters of degree, rather than differences of kind (see in particular Sperber & Wilson, 2015). We will, time allowing, present an elementary conceptualisation of this multi-dimensional space.

3. Experimental methods & results (abridged to avoid prior publication)

The study is composed of two phases. In Phase 1 (Data production) laboratory micro-societies played a Pictionary-like task in one of three conditions: isolation, contact or a control condition, which were simulated by manipulating the degree and structure of interaction between participants. The drawings produced at this stage were then used as stimuli in two surveys run in Phase 2: in one, naïve participants were asked to match the drawings with their meanings; in the other, other naïve participants had to say whether the drawings contained recognisable figures or not. Results clearly show that figurativeness and transparency are much higher exactly and only when the need for communication with outsiders is present.

References

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, *5*(1), e8559.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, *31*(2), 285-309.

McWhorter, J. H. (2011). *Linguistic Simplicity and Complexity: Why Do Languages Undress?* Walter de Gruyter.

Morphy, H., & Layton, R. (1981). Choosing among alternatives: Cultural transformation and social change in Aboriginal Australia and the French Jura. *Mankind, 13*, 56–73.

Shatzmiller, J. (2013). *Cultural Exchange: Jews, Christians, and Art in the Medieval Marketplace*. PUP.

Sperber, D., & Wilson, D. (2015). Beyond speaker's meaning. *Croatian Journal of Philosophy, 15*(44), 117-149.

Trudgill, P. (2011). *Sociolinguistic Typology*. OUP.

Versluys, M. J. (2017). *Visual Style and Constructing Identity in the Hellenistic World: Nemrud Dağ and Commagene under Antiochos I*. CUP.

Verstegen, U. (2012). Adjusting the image–processes of hybridization in visual culture: A perspective from early Christian and Byzantine archaeology. In: P. W. Stockhammer (Ed.), *Conceptualizing Cultural Hybridization* (pp. 67-93). Springer.

Wray, A., & Grace, G. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua, 117*, 543–578.

INTERPRETING PATIENT ANIMACY IN SILENT GESTURE

ROSS TOWNS[1*], MARIEKE SCHOUWSTRA[2]

[*]Corresponding Author: s1667434@sms.ed.ac.uk
[1, 2]Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom

A key finding of studies in the silent gesture paradigm is that people prefer SOV order when describing events in which an action's agent (corresponding to a verb's subject) is animate and the patient (object) is inanimate (Goldin-Meadow, So, Ozyürek, & Mylander, 2008). For so-called reversible events, however, in which agent and patient are both animate (and whose thematic roles may therefore plausibly be reversed), they prefer alternatives like SVO or OSV (Futrell et al., 2015; Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013; Meir, Lifshitz, İlkbasaran, & Padden, 2010).

Three competing explanations for this have been put forward. Under the so-called *ambiguity* hypothesis, participants avoid SOV due to the potential ambiguity caused by positioning two possible subjects before the verb (Meir et al., 2010). The *noisy-channel* account cites the susceptibility of SOV utterances to information loss caused by noise, to which SVO and related orders present a more robust solution (Gibson et al., 2013). Finally, the *role-conflict* account cites the tendency of gesturers to physically embody the roles of animate referents, thereby leading to potential confusion between roles if patients are gestured before actions (Hall et al., 2013).

Here we present an experiment designed to identify whether a relationship between patient animacy and gesture order is apparent in interpretation as well as production. We asked participants to interpret ambiguous gesture sequences, predicting that SVO sequences would be more likely to be interpreted as describing reversible events than would SOV sequences, and vice versa. However, our findings did not support this: gesture order did not seem to influence interpretation of reversible/non-reversible events. This result led us to carefully reconsider the mechanisms behind previous silent gesture production findings. We suggest that our null result is consistent with an account similar to the role-conflict hypothesis described above, in which apparent effects of

reversibility on word order are due to the specific affordances of the gestural modality, and are of limited generalisability to spoken language.

Our study builds on previous investigations of word order alternations in silent gesture in another different semantic domain: that of *intensional* vs *extensional* events (Schouwstra, 2012). By showing participants videos with ambiguous action/verb gestures, Schouwstra (2012; see also Thompson, Schouwstra & de Swart, 2016) found that people tended to interpret SOV gestures as describing extensional events (i.e. actions performed on an entity in the physical world, for example *girl sleeps on book*) but that SVO gestures were interpreted as describing intensional events (i.e. for example mental actions such as *girl dreams of book)*. Participants' biases in interpretation were thus consistent with, though somewhat weaker than, a bias found in gesture production for intensional and extensional events (Schouwstra & de Swart, 2014). A Bayesian computational model fitted to data from the interpretation study was also found to align well with data from production (Thompson, Schouwstra & de Swart, 2016), suggesting that production and interpretation are subject to related cognitive biases. Based on these results, we reasoned that we might find a similar correspondence between production and interpretation in the domain of reversibility.

Basing our experimental design on Schouwstra's (2012) interpretation study, we showed participants ambiguous gesture videos in 12 forced-choice trials. For each trial, we recorded 2 videos of gesture sequences, both consisting of the same gestures recorded in separate shots and concatenated either in SOV or SVO order using video editing software. Crucially, gestures for patients/objects were designed to be ambiguous as to whether they represented an inanimate or animate referent, for example a *guitar-playing* gesture, which might represent either a guitar or a guitarist. Participants watched SOV videos in 6 trials and SVO videos in 6 trials (choice of video randomized between trials), and signaled their interpretation by choosing from an array of two line-drawings, one depicting a reversible event and another depicting a non-reversible event (e.g. *pirate strokes guitarist* or *pirate strokes guitar).* We conducted two runs of the experiment: once using online participants recruited via the CrowdFlower crowdsourcing platform (N = 51, native English speakers) and once with participants in the lab (N = 20, native English speakers) so as to corroborate the results of the online study.

Contrary to our expectations, we found in both runs that SVO videos were marginally less likely to be interpreted as reversible than SOV videos (see Fig. 1). To account for possible variation among participants and trial items, we fit data from each run, respectively, to mixed-effects logistic regressions (specifying by-participant and by-item random slopes and intercepts) predicting

reversible interpretation by SVO gesture order, and used likelihood-ratio tests to compare them with null models featuring random effects only. Inclusion of gesture order
as a predictor was not found to improve model fit for either the online experiment ($\chi 2$ = 1.11, p = .29) or the laboratory experiment ($\chi 2$ = .008, p = .93).

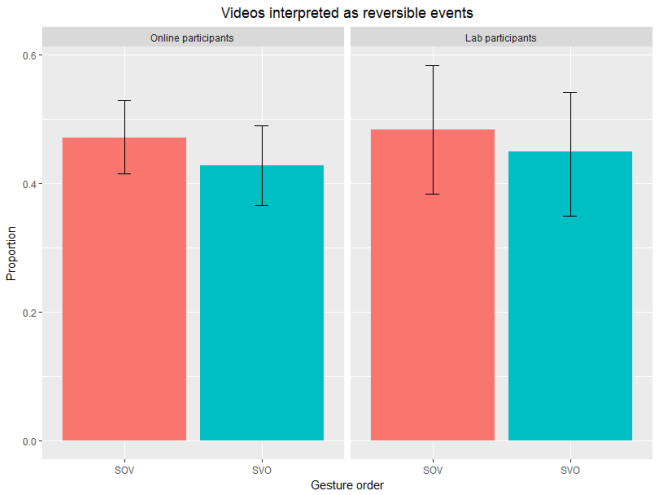

Figure 1. Mean proportions of gesture videos interpreted as reversible events in the online experiment (left) and the lab experiment (right). Error bars represent bootstrapped 95% confidence intervals.

Why, then, should biases in interpretation correspond to production with respect to intensionality, but not to reversibility as in the present study? It may simply be that the stimuli for our experiment simply were not ambiguous enough for any effect to take hold. However, another possibility points to qualitative differences between the mechanisms underpinning production biases in each domain. With respect to intensionality, findings from production and interpretation suggest a bias toward positioning elements in order of abstractness or cognitive accessibility (Goldin-Meadow et al., 2008; Schouwstra, 2012), or an iconic linear sequence in which object-final order mirrors the causal relations between an intensional object on its verb, for example between a thought and the action of thinking (Christensen, Fusaroli, & Tylén, 2016; Schouwstra & de Swart, 2014). Both accounts are thus predicated on properties of SOV and SVO word order *per se*, and can be considered apart from gesture production.

The same may be said of reversibility effects under the ambiguity and noisy channel hypotheses, leading to the expectation that participants might exhibit a bias toward interpreting SVO gestures as reversible. By contrast, the role conflict account is rooted in gesture production, and is therefore less likely to suggest a corresponding bias in interpretation. In support of this production-specific account, Hall and colleagues cite the results of their interpretation study, which found no effect of order on interpretation, either in terms of the purported ambiguity of SOV order or in participants' ratings of SOV and SVO sequences as either more or less appropriate descriptions of visual stimuli (Hall, Ahn, Mayberry, & Ferreira, 2015).

While remaining mindful of the dangers inherent in drawing conclusions from a null result, we suggest that our findings are consistent with an account of reversibility/patient animacy effects on gesture production as being rooted in the gestural modality itself. If this is correct, the extent to which silent gesture findings, as pertaining to reversibility/patient animacy, can be generalised to spoken language is debatable. Hall et al. (2013) suggest that avoidance of role-conflict may shape word-order in spoken language via perspective-taking (MacWhinney, 1977), but this is offered as a tentative suggestion without strong empirical support.

Our interpretation of the present findings does, however, generate two testable predictions, namely that the relationship between intensionality and constituent order in gesture should also be found in both production and interpretation in a different modality, for example visual symbols (Vastenius, van de Weijer, & Zlatev, 2016), whereas effects of reversibility/patient animacy will be confined to production in the gestural modality only. Our results also, in our view, demonstrate the need for caution against interpreting possibly modality-specific effects as general properties of cognition, and for variety in the use of alternative modalities as experimental paradigms.

References

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67–80.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition*, *136*, 215–221.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, *24*(7), 1079–1088.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9163–9168.

Hall, M. L., Ahn, Y. D., Mayberry, R. I., & Ferreira, V. S. (2015). Production and comprehension show divergent constituent order preferences: Evidence from elicited pantomime. *Journal of Memory and Language*, *81*, 16–33.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1–17.

MacWhinney, B. (1977). Starting Points. *Language*, *53*(1), 152-168.

Meir, I., Lifshitz, A., İlkbasaran, D., & Padden, C. (2010). The interaction of animacy and word order in human languages: a study of strategies in a novel communication task. *The Evolution of Language: Proceedings of the 8th International Conference (EvoLang 8)*, 455–456.

Schouwstra, M. (2012). *Semantic Structures, Communicative Strategies and the Emergence of Language*. Utrecht: LOT

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

Thompson, B., Schouwstra, M., & de Swart, H. (2016). Interpreting silent gesture. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).*

Vastenius, A., van de Weijer, J., & Zlatev, J. (2016). The influence of native language word order and cognitive biases in pictorial event representations. *Cognitive Semiotics*, *9*(1), 45.

IMPROVISED WORD ORDER BIASES ARE NOT MODALITY SPECIFIC: EVIDENCE FROM NON-LINGUISTIC VOCALIZATIONS

KATIE MUDD[*1,2], SIMON KIRBY[1], and MARIEKE SCHOUWSTRA[1]

[*]Corresponding Author: kath.mudd@gmail.com
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK
[2]AI-lab, Vrije Universiteit Brussel, Brussels, Belgium

Languages use different word orders (the order of the Subject, Object and Verb) to organize and convey information. The distribution of word orders documented in the world is uneven, favoring SOV and SVO (Dryer, 2013), which may provide a window into biases shaping language. The visual-manual modality has recently been taken up to study basic word order in the absence of a system of language conventions. In silent gesture experiments, in which hearing participants improvise ways of communicating motion events using only their hands and no speech, SOV order is dominant for speakers of various native languages (Goldin-Meadow et al., 2008). However, participants may switch to SVO word order depending on the semantic properties of the event (Schouwstra & de Swart, 2014; Gibson et al., 2013), or in the presence of a vocabulary of conventional signs (Marno et al., 2015). As these findings appear to be robust across different languages (native SOV and SVO speakers in the aforementioned studies), silent gesture is increasingly used as a way to uncover the cognitive biases playing a role in situations of emerging language.

Thus far, these experiments have been carried out using the manual modality alone, leaving open the question whether the findings generalize to other modalities.[1] Are the word order patterns observed in silent gesture essentially a consequence of the fact that participants improvise in the absence of linguistic rules? Or are they (partly) the result of modality specific production constraints (such as suggested in e.g. Hall et al. 2013)? An answer to this question will have

[1] Schouwstra (2017) observes a structural similarity between silent gesture strings and (spoken) utterances of unsupervised second language learners, but this concerns the placement of temporal information; systematic cross-modal investigations of Basic Word Order have, to our knowledge, not been carried out before.

implications for the generality of the evolutionary claims to be made on the basis of silent gesture results.

Here we use a vocal analog of the silent gesture paradigm to test whether biases previously observed in silent gesture generalize to the vocal modality. We do this by testing the word order preferences for *extensional events* (in which both participants in the action are concrete, e.g., 'the pirate throws the guitar') vs *intensional events* (in which the Patient is unspecific or dependent on the action, e.g., 'the pirate thinks of the guitar'). Schouwstra & de Swart (2014) showed a preference for SOV for extensional events and SVO for intensional events in silent gesture. Here, we tested participants in the vocal modality, to see if this meaning-dependent word order pattern replicates outside the manual modality.

In experiment 1, we asked native SVO participants (N=20) to use non-linguistic vocalizations to describe pictures while sitting on their hands. The stimuli were line drawings that consisted of 32 extensional events (e.g. 'a robot drops a drill') and 32 intensional events (e.g. 'a diver loves a car') that were selected to be relatively easy to describe using non-linguistic vocalizations. In each description, the elements were coded as subject, object or verb. For instance, 'evil-laugh' (denoting a witch) was coded as Subject; 'boom' (denoting dropping of an object) was coded as Verb. Sequences with repeated consecutive constituents were recoded as if having only one iteration of each constituent (SVVO was recoded as SVO). Sequences with non-consecutive repetitions remained as they were (VSVO remained VSVO). Word orders other than SVO or SOV (14% of the total data set) were excluded from statistical analysis.

The findings confirm that constituent order in improvised vocalizations is influenced by the semantic properties of the event: a greater proportion of SOV descriptions are given for extensional events than intensional events, and a greater proportion of SVO descriptions are given for intensional events than extensional events (see Figure 1). Event type predicted the proportion of SVO produced (b = 1.965, SE = .737, $p < .01$), mirroring what has been found in silent gesture.

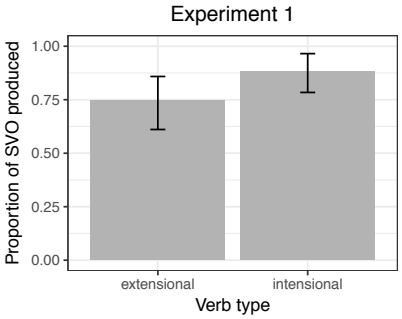

Figure 1. Mean proportions of SVO produced for extensional and intensional events in experiment 1 out of all SVO and SOV produced. Error bars represent 95% confidence intervals.

Strikingly, the results from the vocal modality follow what has been found in the manual modality, namely that semantic properties of the event significantly affect constituent order. However, the results of experiment 1 differ from previous findings, where the preference for SOV for extensional events is much greater, and SVO only becomes dominant for the intensional events (Schouwstra & de Swart, 2014). A possible explanation for the difference in results is that in the present study the vocal channel is used for both language and the improvised vocalizations, and thus may be more susceptible to interference from the native language of participants (SVO in the present study). Despite this possible influence, the vocalizations of participants in the present study were conditioned by meaning, and we thus replicated the basic finding of Schouwstra & de Swart (2014).

In experiment 2, we build on work by Marno et al. (2015) who show an increase in SVO order in silent gesture if participants (native SOV and SVO) are trained on individual lexical items (e.g. 'a girl') prior to complex meanings, akin to the extensional events in the present study (e.g. 'a girl throwing a fish'; Marno et al., 2015). We asked native SVO participants (N=20) to produce vocalizations for individual items first, before going on to complete scenes. Using the same coding scheme as for experiment 1, word orders other than SVO or SOV (9% of the total data set) were excluded from statistical analysis.

Again, our results in the vocal modality reflect those in the manual modality: a higher proportion of SVO is produced when participants first describe individual items before continuing to full scenes. In experiment 2, SVO accounts for a majority of extensional events (87%) and intensional events (99%) while SOV accounts for a minority of extensional events (13%) and intensional events (1%). Overall, more SVO is produced for both event types in experiment 2 than in experiment 1; in a model combining experiment 1 and 2, we find that experiment (b = 3.998, SE = 1.465, $p < .01$) is a stronger predictor of producing

SVO than verb type (b = 1.693, SE = .714, $p < .05$); thus, the high incidence of SVO in experiment 2 is more strongly predicted by the access to the lexicon than the meaning of the event. Following the results of Marno et al. (2015) in the manual modality, the lexicon triggers more SVO, showing similar results in the vocal and manual modalities.

The silent gesture paradigm has been used to make claims about features of language emergence, in the manual and vocal modalities, but could these findings be specific to the manual modality? In the present study, we focused on 2 results from the silent gesture paradigm, and replicated both of the main findings in the vocal modality. In the vocal modality, more SVO is produced than in the manual modality, which may be attributed to interference from the participant's native language, but the overall finding is the same: word order is conditioned by event type in improvisation. These findings provide the first evidence that gestural and vocal improvisation yield similar results for basic word order, demonstrating that the effects observed are modality independent.

References

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, *24*(7), 1079-1088.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, 105(27), 9163-9168.

Hall, M. L, Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition, 129*(1), 1-17.

Marno, H., Langus, A., Omidbeigi, M., Asaadi, S., Seyed-Allaei, S., & Nespor, M. (2015). A new perspective on word order preferences: the availability of a lexicon triggers the use of SVO word order. *Frontiers in psychology*, *6*.

Schouwstra, M. (2017). Temporal Structure in Emerging Language: From Natural Data to Silent Gesture. *Cognitive science*, *41*(S4), 928-940.

Schouwstra, M., and de Swart, H. (2014). The semantic origins of word order. *Cognition* 131, 431–436. doi: 10.1016/j.cognition.2014.03.004

SYSTEMATIC ALLIGNMENT BETWEEN DIFFERENT TYPES OF ICONICITY AND SEMANTIC DOMAINS IN SILENT GESTURE

GERARDO ORTEGA[1], ASLI OZYUREK [1,2,,3]

[*]Corresponding Author: gerardo.ortega@mpi.nl
[1]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[2]Centre for Language Studies, Radboud University, Nijmegen, the Netherlands
[3]Donders Institute for Brain, Language and Cognition, Radboud University, Nijmegen, the Netherlands

1. Introduction

In stark contrast to spoken languages, the manual-visual mode of language is well-suited to create analogies between a referent and the body in that it can represent visible features of a referent through visible articulators. There is mounting evidence that the gestural system is exploited to kick-start language emergence (e.g., Morgan, 2016); however, it is not entirely clear how individuals use their hands and body to iconically express concepts or whether there is any systematic pattern that may lead to linguistic conventions. In a context where interlocutors may want to communicate a given concept without a shared language (e.g., a butterfly), a person may select a prominent feature of the referent that may be familiar to an interlocutor (e.g., the flapping of its wings), and represent it through an iconic gesture (e.g., flapping the arms). Therefore, some of the semiotic resources exploited to communicate in the absence of a common language may be i) shared schemas about actions and objects, ii) the gestural system, and iii) the capacity to describe the physical features of a referent through iconicity.

In this study we examine how gesturers align specific semantic domains with different types of iconic representations (i.e., acting, representing, drawing, personification. See Figure 1) (Hwang et al., 2016; Müller, 2013), and how these different depictions may vary in meaning transparency to an interlocutor. Only by describing gestural representations and how well they are understood by an

interlocutor will it be possible to expand our knowledge on how different manifestations of iconic gestures support comprehension and thus lead to linguistic conventions.

Figure 1. Examples of different modes of representation. 'To smoke' implements the acting strategy because the body re-enacts the action of smoking. 'To go down' implements the representing strategy because two wiggling fingers depict two legs descending. 'House' is depicted by tracing a pentagon. In 'bird' the gesturer uses the personification strategy because the features of the referent are mapped onto his body.

2. Method

In a gesture generation task, 20 native speakers of Dutch were presented with a total of 272 words and were asked to produce a silent gesture that conveyed the same meaning as the concept on the screen. The words belonged to five different semantic categories: actions with objects (e.g., to smoke), actions without objects (e.g., to swim), manipulable objects (e.g., toothbrush), non-manipulable objects (e.g., pyramid), and animate entities (e.g., butterfly). Participants' renditions were glossed, and then, using an existing gesture notation system (Bressem, 2013) we coded the form of each gestural feature (the configuration of the hand, the movement, the orientation, and its placement in gestural space). The gestural features of each concept were compared across participants, and when at least ten participants (50% of the group) produced minimally the same three features for a given concept, this was regarded as a systematic gesture across the population. This resulted in a total of 109 concepts for which at least ten people produced the same physical gestural form. These gestures were then analysed in terms of their mode of representation (i.e., *acting*, *representing*, *drawing*, *personification*).

After the characterisation of the systematic gestures produced in study 1, a different group of 18 Dutch speakers were presented with professionally filmed videos of the systematic gestures by a model and were asked to rate the degree of iconicity in a 7-point Likert scale (1 low – 7 high). These ratings would indicate whether certain semantic domains represented in specific modes of representation were more clearly understood and were more transparent for a different group of viewers.

3. Results

Regarding production, we observed that *acting* was the strategy that dominated all gestural productions. When we look at specific semantic domains we can see that actions with objects (e.g., to smoke), actions without objects (e.g., to swim), and manipulable objects (e.g., toothbrush) implement the *acting* strategy in almost 90% of the cases. Non-manipulable objects (e.g., pyramid) also favour the *acting* strategy but they also have a strong bias towards *drawing*. Animate entities (e.g., butterfly) is a category that stands out from the rest in that it shows an overwhelming bias towards the *personification* strategy followed by *representing*. Results from the iconicity ratings show that actions with and without objects depicted with the *acting* strategy get the highest iconicity ratings. Manipulable and non-manipulable objects implementing the *acting* strategy also get high iconicity ratings but lower than actions. That is, the *acting* strategy leads to better comprehension when it represents actions than objects. The semantic categories involving object manipulation (i.e., actions with objects and manipulable objects) got high ratings when they were depicted through the *representing* strategy. The *personification* strategy was solely used in animate entities and it got relatively high iconicity ratings. *Drawing*, which was primarily used in non-manipulable objects, received overall the lowest iconicity ratings.

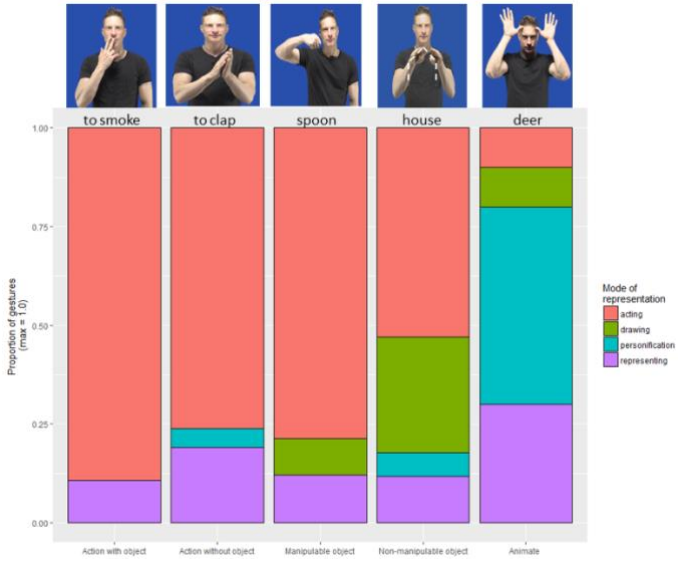

Figure 1. Proportion of gestures showing different types of modes of representations (acting, representing, drawing, personification) per semantic category

4. Discussion

At least for the five semantic categories investigated in this study, individuals converge in the type of iconic depiction (i.e., mode of representation) when producing silent gestures for different concepts. The strong preference for the *acting* mode of representation supports embodied theories of gestural production that claim that gestures originate from action simulations (Cook & Tanenhaus, 2009). Concepts represented with this strategy are more easily understood if they are actions (e.g., 'to smoke') than if they are objects (e.g., 'lighter'). There is a more direct mapping between the *acting* strategy and an action whereas objects depicted with the same strategy require a higher level of abstraction and thus receive lower ratings. *Representing* was produced in similar proportions across different domains but received higher iconicity ratings for actions with objects and manipulable objects. This findings echo the notion of patterned iconicity (Padden et al., 2013; Padden, Hwang, Lepic, & Seegers, 2015) which posits that this strategy results from the need to make noun-verb distinctions. *Drawing* is primarily used for non-manipulable objects which suggests that gesturers may struggle to associate a bodily action to a referent that cannot be hand-held (e.g., pyramid). Interestingly, these representations received the lowest iconicity ratings, arguably because the gesture of an objects' shape or size is too ambiguous and does not transmit clearly enough the intended meaning. The *personification* strategy stands out in that it is primarily used for animate entities and it received fairly high ratings. These data align with previous studies showing that conventioanlised and emerging sign languages tend to use this strategy to represent animals (Hwang et al., 2016).

We observe systematic patterns in participants' silent gestures and also that certain couplings are more transparent than others and it is the interaction between semantic domain and type of iconicity that may be exploited as semiotic tool to allow for referentiality and displacement of absent concepts at the origins of sign language emergence. The patterns reported here bear some resemblance to the form of emerging signed systems (Meir et al., 2017; Tkachman & Sandler, 2013) which suggests that gestural systematicity may be responsible of the remarkable similarity across the lexicons of the sign languages of the world. Our results also echo studies in the spoken modality (Perlman & Lupyan, 2018) which lends further supports that regardless of modality, iconicity may be at the core of language emergence.

References

Bressem, J. (2013). A linguistic perspective on the notation of form features in gesture. In C. Muller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction* (pp. 1079–1098). De Gruyter Mouton.

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: speakers' gestures affect listeners' actions. *Cognition*, *113*(1), 98–104. http://doi.org/10.1016/j.cognition.2009.06.006

Hwang, S.-O., Tomita, N., Morgan, H., Ergin, R., Ilkbasaran, D., Seegers, S., … Padden, C. (2016). Of the body and the hands: patterned iconicity for semantic categories. *Language and Cognition*, 1–30. http://doi.org/10.1017/langcog.2016.28

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES--ELAN system. *Behavior Research Methods*, *41*(3), 841–9. http://doi.org/dx.doi.org/10.3758/brm.41.3.841

Meir, I., Aronoff, M., Börstell, C., Hwang, S. O., Ilkbasaran, D., Kastner, I., … Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207. http://doi.org/10.1016/j.cognition.2016.10.011

Morgan, H. (2016). Language from gesture: a case study from East Africa. In *Paper presented at the 12th International Conference for Theoretical Issues on Sign Language Research*. Melbourne, Australia.

Müller, C. (2013). Gestural modes of representation as techniques of depcition. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An International Handbook on Multimodality in Human Interaction* (pp. 1687–1701). Berlin: De Gruyter Mouton.

Padden, C., Hwang, S.-O., Lepic, R., & Seegers, S. (2015). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in Cognitive Science*, *7*(1), 81–94. http://doi.org/10.1111/tops.12121

Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., & Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, *13*(3), 287–305.

Perlman, M., & Lupyan, G. (2018). People Can Create Iconic Vocalizations to Communicate Various Meanings to Naïve Listeners. *Scientific Reports*, *8*(1), 2634. http://doi.org/10.1038/s41598-018-20961-6

Tkachman, O., & Sandler, W. (2013). The noun-verb distinction in two young sign languages. *Gesture*, *13*(3), 253–286.