LANGUAGE ADAPTS TO SIGNAL DISRUPTION IN INTERACTION

VINICIUS MACUCH SILVA, SEÁN ROBERTS
Language and Cognition Department, Max Planck Institute for Psycholinguistics
Nijmegen, Netherlands
vini.macuch@gmail.com, sean.roberts@mpi.nl

Linguistic traits are often seen as reflecting cognitive biases and constraints (e.g. Christiansen & Chater, 2008). However, language must also adapt to properties of the channel through which communication between individuals occurs. Perhaps the most basic aspect of any communication channel is noise. Communicative signals can be blocked, degraded or distorted by other sources in the environment. This poses a fundamental problem for communication. On average, channel disruption accompanies problems in conversation every 3 minutes (27% of cases of other-initiated repair, Dingemanse et al., 2015). Linguistic signals must adapt to this harsh environment. While modern language structures are robust to noise (e.g. Piantadosi et al., 2011), we investigate how noise might have shaped the early emergence of structure in language.

The obvious adaptation to noise is redundancy. Signals which are maximally different from competitors are harder to render ambiguous by noise. Redundancy can be increased by adding differentiating segments to each signal (increasing the diversity of segments). However, this makes each signal more complex and harder to learn. Under this strategy, holistic languages may emerge. Another strategy is reduplication - repeating parts of the signal so that noise is less likely to disrupt all of the crucial information. This strategy does not increase the difficulty of learning the language - there is only one extra rule which applies to all signals. Therefore, under pressures for learnability, expressivity and redundancy, reduplicated signals are expected to emerge.

However, reduplication is not a pervasive feature of words (though it does occur in limited domains like plurals or iconic meanings). We suggest that this is due to the pressure for redundancy being lifted by conversational infrastructure for repair. Receivers can request that senders repeat signals only after a problem occurs. That is, robustness is achieved by repeating the signal across conversational turns (when needed) instead of within single utterances.

As a proof of concept, we ran two iterated learning chains with pairs of individuals in generations learning and using an artificial language (e.g. Kirby et al., 2015). The meaning space was a structured collection of unfamiliar images (3 shapes x 2 textures x 2 outline types). The initial language for each chain was the same written, unstructured, fully expressive language. Signals produced in each generation formed the training language for the next generation. Within each generation, pairs played an interactive communication game. The director was given a target meaning to describe, and typed a word for the matcher, who guessed the target meaning from a set. With a 50% probability, a contiguous section of 3-5 characters in the typed word was replaced by 'noise' characters (#). In one chain, the matcher could initiate repair by requesting that the director type and send another signal. Parallel generations across chains were matched for the number of signals sent (if repair was initiated for a meaning, then it was presented twice in the parallel generation where repair was not possible) and noise (a signal for a given meaning which was affected by noise in one generation was affected by the same amount of noise in the parallel generation).

For the final set of signals produced in each generation we measured the signal redundancy (the zip compressibility of the signals), the character diversity (entropy of the characters of the signals) and systematic structure (z-score of the correlation between signal edit distance and meaning hamming distance). In the condition without repair, redundancy increased with each generation ($r=0.97$, $p=0.01$), and the character diversity decreased ($r=-0.99, p=0.001$) which is consistent with reduplication, as shown below (part of the initial and the final language):

Gen 0	Outline 1	Outline 2
Shape A	luna	lapi
Shape B	monepilu	nena
Shape C	lunenena	pinenalu

Gen 5	Outline 1	Outline 2
Shape A	kakakakakakakakkaka	kakakakkakakakak
Shape B	lelelelelelelelelele	lalalallalalalalallala
Shape C	mamamamamam	memememe

Linear regressions revealed that generations with repair had higher overall systematic structure (main effect of condition, $t = 2.5$, $p < 0.05$), increasing character diversity (interaction between condition and generation, $t = 3.9$, $p = 0.01$) and redundancy increased at a slower rate (interaction between condition and generation, $t = -2.5$, $p < 0.05$).

That is, the ability to repair counteracts the pressure from noise, and facilitates the emergence of compositional structure. Therefore, just as systems to repair damage to DNA replication are vital for the evolution of biological species (O'Brien, 2006), conversational repair may regulate replication of linguistic forms in the cultural evolution of language. Future studies should further investigate how evolving linguistic structure is shaped by interaction pressures, drawing on experimental methods and naturalistic studies of emerging languages, both spoken (e.g Botha, 2006; Roberge, 2008) and signed (e.g Senghas, Kita, & Ozyurek, 2004; Sandler et al., 2005).

References

Botha, R. (2006). Pidgin languages as a putative window on language evolution. *Language & Communication*, 26(1), 1-14.

Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. Behavioral and Brain Sciences, 31(5):489–508

Dingemanse, M., Roberts, S.G., Baranovaa, J. et al. (accepted) Universal Principles in the Repair of Communication Problems. PLOS ONE.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chemical reviews*, 106(2):720–752.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.

Roberge, P. T. (2008). The creation of pidgins as a possible window on language evolution. *LOT Occasional Series*, 10, 101-138.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2661-2665.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua.*Science*, 305(5691), 1779-1782.