LANGUAGE-SPECIFIC AND UNIVERSAL FACTORS BEHIND MORPHOLOGICAL SIMPLIFICATION: AN AGENT-BASED MODELLING STUDY OF ALORESE

Peter Dekker[*1], Marian Klamer[2], and Bart de Boer[1]

[*]Corresponding Author: peter.dekker@ai.vub.ac.be
[1]AI Lab, Vrije Universiteit Brussel, Brussels, Belgium
[2]LUCL, Leiden University, Leiden, The Netherlands

Large-scale correlational investigations (Lupyan & Dale, 2010) indicate that population size and the degree of contact with other groups correlate negatively with the morphological complexity of the language spoken. To study what mechanisms give rise to morphological simplification and how these surface in the world's languages, we apply agent-based models (Smith, 2014) to a case study with real-world data.

Alorese, an Austronesian language, is spoken on the coasts of the Alor and Pantar islands in Eastern Indonesia, while inward, Papuan Alor-Pantar languages are spoken. Many L1 users of Alor-Pantar languages have learned Alorese as a second language. While Alorese lost most of its morphology, its sister languages, which have not been in contact with Alor-Pantar languages, retained it (Klamer, 2012, 2020). The island geography of this case study presents us with a relatively isolated contact situation between two groups of speakers. We use agent-based models to study the hypothesis that adult language contact caused morphological simplification, focusing on inflectional verb morphology. In our model, a population of agents play a language game (Steels, 1998), in which they try to communicate concepts (verb+person) using inflected verb forms. The model is initialized with phonetic representations of verb forms in Lewoingu Lamaholot (1) (cf. Nishiyama & Kelen, 2007, p. 32), a sister language of Alorese (2) (cf. Klamer, 2011, p. 65), which lost the inflection:

(1) lodo-na   Lamaholot   (2) lodo-Ø   Alorese
    go.down-3SG   go.down-3SG
    'he goes down'   'he goes down'

Model outcomes are compared to Alorese, where simplification has occurred. We evaluate two cognitive mechanisms that we think could give rise to morphological simplification: the procedural-declarative model of L1/L2 language processing (Ullman, 2001) and reduction of word forms based on language-specific phonotactics. According to the procedural/declarative model, in L1 users, grammar is produced by a procedural cognitive system, while the lexicon is memorized

in a declarative system. In L2 learners, linguistic forms which are normally produced in the procedural system by L1 users, such as morphology, are memorized in the declarative system. We chose to represent the procedural system in L1 users by, with a certain probability, generalizing during update: the received signal is added not only to the inferred concept, but to all concepts. In our model with phonotactic reduction, a producing agent drops the affix if adding it would lead to violation of the default CV structure of Alorese. The chosen data representation, staying close to the real word forms in Alorese, allows for this modelling of language-specific phonotactics. It should be noted that the influence of the phonotactic mechanism is more local than that of the procedural/declarative model. Our results show that the standard model (Fig 1a) does not show a substantive decrease in morphological complexity, when the proportion of L2 users increases. The phonotactic reduction model (Fig 1b), which does allow an extra possibility for simplification, does not show this decrease either. The model with generalization for L1 (Fig 1c) shows a decrease in morphological complexity for higher proportions of L2, although the complexity of suffixes stays relatively high compared to the real situation in Alorese. Based on our results, it appears that generalization in the context of the procedural/declarative model may be an important mechanism in explaining the relation between language contact and simplification. In the future, similar data-based agent-based models could be applied to case studies from other languages in the world, to identify which mechanisms are universal and which are dependent on language-specific context.

(a) Basic model. (b) Basic + phonotactic. (c) Basic + generalization L1
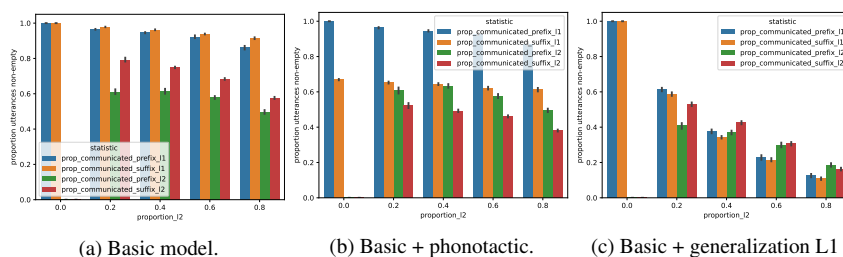
Figure 1.: Morphological complexity (measured by the proportion of utterances for which the affix is not empty) for different proportions of L2 users. Morphological complexity for prefix and suffix for L1 (blue/orange) and L2 (green/red).

References

Klamer, M. (2011). *A short grammar of Alorese (Austronesian)*. Muenchen: LINCOM Europa. (OCLC: ocn760922643)

Klamer, M. (2012). Papuan-Austronesian language contact: Alorese from an areal perspective. *Language Documentation & Conservation, Melanesian Languages on the Edge of Asia: Challenges for the 21st Century. Special Publication No. 5*, 72–108.

Klamer, M. (2020). From Lamaholot to Alorese: Morphological loss in adult language contact. In D. Gil & A. Schapper (Eds.), *Austronesian Undressed: How and why languages become isolating* (pp. 339–367). Amsterdam: John Benjamins Publishing Company.

Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS ONE*, *5*(1), e8559.

Nishiyama, K., & Kelen, H. (2007). *A grammar of Lamaholot, Eastern Indonesia*. Lincom Europa.

Reali, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172586.

Smith, A. D. (2014). Models of language evolution and change. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 281–293. (Publisher: Wiley Online Library)

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In *Approaches to the Evolution of Language* (pp. 384–404). Cambridge University Press.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122.