Simulating morphological change

I have developed an artificial-intelligence computational model in which language change can be simulated. It is assumed that language starts out as a vocabulary with words for actions and objects only (a so-called "protolanguage"; cf. Bickerton 1981, Jackendoff 2002). Agents talk about events in their immediate surroundings (using Steels' 1997 "language game" setup). Depending on the number of distractor events that are simultaneously ongoing, agents have to be more or less specific when selecting referential terms (à la Grice 1975). And depending on the degree to which the hearer's world knowledge is estimated to predict the event to be described, additional role marking may be used to make sure the role distribution of the event participants can be understood properly (Lestrade 2010). In addition to such communicative considerations, frequency and recency of usage play a role in word selection, for both referential items and role markers.

As an important goal of the model is to show that much of language structure emerges spontaneously (Smith and Kirby 2008; Christiansen and Chater 2008), only very general cognitive principles are implemented, such as a desire for communicative success, shared attention, recognition of communicative intention, and desire for economic expressions (cf. Tomasello 2003, Arbib 2015). In the absence of a grammar, agents use proto-principles to communicate about their world, e.g. assuming that things that stand together belong together (cf. Givon 1995 and Jackendoff 2002).

Living in a computer, the world and meaning representation of the agents is necessarily different from ours. Nevertheless, meaning representation should be sufficiently valid for generalization to natural language, at least for present purposes: According to Wierzbicka (1996; cf. also Gärdenfors 2000, Katz and Fodor 1963, Guiraud 1968), natural-language concepts can be decomposed into meaning primitives such a CONCRETE, HUMAN, MALE, etc. Abstracting away from the quality of the dimensions that organize our mental lexicon, this can be modeled as a list of randomly generated forms with values on a number of numerical meaning dimensions (their "vector representations").

Initially word selection involves semantically motivated and fully specified lexical items only. Over time, however, words can *desemanticize* and *erode* as a result of frequent usage (Heine and Kuteva 2007). Desemanticization involves the progressive removal of the meaning dimension of a marker that corresponds to the dimension at which most variation is attested among the objects it has been used for (Bybee 2010); erosion results from sloppy pronunciation and subsequent "wrong learning" (Nettle 1999, Jurafsky et al. 2001). Thus, morphological markers may develop (maximally short forms with maximally general meanings), both for pronominal reference and role marking.

Using my computational model, the development, stability and learnability of different types of morphology can be studied (the pronominal paradigm, for example, turns out to be much less stable than role marking under the present assumptions). The goal of this talk is to discuss the possibilities of such simulations, and learn about phenomena that should be implemented.

Arbib, M.A. (2015). "Language evolution. An emergentist perspective". B. MacWhinney & W. OGrady (Eds.), *The handbook of language emergence*, pp. 600–623. West Sussex, UK: Wiley/Blackwell. **Bickerton**, D. (1981), *Roots of language*. Karoma, Ann Arbor, Mich. **Bybee**, J. (2010), *Language, Usage and Cognition*. New York: Cambridge University Press. **Christiansen**, M. H. & N. **Chater** (2008), "Language as shaped by the brain", *Behavioral and Brain Sciences* 31(5):489-509. **Gärdenfors**, P. (2000), *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT. **Givón**, T. (1995), *Functionalism and grammar*, Amsterdam and Philadelphia: PUB. **Grice**, H. P. (1975), "Logic and conversation". P. Cole and J.L. Morgan (eds), *Syntax and Semantics: Speech Acts* 3, 41–58. New York. Academic Press. **Guiraud**, P. (1968), "The semic matrices of meaning". *Social Science Information* 7(2): 131–139. **Heine**, B. and T. **Kuteva** (2007), *The genesis of grammar: a reconstruction*. Oxford: OUP. **Jackendoff**, R. (2002), *Foundations of language: Brain, meaning, grammar, evolution*, Oxford: OUP. **Jurafsky**, D., A. **Bell**, M. **Gregory** & W.D. **Raymond** (2001), Probabilistic relations between words: Evidence from reduction in lexical production. J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, 229-255. Amsterdam/Philadelphia. John Benjamins. **Katz**, J.J. and J.A. **Fodor** (1963), "The Structure of a Semantic Theory", *Language* 39 (2): 170–210. **Lestrade**, S, (2010), *The space of case*, PhD thesis, Radboud University Nijmegen. **Nettle**, D. (1999), *Linguistic Diversity*. OUP, New York. **Smith**, K. & S. **Kirby** (2008), "Cultural evolution: implications for understanding the human language faculty and its evolution", *Phil. Trans. R. Soc. B* 363:35913603. **Steels**, L. (1997), Constructing and sharing perceptual distinctions. *Machine Learning*: *ECML*-97, 4-13. Springer. **Tomasello**, M. (2003), *Constructing a language: a usage-based theory of language acquisition*, Cambridge, MA: Harvard University Press. **Wierzbicka**, A. (1996). *Semantics: Primes and Universals*. Oxford University Press, Oxford, UK.