

Constructing languages to explore theoretical principles

Guillaume Enguehard (Université d'Orléans, LLL)

Nicola Lampitelli (Université de Nanterre, Modico)

Xiaoliang Luo (Université Paris-Cité, HTL)

Linguistic research is organized into three approaches: inductive, abductive and deductive. Induction consists in proposing general principles based on specific data. Abduction aims to relate observed data to an already existing general principle. Deduction makes it possible to check the abductive step by defining a set of expected but yet unobserved data.

It can be noted that the deductive approach has a rather limited place in this heuristic process. It only serves as a method for validating the abductive step. In other words, the consequences of the principles induced by observation are only ever explored on the scale of a particular phenomenon, never on the scale of the whole system. This limitation of the deductive approach generates another issue consisting in not questioning phenomena that are so regular that a corresponding principle is implicitly accepted as self-evident. We believe that these two issues are due to the fact that the linguist has preconceived views about the object he is studying. He then remains blind to certain connections between linguistic phenomena.

To deal with this issue, we propose a method inspired by experimental archaeology (Bordes, 1947). In archaeology, the causes of an artifact are regularly outside our empirical field and it is not possible to make abstract generalizations as we do in linguistics. Researchers therefore sometimes use speculation and validation of hypotheses through direct experience. The experiment consists in seeing if one obtains an artifact similar to the one observed using this or that construction method. This approach is particularly famous for elucidating questions related to the lithic industry during prehistory. Though it never gives definite answers, it makes it possible to (in)validate hypotheses on the scale of a complete system.

An equivalent approach applied to linguistics would be the generation of linguistic systems based on explicitly formulated principles with a minimum of typological parameters. The origin of these principles does not really matter. They may be the result of data observation, cognitive, behavioral, functional or even a priori speculation. Their value is ultimately determined only by the degree of closeness between the results obtained and the natural languages as a whole. Trying to generate similar systems pushes the linguist to explicitly define the principles that are needed and to explore all the consequences of these principles.

In this talk, we specifically explore three different modules corresponding to basic aspects of word formation: the phonemic inventory, the phonotactics and the wordsize limit. In each case, we propose to implement with spreadsheet formula a unique a priori functionalist principle and we derive various types of data (with or without branching

onsets, etc) from the same formula. This functionalist principle can be summed up in one word: ‘contrast’. Concerning the inventory and the phonotactics, we assume that contrasts between segments must be maximised (Martinet, 1955; Liljencrantz and Lindblom, 1972; Goldsmith, 1976). Thus the same principle applies paradigmatically and syntagmatically (Hjelmslev, 1943). As for the wordsize limit, we suggest that it corresponds to the minimum size necessary for the number of different possible combinations of an inventory to correspond to the size of a lexicon, e.g.:

Words: 200 000
 Segments: 34
 Average word size: 3,461379675

Finally, in order to introduce variation, we use a normal distribution which makes any deviation from these rules all the less probable as it is important.

With simple principles like these, we are already deriving datasets that come very close to natural languages (see Table 1). Of course, these data still contain some quirks sometimes related to practical limitations that we are aware of and will have to overcome in the future. But the most important thing is that we achieve this result without starting from the observation of the data we are trying to reproduce. In this respect, we came across a generated dataset that shows a vowel distribution reminiscent of the phenomenon of vowel reduction, although we have not implemented this notion. As you can see in Table 1, closed vowels are found only in first syllable.

ʃəʔəʔ	jʔə	ʔəʔə	wʔə	mɜpɜp	ɔʔə	həʔə	pɜpɜ	jʔ	həʔə
mɜpɜp	ʊʔə	ɛpɜp	ɪʔəʔ	ʃəʔəʔ	ʔəʔə	ʃɜp	ɪpɜp	mɜpɜ	mɜpɜ
ɛəʔə	ɜpɜ	həʔə	ɔʔə	ɜpɜp	pɜpɜ	ʔə	ʃəʔə	ɜpɜ	ʔəʔə
həʔ	mɜpɜp	jʔ	həʔ	pɜp	ɪpɜ	ʊʔəʔ	pɜpɜ	ɪpɜ	fɜpɜ
ʊʔəʔə	wʔə	mɜp	ɪpɜpɜ	pɜp	pɜpɜp	ɔʔ	ɔʔəʔ	wʔəʔ	ɪʔə
ɜpɜ	həʔə	ɜpɜp	ɪpɜ	pɜpɜ	jʔə	ɪpɜp	ɔʔə	ɜp	ɛpɜp
wʔə	ʊʔəʔə	ʊʔəʔə	ʊʔ	ɜpɜpɜ	ɔʔ	ʔəʔə	ɛpɜp	ɛpɜp	ʃəʔə
ʊʔə	ɪpɜp	pɜpɜp	ɜ	fɜp	pɜpɜp	ɪpɜpɜ	ɔʔ	ɪpɜp	ɛəʔə
ɪʔ	ɪpɜ	ɜpɜp	ɜpɜp	ɪpɜ	ɔʔəʔ	ɛəʔə	ɪpɜ	ɛpɜp	ɪpɜp
jʔ	mɜ	ʊʔəʔə	ɪʔəʔ	jʔə	fɜp	hə	jʔə	ɪpɜp	jʔə

Table 1: Example of initial stress and vowel reduction

To conclude, the exploration and exploitation of linguistic principles in the construction of languages has been present since the beginnings of modern linguistics. The strong regularity of the first international languages illustrates the notion of linguistic system that emerged in the 19th century. Then, the 20th century brought an explosion of fictional languages that explore almost all facets of linguistic diversity. Even today, languages are emerging that are testing the limits of certain linguistic principles, as Toki Pona does with polysemy (Lang, 2014). Most of these initiatives address real scientific questions but from a non-academic point of view. The fact that Saussure’s brother was an Esperantist (de Saussure, 1914) is an amusing illustration of the parallel development of academic linguistics and language construction.

References

Bordes, F. (1947). Étude comparative des différentes techniques de taille du silex et des roches dures. *L’anthropologie*, 51:1–29.
 de Saussure, R. (1914). *La vort-teorio en Esperanto*. Universala Esperantia Librejo.
 Goldsmith, J. A. (1976). *Autosegmental Phonology*. PhD dissertation [ms], MIT.
 Hjelmslev, L. (1943). *Prolegomena to a Theory of Language*. University of Wisconsin Press, Madison, [1969] edition.

- Lang, S. (2014). *Toki Pona: The Language of Good*. CreateSpace, Charleston.
- Liljencrantz, J. and Lindblom, B. (1972). Numerical simulation of Vowel Quality Systems: The Role of Perceptual Contrasts. *Language*, 48(4):839–862.
- Martinet, A. (1955). *Économie des changements phonétiques: traité de phonologie diachronique*. A. Francke, Berne.