Modular algorithmic complexity applied to the Mazatec diasystem

Jean Léo Léonard, Marco Patriarca, Els Heinsalu, Antonella Gaillard-Corvaglia, Pierre Darlu

Dialectology has long been considered as a marginal field in linguistics, mainly concerned with the recollection of empirical facts, with low theoretical expectations. Nevertheless, thanks to quantification of dialectical data (dialectometry, see Goebl, 1981, 1982, 2002)), geolinguistics in particular turns out to be one of the most promising horizon for Complexity Theory (CT) – as much as CT opens new horizons for dialectology.

We'll apply various methods to a Mazatec database from Paul L. Kirk (1966), providing 9000 tokens (750 cognates x 12 locolects): patristic distances (cladistics, see Hennig, 1950, 1966), Levenshtein algorithm (Beijering & al. 2008, Bolognesi & Heeringa, 2002) and dialect intelligibility tests (Kirk, 1970), in order to show multiplicity of prospects on a geolinguistic space. Mazatec as a case study for testing algorithmic complexity has been chosen on several grounds:

i) it once provided the empirical base for a landmark study by Sarah Gudschinsky (1958) on the reconstruction of dialect diversification process (1958),

ii) Kirk's data, with less than 10 000 tokens is easier to process than bigger data available on European languages,

iii) we have thoroughly checked and revisited Kirk's data through fieldwork within the framework of an empirical research project (IUF, MAmP, 2009-14, see Léonard & al. 2012),

iv) phonology and grammar of Mazatec dialects have been formalized within the same project (with declarative phonology and Paradigm Function Morphology).

Conditions for a survey of *algorithmic complexity* are therefore met, allowing a multiplex modeling of Mazatec geolinguistics, from *phylogenetic* (linguistic changer over time and space), *ontogenetic* (typological inner diversity of a diasystem) and *epigenetic* (mutual intelligibility, interpretation) standpoints. We'll compare results from various components of the lexicon (nouns vs. verbs, etc.), grammar (inflectional classes) and phonology (morphophonology inasmuch as diachronic phonology), with both methods: Levenshtein algorithm according to *w* (normalized mean of Lev_Distance, according to three grades or cut-offs) on the one hand vs. patristic distances on the other hand, pondered and unpondered according to structural criteria. This modular approach according to structural components will provide different models of diasystemic diversity through algorithmic complexity (see Léonard & *al.* 2015).

References:

Beijering K. Gooskens C. & Heeringa W. "Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm", Amsterdam, Linguistics in the Netherlands, 2008,), 2008, p. 13-24.

Bolognesi, R. & W. Heeringa, "De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten", in D. Bakker, T. Sanders, R. Schoonen and Per van der Wijst (eds.). Gramma/TTT, Tijdschrift voor taalwetenschap, Nijmegen University Press, Nijmegen, 9 (1), 2002, p. 45-84.

Goebl Hans, "Eléments d'analyse dialectométrique (avec application à l'AIS)", Revue de Linguistique romane, 45, 1981, p. 349-420.

Goebl Hans, Dialektometrie. Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie, Vienne, Verlag der Öst. Akademie der Wissenschaften, 1982.
Goebl Hans, "Analyse dialectométrique des structures de profondeur de l'ALF"• , Revue de Linguistique Romane, 66, 2002, p. 5-63.
Gudschinsky Sarah, "Mazatec dialect history", Language 34, 1958, p. 469-481.
Gribbin, John, Deep Simplicity, Chaos, Complexity and the Emergence of Life, London, Penguin, 2004.
Hennig Willi, Grundzüge einer theorie der Phylogenetischen systematik, Berlin, Deutscher Zentralverlag, 1950.
Hennig Willi, Phylogenetic Systematics. Urbana, University of Illinois Press, 1966.
Kirk Paul L., Proto-Mazatec phonology. PhD dissertation, University of Washington, 1966.
Kirk Paul L., "Dialect Intelligibility Testing: The Mazatec Study", International Journal of American Linguistics, 36-3, 1970, p. 205-211.
Léonard, Jean Léo; Els Heinsalu; Marco Patriarca; Pierre Darlu: "Modeling Regional Variation from EAS: complexity and social aggregates", in Aurrekoetxea Olabarri, Gotzon; Romero Andonegi, Asier; Etxebarria Lejarreta, Aintzane (éds.) Workshop EUDIA-4 : Linguistic variation in the Basque Language & Education, I, 19 June 2015, Bilbao, Basque Country, Spain, 2015, pp.145-172.
Léonard Jean Léo ; dell'Aquila, Vittorio & Gaillard-Corvaglia, Antonella "The ALMaz (Atlas Linguistico Mazateco): from geolinguistic data processing to typological traits", STUF, Akademie Verlag, 6, 2012, p. 78 -94.