Sociolinguistic typology meets historical corpus linguistics

George Walkden, Gemma McCarley, Raquel Montero, Molly Rolf, Sarah Einhaus and Henri Kauhanen
University of Konstanz

Trudgill (2011) has made a powerful case that different types of contact situation may give rise to different types of change according to the sociohistorical circumstances that shaped them: short-term contact involving predominantly adult language acquisition is predicted to yield simplification, while long-term co-territorial contact involving child language acquisition and balanced multilingualism is predicted to yield additive complexification. This paper outlines a programme to test Trudgill's conjecture in the domain of morphosyntax using European and non-European languages that are attested in historical corpora, and presents some preliminary results.

We adopt Walkden & Breitbarth's (2019) proposal that morphosyntactic complexity involves lexical items that bear *uninterpretable features* – features without semantic interpretation that must enter into an agreement relation with interpretable counterparts. Such features are then predicted to be lost or reduced in situations where the diachronic development of the language in question is characterized by extensive adult language acquisition. Walkden & Breitbarth (2019) look at the diachrony of negation; in this paper, we focus on three further areas in which predictions are made.

1) DP-internal agreement. Assuming that agreement in phi-features (person, number and gender) within the DP involves uninterpretable features, we can investigate languages that have lost or reduced agreement in order to see what diachronic trajectory they follow, i.e. whether there are structural determinants of this loss. Here our case study is drawn from the history of English using the PPCME2 and PLAEME corpora, with an eye to determining possible consequences of Norse contact.

2) Case. The nature of syntactic Case and its relationship to morphological case is debated; however, assuming that at least some Case features are uninterpretable syntactic features, we predict that they should be vulnerable in contact situations characterized by adult language acquisition. The Balkan Sprachbund is our testing ground for these predictions.

3) Null subjects. A well-established finding in the adult L2 acquisition literature (e.g. Bini 1993) is that null subjects are difficult for such learners to acquire in a native-like way. Historically this should lead to reduction in or loss of null subjects in certain sociolinguistic scenarios, and here we draw on corpus evidence from varieties of Spanish spoken in South and Central America, especially Afro-Hispanic languages of the Americas (AHLAS).

Trudgill in his book takes a bird's-eye view of the grammar of a language over time. In our case studies we adopt a finer-grained, variationist approach, looking at contact situations for which relatively diverse corpora are available and for which we can track the frequencies of linguistic variants in different dialects and/or texts. A desideratum for establishing that a given change is contact-induced is that the putatively transferred feature is present in the source language (see e.g. Thomason & Kaufman 1988: 111–112). This reasoning cannot hold for cases of simplification, precisely because simplification is not transfer of a feature from one language to another. Instead one must make a plausibility argument based on what is known about the social, demographic and geographical situation at the time of the change. If texts produced in an area where there is known to be more population movement or more adult second-language acquisition also display more simplification, this supports the hypothesis – especially if other instances of contact influence can be demonstrated in these texts on independent grounds.

Overall, our preliminary results indicate that the hypothesis that uninterpretable features are vulnerable in situations involving large-scale adult language acquisition, following Trudgill (2011) and Walkden & Breitbarth (2019), is promising: for instance, i) texts from Bolivia – where historically the proportion of L2 speakers of Spanish was high – show a greater reduction in null subjects over time than texts from the Dominican Republic, and ii) the loss of plural agreement in Middle English is sensitive to syntactic category in a way that would not be expected if the change were purely a morphological one. However, we also find facts and effects that are not obviously captured by this hypothesis, suggesting that more research on the typological effects of contact situations in morphosyntax is needed. Our variationist, corpus-based comparative methodology is one way of conducting such research, and serves to complement

methods based on typological surveys (e.g. Lupyan & Dale 2010) and experimental data (e.g. Raviv, Meyer & Lev-Ari 2019) by adding historical depth and ecological validity.

References

Bini, Milena. 1993. La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro pro-drop. In J. M. Liceras (ed.), *La lingüística y el análisis de los sistemas no nativos*, 126–139. Dovehouse Editions Canada.

Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS One* 5(1), e8559.

Raviv, Limor, Shiri Lev-Ari & Antje Meyer. 2019. Larger communities create more systematic languages. *Proc. R. Soc. B* 286, 20191262.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Los Angeles, CA: University of California Press.

Trudgill, Peter. 2011. *Sociolinguistic typology: social determinants of linguistic complexity*. Oxford: Oxford University Press.

Walkden, George, & Anne Breitbarth. 2019. Complexity as L2-difficulty: implications for syntactic change. *Theoretical Linguistics* 45(3–4), 183–209.