A complex solution to an unsolvable problem?

Comment to *The Dynamics of bilingualism in language shift ecologies*, by Leonore A. Grenoble and Boris Osipov

Roberta D'Alessandro, Utrecht University

One of the foci of the keynote article by Grenoble & Osipov (2023; henceforth G&O) is the long-standing problem of eliciting data from minority speakers and identifying the baseline against which we can draw the comparison for the study of language shift. I will limit my response to this issue, leaving the others to more expert scholars.

Among others, G&O tackle the problem of how to perform a solid quantitative analysis targeting minority and non-standardized languages. This is a cogent question: these varieties often have very few speakers to begin with, and microvariation (a key term that G&O do not use but which in fact they describe all over the place) is wild (see for reference the work of Benincà, D'Alessandro, Poletto, Loporcaro, Ledgeway, *a.o.* for Romance; Van Craenenbroeck, Van Koppen to mention just a few for the Germanic languages). Despite the discouragingly long list of the many issues that linguists are confronted with when trying to work with minority/indigenous languages, the impression I had when reading the article by G&O is that of a step forward for the field of linguistics. I saw convergence of issues, of problems, of methodological attempts: something rare across linguistic subfields. In what follows, I will concentrate on the issues of controlling for geographical variation and establishing a baseline to assess language shift ecologies.

Microvariation and uniformity

Fragmentation of the "traditional" language (following their terminology) and difficulty to access a coherent language system have been described and addressed before, in different subfields. Theoretical studies on language variation very often target endangered languages, usually in the subfield called "dialectology". A large tradition of theoretical dialectological studies, especially in Romance and Germanic (see above) have adopted as a key objective that of trying to identify the general laws underlying the grammars of these languages. Any formal description of a language variety requires some basic uniformity, and here is where what G&O describe comes into play. Theoreticians, like linguists working with quantitative methods, strive to describe and understand language systems, starting from the basic assumption that these systems will be somewhat uniform across all speakers of a given community. Unlike descriptivists, who limit themselves to the observation of variation, theoreticians also want to develop a coherent model of a language system. However, any non-standardized variety, especially when in contact with a dominant language, presents a huge number of variation points, so much so that describing such a variety can seem a hopeless exercise. As the fieldworkers working in the *Microcontact* project[1] (D'Alessandro 2015, 2021) on heritage Italo-Romance varieties in the Americas realized very soon, no two speakers

[1] *Microcontact. Language variation and change from the Italian heritage perspective* (2017-2022) was funded by the ERC-CoG 681959-Microcontact. The ERC is hereby gratefully acknowledged. More information about the project, the interactive atlas for data crowdsourcing, and the publications can be found on the project website: https://microcontact.sites.uu.nl/

speak alike. Microvariation is pervasive, optionality exists in stable grammars, and this has repercussions on the database on which your analysis is built (Andriani *et al.* forthcoming).

The crucial factor is not so much the number of speakers, also very relevant (cfr D'Alessandro, Natvig & Putnam 2021) as the level of standardization of a language: it is easier to find a baseline for a heritage language minority speaking a language that is standardized in another country (like Russian, or Spanish, or Korean in the US) rather than a language that starts out fragmented even before it becomes heritage. In heritage Italo-Romance in the Americas, just like in the case of Even described by G&O in contact with both Sakha (a regional language) and Russian (the dominant language of the area), the fragmentation is present at the baseline, which makes it impossible to identify one system to be compared with the heritage or shifting language (G&O, section 6). When we talk about language shift, or language change, the obvious question is "Shift with respect to what?"

Furthermore, a particular feature that is present in the language might not be present in a specific speaker. Likewise, when we compare the shifted version of a language to its "traditional" version and we find a deviation, we can't be sure that this deviation is the result of change, or shift, or attrition, and that it wasn't already present in the language, microvariationally, in the first place (D'Alessandro, Natvig and Putnam, 2021, Andriani *et al.* forthcoming). Some ongoing work on Turkish heritage varieties in the Netherlands (Tat 2021) shows that what used to be considered a heritage feature is in fact already present in the "traditional" variety. This can in turn offer support to Guardiano *et al*.'s (2016, 2020) *Resistance Principle,* postulating that "in order for syntactic change to occur, it must be triggered by interference data already available" (Guardiano & Stavrou 2021: 3). In other words, no syntactic change is possible if the syntactic feature was not already present, though dormant, in the system.

Generational variation

To solve the problem of microvariation, G&O propose a number of solutions. It seems to me that they are all up to the task, once we accept that there is no perfect description/documentation/model of a language, as the language is a magmatic system. Within *Microcontact*, one of the methods used to establish the baseline was actually to try and find the 1st generation speakers that gave the input to their children: the study was cross-generational but micro-generational, in the sense that we tried to find the actual parents of the heritage speakers, and check whether the feature that their children showed was already present in their grammar or not. This methodology proved very useful but not very easy to operationalize, for two reasons. The first is that 1st generation migrants are not always available, and therefore one needs to rely on a larger set of informants. The second is that language is not always transmitted from parents to children: in many cases the children of the 1st generation immigrants did not learn the language at all, because of the language policies of the host country for instance. The language was then transmitted from the 1st generation speakers to their grandchildren. This means that the classification of 1st, 2nd, and 3rd generation speakers is flawed, and requires extra attention.

Another way to tackle this problem (adopted for instance by Ledgeway 2012), which is also proposed by G&O in a different way, is to consider mainly so-called macroparameters, i.e. elements that are more stable, such as word order, and only then proceed to zoom into variation.

While the word order of the languages in contact can differ, word order within one specific variety is usually stable. This does not solve the problem of understanding language shift entirely, but offers some insights into the direction that grammar might be taking.

Is it worthwhile?

This introduces the last reflection I would like to propose, about the importance of the enterprise of documenting and understanding indigenous/minority languages. While every language is worth investigating, the assumption that only if we have an adequate number of speakers the study will be worthwhile is not on the right track. It is enough for one speaker to produce a phenomenon systematically for this phenomenon to be considered as a possible instance of a language. Investigating the language of one single speaker is worthwhile, just like investigating the language of 2,000 or 2,000,000 speakers: the language production of each speaker is a possible "output" of human language. Obviously, if we only have one instance of a phenomenon by one speaker, we cannot know whether the rest of the community speaks in the same way. But as G&O show very well, even if we have 10, 1000, or 2000 instances of a phenomenon we won't be able to conclude that the rest of the community will speak in the same way. Does this make our investigation less worthwhile? I'd argue that it doesn't: it makes is more important. Rarities can always say more about a species than 'conformities'. This is why I think that this is an important issue, but not a problem.

Selected References

D'Alessandro, R. (2015). *Microcontact. Language variation and change from the Italian heritage perspective*. European Research Council project, https://microcontact.sites.uu.nl/project/

D'Alessandro, R. (2021). Syntactic change in contact: Romance. *Annual Review of Linguistics* 7(7), 309–328. http://doi.org/10.1146/annurev-linguistics-011619-030311

D'Alessandro, R. (in press). Doing Romance Linguistics. In: Corr, A. & Schifano, N. (Eds.), *Romance grammars in context and contact*.

D'Alessandro, R., Natvig, David, & Putnam, M. T. (2021). Addressing challenges in formal research on moribund heritage bilinguals: A path forward. *FrontPsychol.* 12. 700126. http://doi.org/10.3389/fpsyg.2021.700126

Grenoble, L., & Osipov, B. (2023). The dynamics of bilingualism in language shift ecologies. *Linguistic Approaches to Bilingualism*.

Guardiano, C., Michelioudakis, D., Ceolin, A., Irimia, M.A., Longobardi, G., Radkevich, N., Silvestri, G., & Sitaridou, I. (2016). South by south east. A syntactic approach to Greek and Romance microvariation. *L'Italia dialettale* 77, 96–166.

Ledgeway, A. (2012). *From Latin to Romance. Morphosyntactic typology and change.* OUP.