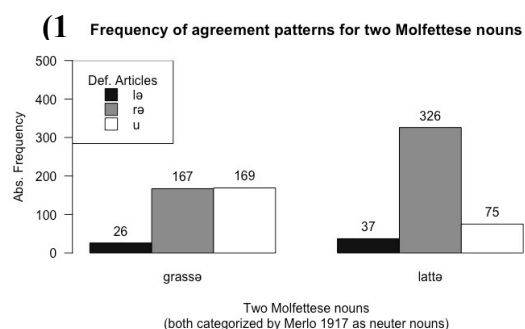# Crowdsourced dialect data for the study of morphosyntactic change
Federica Breimaier

"Progress in linguistics can only be guaranteed by taking into account a wide variety of data from a range of different sources" (D'Alessandro and van Oostendorp 2017: 1). Pursuing this intent, new tools have been provided to carry linguistic inquiries on morphology and morphosyntax. This line of methodological innovation has concerned not only research on standard languages but, especially in the last two decades, also the field of dialectology. Firstly, psycholinguistic procedures have started to be applied to non-standard varieties (e.g. Lundquist et al. 2016 and Lundquist and Vangsnes 2018, on Norwegian dialects, or Fabbro and Frau 2001 on Friulian aphasics); the same can be stated for neurolinguistic paradigms, which have been adopted to analyze case agreement in German dialects (Fleischer et al. 2014) or gender agreement change in an Italo-Romance dialect (Bambini et al. 2018). Secondly, with the advent of the digital humanities, new corpora (e.g. *Das VerbaAlpina-Projekt*) and databases (e.g. the *Zurich Database on Agreement in Italo-Romance*) on non-standard varieties became available. Thirdly, crowdsourcing appears to offer a new and effective possibility to obtain linguistic data (cf. e.g. Zihlmann and Leemann 2018, where data have been gathered on Lucerne German allophones by means of the smartphone app *Dialäky Äpp* and Leemann et al. 2018, on the creation of the *English Dialects App Corpus*). Moreover, the Utrecht-based *Microcontact* project is gathering linguistic data about the Italian varieties spoken by Italian immigrants in America by asking young speakers to record their older relatives and to upload the audio on a dedicated website.

(1 Frequency of agreement patterns for two Molfettese nouns

Two Molfettese nouns
(both categorized by Merlo 1917 as neuter nouns)

My contribution reports the results and the methodology of a dialectological pilot study conducted by means of an online questionnaire on an Apulian variety: Molfettese (province of Bari). When I started this study, my first aim was to examine the gender agreement of this variety by means of face-to-face interviews, but, then, I faced the "one source problem" (Leivada et al. 2019: 9), since Merlo's (1917) study on Molfettese articles is the only source addressing the variety under investigation. Interestingly, according to his analysis, one can assume the presence of a four-gender system (similar to those of other Southern Italian varieties reported in Loporcaro 2018: § 4.5) and my intent was to analyze whether the situation has changed or is changing. Nevertheless, before planning a fieldwork investigation, I decided to organize a crowdsourced pilot study to gain a first impression on how the system has evolved from 1917. In (1), no variation is observed for the 'milk' word, which is consistently reported to select neuter by respondents: *rə llattə*. For *grassə*, on the other hand, the situation is not clearcut: 50% of the participants used a neuter agreement pattern[1], while the remaining 50% chose a masculine target. By looking at the chosen definite articles for *grassə* across age groups this data allow me to say that the variation is patterned[2]: the younger generation is simplifying the opposition between neuter and masculine noun gender.

This methodology proved to be an efficient way to gather exploratory, hypothesis-generating data on gender agreement in Molfettese and the possibility to reach respondents aged between 16 and 90 years allows me to analyze the change in apparent time (Labov 1994). I will report how the

---

[1] Note that *rə* and *lə* are both forms of the neuter article occurring today in free variation, in the same way as has been described for nearby Spinazzolese (cf. Loporcaro 2018: 152f.).

[2] For reasons of space, I am not able to add a second plot.

questionnaire was structured (e.g. which sociological variables have been considered), the advantages that this type of survey can offer to the field of dialectology (e.g. the possibility to reach a higher number of speakers in a small amount of time[3]) and the problems arisen during the structuring of the questionnaire (e.g. the necessity to ask a relatively small number of questions in order for respondents not to abandon the survey) and during data analysis (e.g. the absence of orthographic conventions and the interpretation of the answers).

References

Bambini, V., Canal, P., Breimaier, F., Meo, D., Pescarini, D., and Loporcaro, M. (2018). Brain signature for on-going morphosyntactic change: weaker P600 for a fading gender value in a southern Italo-Romance dialect. Ms. IUSS (Pavia)/Scuola Normale Superiore di Pisa/ Universität Zürich.

D'Alessandro, R. and van Oostendorp, M. (2017). On the Diversity of Linguistic Data and the Integration of the Language Sciences. *Frontiers in Psychology* 8: 1–4. DOI=10.3389/fpsyg. 2017.02002.

Fabbro, F. and Frau, G. (2001). Manifestasions of aphasia in Friulan. *Journal of Neurolinguistics* 14(2): 255–279. DOI= 10.1016/S0911-6044(01)00017-3.

Fleischer, J., Bornkessel-Schlesewsky, I. and Dröge, A. (2014). Alemannische Neurodialektologie: Ergebnisse einer EEG Untersuchung zur Interaktion von Kasus, Wortstellung und Belebtheit im Zürichdeutschen. 18th Arbeitstagung zur alemannischen Dialektologie «Dialekt und Öffentlichkeit», 8-10 October 2014, Tübingen.

Labov, W. (1994). *Principles of linguistic change.* Volume 1: *Internal factors.* Oxford, UK— Cambridge, USA: Wiley-Blackwell.

Leemann, A., Kolly, M-J. and Britain, D. (2018). The English Dialects App: the creation of a crowdsourced dialect corpus. *Ampersand* 5: 1–17. DOI= 10.1016/j.amper.2017.11.001.

Leivada, E., Kleanthes K. G. and D'Alessandro, R. Eliciting big data from small, young, or non-standard languages: 10 experimental challenges. *Frontiers in Psychology* 10, 313. DOI=10.3389/fpsyg.2019.00313.

Loporcaro, M. (2018). *Gender from Latin to Romance: History, Geography, Typology.* Oxford: Oxford University Press. DOI=10.1093/oso/9780199656547.001.0001.

Loporcaro, M., Paciaroni, T., Pescarini, D., Idone, A., Romagnoli, S., Zanini, S., Zakharko, T. (2018). *The Zurich Database of Agreement in Italo-Romance* (DAI). University of Zurich. www.dai.uzh.ch

Lundquist, B., Rodina, Y., Sekerina, I. A. and Westergaard, M. (2016). Gender Change in Norwegian Dialects: Comprehension is affected before Production. *Linguistics Vanguard* 2(1). DOI=10.1515/lingvan-2016-0026.

Lundquist, B. and Vangsnes, Ø. A. (2018). Language separation in Bidialectal Speakers: Evidence From Eye Tracking. *Frontiers in Psychology* 9: 1–16. DOI=10.3389/fpsyg.2018.01394.

Zihlmann, U. and Leemann, A. (2018), l/-vocalisation in Lucerne Swiss German dialects: a sociophonetic analysis using big data. In M. Belz, C. Mooshammer, S. Fuchs, S. Jannedy, O. Rasskazova and M. Zygis (eds), *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13).* Leibniz-Zentrum Allgemeine Sprachwissenschaft: 201–204. DOI= 10.18452/18805

Das VerbaAlpina-Projekt. https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=133&db=181

Microcontact. Language Variation and Change from the Italian heritage perspective. https:// microcontact.sites.uu.nl

─────────────────────

[3] The online questionnaire managed to reach 450 participants, balanced across age and gender.