South Bavarian rhotics in crowdsourced linguistic data from Northeastern Italy: a diachronic and qualitative comparison

Abstract for *Beyond Borders: German-speaking Minorities in Italy and around the World* (30.06.22)

[anonymised author name(s)]

Crowdsourcing offers the advantage of gathering large quantities of linguistic data from a wide range of localities, without the demanding technical means an equivalent in-person data collection would require. Especially for lesser-spoken minority languages, overcoming geographical and technological barriers to reach speakers is a crucial challenge addressed by crowdsourcing (Leivada et al. 2019). Nevertheless, crowdsourcing comes with the disadvantage of a potentially lower data quality caused by the restricted ability of the researcher to control the data at the source (e.g. noisy or empty recordings, misunderstood instructions...).

Despite the inability to control *who* exactly is filling an online questionnaire, the question whether crowdsourced data can constitute a reliable and valid source for research on dialects and minority languages has been answered mostly positively in the literature (see e.g. Munro et al. 2010; Braunger et al. 2018; Kruijt 2022; Hilton in press). In this contribution, we aim to compare the realisation of rhotics in South Bavarian dialects spoken in Italy, contrasting two methods (traditional fieldwork vs. crowdsourcing) as well as three moments in time (the late 1960s vs. the late 2000s vs. the 2020s). The former method is represented by the *Tirolischer Sprachatlas* (Klein et al. 1965–1971) and the *Insre Sproch* dialect atlas (Scheutz 2016), the latter by the *VinKo* crowdsourcing project (Rabanus et al. 2014–; Cordin et al. 2018). Since the Tyrolean *VinKo* data predominantly includes contributions from young (female) informants (born around 2000), we can compare the three moments in a diachronic perspective.

The **variables** compared are:

1. The alveolar vs. uvular realisation of /r/ in onset position, e.g. *learerin*

2. The realisation of /r/ in unstressed coda position after a short vowel, e.g. *wosser*

3. The realisation of /r/ in stressed coda position after a long vowel, e.g. *johr*

4. The realisation of /r/ in stressed coda position before a coronal stop/affricate, e.g. *wirt*/*wird*

Based on the *Tirolischer Sprachatlas* (Klein et al. 1965–1971), one would expect to find:

1. Uvular /r/ in *Ahrn-*, *Passeier-*, *Ulten-* and *Etschtal* (Plaus ↔ Leifers), alveolar /r/ elsewhere

2./3. Vocalisation to /o/ in *Ahrn-* and *Pustertal*, audible alveolar/uvular (cf. 1) realisation elsewhere; syllable-bearing devoiced /r̥/ in *Vinschgau*

4. *Idem*, but with rhotic assibilation before coronal stops/affricates in *Wipp-* and *Pustertal*

If the crowdsourcing method is valid, we expect to find all these previously attested realisations and their geographical distribution reflected faithfully in the crowdsourcing data. A possible 'invalidity' of the data is that speakers could tend to 'standardise' their dialect input because of 'standard-German-prompting' biases (e.g. *VinKo* website in standard German). On the other hand, one could expect data comparable to that of traditional fieldwork methods, to the extent that the prompt words are written in the dialect and the whole purpose of the task is explicitly stated in the instructions to be about dialects.

From a systematic examination of the data, it **results** that the realisations are indeed found where expected in the crowdsourced data. With respect to diachrony, we observe in accordance with Scheutz (2016) that specific linguistic variables have gained/lost ground compared to their 1960s distribution, and we additionally observe the further expansion of some developments he reports:

1. Spread of uvular /r/, with alveolar /r/ only found in 4 *VinKo* informants out of 76 (Kardaun, Leifers, Graun and Mals), whose average age is 1.4x higher than that of the other informants

2./3. Vocalisation to /o/ lost ground compared to *TSA* data but resurfaces as [ɔ] at places in *Pustertal* where Scheutz reports its disappearance; vocalisation in general extended across the entire East of South Tyrol (Sterzing ↔ Bozen), with *VinKo* findings overlapping accurately with what Scheutz reports but with additional incursions (e.g. Meran); /ɽ/ as expected in Graun and Mals

4. In accordance with Scheutz, loss of assibilated rhotics (/rʃt/) in Brenner, Bruneck, Lüsen, Olang, Sand, Ratschings but in contrast with Scheutz, further loss of /rʃt/ in Terenten and complete disappearance of /rʃt/ outside *Pustertal* (except in Graun)

References

Braunger, P., W. Maier, J. Wessling & M. Schmidt. 2018. Towards an Automatic Assessment of Crowdsourced Data for NLU. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association.

Cordin, P., S. Rabanus, B. Alber, A. Mattei, J. Casalicchio, A. Tomaselli, E. Bidese & A. Padovan. 2018. VinKo, Versione 2 (20.12.2018, 09:20). In *Lo spazio comunicativo dell'Italia e delle varietà italiane. Korpus im Text*, 1–15. München: Ludwig-Maximilians-Universität.

Hilton, N. in press. Validity of crowd-sourced minority language data: Observing variation patterns in the Stimmen recordings. In M. Coler & A. Nevins (eds.), *Contemporary research in minority and diaspora languages of Europe* Contact and Multilingualism, Berlin: Language Science Press.

Klein, K. K., L. E. Schmitt & E. Kühebacher. 1965–1971. *Tirolischer Sprachatlas (3 Bände)* Deutscher Sprachatlas. Regionale Sprachatlanten Nr. 3: Tirolischer Sprachatlas. Marburg: N. G. Elwert Verlag.

Kruijt, A. 2022. *Crowdsourcing language contact: pronoun and article morphology in Trentino-South Tyrol and Veneto*: University of Verona dissertation.

Leivada, E., R. D'Alessandro & K. Grohmann. 2019. Eliciting Big Data From Small, Young, or Non-standard Languages: 10 Experimental Challenges. *Frontiers in Psychology* 10.

Munro, R., S. Bethard, V. Kuperman, V. T. Lai, R. J. Melnick, C. Potts, T. Schnoebelen & H. Tily. 2010. Crowdsourcing and language studies: The new generation of linguistic data. In *Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Proceedings of the Workshop*, 122–130. Stroudsburg, PA: Association for Computational Linguistics.

Rabanus, S., A. Kruijt, M. Tagliani, S. Bertollo, A. Tomaselli, A. Padovan, B. Alber, P. Cordin, R. Zamparelli & B. M. Vogt. 2014–. VinKo (Varietà in Contatto). Joint project of the universities of Verona, Trento and Bozen-Bolzano.

Scheutz, H. 2016. *Insre Sproch. Deutsche Dialekte in Südtirol. Mit dem ersten "sprechenden" Dialektatlas auf CD-ROM*. Bozen: Athesia.