

Change – The transformative power of citizen science

Old dialect words through the ages – the ABCs of dialect project

Amelie Dorn* (a), Rebecca Stocker (b), Philipp Stöckle (a)

(a) Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH),
Austrian Academy of Sciences, Vienna, Austria

(b) Department of German Studies, University of Vienna, Vienna, Austria

Abstract

Language is subject to constant change. The use of words can change across generations, just as the meaning of words can change over the years. The FWF Top Citizen Science project “The ABCs of Dialect” aims to transcribe and reflect old dialect words contained on the more than 100-year-old paper slips of the Dictionary of Bavarian Dialects in Austria (WBÖ). On the one hand, historical data is digitally processed, analysed and made accessible via the Zooniverse platform; on the other hand, interested citizens should be inspired to rediscover their linguistic and regional heritage. A crowdsourcing approach enables more extensive processing and transcription of the materials, while in a special reflection section citizens simultaneously reflect on and discuss the use and meaning of the classified words.

Keywords: citizen science, dialects, language, cultural heritage.

© 2024 Amelie Dorn, Rebecca Stocker, Philipp Stöckle

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published by NHM, BOKU and ECSA and peer-reviewed under responsibility of ECSA-ÖCSK-2024 (Change – The transformative power of citizen science)

Over time, languages are subject to constant change. This affects various linguistic levels. Changes in the overall use and meaning of words, i.e. the lexicon, are typically those that are most prominently perceived by people. While the use of words can change over the course of generations, the meaning of words can also evolve over time. The German lexeme *Ampel*, for example, was borrowed from the Latin word *ampulla* and originally referred to a small container for oil, before it then took on the meaning “lamp” and finally “traffic light”.

* Corresponding author. E-mail address: amelie.dorn@oeaw.ac.at

Dialects are considered to carry a particularly high level of cultural content of a language. Changes in the lexis of this typically spoken register of languages are particularly interesting to capture and have been the subject of a wide range of diachronic and synchronic studies across various languages.

As regards the German language, historic projects from the 19th and early 20th century, e.g., the Wenker project (Fleischer 2017) or the Dictionary of Bavarian Dialects in Austria (WBÖ) (Stöckle 2021) have aimed at capturing the regional variation of local dialects by making an attempt at documenting the local speech of the population. Nowadays better known as citizen science, i.e., the participation of lay persons at different stages of the research process (Pettibone and Ziegler 2016), these historic projects have already then seen the participation of local populations as a key aspect to understand language usage and practiced early approaches to what is now known as citizen science. The Top Citizen Science (<https://www.fwf.ac.at/en/funding/portfolio/communication/top-citizen-science>) project “The ABCs of Dialect” (FWF TCS134) follows up on this and brings a historic non-standard language resource into the 21st century together with citizen science methodologies and digital technologies.

The project aims to transcribe, reflect and analyze the old manuscripts and dialect words found on the over 100-year-old paper slips of the WBÖ. At the basis of the project lies a large and rich collection, mostly collected in the first half of the 20th century. Founded in 1913, the aim of the dictionary project was to comprehensively document the richly structured Bavarian dialects of (historical) Austria. To this end, language material was collected over the following decades on the basis of 109 questionnaires with the help of volunteer collectors, which was to serve as an empirical basis for the dictionary. The questionnaire-based collections were supplemented by direct surveys carried out by trained dialectologists in the form of field trips and questionnaire surveys. In addition to the empirically collected data, excerpts were taken from dialectological literature and other written sources. In this way, it was not only possible to fill gaps in the lexicographic material, but older stages of Bavarian dialects were also included in the data collection. Fig. 1 shows an example of a paper slip from the data collection, depicting the lemma *abfetznen* (“to brawl”) (Wahlmüller 2020).

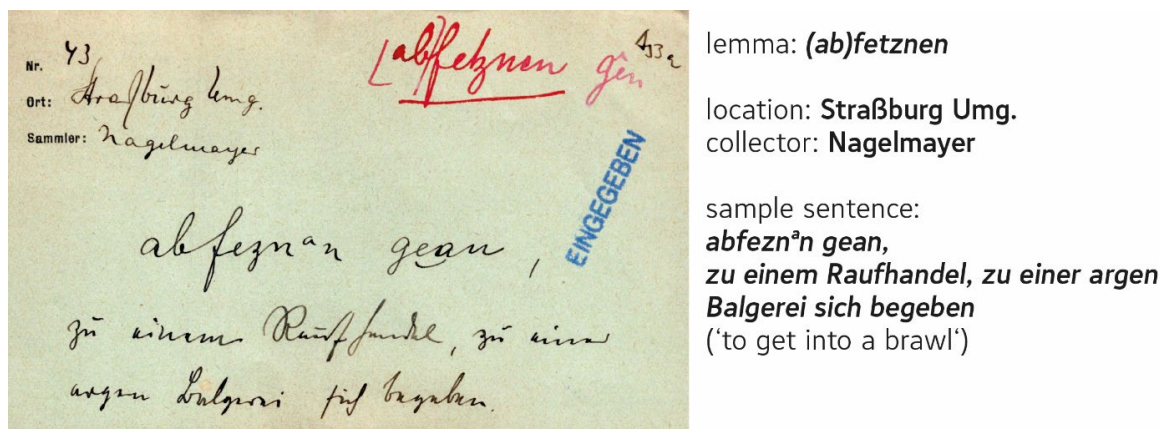


Figure 1. Example of a WBÖ paper slip on the lemma *abfetznen* “to brawl”

In addition to the lemma and information on the origin, this paper slip also contains a sample sentence illustrating the use of the lexeme. However, due to the empirical method which involved non-professional collectors, the type of information on the paper slips may vary considerably. While information on the origin and the lemma is available for most of the data, only a smaller proportion contains sample sentences. In addition, there is often information on the meaning, pronunciation or grammar.

In order to facilitate the work with the data and at the same time speed up the lexicographical work on the dictionary articles, a digital database was set up in the 1990s. As the first three volumes of the dictionary, covering the initial letters A, B/P and C, had already been published at this time, digitization (i.e., manual transcription) was started from the letter D onwards (Bowers and Stöckle 2018). Since December 2018, this database has been publicly accessible via the Lexical Information System Austria (LIÖ) (<https://lioe.dioe.at/>).

The project's main focus is to make a part of this rich and diverse non-standard language resource openly available to society and to fill people with enthusiasm for their own linguistic and regional heritage. In particular, the project aims to transcribe and reflect the paper slips ranging from letters A, B/P to C (hence the name of the project). This part of the collection comprises of over 400.000 paper slips which have to-date not been digitised. The historical language data is digitally processed, analyzed and made accessible via the Zooniverse platform, which is a popular crowdsourcing platform for a diverse range of citizen science projects (Van Hyning and Jones 2021), including projects on cultural heritage (Ridge 2020). A crowdsourcing approach enables more extensive editing and transcription of the materials, while in a special reflection section, citizens simultaneously reflect on and discuss the current use and meaning of the classified words. This combined approach of digital citizen science methods and reflections on individual language use makes it possible to demonstrate and analyze language change using old dialect words. To-date, the project counts over 1900 transcriptions and over 120 reflections of which over 60% show that the participants encountered the words in question for the first time.

One of the main challenges of the project lies in the range of diverse old handwriting systems used to note the German dialect words on the paper slips, including Kurrent script (*Kurrentschrift*) and Sütterlin, that are not in use any longer and that people nowadays also find increasingly difficult to read. Therefore, our project offers a range of supporting materials and aims to enhance data quality by having each paper slip transcribed by three participants. It also aspires to foster cross-generational participation, where younger persons who are typically more versed in handling digital tools like the transcription platform, and older persons who are typically more familiar with old handwriting systems, come together in analyzing and reflecting on their linguistic heritage. Additionally, pupils have also enjoyed learning new skills and old vocabulary in the context of the annual *Citizen Science Award*, organised by Austria's Agency for Education and Internationalisation (OeAD) (<https://youngscience.at/de/awards-und-guetesiegel/citizen-science-award>) where school classes, families and individual persons can participate in a diverse range of citizen science projects. While the pupils' (ages 15 and up) transcriptions tend to be less accurate – as they are not used to reading a lot of different handwritings anymore – they usually contribute more data to the project than other people by motivating and/or challenging each other.

References

- Bowers J, Stöckle P (2018) TEI and Bavarian dialect resources in Austria: updates from the DBÖ and WBÖ. In: Frank AU et al. (eds.): Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2 25-26), January 2018 Vienna, Austria. Wien: Gerastree proceedings, pp. 45–54.
- Fleischer J (2017) Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen: Dokumentation, Entdeckungen und Neubewertungen. Hildesheim/Zürich/New York: Olms. (Deutsche Dialektgeographie. 123) 201 S.
- Lenz AN (2018) The Special Research Programme “German in Austria. Variation – Contact – Perception”. In: Yearbook Sociolinguistica 32, 269–277.
- Pettibone L, Ziegler D (2016): Citizen Science: Bürgerforschung in den Geistes- und Kulturwissenschaften. In: Oswald K, Smolarski R (eds.): Bürger Künste Wissenschaft: Citizen Science in Kultur und Geisteswissenschaften. Gutenberg: Computus Druck Satz & Verlag, 57–69.
- Ridge M (2020) Crowdsourcing in Cultural Heritage: A practical guide to designing and running successful projects. In: Schuster K, Dunn S (eds.): Routledge International Handbook of Research Methods in Digital Humanities. London: Routledge, 363–383.
- Stöckle P (2021) Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In: Lenz AN, Stöckle P (eds.): Germanistische Dialektlexikographie zu Beginn des 21. Jahrhunderts. (Zeitschrift für Dialektologie und Linguistik - Beihefte 181). Stuttgart: Steiner, 11–46. [doi:10.25162/9783515129206].
- Van Hyning V, Jones M (2021) Data’s Destinations: Three Case Studies in Crowdsourced Transcription Data Management and Dissemination. <https://doi.org/10.5281/ZENODO.5750691>
- Wahlmüller E (2020) fetzen. Mit einem Lautungsüberblick von Fabian Fleißner. In: Wörterbuch der bairischen Mundarten in Österreich (WBÖ). Published via Lexikalisches Informationssystem Österreich (LIÖ). URL: https://lioe.dioe.at/articles/fetzen_Verb [accessed 29.04.2024].