Philipp Stöckle / Sabine Wahl

LEXICOGRAPHY AND CORPUS LINGUISTICS
The case of the Dictionary of Bavarian Dialects in Austria (WBÖ) and its database

Keywords Lexicography; corpus linguistics; Bavarian dialects; diminutives; digital humanities

The "Dictionary of Bavarian Dialects in Austria" (WBÖ) is a long-term project at the Austrian Academy of Sciences (ÖAW). Its main goal is the lexicographical documentation of the richly structured (historical) Bavarian dialects in Austria and South Tyrol.

The data for the WBÖ was collected in the first half of the 20th century, mainly with the help of dialect speakers who volunteered to support the dictionary by filling in numerous questionnaires. In addition, dialect literature as well as dialectological studies were extracted. This led to a data collection of about 3.6 million hand-written paper slips, the so-called "Hauptkatalog" (main catalogue), with often very detailed information on a huge number of dialectal words, including their meanings, their pronunciation as well as example sentences with which the collocations and the usage of the words can be studied. From the 1960s until 2015, the letters *A*, *B/P*, *C*, *D/T* and *E* were published in five volumes. From the letter *F/V* onwards, the Dictionary of Bavarian Dialects in Austria is published online via the platform "Lexikalisches Informationssystem Österreich (LIÖ; Lexical Information System Austria)".

Parallel to writing dictionary articles and with the intention of facilitating and speeding up the lexicographical work, the hand-written paper slips were manually transferred into the electronic database system TUSTEP (starting in the 1990s with the letter *D*). From 2014 to 2019, this data format was changed into a standard XML/TEI format (cf. Bowers/Stöckle 2018). This new database contains about 2.4 million entries and is also available online and open access via LIÖ. By exploiting the full potential of digital publications, LIÖ offers direct linking between the dictionary articles and the data as well as a mapping tool. Moreover, the database (although not being annotated systematically) can be used for linguistic research beyond lexicography (e. g., Stöckle/Hemetsberger/Stütz 2021).
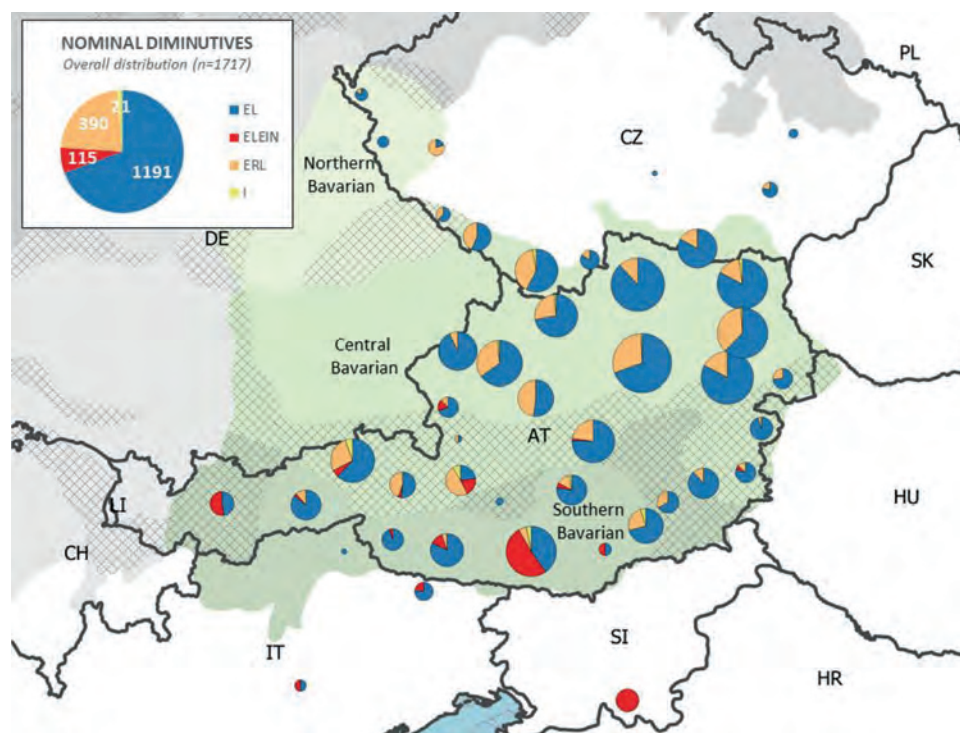
In our contribution, the functionalities of LIÖ are demonstrated as well as its potential to answer research questions using the example of nominal diminutive suffixes, which appear in many different variants in Bavarian dialects (e. g., diminutives of *Fleisch* 'meat': *Fleischl*, *Fleischerl*, *Fleischele* or *Fleischi*). While variation in nominal diminutives has been studied for contemporary (Standard) German in Austria (cf. Korecky-Kröll in print; Schwaiger et al. 2019), the WBÖ material represents a huge corpus of historical base dialects.

In the lexicographic process, diminutives can be dealt with in various ways. In the case of the WBÖ, the diminutives are included in the dictionary article of the base noun. All phonetic variants of the diminutives in the WBÖ database are subsumed under the suffix types EL (for diminutives ending in *-el* or syncopated *-l*), ERL, ELEIN (for *-ele/-le*), and I.

From a variationist point of view, corpus-based analyses were conducted using the WBÖ database to detect geographic patterns of diminutives in the Bavarian dialect area in Austria and South Tyrol.

For this analysis, 20 lemmas from recently published articles on nouns starting with *F* were selected. In the underlying database entries (n = 9,055), 1,717 diminutive forms were detected which could be clearly assigned to one of the diminutive types mentioned above.[1]

The following map shows the distribution of the nominal diminutive variants:

Figure 1: Geographic distribution of diminutives

As shown in Figure 1, type EL (*Fleischl*) is the most common variant, and it is found in all but a few regions. Almost a quarter of the diminutives belong to type ERL (*Fleischerl*), but they appear to form a geographic pattern with type ELEIN (*Fleischele*), which almost only appears in Southern Bavarian. The diminutives of type I (*Fleischi*) are generally rare, and their geographic distribution does not seem to follow a clear pattern.

In future research, our results on the morphology of diminutives in the WBÖ and – in a further step – on semantic and pragmatic aspects will be compared to findings on contemporary nominal diminutives in a variety of electronic corpora (of Standard German) in Austria (e. g., Schwaiger et al. 2019).

References

Bowers, J./Stöckle, P. (2018): TEI and Bavarian dialect resources in Austria: updates from the DBÖ and WBÖ. In: Frank, A. U. et al. (eds.): Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CHR-2), 25–26 January 2018 Vienna, Austria. Vienna, pp. 45–54.

Korecky-Kröll (in print): "Ma tuat net so vüü vernledlichen!" – oder doch? Verweigerung und Hinzufügung von Diminutiven als Schnittstellenprobleme von mündlichen "Wenker"-Übersetzungsaufgaben. In: Zeitschrift für Dialektologie und Linguistik.

---

[1]   We would like to thank Theresa Eiweck for her help with data preparation.

Lexikalisches Informationssystem Österreich (LIÖ). https://lioe.dioe.at (last access: 25-03-2022).

Schwaiger, S./Barbaresi, A./Korecky-Kröll, K./Ransmayr, J./Dressler, W. U. (2019): Diminutivvariation in österreichischen elektronischen Korpora. In: Bülow, L./Fischer, A. K./Herbert, K. (eds.): Dimensions of linguistic space: Variation – multilingualism – conceptualisations/Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung. Berlin, pp. 147–162.

Stöckle, P./Hemetsberger, C./Stütz, M. (2021): Die WBÖ-Belegdatenbank als Quelle für syntaktische Analysen – Möglichkeiten, Grenzen, Perspektiven. In: Wiener Linguistische Gazette 89, pp. 579–626. https://wlg.univie.ac.at/fileadmin/user_upload/p_wlg/892021/FSLenz_Stoeckle_etal.pdf (last access: 25-03-2022).

Contact information

Philipp Stöckle
Österreichische Akademie der Wissenschaften
philipp.stoeckle@oeaw.ac.at

Sabine Wahl
Österreichische Akademie der Wissenschaften
sabine.wahl@oeaw.ac.at