Methods in heritage language research: Reflections from the 14th Workshop on Immigrant Languages in the Americas (WILA)

Karoline Kühl & Samantha M. Litty

Abstract. This introductory paper by the WILA 14 conference organizers introduces the WILA 14 conference theme, "Methods", and serves as an overview of methods currently used in heritage language research, a discussion of methodological considerations necessary for conducting research, and includes suggestions and food for thought for both novice and experienced heritage language researchers. Discussions regarding methods and methodological choices from WILA 14 have been incorporated into this paper, and examples are drawn from the papers contained in this volume.

Keywords. heritage language; methods; research methods; immigrant language

1. Introduction. Since the early days of the Workshop on Immigrant Languages in the Americas (WILA), these annual meetings have invited research from translocated languages, immigrant languages, minority languages, and heritage languages (HLs) that are or were spoken in North, Central, or South America. Although WILA's initial focus was on Germanic heritage languages—simply due to the research areas of the original founders—other languages were included and are becoming increasingly more common (see also previous WILA proceedings). The time where the vast majority of projects presented at WILA focused on historical settings or data has also passed. Once, where small data sets reigned, these are now growing, so too are the many research questions and foci of the HL research community at large.

Not only has our knowledge on an impressive number of immigrant languages in the Americas grown since the first WILA in 2010, it has also become increasingly clear that methodological decisions at every stage of research—data collection, selection of specific data sets and the subsequent processing and analysis—have a decisive influence on the research results. For this reason, the 14th WILA (2023) was organized under the collective theme of 'Methods'. In this short introduction, we reflect on the findings that have emerged from our long engagement with immigrant languages in the Americas and, more specifically, from the contributions and discussions within WILA 14.

As can be seen from the broad range of HLs and projects presented in this volume, the research interests of WILA participants span a wide range across (sociolinguistic) studies of language variation and change to social factors that have consequences for language maintenance and shift. The approaches toward data sets are both synchronic and diachronic, and they rely on data sets of very different sizes and very different types, each of which have their own particular methodological challenge(s).

2. Choice of data points. As researchers we must admit to a general bias in that we prefer to work with communities or data sets where the HL is well represented. Historical or inherited datasets made with the intention of documenting certain immigrant varieties as, for example, Lester W. J. "Smoky" Seifert's Wisconsin German Questionnaire from 1946 (Bousquette this volume), Einar Haugen's recordings of North American Norwegian around 1950, or the North American parts of the *Corpus of American Danish* obtained by different Danish and Canadian linguists between 1963 and 1990 (used by Kuhlmann & Heegård this volume) present us with a selection of consultants deemed worthy of participation by the researchers. We mostly do not know, however, on which grounds they were chosen at the time. Working with previously collected

 Bergen Language and Linguistics Studies 15(1), 1–6.
DOI: 10.15845/bells.v15i1.4546.

datasets means we are reliant on previous researchers to not only have documented the language or dataset in use, but also their own methods and sometimes even thought processes. Without this information, we might be left with the following questions remaining unanswered:

- Were the consultants the ones most proficient in the HL or were they the only ones left who were able to have a conservation in the HL?
- Were the interviewed consultants the only ones home when the interviewer stopped by, or were they the ones who happened to answer the phone when then researcher tried to set up the interview?
- Were they the only ones willing or able to do the tasks or to do an interview?
- Were they just average people, randomly selected out of a huge community of fluent HL speakers? Or were they the HL speakers who are already active in their language community, or in language education, or language revival or activism?

The bias persists in recent data collection efforts as, for example, stated by Ladilova (this volume) who, in a thorough presentation of her selection criteria, describes how she chose smaller rural settlements for data collection as it was reported that Volga German and the original traditions were better preserved here than in the urban settlements.

As reasonable as the wish for 'good data' may be, this bias implies that we may get and perpetuate the impression of much more maintenance of the HLs than was actually the case, as we only get to see, or choose for ourselves, certain parts of a population. In this vein, when targeting a community of HL speakers (if one as such is known to exist), what considerations should a researcher focus on when determining who to include? If the HL speaker community is moribund or if there is no longer active (im)migration to the HL speaker community, the considerations will likely be different than in communities where HL speakers are plentiful or where continued (im)migration means the number of speakers and subsequent generations is likely to continue to grow. In this instance, a researcher intending to focus on a specific generation or social phenomenon may choose to attempt to document speakers across different generations (age-wise and according to immigration generation). Depending on the available data or potential for gathering more data researchers may also consider including data from both L2 speakers and monolinguals (or neither), or they may consider the use of different types of bilinguals (i.e., sequential vs. simultaneous, or bilinguals from a different contact setting, or with a different configuration of languages spoken).

As an insight for future data collection, we emphasize:

- It is crucially important to document the data collection process, including decisions about the consultants that are included in the data collection.
- When choosing consultants ourselves, it may be worthwhile to include consultants who represent many points on the continuum of HL competence, so from the most fluent speakers to the weakest speakers.

3. Mining data: Discovery. Recent work has shown that not only recordings of spoken language represent valuable data for research into immigrant languages. Data that was never intended as a basis for language research in the broadest sense has also proved to be fruitful, e.g. census data or cookbooks. Examples in this volume appear in the work by Hoffman who includes church records, college academic catalogs, college and high school yearbooks, and books written by members of the community in Lindsborg, Kansas, in the mapping of language shift from Swedish to English. Another example of how data that were not intended for linguistic research can be put to use are the so-called ego documents, for example the personal immigrant letters approached by Hjelde & Bjørke (this volume) in research on syntactic and morphological features.

It is clear that the choice of a certain data type will define the linguistic register, as for example personal letters are closer to "language of proximity" (Koch & Oesterreicher 1986,

2007) than planned and edited church records (“language of distance”, ibid.). Nevertheless, Hjelde & Bjørke (this volume) show convincingly that an identical text type does not necessarily yield the same outcome: The personal letters from Norwegian immigrants in the early 1900s who were acquainted with the process of writing personal letters, clearly differ from the letters of earlier immigrants (those who emigrated before the mid-1800s), who were much less proficient writers and thus relied more on formulaic language from religious texts. The letters of later American-born generations are different again, as they had become literate in the majority language, but only to some degree in Norwegian.

As an insight for future data collection, we highlight:

- Data types not intended for linguistic research should be recognized as useful for the investigation of patterns of language shift and, as long as the data do not simply represent standard language of distance, also for linguistic analyses of language variation and change.
- The same text/data type may show diachronic changes, caused by changes in how the register is formed, but also due to changes in the linguistic competences of the writers/speakers.

4. Born data: Elicited. Data sets elicited specifically for linguistic research are often built from spoken language production data, but the methods by which they are obtained vary widely. It goes without saying that different methods may have different task effects which in all likelihood influence the outcome. They are therefore relevant to acknowledge and discuss.

4.1. INTERVIEWS. **Semi-structured** interviews are often used as a method for recording free or at least naturalistic speech (Kuhlmann & Heegård and Jóhannsdóttir this volume). They are mostly based on the long tradition of sociolinguistic or ethnolinguistic interviews, addressing the consultant and their lives. A number of questions are often predetermined, not least in order to collect biographical details or to address certain experiences that are of interest to the researchers. However, most space is usually given to the free dialogue, in which the consultant determines the topics and the course of the interview. By doing this, researchers aim at documenting authentic linguistic strategies, also concerning avoidance of certain topics or construction. Still, as pointed out by van Baal (this volume) in favor of acceptability judgement tasks, the non-occurrence of a construction does not prove its non-existence in the repertoire of a speaker. The same holds for low frequency forms that mostly need to be elicited.

The **length** of an interview or interaction with consultants, in terms of activation of the HL, can be crucially important as these naturalistic conversations hold a lot of potential, particularly if they are allowed to last as long as possible. While interviews in the historical datasets are often only limited to a rather quick exchange of questions and short answers due to the limited technological capabilities of the recording devices of the time, these limitations no longer exist today (see Section 5 for a discussion of the challenges of processing data into corpora).

4.2. ELICITATION TASKS. **Guided narration** tasks are widely used in heritage language research, often in the form of Mercer Mayer’s *Frog, where are you?*, widely known in linguistic research simply as “the frog story” (Hietpas, Bousquette, and Jóhannsdóttir this volume) or the *Pear Film* (Jóhannsdóttir this volume). One obvious advantage of choosing these established works or task types for eliciting speech is that they will yield speech that is comparable across individual speakers of the same heritage language, for example in terms of motion verbs or lexical items, but also across different heritage varieties. They thus create the option of adding new, but still comparable data to an already existing data set when possible. It should be borne in mind, though, that not all HL speakers are necessarily familiar with the concepts required for the narrative in their HL. This may be the case either because they do not occur in the region (Bousquette this volume for animal terms in Wisconsin German) or because they are not part of the speakers' everyday lives. This can lead to the lexical terms being either unknown or so unfamiliar that they are only produced with a substantial delay.

Some studies develop **visual or oral stimuli** for different purposes, e.g., for eliciting past tense in North American Icelandic (Jóhannsdóttir this volume), non-subject initial sentences in Latin American Norwegian (Melvær et al. this volume), or acceptability judgements of definite forms in North American Norwegian (van Baal this volume). It is fundamental to recognize and discuss the extent to which these stimuli have led to their goal and what challenges may have arisen in their use as the contributions mentioned above do in an exemplary manner. For the advancement of heritage language research, it seems crucial to make newly developed stimuli (freely) available to other researchers, including the dissemination of findings regarding their applicability. This in turn can facilitate the production of comparable data sets across languages and the replication of studies at a later point of time.

Some tasks, such as **translation** tasks as employed by, e.g., Seifert's Wisconsin German Questionnaire (1946) (see Bousquette this volume), seem to ebb and flow in their popularity. The changes in popularity of a set of tasks may be due to researchers shifting toward other data elicitation methods, or from gaining experience and learning for example, that heritage language speakers who may not often use the language or only use it in specific domains may struggle to produce translations as this act of translating specific sentences or phrases is likely quite far from their regular usage. Another aspect that makes this type of data collection problematic is the trigger effect that a translation task can potentially cause due to the simultaneous high degree of activation of both languages. For inexperienced translators the processing of translation tasks can lead to an increasing occurrence of cross-linguistically analogous structures and also of transfer occurring more frequently, neither of which would then be representative of naturalistic, "genuine" language use by the informants. On the other hand, the advantage of these tasks is, of course, that they provide fast and clear data on lexical items and constructions. Additionally, as translation tasks were quite popular at one time, this type of task may also be seen as a way to attempt to keep some similarity in the data across datasets collected years apart or in different heritage language communities.

4.3. CONSULTANT CONSIDERATIONS. A main methodological consideration in all kinds of HL data elicitation is the research setup in terms of **language choice**. In cases of a moribund HL or where the use of the HL is restricted to very specific settings and participants, a researcher's linguistic persona becomes a very significant factor. If a researcher speaks only, for example, a homeland variety of the HL and does not speak the dominant language (or pretends not to do so), then the use of the HL by the consultants becomes communicatively plausible and necessary. If this is not possible (as is mostly the case for North America), then researchers need to decide on strategies on how to act if the consultants switch to the majority language (see Ladilova this volume). Their decisions as to prompt the switch back to the HL will shape the outcome in terms of data. Thus, this is a methodological issue to be considered and documented as part of the research endeavor.

When collecting data, the **activation of a HL** (the language mode, cf. Grosjean 2021) should also be included in the methodological considerations. In cases where we are dealing with people who use their HL rather infrequently in everyday life and almost only with familiar people, an encounter with an unknown researcher who speaks a form of the HL fluently and who starts carrying out tasks relatively suddenly can be overwhelming. Taking the time to talk to the consultants before the data elicitation starts is important. It is also important to ensure that the consultants do not feel under time pressure and that the tasks do not place an undue burden on them. If possible, the same consultants should be visited several times across various intervals. On the one hand, this enables real time-panel studies of groups and individuals (see Hietpas this volume), but it also means that the consultants are better able to grasp the data collection situation and, through cognitive activation of their HL, may show more fluent speech production on subsequent visits rather than by simply having only one data collection interview.

One concern that has not been given much or any attention in previous methodological discussions is the **well-being of the consultants** in a data collection situation. Especially when working with older consultants, longer tests can be demanding, possibly even quite stressful. As van Baal (this volume) discusses, it is critical to make sure that the consultants have actually understood the instructions and stimuli, perhaps by adding a test trial. This must be taken into account in methodological considerations and reflections.

4.4. TRIANGULATION. It is of course possible to combine different forms of data elicitation for very specific purposes and get the best out of everything. A good example of this is Melvær et al. (this volume), who employ a retelling task (of Chaplin's film, *Modern Times*) for eliciting non-subject initial sentences in Latin American Norwegian. The task was to retell the story to a researcher who had left the room during the film and who would then draw a comic according to their stories, thereby creating a plausible communicative need leading to relatively free speech that still yielded results for a specific construction. An example of a successful triangulation of data is the data collection by Ladilova (this volume) in several Volga German communities in Argentina that has included both quantitative and qualitative methods, such as interviews including follow up-conversations, questionnaires, observations and photographs, picture stories and word lists (other projects such as the Icelandic in America-project have also employed different triangulation approaches, see Jóhannsdóttir this volume).

5. Data processing. A main methodological consideration regards the decisions embedded in the processing of data after they have been collected: This concerns the possibilities available to a researcher or a research group in terms of time and financial resources. Naturalistic HL data are often characterized by variation in terms of non-standard features such as dialectal variation, multilingual language use, and perhaps fossilized features. The variation is, of course, neither problematic in itself nor unexpected, but it presents a methodological challenge with regard to e.g., corpus tools, transcription standards, and so on. Transcribing spoken multilingual language is an extremely time-consuming (and therefore costly) task, and the variation in HL data adds to this. The efforts grow even more, if annotations are to be added to the data. It goes without saying that the processing of many consultants' speech into (searchable) HL corpora is only possible for longer-term research projects that are able to fund stable data processing logistics. Examples of this are the *Corpus of American Nordic Speech* (CANS), the corpus of North American Icelandic created within the project *Heritage Language, Linguistic Change and Cultural Identity*, and the *Corpus of American Danish* (CoAmDa). Often, individual researchers will need to select which data to process further according to their available options and capabilities. Crucially, these decisions need to be documented and discussed, as they create datasets that are then seen as representative for a certain speaker community (see Ladilova this volume, and previous discussion of the bias of choosing the most fluent speakers for data sets).

Spoken data elicited in a narrower way, targeting very specific features, may be handled more easily, but still the processing is time-consuming and demanding. In the future, AI may facilitate data processing. However, naturalistic HL data which are often multilingual and contain non-standard or perhaps fossilized features will probably still be challenging for LLMs trained on monolingual standard language data.

Decisions regarding transcription standards are another major consideration in the process of data processing. It requires decisions as to how the linguistic variation and the non-standard features that are inherent to a lot of naturalistic HL data are to be transcribed, whether the orthographic form should reflect the variation (which would make searches more difficult, unless the principles for the orthographic realization are documented carefully and comprehensively), or whether this happens in another form (e.g. through annotations). In other words: The transcriptions, e.g., the decision for or against standardization, determine the conditions for e.g., search processes in the data. Hjelde & Bjørke (this volume) discuss the challenges and limitations of searches in their data set regarding orthographic variation in standard spelling across time (from Dano-Norwegian to Bokmål) and also regarding individual orthographic variation produced by untrained writers which is impossible to grasp in search strings. Nevertheless, the individual variation is, in itself, an important insight which would have been lost if the letters in their dataset would have been normalized to standard spelling.

As the most important point for future data processing, we maintain that choices of which data/speakers out of a bigger set are processed further as well as decisions regarding transcription and annotation need to be documented carefully and comprehensively and stored along with other metadata of the specific data set.

6. Data Analysis. Means of comparison. A key methodological decision in HL data analysis is the standard of comparison that researchers apply to their data when assessing language variation and change. There is a general consensus in linguistic research on heritage languages that the best basis for comparison is the actual baseline, i.e. the variety that the speakers acquired as children. This baseline does not have to correspond to the (written) homeland standard language of the present, but may show features of "languages of proximity", dialectal and social variation, fossilized features of historic language and also be characterized by multilingual language use. If, however, the baseline cannot be established, the comparison to the homeland variety may be the most appropriate choice – at least if this does not imply that the language variation and change in the HL data are not categorized as wrong, but simply as different. Often, though, triangulation through different sources, for example recordings or descriptions of homeland dialects, historical language descriptions and findings from contact linguistic studies in general can provide a fairly good means of comparison, even if the actual baseline is not available.

7. Conclusion and inspiration. In this brief introductory paper to the 14th Selected Proceedings of the Workshop on Immigrant Languages in the Americas, we present some common methodological reflections which we consider most relevant and important in HL linguistic research, showcasing some of these considerations as they appear in the research in the rest of this volume, and providing brief suggestions we hope may help future researchers in the field. Depending on whether a researcher is focusing on a project with historical or modern data, with born or discovered data, if the data is intended for linguistic research or other purposes, the researcher must carefully consider how to approach their research question and goals, and also how these may be used in the future.

Moving toward the future we might also consider as Salmons (this volume) does, in using a set of previously explored data and completed research from other HLs, HL communities, or other researchers to answer a research question too broad to be answered by only one HL or HL community alone. As Salmons' paper closes this volume with food for thought for future WILAs, we too would like to conclude this paper with a plea for future HL researchers to spend some time considering the methodological considerations we've presented here.

References

Koch, Peter & Wulf Oesterreicher. 1986. Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36(85). 15–43.

Koch, Peter & Wulf Oesterreicher. 2007. Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik* 35(3). 346–375.

Grosjean, François. 2021. Life as a bilingual. Knowing and using two or more languages. Cambridge: Cambridge University Press.