Language (2026), **102**:2, 469–496
doi:10.1017/S0097850726000081

Linguistic Society of America

TEACHING LINGUISTICS

The Language Profiles Project

Avery Ozburn

University of Toronto Mississauga, Canada

Email: avery.ozburn@utoronto.ca

Received: 05 June 2023; **Revised:** 22 December 2025; **Accepted:** 26 January 2026

Keywords: linguistic pedagogy; underrepresented languages; data sets; community engagement; inclusion

Abstract

This article introduces a novel linguistics pedagogy resource called the Language Profiles Project (LPP), an open access resource for linguistics instructors and students. The goal of the LPP is to make it easier and more engaging for instructors in North American institutions to incorporate underrepresented languages into undergraduate linguistics courses. More concretely, a language profile combines data sets for use in linguistics courses with contextual information about the language and culture. In this article, I describe the resource, the motivations for creating it, and some ways in which it could be used.

1. Introduction

This article introduces a novel linguistics pedagogy resource called the Language Profiles Project (LPP), an open access resource for linguistics instructors and students. The goal of the LPP is to make it easier and more engaging for instructors in North American institutions to incorporate underrepresented languages into undergraduate linguistics courses. More concretely, a language profile combines data sets for use in linguistics courses with contextual information about the language and culture. In this article, I describe the resource, the motivations for creating it, and some ways in which it could be used.

The foundational pedagogical assumption behind this resource is that diversity of language representation in linguistics courses helps students to gain a more balanced perspective about language. I am a junior professor in a linguistics program at a North American institution (University of Toronto Mississauga), and my personal pedagogical aspiration as related to this project is to develop a linguistics curriculum at my institution that is not focused on majority languages, but instead presents students with a more balanced perspective. An additional aspiration is to have some fundamental knowledge about each language discussed in my classes, including how to properly pronounce the language name, what the status of the language is, where it is spoken, and a little bit about the people, culture, and history. The goal is to be able to bring that information to students, such that they see the data sets in class as real languages, not just problems to solve.

This article is organized around several themes central to the motivations, content, and applications of the LPP. Section 2 situates the project in the context of my institution and research team. Sections 3 through 5 then discuss the motivations, content, and applications of the LPP with respect to three core themes: data sets, contextualization, and community collaboration. Specifically, I discuss data sets in linguistics classrooms, from the perspective of the literature, my own experiences, and the LPP in Section 3, and then move into contextualization, including the reasons why we might want to contextualize languages in linguistics classrooms and how the LPP aims to accomplish this goal, in Section 4. I then discuss

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Linguistic Society of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

collaboration with communities in Section 5, bringing in the literature on community engagement and explaining the LPP framework. Section 6 talks about the long-term goals of the LPP and concludes.

2. Situating the project

I begin by explaining the general context in which the LPP is situated, including the research team, the language profiles currently being developed for the website, and the choice of languages.

2.1. Project team, context, and assumptions

The project is run from a North American university in which instruction takes place predominantly in English. My university is on the traditional lands of the Huron-Wendat, the Seneca, and the Mississaugas of the Credit. I am a junior professor in a Linguistics program, teaching phonology, phonetics, and fieldwork courses from introductory to advanced levels, and I began this project out of a desire to have more easily accessible diverse data sets, complete with information on how to pronounce the language names, where the language is spoken, and other such contextual details. Besides me, the project team consists of a large number of students, both undergraduate and graduate, who are helping to build the resources in various capacities. About twenty-five undergraduate students and five graduate students have been involved over the course of the project. The team also consists of community-member collaborators, some of whom are students and some of whom are not. A few other professors, both at my institution and elsewhere, have also chosen to be affiliated in some capacity, such as collaborating on languages they speak and/or work on.

In the context of my institution, it is traditional for introductory linguistics to focus almost entirely on dominant varieties of English. For example, in the sound introductory course, it is typical for only the IPA of English to be introduced, with no content on IPA symbols that are not used for English. Similarly, we have a minor program specifically for English Language Linguistics, in which all of the students' linguistics exposure comes from English. Going beyond English, we have an upper-year course on Romance linguistics that is regularly offered, but a class on African linguistics that is on our list of courses has never been offered. It is from this context, of a North American, primarily English-language institution with a focus on major European languages, that I use the term 'underrepresented' throughout this article. For example, while Yorùbá is not underrepresented within Nigeria, it is within the linguistics courses offered at my institution. The word 'underrepresented' can have a wide range of meanings, but for the purposes of the LPP, the languages are underrepresented in the sense that (i) they are not often seen in linguistics classes in my North American institution, (ii) they have been historically marginalized and/or disadvantaged in some way, such as through genocide, colonization, or language policies, and (iii) the effects of this marginalization continue to this day.

The project has been funded through a small institution-internal teaching grant for undergraduate courses, which provides funding for projects that reinvent or enrich undergraduate education. This funding supports the work of student researchers and community-member collaborators. Since the particular grant is designated to support undergraduate education at my institution, the primary audience of the LPP is meant to be undergraduate students at my institution, from the perspective of the funders. This fact places certain limitations on what is currently possible for the LPP. Specifically, while the goal of the LPP is to move away from Eurocentricism, the context of the project means that the materials are all in English, so the audience remains Anglocentric. Translation of some of the materials into French is also in progress, for the purposes of linguistics courses offered in French. However, we are unfortunately not currently able to create materials for each language in the language itself, due to the funding structure and its limitations on the primary audience for the materials.

With a project of such a large scope, quality control is potentially an issue. I act as an editor for the resource, but we have several methods of quality control, with many members of the team involved in checking materials. Community members, students, and I go through all of the content before putting it

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

online. This includes data sets that are being cited from the literature, to ensure that they accurately reflect the variety of the language being represented in the profile. My graduate students in particular help to check data sets outside of my main area. Of course, it is still possible that there could be issues that might not be caught, especially because there may be disagreement among community members on the best representations. We thus have a feedback form on the website for users to notify us of any issues, as an additional fail-safe for quality control. While no one has raised issues as of yet, we will address any that arise.

2.2. Current and in-progress languages

The LPP currently has five available languages online: Ilocano, Kanien'kéha, Pangasinan, Yiddish, and Yorùbá, and we are in the process of finalizing and uploading eight others: Armenian, Dagaare, Kalenjin, Mongolian, Oromo, Q'anjob'al, Scottish Gaelic, and Zulu. Another eight languages are in progress, in varying stages of completion. I refer to 'languages' when I refer to the LPP, but as a referee points out, it would perhaps be more accurate to think of these as language varieties, specific to the language users consulted in the development of each profile. For example, the Oromo profile primarily represents the Eastern Oromo variety, along with the more specific location, age, and gender of the primary Oromo-speaking collaborator. This has potential implications in that other community members may feel that the profile is misrepresenting them if there are differences from their own varieties. In a project with limited scope and funding, it is impossible to fully resolve this issue, but the website encourages feedback in order for us to incorporate other varieties and perspectives where applicable.

As part of the purpose of this project is to contextualize these languages, I briefly provide some background on the thirteen languages for which profiles are currently or nearly available. Table 1 lists the languages, their language family (with more specific families where widely agreed upon in parentheses), and their region. Readers are referred to the website, https://languageprofiles.ca/, for further information about each language.

2.3. How languages are chosen

There are two primary factors behind the decision of which languages to include in the LPP. The first criterion is that the language is underrepresented, as defined in Section 2.1 above. Some of the chosen languages are endangered, while others have millions of speakers but are nonetheless underrepresented

Language	Family	Primary region
Armenian	Indo-European	Armenia
Dagaare	Niger-Congo (Gur/Mabia)	Ghana
Ilocano	Austronesian	Philippines
Kalenjin	Nilotic	Kenya
Kanien'kéha	Iroquoian	Canada, United States
Mongolian	Mongolic	Mongolia
Oromo	Afro-Asiatic (Cushitic)	Ethiopia
Pangasinan	Austronesian	Philippines
Q'anjob'al	Mayan	Guatemala
Scottish Gaelic	Indo-European (Celtic)	Scotland
Yiddish	Indo-European (Germanic)	Central/Eastern Europe (historically)
Yorùbá	Niger-Congo	Nigeria
Zulu	Niger-Congo (Bantu)	South Africa

Table 1. *Languages currently available or soon to be available in the LPP.*

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

according to these criteria. Language profiles all contain some historical and contextual information related to how and why the language is underrepresented.

Two examples of languages included in the LPP that are underrepresented in distinct ways are Yiddish and Oromo. While many students at my institution have heard of Yiddish, the Yiddish-speaking community has suffered immensely due to genocide: millions of speakers were murdered in the Holocaust, which led to global decline in the use of the language (see e.g. Weinreich 1992). By contrast, many people in North America have never heard of Oromo, even though it is one of the most widely spoken languages in all of Africa (Walga 2021), and it has been subject to marginalization through language policies, as it was not permitted to be used in official capacities in Ethiopia for many decades (Baxter 1978).

The creation of language profiles only for marginalized languages is potentially controversial, as some literature (e.g. Bender 2019) has argued that special focus on minority languages may inadvertently support the treatment of majority languages as 'default' and minority languages as 'exotic'. However, the funding resources available to build the LPP are extremely limited, and I feel that they are best directed at raising the profile of languages that have been underserved, rather than at providing even more space for languages like English or French. It is my eventual goal to be able to include nonminoritized languages as well, for exactly the reason of not treating minority languages as 'exotic'. However, with the current funding situation, I feel that the aims of the project are compromised if I start with majority languages, rather than with as many minoritized languages as possible. As such, for now, the LPP focuses on minoritized and marginalized languages.

The second criterion for inclusion in the LPP is that the team is connected in some way to members of the language community. Team members may be part of the community themselves, may know members of the community directly, or may be put in touch with the community through word of mouth. This criterion was established so that the project team can work directly with community members (both within and outside of the project team) to ensure that they are happy with how the profile represents them. Particularly with underrepresented and marginalized languages, it is critical that the LPP accurately reflects the language and culture, and the involvement of community members is the best way to make sure that a language profile is done properly and respectfully. This choice also allows language profiles to be much richer than if they were created without community input. The community can be involved in the construction of the profile and can guide the LPP team on how they want their language represented to linguists, as well as on how to ensure that the profile gives back to the community. This process follows the community-engaged fieldwork theories and practices elaborated on in Section 4.1 and Section 5.1.

2.4. Target audiences

The LPP occupies what Bucholtz (2021) calls the 'contact zone' between academia and communities, where anyone is a collaborator and there is no notion of 'the field' as being outside of or in any way less than the academic field.[1] As such it is targeted to both audiences, with the view that these traditionally different audiences should not be separated. Within these larger categories, however, the LPP is targeted primarily at a subset, particularly when it comes to the academic audience of linguistics instructors. As pointed out by referees, not all linguists believe that incorporating social context into linguistics is a good thing, or that incorporating underrepresented languages is good; some may believe that the former, in particular, compromises the scientific nature of the discipline. The LPP is not aimed at these linguists, although I provide arguments in Section 4.1 on why contextualization is, in fact, crucial for linguists. Instead, the pedagogical aim of the LPP is to make incorporating underrepresented languages and their contexts into linguistics classrooms easier for the many linguists who already would like to do so. That

[1] In this vein, I use the terms 'field' and 'fieldwork' throughout this article when referring to the linguistic subdiscipline involving primary documentation together with language users. I use this term in a modern, community-engaged sense, to refer to collaborative work between academics and language users, rather than in the traditional extractive sense of the word.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

said, in each section, I provide motivations for the resource that argue for the linguistic, ethical, and social importance of context and of underrepresented languages in linguistics classrooms.

3. Data sets in linguistics classrooms

Now that the project has been situated, I move on to discussing one of the core elements of the LPP: the data sets. As noted, one of the fundamental goals of the LPP is to create data sets for use in undergraduate linguistics classrooms, with a current focus on underrepresented languages. In this section, I outline the motivations for creating these data sets, the LPP content as it relates to data sets, and possible applications of these LPP data sets.

For the purposes of this article and the LPP, a 'data set' refers to a set of curated language examples, presented using IPA and/or a form of standardized orthography, for instructors and students to use in linguistic analysis so as to examine a specific point about linguistic structure. The LPP includes a variety of data sets for each language, spanning linguistic subfields as well as topics within them. Currently, we have data sets for phonetics, phonology, morphology, syntax, sociolinguistics, and historical linguistics, with the intention of adding semantics and other areas as the project continues to grow. The data sets range over undergraduate levels: some are straightforward illustrations of basic concepts, while others are more complex. For example, one of the Yorùbá syntax data sets simply illustrates SVO word order, while one of the Oromo phonology data sets is a complex counterbleeding problem. Each data set provides the data and a basic description of what it shows, and is not couched within any specific formal theories. Instructors are therefore free to adopt them in any way that works in their courses. Data sets could be used for concept illustrations in lectures, for exercises in tutorials, for extra practice problems, or for quizzes, assignments, and tests. However, they are primarily intended for lectures/tutorials and extra practice exercises, because the publicly available descriptions of what the data is showing may make them unsuitable for assignments.

Examples in each data set may be directly from community members who are part of the project, or in some cases may come from the linguistic literature on the language. In all cases, however, they have at minimum been confirmed with language users. Many of the data sets, particularly for phonetics and phonology, include recordings from a language user, which helps directly tie the data to the people behind the language. As much as possible, culturally relevant examples are included, to additionally help connect the data to the community. That said, data sets are presented primarily in writing (often IPA) in columns or rows in the Western knowledge tradition; while the LPP represents a first step toward inclusion, incorporating other kinds of knowledge systems into the way linguistic data sets are presented is unfortunately left to future work. While not ideal, using traditional linguistic data sets creates a lower barrier for instructors to adopt the LPP in their courses.

3.1. Background and motivations

The creation of a database for linguistics data sets on underrepresented languages is motivated by several factors from both the literature and my own personal experience. In this section, I overview barriers to including underrepresented languages in linguistics courses (Section 3.1.1), arguments for moving beyond the historical focus on majority languages in the field (Section 3.1.2), and the use of diverse data sets in promoting greater learning and belonging in students (Section 3.1.3).

3.1.1. Lowering barriers to inclusion of underrepresented languages

There are several barriers to including underrepresented languages in undergraduate linguistics classrooms. Including less-represented languages can greatly enrich student education, as discussed further below, but doing so can be difficult for instructors, for reasons ranging from the challenge of finding a relevant data set to the problem of not knowing how to pronounce the name of the language. This section addresses some of these issues and describes how language profiles can lower these barriers.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

One of the major reasons why linguistics instructors might use examples from majority languages, even when other languages also show the same pattern, is simply availability of the data. Instructors are busy, and it can be difficult to hunt through grammars and journals to find examples of other languages with a specific pattern from which to construct a data set. Especially for linguists who speak and work on majority languages themselves, it can be hard to even know where to start to look for data sets on underrepresented languages. Resources like *WALS* (Dryer & Haspelmath 2013) can help identify relevant languages, but it remains time-consuming and challenging to sift through the results for a language that fits the purpose, in terms of both the pattern and the availability of sufficient data. Even after finding an appropriate language, instructors then need to track down sources to compile a list of examples to form a data set. Some of these sources are not available online, are not possible to copy/paste from if they are electronic, are not written in IPA, are missing glosses, or otherwise are difficult to use as-is. As a result, it is understandable why instructors might default to talking about languages they know more about or languages with easily available data sets, such as from textbooks: doing so can save a great deal of preparation time.

The LPP comes out of a desire to mitigate this issue by compiling data sets on underrepresented languages for easy adoption by instructors. The collection of these data sets in one place, with searchable topic tags (e.g. 'assimilation', 'word order'), can vastly reduce the time commitment of finding underrepresented language data sets to include in a course. The data sets also are already prepared and ready to be adapted to the level and analytical choices of the course, and they are available in both online and Word document format. Thus, the LPP directly addresses a major obstacle to the inclusion of underrepresented languages in undergraduate linguistics teaching: the amount of time and energy that instructors must spend to include them.

Another barrier to using underrepresented languages in classrooms is that instructors may not be comfortable discussing a language they feel they have insufficient knowledge about, particularly if they are unsure about fundamental information like how to pronounce the language name. I have had this experience myself and have heard of this from many colleagues, many of whom admit to using pronunciations that they make up and later find out are incorrect. Obviously, this sort of situation is not ideal. The LPP can mitigate such issues by providing instructors with pronunciations of the names of underrepresented languages by language users themselves, so that instructors can feel confident mentioning the language. Beyond the issue of pronunciation, some underrepresented languages have historically been referred to in linguistics by names that are now considered derogatory or pejorative. Instructors creating data sets from older resources may not have the background to know that, and they may not even think to look for a preferred term, leading to perpetuation of the use of these names. Again, the LPP can address this concern by providing easy access to preferred language names.

In addition to language names, instructors may not feel comfortable discussing a language without knowing other information, such as where it is spoken, how many people speak it, what its endangerment status is, and so on. Students sometimes ask about this information, and even when they do not, it can be important information for instructors to be aware of. As an example, there have been occasions in my classes when student questions raise the issue of what types of other data might be interesting to see for a particular phenomenon in a particular language. Knowing how feasible it is for a linguist to eventually learn that information depends directly on knowing about the language status. The LPP compiles this information for the languages included, which can increase instructor confidence in discussing them. Particularly in classes where it is relevant to give language examples out loud, not knowing anything about how a language sounds can also be a barrier. Instructors may not want to mispronounce example words and sentences, making it challenging to discuss less familiar languages. The LPP's inclusion of language samples, as well as recordings for many of the data sets (especially phonetics and phonology ones), reduces this barrier as well.

Overall, language profiles aim to make the incorporation of underrepresented languages into classrooms easier by collecting data sets for instructors to adopt, along with basic information about the language (pronunciation of the name, where it is spoken, sample recordings, etc.) that will allow instructors to feel more comfortable discussing a language that they may not know anything about or

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

even have heard of otherwise. The goal is to lower these common barriers so that more underrepresented languages can be included.

3.1.2. Historical focus on majority languages

The Linguistics Equity, Diversity, and Inclusion Repository (LEDIR) provides a guide for inclusive linguistics teaching, including through a proceedings paper (Sanders et al. 2020), an online handbook (Umbal et al. 2021), and a book chapter (Sanders et al. 2024). While the focus differs from that of the LPP, these resources nonetheless discuss similar considerations. Among their recommendations for linguistics instructors are to avoid focusing on English and other major European languages, to develop problem sets from languages that are not major standardized European languages, to expose students to a wider range of voices, and to use lesser-studied languages to exemplify linguistic phenomena. The LPP's data sets from marginalized, underrepresented languages, including content developed by community members, make it easier for instructors to target all of these goals. The LEDIR authors argue that instructors will pass language-based biases to students, even unintentionally, if instructors fail to challenge biases in teaching the next generation. The LPP answers this call to action, tackling underrepresentation in linguistics courses by creating a detailed resource from which instructors can obtain data sets.

Beyond education, there are wider discussions within the field about the dangers of focusing only on majority languages. Related to the European colonial history of the field of linguistics, Gil (2001) notes that the current consensus in the field is that linguists must overcome Eurocentrism and increase investigations of non-European languages.[2] He discusses numerous examples in which Eurocentric biases negatively affect fieldwork on other languages, such as researchers missing important generalizations due to the failure to tone-mark data. He also discusses cases in which Eurocentrism can lead researchers toward questions that are not relevant to the grammar of a particular language, or away from hypotheses that are worth considering. He argues that a critical part of linguistic work, particularly linguistic fieldwork, is unlearning the Eurocentrism embedded in the field, including in linguistics education. The LPP addresses these concerns by facilitating a move away from Eurocentrism in the earliest years of linguistics education by making it easier for instructors to incorporate a variety of languages, rather than focusing on majority European ones.

Relatedly, Harrison (2007) notes that the loss of languages is about more than just language;[3] it is also about ways of knowing the world and about human cultural heritage. He argues that language death starts with discrimination of some form, and that people will keep their knowledge and their language if they feel its place is affirmed, but that they may not if they are led to believe it is useless. The LPP affirms the place of minority and historically excluded languages in linguistics classrooms. In addition to the importance of preserving these languages for the communities themselves, Harrison (2007:19) notes the importance of language preservation for linguists, saying that '[l]inguists sorely need the oddest, quirkiest, and most unusual languages' and that linguists cannot understand human cognition just from majority languages. It is worth noting that 'odd' and 'quirky' refer to perception by linguists, not by speakers, and that these are potentially problematic descriptions, in that they may suggest that patterns a linguist has not seen before are 'quirky'. The point of the LPP is in fact to reduce this perception of nonmajority languages as 'quirky' by giving such languages equal representation in linguistics training from the beginning. While Harrison may not have meant these terms in this way and may not have been critically challenging these traditional views of 'odd' and 'quirky', it is clear that including a wider variety of languages beyond majority ones broadens linguistic understanding, making this important for linguistics students.

[2] Note that there are also other views of what constitutes Eurocentrism, such as Eurocentric views of what constitutes language, which are discussed further in Section 4. See also work such as Mellow 2015.

[3] I use the word 'loss' rather than 'disappearance', as used by Harrison, due to work on language oppression that has pointed out the problems with unaccusative verbs like *disappear* being used to refer to language. See, for instance, work by Leonard (2019, 2021a).

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

3.1.3. Inclusion

Language profiles can also help students from underrepresented communities feel represented in the classroom, by increasing the likelihood that their linguistics instructors will talk about their language or the language of their heritage community.[4] As noted, the languages represented in the LPP are ones whose users have been historically marginalized or disadvantaged through events like genocide and colonization, and by language policies. Including data sets from these languages in linguistics classrooms gives the users of these languages a voice that they have not historically had, instilling important values in students with regard to diversity and inclusion.

Pedler et al. (2022) argue that a sense of belonging at university is a crucial factor for student success. In their study, students with a greater sense of belonging are found to have better levels of motivation, self-confidence, engagement, achievement, and enjoyment at university. They also find that a sense of belonging reduces drop-out rates. The LPP's specific focus on marginalized languages increases the chances that students from those communities will have their languages and cultures represented in linguistics classrooms. Since many of the profiles are authored by community members themselves, students are also exposed to work credited to members of underrepresented communities. Both of these factors can make students feel a greater sense of belonging in the classroom. Additionally, the LPP's engagement with community members to ensure that their language and culture is represented in the way they feel best again fosters the sense of belonging that is crucial to student success, since misrepresenting a language could have a severe negative impact on students from that community.

As a concrete example, consider the discussion of verb-second languages in the classic textbook *Syntax: A generative introduction* (Carnie 2021). The title and description of this section mention Germanic as an entire family, but all of the examples come from German, and the German pattern is not found in all Germanic verb-second languages. In Yiddish, unlike in German, verbs continue to come second in embedded clauses (Jacobs 2005). Yiddish and German are closely enough related that a Yiddish speaker seeing the German embedded clause examples could easily understand them, but the word order is incorrect for Yiddish. The fact that this data is presented as representative of all Germanic verb-second languages may therefore lead to feelings of exclusion among Yiddish-speaking students, or to such students not trusting the material. The LPP's Yiddish verb-second data set provides richer data for instructors to use, while also helping to build more inclusive classrooms for students from underrepresented language communities.

An example from personal experience showing the value of this inclusivity comes from my interactions with Tamam Youssouf, the Oromo consultant who worked with us on the Oromo language profile, who is also a linguist. Even before beginning the LPP, I would sometimes talk to him about Oromo data sets for my courses, and I was surprised at first by how much he would thank me for including his language, even though he was the one helping me to check the information and create the data set. I gradually grew to understand how much it meant to him to see his language represented in linguistics courses, especially as someone who had taken linguistics courses himself without seeing much, or perhaps any, representation of his language. As the LPP grows to incorporate a large database of underrepresented languages, it will contribute to changing this experience for more students from underrepresented communities. Despite its status as one of the most widely spoken languages in Africa (Walga 2021), there is comparatively little linguistics literature about Oromo, and many of my students had never even heard of it previously. Another reason that the Oromo consultant has thanked me for representing his language has been simply for exposing students to its existence, which he hopes will lead to increased interest in future linguistic work on it.[5] If new linguists are never exposed to underrepresented

[4] Note that the project still has value even if no students in the classroom come from the communities of the languages discussed, because it is still educating other students about these languages, cultures, and communities and still fulfilling many other goals discussed throughout this article.

[5] As a referee points out, linguist-speakers may or may not be representative of community norms with respect to language or the desire for linguistic research on a language. This is why the project team always aims to consult with multiple members of a community where possible, often with a linguist-speaker connecting with other community members to get their perspectives.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

languages, then these languages are far more likely to remain underrepresented in linguistic contexts. Introducing students to more underrepresented languages while they are undergraduates could open their horizons to future work on these languages, eventually contributing to increasing representation in the field as a whole.

The meaningfulness of including underrepresented communities in classrooms is also emphasized in the academic literature. Ambrose et al. (2010) discuss the motivational and socio-emotional aspects of learning and how helping students feel connected to the content contributes to their academic success. They particularly emphasize using 'multiple and diverse examples' (p. 183) and ensuring that 'course content does not marginalize students' (p. 184). Specifically, multiple examples help students recognize that concepts apply in diverse contexts, increasing the chances of students relating to some of them. They recommend examples that speak to a variety of social groups, including cultures, which is an explicit goal of the LPP, and note that this strategy helps students connect to the content and feel like they belong in the discipline. They further point out that systematically underrepresenting some perspectives in a course can alienate students. The LPP aims to change systematic underrepresentation of certain languages and communities in linguistics courses, further addressing this issue.

3.2. LPP content: data sets

As shown in Figure 1, the data sets section of a language profile consists of an expandable accordion in which the first section lists available data sets and each subsequent section contains one data set. Each data set is tagged with a linguistic subdiscipline and is available online as well as in downloadable format.

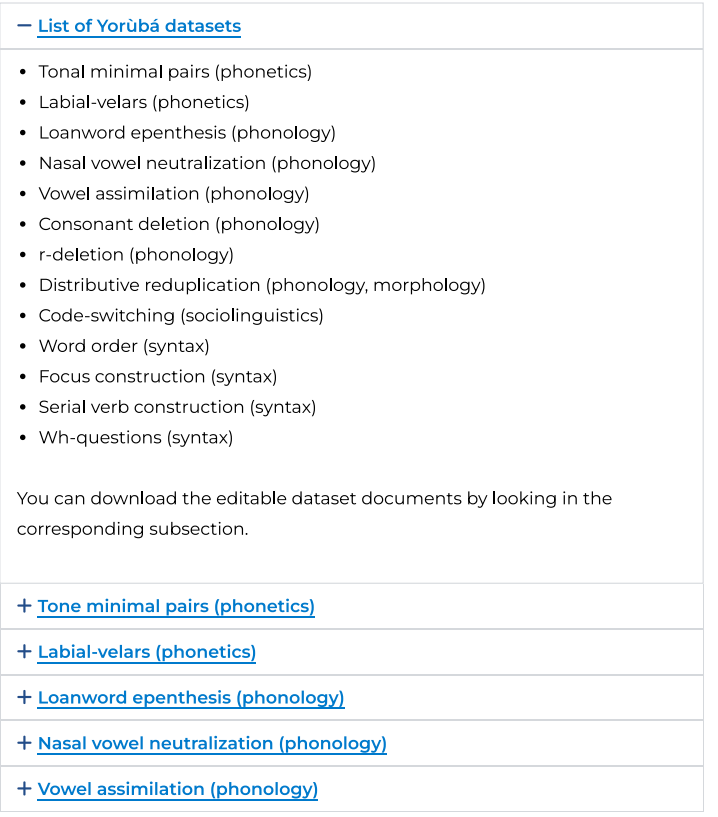


Figure 1. *Yorùbá profile data sets.*

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

Currently, downloadable data sets are available as .docx files, plus .zip files for audio, but we are working on also making all data sets available in .pptx format for easier incorporation into lecture slides.

The presentation of the data depends on the subdiscipline. The language's own orthography is usually included for both p-side (phonetics, phonology) and s-side data sets (syntax, semantics). There is IPA for p-side data sets, but typically not for s-side ones, which are generally given in orthography and, where relevant, in Roman transliteration. Morpheme-by-morpheme glosses are presented for s-side data sets, while p-side data sets are generally given in columns. Choices were based on existing data sets for various courses, my own preferences for courses that I teach, and requests from colleagues for other courses.

An excerpt from the Yiddish umlaut and diminutives data set is shown in Figure 2. The data set provides Yiddish orthography, IPA, and translations; it is also accompanied by recordings. For space reasons, not all of the data is shown. Following the data is a brief explanation of what it shows, together with tags for subdiscipline and topic. Data sets also include the source of the information, which in this case was my own knowledge of Yiddish.

Similarly, Figure 3 shows a syntax data set, on basic Yorùbá word order. The data is in orthography, followed by morpheme-by-morpheme glosses, followed by translation. Again, there are tags to indicate the topic, and the source for this data set was Samuel Akinbo.

3.3. Applications

3.3.1. In undergraduate linguistics courses

The primary intended use of language profiles is to enrich undergraduate linguistics education. Instructors can search the LPP for a data set to include in their class, either by topic or by language or both. They are as theory-neutral as possible, so they can be adapted in any way that suits an instructor's purpose. I have personally used data sets from the LPP in lectures to illustrate concepts, as tutorial exercises, as assignments, as extra practice problems for students, and even in quizzes. Instructors could replace a more commonly used language in their course materials with an underrepresented one with the same pattern (e.g. replace German with Yiddish), or could use the LPP data sets to create new materials or

Yiddish diminutives and umlaut

Noun (orthography)	Noun (IPA, Standard)	Noun (IPA, Poylish)	Diminutive (orthography)	Diminutive (IPA, Standard)	Diminutive (IPA, Poylish)	Translation
קאַץ	kats	kats	קעצל	kɛtsl	kɛstl	Cat
הונט	hʊnt	hɪnt	הינטל	hɪntl	hɪntl	Dog
הויז	hɔjz	hɔz	הײַזל	hajzl	hazl	House
שטאָט	ʃtɔt	ʃtɔt	שטעטל	ʃtɛtl	ʃtɛjtl	Town
ליכט	lɪxt	lɪxt	ליכטל	lɪxtl	lɪxtl	Light
פֿערד	fɛrd	fɛrd	פֿערדל	fɛrdl	fɛrdl	Horse
ביין	bɛjn	bajn	ביינדל	bɛjndl	bajndl	Bone
שוך	ʃʊx	ʃɪx	שיכל	ʃɪxl	ʃɪxl	Shoe
לאַנד	land	land	לענגל	lɛndl	lɛndl	Country
פֿיַער	fajər	fajər	פֿיַערל	fajərl	fajərl	Fire

- Yiddish forms diminutives by adding -l and sometimes changing the vowel
- Back vowels in the base form generally become front in the diminutive
  - ʊ→ɪ, ɔ→ɛ, a→ɛ, ɔj→aj/ɛj in Standard Yiddish
- Whether Standard Yiddish oy becomes ay or ey in the diminutive corresponds to whether the Poylish dialect has o or oy in the base form

Morphology: diminutives
Phonology: umlaut
Historical: mergers, vowel shifts

Figure 2. *Excerpt from Yiddish diminutives and umlaut data set.*

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

Yorùbá word order

(1) a. Adé yá bàtà
Ade borrow shoe
"Ade borrowed a shoe."

b. Adé wọ bàtà
Ade wear shoe
"Ade wore a shoe."

c. Ó wọ bàtà
3SG.SUBJ wear shoe
"(S)he wore a shoe."

d. Ó wọ̀ ọ́
3SG.SUBJ wear 3SG.OBJ
"(S)he wore it."

- Yoruba has SVO word order

Syntax: word order, SVO

Figure 3. *Yorùbá word order sample data set.*

supplement existing ones. We eventually intend to create a section of the website where we can share course activities created using the LPP data sets and other LPP materials.

Upon choosing a data set, instructors can incorporate it into their course based on the specifics of what they need. For example, a given phonology data set could be used in a lower-level class that uses rules (e.g. SPE; Chomsky & Halle 1968) or in an upper-year one that uses Optimality Theory (OT; Prince & Smolensky 1993), and a given phonetics data set could be used in an introductory course that overviews language sounds with minimal pairs or in a higher-level course that analyzes the spectrograms of those sounds. Similar examples apply across other subfields as well. While this underspecification of data set level and theory does mean that instructors must create their own questions, it also means that the data sets are adaptable to a wide range of courses, with suggestions from the data set creators about the relevant courses. Furthermore, the fact that the data sets are already compiled saves instructors a lot of time in finding data that illustrates their desired concepts, as discussed above (Section 3.1.1).

3.3.2. As a field methods exercise

The Oromo language profile was created in part in the undergraduate field methods course I taught in Winter 2023. Some parts of a language profile, such as the pronunciation of the language name, the basic linguistic typology, and some basic phrases, are core components of most field linguistics courses. Many such courses also spend some time on the culture, particularly in the context of culturally sensitive elicitations. Much of a language profile thus already comes together from natural class activities.

In discussions with the Oromo language consultant, I developed two novel projects for this field methods course, both of which also helped contribute to the language profile. One was the 'data set assignment', for which students were required to use the data they elicited to create a data set for another linguistics course. These data sets form the basis of a number of those now incorporated into the Oromo language profile. The other was the 'community assignment', for which students were required to do something linguistic for the benefit of the community. There is a large population of Oromo speakers in the city where my university is located, and the consultant is very involved in the community and in teaching the language to heritage speakers. Many students chose to use traditional stories or descriptions of traditional ceremonies as the text to illustrate linguistic concepts like Oromo word order to heritage speakers. Some of the elicitations for this project, particularly the ones focused on cultural stories and traditions, also helped build the language profile.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

Since the LPP is essentially designed to provide a mini documentation-like experience in a wider range of linguistics courses, creating a language profile through the contributions of field methods students is very natural. It is an excellent way to help field methods students think about how language is viewed in other linguistics courses and where the data in their other courses comes from. The language profile template can also give them a roadmap for learning about culture and context when working with language speakers, although of course, every language and community is unique.

The students in the field methods course were very enthusiastic about using what they learned to create a language profile, with anonymous comments such as:[6]

- 'In addition to the linguistic data, we've gathered a large amount of cultural data as well that would be great to share with others.'
- 'I like to study language in context, so I do like the cultural parts. I also like the idea of a systematic development of language profiles in field methods courses. It's a natural fit.'
- 'I love the idea of building language profiles, also love how this course will bring so many wonderful works to the community and contributions to the faculty in the linguistics department.'

3.3.3. As inspiration for a research project

Finally, language profiles data sets can serve as inspiration for research projects at all levels, from undergraduates to career researchers. These languages are underrepresented, and many of the LPP data sets are of phenomena that have not previously been theoretically analyzed, or in some cases even described. Many of them would therefore make excellent topics for student projects, or even for larger-scale research projects with the goal of publication. The profiles themselves also contain additional linguistic information that can provide further inspiration. As an example, the Pangasinan language profile includes sociolinguistic data sets compiled by a student who grew up speaking the language.[7] He has intuitions about the variation being age-related, but there are no sociolinguistic studies of Pangasinan to cite as confirmation of these intuitions. The data he compiled for the language profile can form the basis of a future full-scale sociolinguistic study on Pangasinan. Similarly, many of the Oromo data sets were also used by students for their final projects in the previously mentioned field methods course, and several of these students are currently pursuing these papers further for eventual publication. Unlike the Pangasinan example, many of these Oromo patterns have been at least briefly described in existing literature, but few of them have been explained in depth or analyzed in the context of their theoretical implications. As a final example, when I spoke with a colleague who teaches historical linguistics about whether the course could make use of the language profile information about Slavic influences on Yiddish, her comment was that it would make 'a wonderful prompt for a research paper/project'. Overall, then, the data sets compiled as part of the LPP can also be used as a resource to inspire research projects on underrepresented languages, both for undergraduate course projects and for larger-scale publishable research.

4. Contextualization

Moving beyond data sets, a second core component of the LPP is contextualization for each language. Rather than simply present data sets with little to no information about the language, the LPP includes recordings, images, and other content to help bring the language and community to life for students (and instructors) who may not have previously been familiar with it. In this section, I discuss the motivations for including this contextual information, the content of the LPP with respect to contextualization, and then the practical applications of it.

[6] Student comments were collected anonymously, and Research Ethics Board review is not required for secondary use of anonymous information, under TCPS-2, Article 2.4 (https://ethics.gc.ca/eng/tcps2-eptc2_2022_chapter2-chapitre2.html#a).

[7] Following Cheng et al. (2021), I avoid the use of the term 'native speaker'. See also other work by the ROLE Collective.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

4.1. Background and motivations

4.1.1. Linguistic extraction and colonialism

In the core linguistics courses at institutions where I have been a student and instructor, including previously in my own classes, languages have traditionally been treated as wholly disconnected from the people who speak them. Of course, this generalization is not universal, as this connection is fundamental to certain types of courses, like field methods or sociolinguistics. However, overall, in most of the undergraduate classes that I took as a student, was a teaching assistant for as a graduate student, taught in my first few years as a professor, and have heard about from other students and colleagues, students are exposed to languages simply as problems to solve. This view is also promoted by textbooks commonly used in undergraduate linguistics classes. As an example, consider the classic phonology textbook *Introducing phonology* (Odden 2005). This book has many amazing data sets on a wide variety of languages, but there is little to no information about the languages themselves. In the exercises at the end of every chapter, data sets are titled with the language name and then launch immediately into the phonology questions, without any further information about the language. Students going through this textbook would need to do their own research to find out even basic information like where the language is spoken or how to pronounce the language name. This format is the most common one I have seen at my institution. Furthermore, unlike the related field of linguistic anthropology, which focuses on the relationship of language to communities, linguistics has traditionally been a field where this type of detachment has been actively advocated for (see e.g. Hymes 1963 on linguistics versus linguistic anthropology), with the idea that it makes the field more 'scientific' (e.g. de Saussure 1986 [1916]; see also Nakassis 2016). Similarly, Leonard (2021b) notes how publications in linguistics often do not mention the associated language communities, while also discussing how this contextualization has become commonplace in linguistic anthropology. Various strategies have been employed to integrate more context into linguistics courses. For example, Amberg and Vause (2018) discuss integrating it into introductory courses by having students follow a language through the course with some cultural information. While the LPP takes an alternative approach in which students are not doing fieldwork themselves, both strategies are aimed at including more than just data, making students more engaged, giving them a broader perspective on a language, and helping them apply concepts to underrepresented language(s) throughout a course as a way of integrating knowledge and scaffolding.

Given the existing context, undergraduate students in linguistics can easily start thinking that linguistics data is just words on paper or problems to solve, and they can forget that it in fact comes from real languages spoken by real people. Such a perspective is problematic in the broader context of the values that linguistics instructors instill in students. As noted by Errington (2001, 2008), linguistics was historically used as a tool for colonialism; language documentation was an extractive process of colonizers taking language from communities. With recent broader global movements of decolonization and the movement within linguistics toward ethical, community-engaged, and ethnographically informed linguistics fieldwork, the fact that undergraduate courses still typically treat language data as independent from the people, cultures, and communities behind the languages is inherently troublesome and behind the times, for the reasons to be elaborated on throughout this section. Most prominently, as discussed below, doing so ties into the historical colonialism that characterized the earliest linguistic documentation and can negatively affect student learning by causing a feeling of nonbelonging. All nonfabricated data sets were ultimately, at some point, collected from language users, but by the time they reach undergraduate students, that context has often been removed. Bjorndahl et al. (2024) address this issue directly, calling for measures such as adding context to course materials and data sets, introducing information like preferred community names, not treating language as 'disembodied data', and introducing students to the communities from which data were obtained.

Davis (2017:37) defines linguistic extraction as 'the process of discussing languages and language reclamation movements removed from the personal lives, communicative practices, and embodied experiences in which they inherently are embedded'. Linguistic extraction is the treatment of languages as objects, without the human element and the historical, social, and personal contexts. Davis, along with

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

other authors such as Whaley (2011), argues that treating languages this way is a form of colonial violence and that it suggests that linguists value linguistic data over the communities, speakers, and contexts from which the language comes. Even in discussing simple information like numbers of speakers, Davis suggests discussing the historical and current contexts, because issues like language endangerment do not occur outside of sociopolitical causes; talking about languages removed from this context minimizes the root causes of intergenerational language disruption. Davis proposes that the way to move past colonial linguistic practices and linguistic extraction and erasure is to discuss languages within their social, historical, and current contexts. The LPP aims to provide such context, helping instructors to avoid extractive linguistic discussions in classrooms.

Similarly, Leonard (2018) discusses how practices in linguistics can reinforce colonial norms. He argues for a model of language documentation that is embedded in the views and voices of the community that uses the language. Like Davis, he proposes a decolonial approach that looks not only at the current status about facts like number of speakers, but also at the historical context that led to it. He also notes how viewing languages as the object of scientific inquiry tends to be based in colonial norms, where language is analyzed in a way that is disconnected from the people who use it. This scholarly tradition contrasts with Indigenous ways of describing the world, which place greater value on interrelatedness. He specifically discusses linguists who present grammatical issues as 'a puzzle to be solved' (Leonard 2018:58), using 'language data isolated from cultural context' (ibid.); such an approach fails to acknowledge the people behind the language and the importance of the language to them. The result is linguistic materials that are not aligned with the needs of communities and that perpetuate colonial practices of exploitation of people for resources. Based on his own experiences in Indigenous communities, Leonard says that language users perceive a distinction between language and culture as strange; the communities typically define language with reference to culture. He suggests that linguistics courses, including introductory ones, should include these Indigenous ideas about the interrelatedness of language and culture, and that linguistics training should be complemented by ethnography training. The LPP represents a step in this direction, of rooting the view of the language presented to linguistics students in community understandings of what the language means. In later work, Leonard (2021b) also discusses the need to redefine academic fields with Indigenous views of knowledge incorporated, focusing on avoiding colonial norms that have brought harm to Indigenous communities. Talking about languages as detached from their communities is an example of this type of colonial way of knowing.

While Davis 2017 and Leonard 2018 are focused on language reclamation contexts, their points extend to linguistic classrooms, where languages typically continue to be discussed in the extractive way they describe, as has also been discussed in Leonard 2020, among others. The LPP aims to directly address this issue by providing crucial cultural and historical context for languages alongside the traditional linguistic data sets, thereby acknowledging the connection between language and community in the way these authors argue is necessary. Training students in this viewpoint from the earliest undergraduate linguistics courses can help to create a new generation of linguists who respect the language–culture connection and focus on nonextractive linguistic methods.

Holden (2018) more explicitly discusses these issues in the context of linguistic teaching. Holden describes using Indigenous frameworks for teaching introductory linguistics, specifically arguing that doing so does not mean rejecting the conceptualization of linguistics as a science. Instead, it is just broadening the concept of language and ensuring that linguistics represents authentic knowledge. He suggests a paradigm in which students are given information about culture and communities and in which there is a relationship between linguistics and language pedagogy, such that linguistics is relevant to communities. Holden also draws directly from community knowledge in his approach to teaching linguistics, seeking and validating community perspectives. This same framework is developed through the LPP, with a focus on those kinds of introductory linguistics courses.

Similarly, Tsikewa (2021) also focuses on these issues in linguistics classrooms, specifically field methods classes. Starting from the history of linguistics in colonial enterprises (e.g. Errington 2008, Hale 1972), Tsikewa discusses the power relationships and privilege inherent in linguistics historically. She then moves toward an argument that the linguist is only one of the experts, and that community members

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

are co-directors, with their own expertise. Similar ideas outside the context of the classroom have been discussed by Czaykowska-Higgins (2009). Tsikewa suggests a field methods paradigm in which historical and cultural texts of the area and people are read critically by students, and where outside researchers do a 'cultural boot-camp' to better understand the community. She places particular emphasis on resources created by community members. The LPP focuses on similar ideas but with greater breadth and less depth, in order to reach a broader range of undergraduate classes.

All of these works, among other recent discussions of decolonization in linguistics (e.g. Charity Hudley et al. 2024), suggest a need for linguistics to better emphasize Indigenous and decolonial research methods, in which a less extractive, more contextualized approach is adopted. This idea connects to what Czaykowska-Higgins (2018) discusses as 're-humanizing linguistics'; she advocates for thinking about what we mean by linguistic knowledge and by 'language', bringing together different views and knowledge sources and avoiding adopting the Western notion of 'language' uncritically. This is the viewpoint adopted by the LPP, in the specific context of how linguistics is being taught in North American classrooms. Adopting non-Eurocentric models of language in classrooms, including both in the languages discussed (i.e. diversity of data sets) and in the way in which these languages are discussed (i.e. with contextualization), ties into Culturally Responsive Pedagogy (CRP), as discussed, for instance, by Ragoonaden and Mueller (2017), who suggest that validating cultures and incorporating a holistic perspective into the classroom can nurture student well-being and support positive educational outcomes.

Relatedly, with specific discussion of endangered languages, Hill (2002) discusses the problematic ways in which linguists have talked about endangered languages, which can alienate these languages from their communities and treat them more like objects for preservation by elite academics than as connected with everyday life. The LPP idea of contextualization aims to show students the vibrance and vitality of the cultures and communities behind each language, with the view that illustrating language in context from the initial training in linguistics can help prevent a new generation from thinking of languages as disconnected from the communities that use them.

The LPP is inspired in part by the level of engagement with languages that students who have participated in documentation-based projects with me have shown. In general, students who do research-based courses often get a much deeper, more nuanced view of language that they are not afforded in other linguistics classes. Hema Ramnarine, an upper-year student in a Research Opportunity Program (ROP) course, wrote in her final report:

> In researching these languages, I've discovered that a great deal of them … have become extinct, endangered, or obsolescent. In my view, this is incredibly damaging, as this ROP has allowed me to realize that in a very real sense, a language is not just a form of communicating, it represents the history and the modes of thinking of its speakers. It is a representation of their lifestyle and values, and it has reinforced for me the importance of creating corpora for these languages.[8]

Through this exposure to speakers and to cultural aspects of languages and communities, this student developed a new perspective on the connection between language and community, as well as on the importance of addressing issues of language endangerment and revitalization. This sort of shift in how students think about language is precisely the motivation behind the LPP. The connection between languages and the people who speak them is a core aspect of language, so this should be represented in as many courses as possible, to reach as many students as possible, not just the few who have the opportunity to do documentation-based research courses.

Clearly, it would be impossible to introduce documentation for every language discussed in every linguistics course. However, language profiles allow us to bring a miniature form of some documentation-like elements into general linguistics classes. While a language profile involves no direct contact with speakers, it includes recordings of the language, cultural information, and multimedia representations of

[8] Quote used as personal communication with permission from the student, Hema Ramnarine. Personal communication does not require REB approval.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

the community that students traditionally experience only in documentation scenarios. These components can engage and interest students, to change the embedded mindset of language as a disconnected entity and to show students the richness of the communities and cultures behind the data sets they see. The historical and current language context related to marginalization can additionally help enrich student perspectives on core issues with which all ethically minded linguists should be concerned.

4.1.2. Interdisciplinarity

Turning to ethnography, Harrison (2005) argues that ethnographic methods, particularly inclusion of cultural content, are critical to language documentation. The ethnographic approach involves a focus on culturally embedded language. Harrison notes that many endangered languages are Indigenous languages, with strong ties to specific areas, and that the full richness of these languages cannot be properly understood when they are separated from their environment. He argues for a holistic approach to language that pays attention to the interaction between speech and cultural knowledge. He provides the example of the verb 'go' in Tuvan, which is influenced by local topography and river directions; without that information, any linguistic description of the verb would be incomplete. Overall, he suggests that cultural contexts and knowledge systems should be viewed as part of language and as relevant to the core linguistic subfields like phonology, morphology, and syntax with which linguists are traditionally concerned. Relatedly, Nagy (2000:147) discusses the connectedness of language with other disciplines, noting that 'truly adequate preparation for fieldwork would include work in psychology, sociology, and anthropology'. Similar points have been made, for instance, by Shulist and Rice (2019).

Historical context can also help to situate linguistic facts. For example, Kahn (2017) notes that the stress rules of Yiddish differ depending on the language origin of individual words (Germanic versus Hebrew versus Slavic), and that understanding the historical development of Yiddish and the factors behind its extensive language contact can help students to better understand this phonological pattern. Thus, there are clear linguistic arguments for discussing geographic, historical, and cultural context in order to better understand linguistics itself.

The LPP emphasizes all of these ideas to students early in their linguistic journeys. By including substantial contextual information, the project emphasizes that language is more than just words, and that it is intricately related to culture. Moreover, as addressed by much of the literature, power is a highly relevant construct in how languages are described and discussed, and the LPP brings out these historical and current contexts. By showcasing marginalized languages, it emphasizes to students and community members that those languages have a place in linguistics classrooms, while also making available a wider range of data sets from which we can learn what is possible in human language. The project exposes students to the interconnectedness of linguistics with other fields and with communities in a way that is often absent in current courses, especially courses in 'core' linguistic areas like phonology and syntax.

Beyond linguistics, Chamany et al. (2008) discuss the importance of social context in biology classrooms. They note that students copy the behavior of their instructors, and that integrating social context into the curriculum demonstrates social responsibility for students. The same reasoning is applicable to linguistics. Even if many linguistics instructors know better than to think of languages as problems to solve, treating languages in classrooms as disconnected from the communities that use them will model that mindset for students. Instead, integrating context into linguistics will reinforce for students the social responsibilities of linguists toward the languages we discuss.[9] Additionally, this sort of interdisciplinary approach that is promoted by the LPP, integrating linguistics with contextual information that intersects with other disciplines like history, anthropology, sociology, music, art, literature, and so on, has been proven to positively affect student learning outcomes. As discussed by Lattuca et al. (2004), for instance, interdisciplinary learning has been referred to as more engaging, helping to capture students' interest, to develop their higher-order cognitive abilities, to improve their recall, and to promote self-

[9] Note that the LPP does not currently contain a guide for leveraging the context for social responsibility, though we would eventually like to add something to that effect.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

motivation in learning. Similarly, Ambrose et al. (2010) discuss seeing students holistically, and they argue that doing so can shape the climate of courses and therefore also student learning. The LPP does so in both a broad sense, providing a much more holistic view of language than is typical of linguistics classrooms, but also in the more specific sense of diversifying examples and avoiding marginalization of certain perspectives.

4.1.3. Previous approaches to contextualization in linguistics classrooms

One of the few linguistics pedagogical papers arguing for an approach like the LPP is Zuraw 2022. Zuraw discusses how she goes beyond basic information when talking about languages in her classes, including language-dependent content such as the complexities of language shift, why a current name is used, and images related to the language and its users, including notable speakers, events, writing systems, art works, book covers, flags, and so on. Zuraw notes that it takes only a few minutes during class to incorporate this kind of information, though it takes her a great deal of time to compile it.

The LPP builds on this work by Zuraw (2022), with similar types of information used to contextualize the languages, and aims to make it easier for other linguistics instructors to follow her lead, since many instructors may not be willing or able to put in the time commitment needed to gather this information. As discussed above, the goal of the LPP as a publicly available resource is for instructors to have access to this sort of contextualizing information without needing to compile all of it themselves. Zuraw also notes how images created by language users are more useful because they avoid the tourist perspective. The LPP builds on this goal by involving community members in the entire material-creation process, including recordings, images, and a larger selection of information. The involvement of community members, discussed further in Section 5, helps to ensure that the material selected to represent a community is accurate and appropriate, and connects to recent discussion of the harm that can be done to language communities when language materials are created and curated by non-community members without engaging with problematic history in the field (e.g. Rice 2022). While the LPP builds on Zuraw's points in these ways, it takes a more bottom-up approach to language choice: Zuraw talks about compiling this information for languages she already discusses in her classes, while the LPP develops data sets from scratch, building from the choice of language rather than from existing data sets (see Section 3.2). Further, as noted above, the LPP does not currently have the resources to compile this information for majority languages, while Zuraw describes using this approach for all languages discussed in her courses. Overall, the LPP differs from Zuraw's (2022) approach, but builds upon the linguistics pedagogy framework she proposed, in order to make going beyond basic language information more accessible to more instructors and for more languages, as well as to include communities in the choices of how linguistics classes represent them.

4.2. LPP content: contextualization

The contextual information for each language in the LPP includes (but is not limited to): transcriptions and recordings of the language name pronunciation, maps of where the language is (or was) spoken, numbers of language users (including learners, where applicable), historical information, sample texts in the language, basic greetings and fun phrases, lists of well-known community members, and cultural information about things like food and music. Pictures, recordings, and links to videos are provided where possible.

The two components, data sets and contextual information, are also integrated where possible. For example, some data sets use culturally relevant words and phrases, and some include the language's orthography in addition to the IPA. These choices also make it easier for the communities to make use of the language profiles themselves (see Section 5). Upon clicking a language name, either from the homepage or from one of the search pages, the user is taken to a page that looks like Figure 4, which is the landing page for the Yorùbá profile. The buttons 'gallery', 'overview', and 'data sets' all link to the relevant sections of the profile.

Each profile begins with a gallery that intersperses language information, contextual information, and cultural information. The galleries cycle through eight to ten images, with brief captions and links to the

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

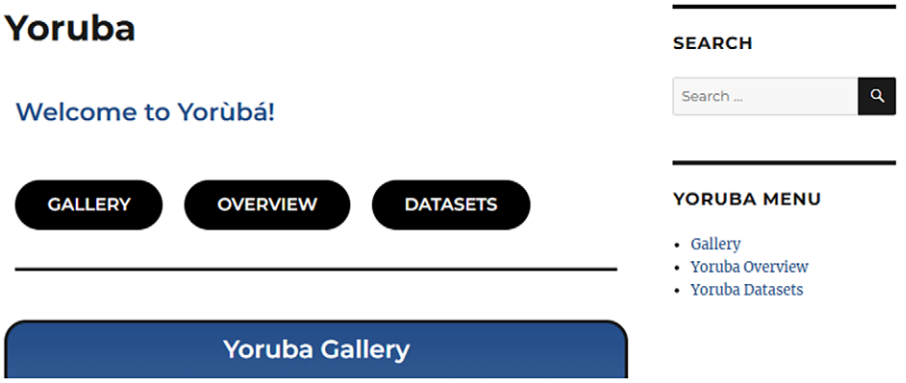


Figure 4. *Yorùbá profile landing page.*

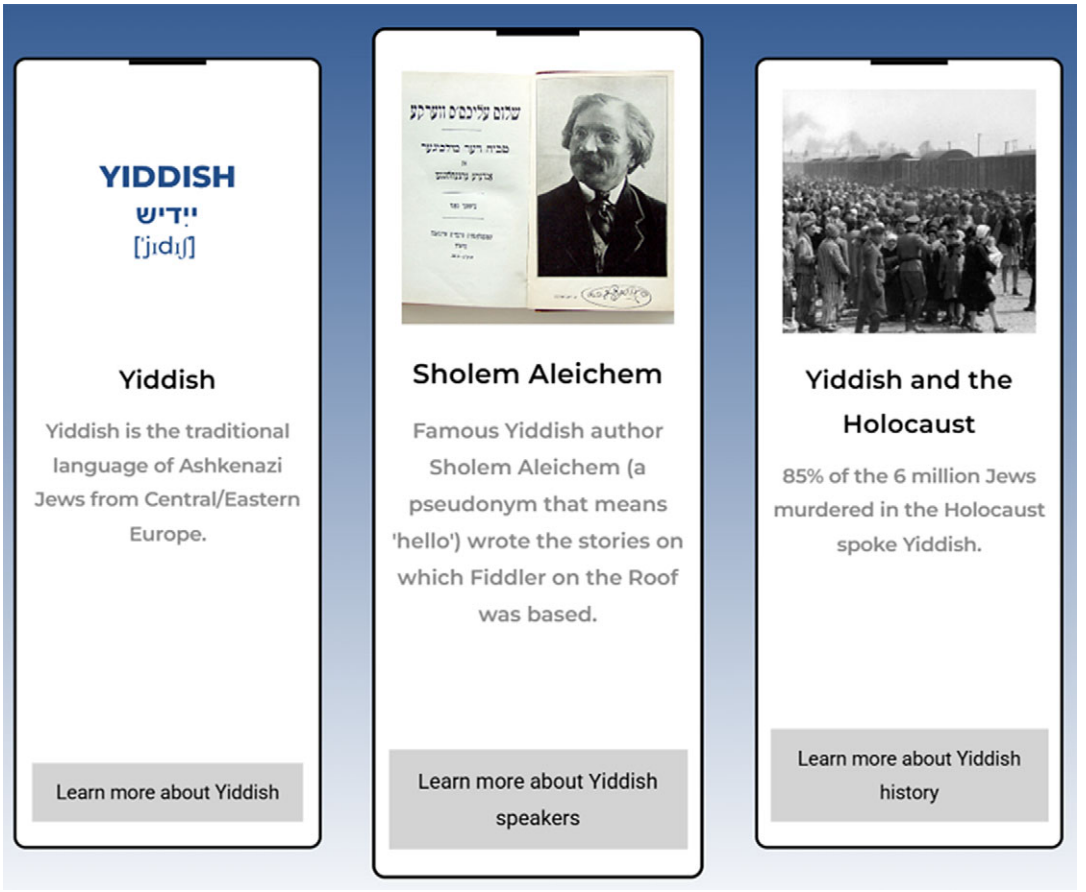


Figure 5. *One view of the Yiddish profile gallery.*

relevant sections of the profile. One view of the Yiddish gallery is shown in Figure 5. All images are public domain, under a creative commons license, or used with permission of the copyright holders, and credits are provided immediately below the gallery. Images are also all chosen in consultation with members of the community.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

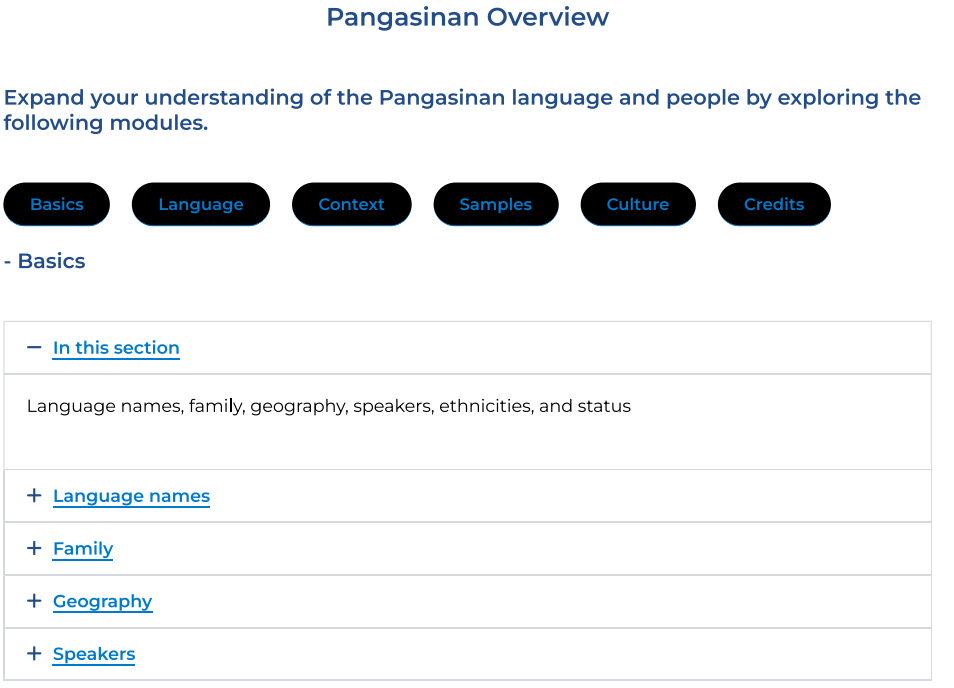


Figure 6. *Pangasinan profile overview.*

The overview section is divided into six subsections: basics, language, context, samples, culture, and credits. Buttons at the top of the overview section link to expandable accordions for each subsection. 'Basics' includes an overview of regions, numbers of speakers, and pronunciation of the language name. 'Language' includes basic information about linguistic typology, as well as about the writing system if there is one. 'Context' contains key points about the historical and current situation for the language, including about marginalization and language shift. The 'samples' section has words, phrases, and short texts from the language. 'Culture' includes information about food, music, festivals, literature, and so on. The 'credits' section shows the references and image credits for everything in the profile, along with information about who contributed to creating it. The top of the overview section for Pangasinan is shown in Figure 6.

In addition to the information on the website, there are downloadable sample slides, with a subset of the language information. These are designed such that they could be included as-is in lecture or tutorial slides. They typically include language name pronunciation (with recording), a special phrase from the language (with recording), a map of the language area, information about the language family and number of speakers, a picture and brief description of a cultural symbol, a picture and a brief description related to an important historical fact, and a word cloud of relevant linguistic terms. A sample slide for Oromo is provided in Figure 7.

4.3. Applications

4.3.1. In undergraduate linguistics classes

Like the data sets, the contextual information is primarily designed to be used in undergraduate linguistics courses. Instructors can present information from a language's profile along with a data set. For example, instead of pronouncing the language name themselves, instructors have the option to play the recording from the language profile. They can also use the profile to show students a map of where the language is spoken, share samples of text and recordings, show culturally relevant pictures, discuss the status and history of the language, and much more.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

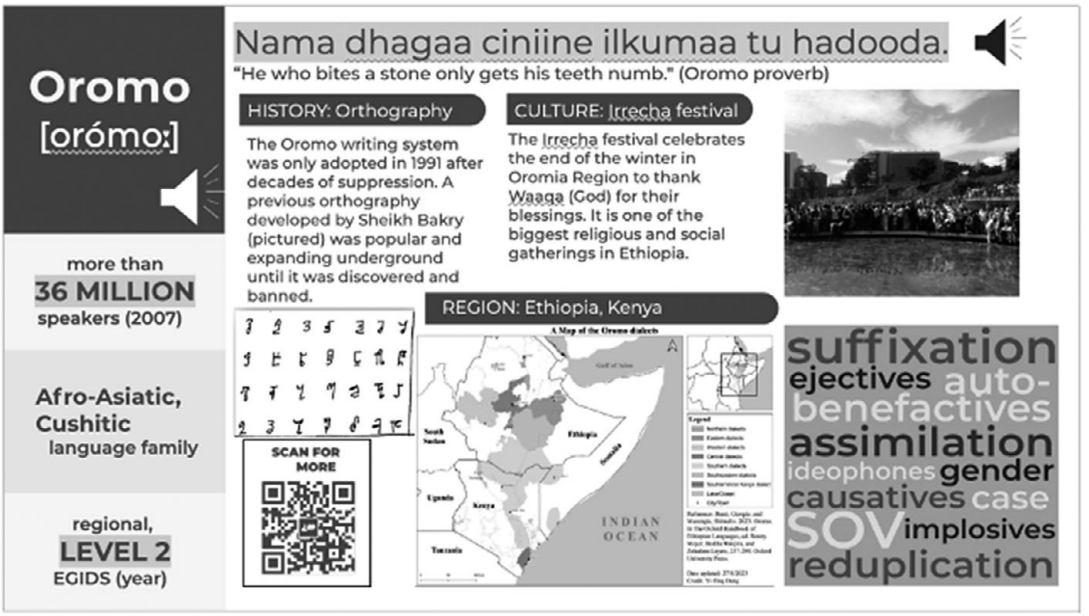


Figure 7. *Oromo sample slide.*

Language profiles are very rich, and in most linguistics courses, it would be impossible for an instructor to have time to present everything. The intention is for instructors to present a subset and then to direct interested students to the LPP website for more information. The top of each language profile webpage shows a sampling of the information included in the profile, and this can be directly included in course materials, making it easy for instructors to include some of the contextual information in their courses without too much work. For instructors who wish to explore further and decide themselves what to include, the online gallery links to other sections of the profile with further information, so that it can also be used as a jumping off point.

I tested a preliminary version of the Yorùbá language profile in a second-year undergraduate phonology class. Students were given the loanwords data set as an assignment, and when we discussed the assignment as a class afterward, I shared the preliminary language profile with them. Among other things, the students got to hear a Yorùbá speaker pronounce the language name, learn a bit about other linguistic properties of Yorùbá, hear a recording of a Yorùbá proverb and see it written in Yorùbá orthography, and watch a short video clip of a child playing the Yorùbá talking drum. We spent only two to three minutes looking at the profile together, which in the scheme of the class is not very much time, but it was enough for students to get a taste of Yorùbá language and culture far beyond what they usually see in a North American linguistics course. After this class exercise, many students told me that the language profile helped them to understand languages as more than just data and that it increased their interest in understudied languages and their investment in issues like language equity.[10] For instance, one student told me: 'I think it's important to treat these languages as more than just words on paper and recognize that real people speak/spoke them. It helps bring them to life'.[11] The test version therefore fulfilled the goal of instilling an understanding of language as connected to people, cultures, and communities.

[10] While Yorùbá has millions of speakers and is reasonably well studied compared to other African languages, it remains an understudied language compared to European languages. As an example, while Gordon (2005) indicates that a third of the world's languages are spoken in Africa, Kochetov (2020) finds that only 6% of published articulatory research is on African languages, and Ozburn et al. (2023) note that there are only a handful of papers in the past few decades that explore perception of African languages. Thus, African languages in general, including Yorùbá, count as understudied from at least some perspectives, despite the large number of speakers.

[11] Comment from an anonymous survey.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

The intended impact of language profiles on undergraduate education is to make this experience possible on a much larger scale, with more languages and in more courses. Used in the way described here, language profiles can fill a critical gap in fulfilling the goals described in Section 4.1, thereby greatly enriching undergraduate linguistics education. As noted in Section 2.4, not all linguists are supportive of contextualization in linguistics, but the target audience of this aspect of the LPP is those who, for reasons outlined in Section 4.1 or their own reasons, would like to incorporate context for the languages they discuss, but have lacked the resources or time to compile the information.

4.3.2. As an outcome to linguistic fieldwork

Just as a language profile is a natural fit with a field methods course (see Section 3.3.2), it is similarly a natural outcome for higher-level linguistic fieldwork. Many of the languages in the LPP, both those currently available and those coming soon, are ones on which LPP team members have done fieldwork research. Much of the information that goes into a language profile is naturally learned by researchers over the course of linguistic fieldwork. However, often only the linguistic information is shared more broadly in the knowledge-dissemination component of the research. Additionally, at least in my experience with seeing the course materials of other linguists, many fieldworkers already create data sets for their undergraduate courses based on the languages they work on. However, few resources are available for sharing those data sets more widely with the linguistics community. Instead, most publications resulting from fieldwork are research-oriented.

The LPP essentially provides fieldworkers with a way to publicly share more from their fieldwork experiences, including what they have learned about the culture and broader context of the language and any undergraduate data sets they create. It is a natural fit as an outcome to linguistic fieldwork, in the same way as discussed for a field methods course in Section 3.3.2. Discussions of how linguists can work in ways that are useful for language communities often focus on very direct ways, such as writing a grammar for nonlinguists. Language profiles are another way in which linguists can use their fieldwork to create something useful for the community, in both indirect and direct ways. I elaborate on this point in the next section, which focuses on the community collaboration aspect of the LPP.

5. Community collaboration

The final major theme of the LPP is community collaboration. While it is often unclear where data sets in linguistics classes come from and to what extent the community was involved, the LPP puts a strong focus on bringing community-engaged research into the classroom. In this section, I discuss the motivations for community collaboration, the content of the LPP with respect to collaboration, and then practical applications.

5.1. Background and motivations

Recent literature on fieldwork methods is heavily focused on the responsibilities of linguists toward communities. For example, Florey et al. (2009) argue that language activism is fundamental to modern linguistics and that it is the best way to preserve linguistic diversity; this theme is also taken up by work such as Woods 2023. Similarly, Clyne (2007:§§03.1–03.2) argues that linguists 'have a social responsibility to extend the linguistic horizons of the wider community', because language is fundamental to individual and social well-being and to all fields. Similarly, Nagy (2000:147) discusses how including language speakers in linguistic fieldwork results in 'better understanding of the cultural context in which the language is embedded'. Leonard (2018) likewise proposes that a decolonial approach to language documentation considers community needs and involves active engagement with the community at all stages.

Rice 2006 and 2011 provide excellent overviews of these positions. Within the activist framework, Rice 2006 focuses on the responsibilities of linguists toward language revitalization, as well as toward the

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

communities and knowledge systems in which languages are embedded. Rice argues that language cannot be viewed in isolation, and she discusses the shift toward empowerment research, namely, research that is not only on, but also for and with language communities. This approach contrasts with the objectification of linguistics research in the past, which was characterized by Western ideologies of contempt for other intellectual heritages (e.g. Errington 2001, 2008; see also discussion in Section 4.1.1). It is increasingly recognized that language belongs to the community, and that fieldworkers need to know about local culture. Empowerment research brings in the goals of the communities related to their languages; linguists are responsible not only for their personal goals, but also for doing work to benefit the community and for respecting their intellectual traditions. Rice (2006) argues that this model ultimately leads to better understanding of languages.

Rice 2011 looks at the relationship between community-based work and traditional linguistic research. Rice notes that there is now an extensive literature about researchers' responsibility to communities (see e.g. D'Arcy & Bender 2023 for a summary), which has become increasingly recognized in recent years. According to the view she adopts, when possible, linguistic research should be relevant to the community, involve community members, and be oriented toward promoting positive changes. Such collaboration between linguists and communities often results in greater appreciation of the link between language and culture. She notes that in recent years, there has been a shift from looking only at the structure of language to looking more broadly at language ecology: linguists have become more interested in the speakers behind languages. That said, linguists and speakers often have different ideas about basic issues like what is important to know about a language. Rice (2011) argues that community-based research serves both society and the discipline by providing greater insight into language and culture. She also talks about 'giving back' as part of community-based research: colonizers have taken away from communities, and linguists are sometimes seen as trying to take away a language through Eurocentric research paradigms.

The notion of 'giving back' and of the 'empowerment' of communities can be viewed as problematic due to the perpetuation of uneven power dynamics, with Leonard (2017) arguing that this notion itself maintains some of the issues of colonial linguistic practices. The term 'accompaniment' (Bucholtz et al. 2017) for community-engaged research practices can be used instead, and the goal of these models is that linguistic research should be done in partnership with communities such that it is relevant to the communities. As also noted by Battiste and Henderson (2000), communities should have control over information related to their heritage; otherwise we continue to perpetuate colonial paradigms of making decisions for marginalized groups. Instead, the decolonial paradigm situates linguistic documentation in what Bucholtz (2021) calls the 'contact zone', where anyone can be a collaborator.

The LPP is intricately connected to this framework and these issues. In terms of language activism, the LPP is a way to raise awareness among linguists about marginalized languages and the reasons for their marginalization, to increase representation of these languages in linguistics course data sets, and to connect with communities to learn how to support their languages and cultures. The LPP provides linguists-in-training with some critical insights early in their studies, by (i) showing how language is embedded in anthropological, sociological, historical, and cultural context from which it cannot be separated, (ii) emphasizing the importance of including community members in decisions about how their languages are represented, (iii) modeling sensitivity toward speakers and communities, instead of the data-only focus usually adopted in linguistics courses, and (iv) illustrating ways in which linguistic materials can be made useful for communities. In other words, the LPP sets an example for students in these crucial aspects of community-based, accompaniment linguistics.

5.1.1. Intersection between 'community member' and 'academic'

As noted, the LPP is in the 'contact zone' between academia and communities, building a place where these traditionally separate domains are treated as tightly interconnected and overlapping. As previously noted, many of the language profiles also have primary authors who are from the communities themselves, and showcasing their work in linguistics courses can contribute to a sense of belonging for students who may themselves be from marginalized communities. Supporting the work of speaker-

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

linguists and showing students that one can be both a community member and a linguist is a critical element of the LPP.

In this vein, Gaudry and Lorenz (2018) discuss various models of Indigenization of the academy, arguing that Canadian academics thus far has focused primarily on inclusion only, rather than reconciliation or decolonization. The LPP is situated in the reconciliation framework, aiming to find a common ground between Western and Indigenous (and other non-Western) ideals, changing the academic definition of language, and specifically reintegrating the perspectives of the communities that use each language. Since each language profile is directed by community members of that language, the LPP creates a community-led process for reintegrating other worldviews and perspectives into linguistics classrooms. The importance of attribution of material to community members has also been discussed in relation to the decolonization of archiving by Anderson and Christen (2019), further emphasizing the recognized need across disciplines to ensure recognition for community members for their data.

5.1.2. Outcomes for language communities

Finally, an explicit goal of the LPP is to produce linguistic outcomes that are also of use to language communities. Any time we use a data set in a linguistics course, that data must ultimately have come from language users. That said, it is easy to lose sight of this fact in the context of many linguistics classes. The philosophy of being mindful of where language data comes from, as well as the practice of doing research in a way tied to the desired outcomes of communities that share their languages with linguists, has not yet made it into most undergraduate linguistics classrooms. The LPP assists with this mindfulness through its many data sets curated by community members, recordings of data from the community, and information about the culture, history, and community that help instructors and students to connect the data to those aspects of the language.

In cases of data from majority languages like English or French, it may not be as critical to think about these issues as it is for the marginalized and underrepresented languages included in the LPP. Linguistics instructors who benefit from having data sets on these languages to illustrate points in our classes, as well as to increase diversity and inclusion, have responsibilities to these communities, in the same way that researchers do (as discussed in Section 5.1.1). The minimum responsibility should be to raise awareness about these languages among students, and to be aware about them personally as well. Beyond basic information like where the language is spoken, the language profiles also all contain critical historical and other contextual information related to the status of the language, such as discussion of language policy in the Philippines for Pangasinan, of the Holocaust for Yiddish, and of colonization for Yorùbá.

The LPP is also working toward going beyond this minimum to more concretely produce collaborative outcomes aligned with the goals of language communities. First, members of the community are involved in deciding whether and how they would like to see their language represented in the LPP. Second, there is a natural relationship between linguistics data sets and language-learning materials, even though this relationship is rarely pursued in practice, at least by theoretical linguists. Many linguistic data sets are on topics that a language learner will need to learn, such as which sounds are contrasted in the language or what order its words go in. Of course, the analytical techniques and terminology that students learn in linguistics may not be relevant to language learners, but the language examples that linguistics courses use very often are. Minor tweaks like including the language orthography and including explanations without specialist language are straightforward ways to transform linguistics data sets into language-teaching materials, but this issue is rarely considered in linguistics teaching. Where the language community expresses interest in doing so, the LPP aims to change this paradigm, viewing data sets as dual-purpose materials for both linguistics and language teaching.

The importance of such community outcomes has been emphasized by, for instance, Hermes et al. (2012) and Perley (2012). Hermes et al. (2012) discuss the tensions in priorities between academia and materials that are usable for learners, where more time prioritizing academic linguistics can make documentation less accessible for the community and result in less time for practical resources. They argue for acknowledging this issue, finding opportunities for collaboration, and taking a perspective that emphasizes shorter documentation in a user-friendly format that can be distributed immediately.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

Similarly, Perley (2012) discusses how linguistic documentation should not come at the expense of opportunities for the community to learn the language, and advocates for a model in which linguists are more interested in the people and include communicative practice as part of a collaborative method of documentation. The LPP is aiming to ensure that the materials are useful not only for linguists, but also for these types of community needs, when desired by the community. In this vein, the LPP hopes to further explore the connections between linguistics, language reclamation, and second language acquisition, as discussed for instance by McIvor (2020).

Overall, creating linguistic outcomes in support of the goals of language communities is a fundamental goal of the LPP, one that is important in current philosophies about use of linguistic data, yet not traditionally considered in undergraduate teaching (although there are increasing movements to change this tradition and consider such goals). More details on ways in which language profiles can be used in communities are discussed in Section 5.3.

5.2. LPP content

As previously noted, all of the content of the LPP, described in Section 3 and Section 4, has been built in collaboration with communities, so the collaborative theme of the LPP runs throughout the data sets and contextual information. In the data sets, the community component can be seen through choice of examples and topics, inclusion of language orthography and recordings instead of only IPA, and descriptions of phenomena that are useful to nonlinguists. In the contextual information, the final choice about what constitutes representative information and images related to the culture and history is left to community members.

As with the plans to create a place in the website for activities developed with the LPP in the future, we also plan to deepen the community content, with specialized content based on the requests and goals of each community. This could include content specially designed for language learners, such as language lessons, storybooks, minimal-pair recordings, and other such content. Given the current funding structure as described in Section 2.1, this has not yet been possible to pursue, but it remains a goal to develop the LPP materials further in this direction, just as they will be developed further for linguistic classroom purposes.

5.3. Applications

Since the LPP is embedded in community-engaged research, though a linguistics pedagogy project at its core, it is perhaps most crucial that the language profiles can be used in language courses and other community contexts. The ways in which a profile can be used in the community depend greatly on the context of each individual language, but in this section, I outline some possible ways that language profiles can benefit communities rather than just linguists.

I discuss Yiddish as a case study for the use of language profiles in language courses. Yiddish is offered as a language course at my university, and I discussed the development of the profile with some of the Yiddish language instructors. Somewhat surprisingly, they told me that some students in their courses have no background knowledge of Yiddish culture, but simply come across the class when looking for a course to fulfill humanities breadth requirements. The Yiddish language profile thus gives instructors a starting point for sharing these cultural elements with students. More significantly, we have worked to also make the data sets usable for language teachers. Many of the Yiddish data sets include nonspecialist explanations of the patterns, as well as Yiddish orthography and recordings, allowing them to be easily adapted for language teaching. Including these elements is a natural fit for a project designed to show language in context, given that what the language looks and sounds like is a foundational element of a language profile.

As a more specific example, one concept sometimes taught in Yiddish courses, though not typically using this terminology, is that syllabic nasals assimilate in place to immediately preceding consonants. This assimilation is not indicated in the Yiddish orthography or even generally in transliteration, but it is in IPA transcription. This feature of Yiddish is sometimes being lost among those who are participating in the

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

revitalization of the language, partly because it is not always easy to explain to students without talking about consonant place in a slightly more linguistic way. The nasal assimilation data set can be used in Yiddish courses to provide an explanation for students, even those unfamiliar with IPA, of the fact that the 'n' sound is changing to be pronounced in the same place in the mouth as the preceding sound. It can therefore help instructors to explain and preserve this property of Yiddish, which is highly characteristic of traditional pronunciations of the language. As part of the LPP, a proof-of-concept was created about this Yiddish phenomenon via a presentation in Yiddish about this nasal assimilation, and it was very well received, with one Yiddish graduate student and instructor, Miriam Borden, commenting that this type of resource is something that has not previously existed but should be incorporated into every Yiddish class.

Given that the languages in the LPP are underrepresented, it can be difficult to find language courses in which the profiles can be used. But even when there are no courses at the university, these conversations can happen within the community. As an example, there is a large population of heritage speakers of Oromo in my city, and the Oromo speaker-linguist working with us, Tamam Youssouf, is heavily involved in the community, including in Oromo classes for adult heritage learners. As such, he and I are working together to ensure that the language profile is of use in this context as well, with the goals of building interest in the language among the community and making materials usable for teaching. Both goals can be fit naturally to the LPP, simply by ensuring that data sets also can be understandable to nonlinguists.

Overall, beyond simply building awareness about underrepresented languages among students, the LPP can concretely be useful to communities, depending on their needs and desires. While rarely done in most linguistics courses, it is natural to make linguistic data sets for undergraduate classes double as language-teaching materials. This approach is even more appropriate in a situation like the LPP, where language orthography and recordings of speakers are part of the data set regardless, and where data sets are intentionally left open to instructor decisions about analytical approach and the level of the course. One of many future plans of the LPP is to work even further with instructors of language courses, both in the university setting and in the community, to find ways to make these profiles as useful as possible in those contexts.

6. Long-term goals and conclusions

Language profiles are currently in their infancy, but the feedback thus far has been overwhelmingly positive, from instructors, students, and community members. That said, particularly because the LPP is so new and the funding to create it is quite limited, we are currently working to address various shortcomings. One difficulty is balancing the amount of contextual information, between an amount that is possible to discuss in a few minutes in a linguistics course and an amount that does justice to the complex histories and cultures of these languages. We are addressing this through extensive discussions among the entire team, including both instructors and community members. Another issue, as mentioned previously, is whether the LPP might make marginalized languages seem more 'exotic', due to the lack of resources to include English and other majority languages. Additionally, due to the source of funding being oriented toward undergraduate courses at my institution, the community resources that are an eventual goal of the LPP are currently undervalued; the funding model values only the output for academic audiences, not for communities. For both of these issues, I am working on finding alternative and additional sources of funding.

In addition to addressing the above issues and improvements for the language profiles mentioned in Section 2, the LPP team is working on around a dozen additional languages that should be made available within the coming months, and we will continue to create language profiles for as long as we have community members willing to work with us and the funding necessary to create them. My hope is that the project will also continue to grow through contributions of linguists and communities around the world, and that this model of discussing languages in classrooms will become widely adopted. If, as a field, we want students to develop into ethically minded linguists who treat languages as the crucial component of

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

culture and communication that they are, who are invested in issues like linguistic diversity and language revitalization, and who see languages as more than just data on paper, then we should start training them from the beginning of their studies. Language profiles fill a critical gap in our ability to do so.

Data availability statement. Data is available on the project website at https://languageprofiles.ca/.

Acknowledgments. Thank you to the many people involved in this project, including those who shared their languages, those who helped with website design, and those who helped write and edit information for the website content. [Full editorial history: Received 05 June 2023; revision invited 25 August 2023; revision received 25 June 2024; revision invited 01 March 2025; revision received 14 May 2025; accepted pending revisions 09 July 2025; revision received 22 December 2025; accepted 26 January 2026.]

Funding disclosure. This project received funding from the University of Toronto Learning and Education Advancement Fund (LEAF).

Competing interests. The author declares no conflicts of interest.

Ethics statement. The research for this project was approved by the University of Toronto Research Ethics Board, Protocol #43963. All quotes from students and community members are used with permission. Quotes from anonymous survey responses do not require ethics approval.

References

Amberg, Julie S., and Deborah J. Vause. 2018. Integrating fieldwork: A scaffolded curriculum for the introductory linguistics class. *American Speech* 93(2).324–34. https://doi.org/10.1215/00031283-6926168.

Ambrose, Susan A.; Michael W. Bridges; Michele DiPietro; Marsha C. Lovett; and Marie K. Norman. 2010. *How learning works: Seven research-based principles for smart teaching*. San Francisco: John Wiley & Sons.

Anderson, Jane, and Kimberly Christen. 2019. Decolonizing attribution: Traditions of exclusion. *Journal of Radical Librarianship* 5.113–52. https://journal.radicallibrarianship.org/index.php/journal/article/view/38.

Battiste, Marie, and James (Sa'ke'j) Youngblood Henderson. 2000. *Protecting Indigenous knowledge and heritage: A global challenge*. Saskatoon: Purich.

Baxter, Paul T. W. 1978. Ethiopia's unacknowledged problem: The Oromo. *African Affairs* 77(308).283–96. https://doi.org/10.1093/oxfordjournals.afraf.a096985.

Bender, Emily M. 2019. The #BenderRule: On naming the languages we study and why it matters. *The Gradient*, September 15, 2019. https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/.

Bjorndahl, Christina; Lynsey Wolter; Reed Blaylock; Evan D. Bradley; Ann Bunger; Kristin Denham; Jessica A. Grieser; Kazuko Hiramatsu; Wesley Y. Leonard; and Michal Temkin Martinez. 2024. Increasing justice, equity, diversity, and inclusion in linguistics through small teaching. *American Speech* 99(2).238–60. https://doi.org/10.1215/00031283-11255052.

Bucholtz, Mary. 2021. Community-centered collaboration in applied linguistics. *Applied Linguistics* 42(6).1153–61. https://doi.org/10.1093/applin/amab064.

Bucholtz, Mary; Dolores Inés Casillas; and Jin Sook Lee. 2017. Beyond empowerment: Accompaniment and sociolinguistic justice in a youth research program. *Sociolinguistic research*, ed. by Robert Lawson and Dave Sayers, 25–44. London: Routledge. https://doi.org/10.4324/9781315671765.

Carnie, Andrew. 2021. *Syntax: A generative introduction*. London: John Wiley & Sons.

Chamany, Katayoun; Deborah Allen; and Kimberly Tanner. 2008. Making biology learning relevant to students: Integrating people, history, and context into college biology teaching. *CBE—Life Sciences Education* 7(3).267–78. https://doi.org/10.1187/cbe.08-06-0029.

Charity Hudley, Anne H.; Christine Mallinson; and Mary Bucholtz (eds.) 2024. *Decolonizing linguistics*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197755259.001.0001.

Cheng, Lauretta S. P.; Danielle Burgess; Natasha Vernooij; Cecilia Solís-Barroso; Ashley McDermott; and Savithry Namboodiripad. 2021. The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.715843.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clyne, Michael. 2007. Are we making a difference? On the social responsibility and impact of the linguist/applied linguist in Australia. *Australian Review of Applied Linguistics* 30(1).3.1–3.14. https://doi.org/10.2104/aral0703.

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3.15–50. http://hdl.handle.net/10125/4423.

Czaykowska-Higgins, Ewa. 2018. Reflections on ethics: Re-humanizing linguistics, building relationships across difference. *Reflections on language documentation 20 years after Himmelmann 1998* (Language Documentation & Conservation special publication 15), ed. by Bradley McDonnell, Andrea L. Berez-Kroeker, and Gary Holton, 110–21. Honolulu: University of Hawai'i Press. http://hdl.handle.net/10125/24813.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

D'Arcy, Alexandra, and Emily M. Bender. 2023. Ethics in linguistics. *Annual Review of Linguistics* 9.49–69. https://doi.org/10.1146/annurev-linguistics-031120-015324.

Davis, Jennifer L. 2017. Resisting rhetorics of language endangerment: Reclamation through Indigenous language survivance. *Language Documentation and Description* 14.37–58. https://doi.org/10.25894/ldd147.

de Saussure, Ferdinand. 1986 [1916]. *Course in general linguistics*. Trans. by Roy Harris. New York: Open Court.

Dryer, Matthew S., and Martin Haspelmath (eds.) 2013. WALS Online (v2020.3). [Data set.] https://doi.org/10.5281/zenodo.7385533.

Errington, Joseph. 2001. Colonial linguistics. *Annual Review of Anthropology* 30(1).19–39. https://doi.org/10.1146/annurev.anthro.30.1.19.

Errington, Joseph. 2008. *Linguistics in a colonial world: A story of language, meaning, and power*. Malden, MA: Blackwell. https://doi.org/10.1002/9780470690765.

Florey, Margaret; Susan Penfield; and Benjamin Tucker. 2009. Towards a framework for language activism. Paper presented at the International Conference on Language Documentation & Conservation, University of Hawai'i at Manoa. http://hdl.handle.net/10125/5014

Gaudry, Adam, and Danielle Lorenz. 2018. Indigenization as inclusion, reconciliation, and decolonization: Navigating the different visions for indigenizing the Canadian Academy. *AlterNative: An International Journal of Indigenous Peoples* 14(3).218–27. https://doi.org/10.1177/1177180118785382.

Gil, David. 2001. Escaping Eurocentrism: Fieldwork as a process of unlearning. *Linguistic fieldwork*, ed. by Paul Newman and Martha Ratliff, 102–32. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511810206.006.

Gordon, Raymond. 2005. *Ethnologue: Languages of the world*. 15th edn. Dallas: SIL International.

Hale, Kenneth. 1972. Some questions about anthropological linguistics: The role of Native knowledge. *Reinventing anthropology*, ed. by Dell Hymes, 382–400. New York: Pantheon.

Harrison, K. David. 2005. Ethnographically informed language documentation. *Language Documentation and Description* 3.22–41. https://doi.org/10.25894/ldd273.

Harrison, K. David. 2007. *When languages die: The extinction of the world's languages and the erosion of human knowledge*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195181920.001.0001.

Hermes, Mary; Megan Bang; and Amanda Marin. 2012. Designing Indigenous language revitalization. *Harvard Educational Review* 82(3).381–402. https://doi.org/10.17763/haer.82.3.q8117w861241871j.

Hill, Jane H. 2002. 'Expert rhetorics' in advocacy for endangered languages: Who is listening, and what do they hear? *Journal of Linguistic Anthropology* 12(2).119–33. https://doi.org/10.1525/jlin.2002.12.2.119.

Holden, Josh. 2018. Indigenous universities and language reclamation: Lessons in balancing Linguistics, L2 teaching, and language frameworks from Blue Quills University. *Collaborative approaches to the challenge of language documentation and conservation: Proceedings of the 2018 Symposium on American Indian Languages (SAIL)*, ed. by Wilson de Lima Silva and Katherine Riestenberg, 20–37. Honolulu: University of Hawai'i Press. http://hdl.handle.net/10125/24930.

Hymes, Dell H. 1963. Notes toward a history of linguistic anthropology. *Anthropological Linguistics* 5(1).59–103. https://www.jstor.org/stable/30022398.

Jacobs, Neil G. 2005. *Yiddish: A linguistic introduction*. New York: Cambridge University Press.

Kahn, Lily. 2017. Yiddish. *Handbook of Jewish languages*, revised and updated edn., ed. by Lily Kahn and Aaron D. Rubin, 642–48. Leiden: Brill. https://doi.org/10.1163/9789004359543_025.

Kochetov, Alexei. 2020. Research methods in articulatory phonetics II: Studying other gestures and recent trends. *Language and Linguistics Compass* 14(6):e12371. https://doi.org/10.1111/lnc3.12371.

Lattuca, Lisa R.; Lois J. Voigt; and Kimberly Q. Fath. 2004. Does interdisciplinarity promote learning? Theoretical support and researchable questions. *The Review of Higher Education* 28(1).23–48. https://doi.org/10.1353/rhe.2004.0028.

Leonard, Wesley Y. 2017. Producing language reclamation by decolonising 'language'. *Language Documentation and Description* 14.15–36. https://doi.org/10.25894/ldd146.

Leonard, Wesley Y. 2018. Reflections on (de)colonialism in language documentation. *Reflections on language documentation 20 years after Himmelmann 1998* (Language Documentation & Conservation special publication 15), ed. by Bradley McDonnell, Andrea L. Berez-Kroeker, and Gary Holton, 55–65. Honolulu: University of Hawai'i Press. http://hdl.handle.net/10125/24808.

Leonard, Wesley Y. 2019. Indigenous languages through a reclamation lens. *Anthropology News*, September 19, 2019. https://doi.org/10.1111/AN.1266.

Leonard, Wesley Y. 2020. Insights from Native American Studies for theorizing race and racism in linguistics (Response to Charity Hudley, Mallinson, and Bucholtz). *Language* 96(4).e281–e291. https://doi.org/10.1353/lan.2020.0079.

Leonard, Wesley Y. 2021a. Contesting extinction through a praxis of language reclamation. *Contesting extinctions: Decolonial and regenerative futures*, ed. by Suzanne M. McCullagh, Luis I. Prádanos, Ilaria Tabusso Marcyan, and Catherine Wagner, 151–67. New York: Bloomsbury.

Leonard, Wesley Y. 2021b. Toward an anti-racist linguistic anthropology: An Indigenous response to white supremacy. *Journal of Linguistic Anthropology* 31(2).218–37. https://doi.org/10.1111/jola.12319.

McIvor, Onowa. 2020. Indigenous language revitalization and applied linguistics: Parallel histories, shared futures? *Annual Review of Applied Linguistics* 40.78–96. https://doi.org/10.1017/S0267190520000094.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press

Mellow, J. Dean. 2015. Decolonizing Western science, research, and education: Valuing linguistic diversity. *Honoring our elders: Culturally appropriate approaches for teaching Indigenous students*, ed. by Jon Reyhner, Joseph Martin, Louise Lockard, and Willard Sakiestewa Gilbert, 45–60. Flagstaff: Northern Arizona University. https://jan.ucc.nau.edu/jar/HOE/HOE4.pdf.

Nagy, Naomi. 2000. What I didn't know about working in an endangered language community: Some fieldwork issues. *International Journal of the Sociology of Language* 144.143–60. https://doi.org/10.1515/ijsl.2000.144.143.

Nakassis, Constantine V. 2016. Linguistic anthropology in 2015: Not the study of language. *American Anthropologist* 118(2).330–45. https://doi.org/10.1111/aman.12528.

Odden, David. 2005. *Introducing phonology*. Cambridge: Cambridge University Press.

Ozburn, Avery; Gianna F. Giovio Canavesi; and Samuel K. Akinbo. 2023. Perception of ATR in Dàgáárè. *Laboratory Phonology* 14(1). https://doi.org/10.16995/labphon.8919.

Pedler, Megan Louise; Royce Willis; and Johanna Elizabeth Nieuwoudt. 2022. A sense of belonging at university: Student retention, motivation and enjoyment. *Journal of Further and Higher Education* 46(3).397–408. https://doi.org/10.1080/0309877X.2021.1955844.

Perley, Bernard C. 2012. Zombie linguistics: Experts, endangered languages and the curse of undead voices. *Anthropological Forum* 22(2).133–49. https://doi.org/10.1080/00664677.2012.694170.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. New Brunswick, NJ: Rutgers University, and Boulder: University of Colorado, MS.

Ragoonaden, Karen, and Lyle Mueller. 2017. Culturally responsive pedagogy: Indigenizing curriculum. *Canadian Journal of Higher Education* 47(2).22–46. https://doi.org/10.47678/cjhe.v47i2.187963.

Rice, Keren. 2006. Ethical issues in linguistic fieldwork: An overview. *Journal of Academic Ethics* 4(1–4).123–55. https://doi.org/10.1007/s10805-006-9016-2.

Rice, Keren. 2011. Documentary linguistics and community relations. *Language Documentation and Conservation* 5.187–207. http://hdl.handle.net/10125/4498.

Rice, Mskwaankwad. 2022. Power and positionality: A case study of linguistics' relationship to Indigenous peoples. *Proceedings of the Linguistic Society of America* 7(1):5295. https://doi.org/10.3765/plsa.v7i1.5295.

Sanders, Nathan; Lex Konnelly; and Pocholo Umbal. 2024. An action-based roadmap for equity, diversity, and inclusion in teaching linguistics. *Inclusion in linguistics*, ed. by Anne H. Charity Hudley, Christine Mallinson, and Mary Bucholtz, 357–80. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780197755303.003.0018.

Sanders, Nathan; Pocholo Umbal; and Lex Konnelly. 2020. Methods for increasing equity, diversity, and inclusion in linguistics pedagogy. *Proceedings of the 2020 annual meeting of the Canadian Linguistic Association.* https://cla-acl.ca/pdfs/actes-2020/Sanders_Umbal_Konnelly_CLA-ACL2020.pdf.

Shulist, Sarah, and Faun Rice. 2019. Towards an interdisciplinary bridge between documentation and revitalization: Bringing ethnographic methods into endangered-language projects and programming. *Language Documentation & Conservation* 13.36–62. http://hdl.handle.net/10125/24798.

Tsikewa, Adrienne. 2021. Reimagining the current praxis of field linguistics training: Decolonial considerations. *Language* 97(4). e293–e319. https://doi.org/10.1353/lan.2021.0072.

Umbal, Pocholo; Lex Konnelly; and Nathan Sanders. 2021. Handbook for inclusive linguistics teaching. The Linguistics Equity, Diversity, and Inclusion Repository. [Updated 2023.] https://ledir.ling.utoronto.ca/?page_id=33.

Walga, Tamene Keneni. 2021. Prospects and challenges of Afan Oromo: A commentary. *Theory and Practice in Language Studies* 11(6).606–12. https://doi.org/10.17507/tpls.1106.03.

Weinreich, Uriel. 1992. *College Yiddish*. 5th revised edn. New York: Yivo Institute for Jewish Research.

Whaley, Lindsay J. 2011. Some ways to endanger an endangered language project. *Language and Education* 25(4).339–48. https://doi.org/10.1080/09500782.2011.577221.

Woods, Lesley. 2023. *Something's gotta change: Redefining collaborative linguistic research*. Canberra: ANU Press. http://doi.org/10.22459/SGC.2022.

Zuraw, Kie. 2022. Four inclusive practices for the phonology classroom. *Proceedings of the 2021 Annual Meeting on Phonology.* https://doi.org/10.3765/amp.v9i0.5152.

Avery Ozburn is an assistant professor of Linguistics in the Department of Language Studies at the University of Toronto Mississauga. Email: avery.ozburn@utoronto.ca.

Cite this article: Ozburn, Avery. 2026. The Language Profiles Project. *Language* 102, 469–496. https://doi.org/10.1017/S0097850726000081.

https://doi.org/10.1017/S0097850726000081 Published online by Cambridge University Press