Ethical concerns for theoretical research in linguistics: Issues and best practices

Pritha Chandra, Roberta D'Alessandro, & Michael T. Putnam

June 11, 2023

1 Introduction

Traditionally, theoretical linguistics, a.k.a. *formal* and/or *generative* linguistics (GL), is a branch of research with a focus on structural patterns underlying the mental grammars/I(individual) languages of speakers. It is often assumed that because of the nature of its inquiry, generative grammarians tend to neglect the geo-political and cultural aspects shaping languages and linguistic communities. While the language communities are not the primary object of the study of GL, more and more data are being collected from different languages with the aim to come to solid generalizations. This enlargement of the language pool on which GL works translates into an increase of theoretical work on minorities and non-standardized varieties, which in turn has brought about the necessity to reflect on the standards and ethical methods for data collection, analysis, and the distribution of findings stemming from this research. The further away GL gets from English and other standard languages, the more stringent these ethical concerns become: Who is the owner of the data? Is *speaker* the right way to name those who provide the data? How much of the feelings and attitudes of the speakers towards their data should be reported in the scientific articles? Our chapter sets to offer an overview of the main issues and best practices in data collection, with a focus on heritage and minority varieties of Indo-Aryan, Romance, and Germanic. We start by challenging the statements according to which GL practice is oblivious of ethical concerns; we show how in fact the generative methodology is in fact sometimes more "ethically oriented" than other methodologies. We move on to present three case-studies, showing how important it is that researchers become more cognizant of

the socio-historical events that have shaped the various language communities and given rise to certain linguistic attitudes. Finally, we touch upon the importance of involving minority speakers in the documentation of their own language (when possible/applicable).

1.1 Generative Grammar - Basic assumptions

The Generative Linguistics (GL) tradition started out with Chomsky's early work in the '50s and '60s, and especially in its first decennia of life, was focused on finding ways to formalize natural language Chomsky (1957, 1965). Chomsky's first works, still in use in informatics courses, reduced natural language (or rather, natural languages) to a sequence of rules on syntactic items, which were to be implemented in formal computational notation (i.e., Natural Language Processing (NLP) (Berwick & Stabler, 2019).

This formal approach to languages, known as *transformational grammar*[1], based on rules taking as their input the output of previously applied rules to produce sentences in a given language (often, but not only, English - see Postal (1966) on agreement and pronouns in Spanish, for instance) focused on the sentences themselves and their internal structure. The assumption of a uniform underlying structure was necessary in order to reduce natural languages to a set of strings obtained by combining elements from a lexicon through the use of grammatical rules ("By a *language* we will mean simply a set of strings in some finite set V of symbols called the *vocabulary* of the language. By a *grammar* we mean a set of rules that give a recursive enumeration of the strings belonging to the language." (Chomsky & Schützenberger, 1963: 118). No space for variation was allowed, as that would complicate the rules considerably. It wasn't until Principles & Parameters (Chomsky, 1981) that language variation took a relevant place within GL.

Since Aspects, it was recognized that the extant forms of language are a subset of theoretically possible structures, though a theoretical way to capture this subset was not immediately evident. Government and Binding (GB) Theory (Chomsky, 1981) gave a workable solution to this central problem by postulating a set of universal principles of grammar construction (aka Universal Grammar), each of which had an open parameter, whose value was set on the basis of the available linguistic data. Because of its accuracy, GB opened the

[1] We recognize that there are a separate family of *generative* approaches that are declarative (rather than transformational) in nature; see e.g., Borsley and Börjars (2011) for an overview of the core desiderata of these frameworks. We do not discuss these frameworks further in this chapter; however, the concern for ethical research in formal linguistics also holds for these non-transformational ones.

path to a worldwide exercise in the formalization of languages; many generalizations were drawn on the basis of languages that were very distant from English (Huang 1982,Fukui 1986), the work by Jim McCloskey, Ken Hale, Luigi Rizzi, and many others).

In 1993/1995, Chomsky (1995) outlined a new research project, called the Minimalist Program (MP) with the aim to investigate language from a more biological viewpoint, as well as understand the reason for language design and the relation that our human language faculty (narrow) may have evolved from primate cognition and communication systems (broad).

2 GL Methodology and Ethical Concerns

Notwithstanding the visible shifts in the focus of GL, from being a pure combinatorial approach in Syntactic Structures to a cognitive one in Aspects and GB and finally a minimalist take in recent years, driven mostly by questions of optimal derivations and design (cf. Boecks and Hornstein 2010), its fundamental goals remain the same and that is to formulate theories or grammars of human language that meet certain adequacy criteria.

The grammar must be, first and foremost, **observationally** adequate. i.e. it should describe which sentences are good and which are bad in a language. It should also be **descriptively** adequate, and model all possible sentences in a language, and only them, and, most importantly for our purposes, correctly describe the intuitions of an idealized native speaker. Setting further goals for linguistic theory, Chomsky states that the grammar should additionally meet the condition of explanatory adequacy, i.e. "offer an explanation for the intuition of the native speaker on the basis of an empirical hypothesis concerning the innate predisposition of the child to develop a certain kind of theory to deal with the evidence presented to him"(Chomsky, 1965: 24-26). Once the grammar meets these conditions, especially with respect to descriptive and explanatory adequacy, we may then evaluate it in terms of design, elegance and economy that also define the rest of cognition and biology.

The adequacy conditions inform GL methodology. The primary objectives are to build an acquisition model and reveal abstract linguistic representations underlying the linguistic competence of native speakers. Central to GL methodology is therefore the idea of an idealized speaker, abstracted from the variations and oscillations that characterize a community. Linguists rely on the intuitions of native speakers, assuming that these are reflections of

their underlying linguistic knowledge.

There are ethical questions that obviously arise with adopting these notions and methods, including the question of whether there can ever be an 'idealized speaker', insulated from their socio-cultural context and community, and whether language can be abstracted away from its performance and social use. These questions become very important especially in the current stage of GL, where there is an increased interest in collecting and analyzing linguistic data of marginalized and heritage languages. Since the identities of these languages are closely tied to the identities of the underrepresented communities, one must ask if it is still appropriate to embrace the fundamental principles of GL or if they need to be critiqued and dropped. Should new methodologies be adopted, instead, while collecting data from a large number of speakers? In what follows we will reflect on some of the main methodological issues often contested to GL, observing how they are not always as harmful and unethical as they might look at first sight.

2.1 The speaker and their language

2.1.1 The idealized speaker

As mentioned above, the speaker is considered by GL mainly as a sort of language machine operating a grammar on a vocabulary; intuitions about language are considered in the formal, aristotelian sense, as abstract rules that apply to some matter, independently of the matter itself (e.g., rules applying to a set of words, like adjectives, independently of the adjectives themselves). With the spotlight remaining solely on the abstract grammar of individual speakers, the researcher is able to sidestep the complex identity issues defining both the speakers and their languages. It is a recognized fact that individuals have their social and community affiliations; neither they nor the languages they speak are disembodied, abstract constructs. However, when the objective is to examine the mental representations underlying linguistic knowledge in speakers, these complex socio-cultural factors act to impede rather than facilitate such inquiries. While the conceptions of idealized speakers, speech communities and linguistic forms have their own challenges, it helps in maintaining scientific objectivity, to a certain extent, in linguistic research.

Researchers also have their own biases, cultural viewpoints, and value judgements. If the socio-cultural identities of consultants and their languages are included in linguistic analysis, it is very likely that scientific objectivity, to some degree, will be compromised.

Abstracting away from the social milieu in which the data are used by speakers is useful because it creates the right conditions where the researcher is trained and expected to treat the data in the same unbiased manner as they treat their own linguistic judgments. In fact, one of the primary goals of generative linguistics since its inception has been to describe intuitions, not to prescribe rules to its speakers; see for instance, Chomsky (1954)'s criticism of the correctional approach adopted by Reiger (1953)in his textbook of Modern Hebrew. Language description, albeit within the confines of a theoretical framework, is a norm in generative studies, and it is very rare that generalizations and theoretical predictions are formulated based on speakers' socio-cultural status or racial and gender identities.

2.2 The speaker's intuitions

GL aims at devising a set of procedures that enable linguists to deduce statements about a 'linguistic grammar' from a fixed sample of linguistic material from any given language. It does so by trying to identify the "limits" of grammar, by investigating borderline constructions, the grammaticality of which can show whether a given grammar can or cannot generate them. Investigating the limits of variation entails checking for the grammaticality of sentences that would not be normally produced spontaneously, because of their complexity. The only possibility that is available to understand what the grammar of a speaker can or cannot produce is to ask for speakers' intuitions or acceptability judgments about given sentences. When it was first introduced, this method was a radical departure from the method adopted by American structuralists, who extrapolated regularities from a corpus of data, often of 'exotic' languages of the Americas.

Asking a speaker for their intuitions might not be the optimal way to ascertain whether their grammar can actually generate a particular structure. From an ethical viewpoint, however, it is remarkable, as it puts the speaker's own opinion as central to the linguistic investigation. Generative linguists do not collect data, for instance on fieldwork, disregarding what the speakers think about them. They mostly (but not always) might disregard what the speakers think of the social situation in which the languages are used, but they certainly do not replace the speakers in establishing what can or cannot be said in a language. The speakers are, for GL, the ultimate authorities and the owners of their data, which is ethically speaking a very inclusive strategy.

2.2.1 One speaker is enough

As stated above, the data in GL are collected from speakers through observation, though in most cases, linguists elicit data through questionnaires and one-on-one interviews; there are also no specific guidelines for the number of consultants required for any study. It is understood that an entire corpus cannot capture the actual number of sentences that can be potentially generated by a single speaker or a group of speakers, since they are, in principle, infinite in number.

As long as one structure is consistently used by a speaker, that structure is a possible example of something generated by human language, which makes it an interesting object of study. In this respect, GL investigation does not have any issues with qualitative analyses of small data samples. Conversely, the general trend nowadays is towards collecting big sample of data, which can be then analyzed statistically: if the sample is not big enough, the analysis will not be possible, and therefore the small amount of data negligible. While this is a perfectly acceptable viewpoint when large standardized languages are involved, it becomes problematic when investigating minority languages, languages with a very few speakers. Paradoxically, GL, which is traditionally less interested in the study of variation, is perfectly at ease with the work on languages spoken by 3 speakers, in that it assumes that language is a psychological object, and therefore present in the mind of the speaker. Any insight in the products of human cognition can and should be investigated to gain a better understanding of how language works. As explicitly stated in D'Alessandro, Natvig, and Putnam (2021) and Leivada, D'Alessandro, and Grohmann (2019), small and endangered languages can reveal a great deal about possible and impossible structures, something that large standard languages can fail to show. Therefore, with respect to the importance attributed to small languages, GL is perhaps a more equal framework than many quantitative ones, as the number of speakers is not a relevant factor for selecting structures to examine. We could summarize this point by saying that, exaggerating slightly, the *one speaker is enough*[2] approach to linguistic study is an implicit recognition of the importance of minorities, which we will discuss more extensively in the next section.

Finally, we will also like to note that the inclusion of non-standard linguistic vari-

[2]To be accurate, though, *one speaker is enough* approach is, in practice, a relic of the past: the usual practice nowadays is to look for a small number of speakers sharing a grammar and relying on their intuitions. If a structure is produced systematically by only one speaker, it could be due to some idiolectal quirk. The habit according to which the researcher's intuitions suffice for the inquiry has also been abandoned.

eties, dialects, heritage and minoritized languages in generative studies is also motivated by theory-internal quests. Since the 1980s, with the advent of the GB theory and the the Principles and Parameters architecture defining the contours of the Faculty of Language (Chomsky, 1981), tackling variation within the limited hypothesis space allowed by Universal Grammar becomes one of the primary tasks. It is of paramount interest to linguists that they examine how the proposed parameters unfold with limited and insufficient data, as often happens in heritage and bilingual contexts.

2.2.2 All languages are equal

The first observation that comes to mind is that, if one speaker is enough (hyperbolically speaking), no large groups of speakers are needed to investigate a specific phenomenon. If two or three speakers producing a structure are enough to assume that that's a possible structure in a language, then the community of speakers does not need to be extremely large, as we argued above. This has immediate repercussions on the study of minority languages, or languages spoken by a handful of speakers, which must be excluded by experimental investigations, for instance, as they do not reach the minimum amount of controllable uniform stimuli that are necessary for statistic-based generalizations.

Minorities are also relevant in a different way, as we saw before: it is a well-known fact that standardized languages try to get rid of discrepancies, especially when they are influenced by the reflection of grammarians (see Jespersen 1925, Haugen 1997, Milroy and Milroy (2012) and many others); standard varieties avoid, for instance, reduplication, or redundancies, which are very common in substandard varieties and dialects (Barbiers, Koeneman, Lekakou, & van der Ham, 2008). GL, as stated in section 2.2, is instead interested in the "fringe" of grammar, and therefore has as its first target non-standardized varieties. While the role of standardization and education on I-language is not completely clear, it is no doubt that the language of illiterates or minorities is less polished, less artificial, and therefore more telling regarding what grammar can or cannot generate. In this sense, inclusion is a requirement of GL. In the past, this aspect has been admittedly overshadowed by the fact that the large majority of studies in GL was on English [3]. However, the role of minority and non-European grammars has increased noticeably in the last 30 years, as we

[3]The reasons for that are many, the main one possibly being that GL is an American tradition (cfr. D'Alessandro 2019)

will see below.

Looking at the issue from a different perspective, we notice that the focus on subparts of grammar and on the speaker's intuitions means that linguists can study languages without considering whether they are standard or prestige language or just dialects of another language. Language being an abstract object of inquiry, it does not need to be studied along with any social (mis)conception surrounding it. This has always been a very appealing feature of GL, but it also meant that, in the initial years when language variation was not a central concern, linguists could get away with limiting their studies to only a few, mostly standard, languages. The idea of language as a scientific object, divorced from social perceptions, has changed in the last decades, when the number of varieties taken into account has increased noticeably. In fact, this has been one of the enabling factors for why many non-standardized varieties have been included in GL studies; new dimensions to the formal study of language have been introduced, like the diachronic and the contact one. Variation and microvariation in languages now have a central place in formal analyses, and the field is in full swing with the study of all sorts of minority languages (see for instance D'Alessandro et al. 2021; Nevins 2022; Polinsky and Putnam 2023).

Regarding microvariation, the Syntactic Atlas of Dutch Dialects (SAND), the ASIt project (`http://asit.maldura.unipd.it/`), the Microcontact Atlas (`https://microcontact.hum.uu.nl/#home`) are just some illustrative examples of projects that have successfully combined large-scale dialectal data collection with theoretical inquiries on variation. Research on minoritized and heritage languages has also gained pace in recent years Cummins (2005), Benmamoun, Montrul, and Polinsky (2013)Polinsky (2018a) among many others).

This theory-internal inquiry, along with the rise of cross-border migration, inter-mixing of populations and languages and the need for greater assimilation of communities, have opened up immense possibilities for generative linguistics to fruitfully combine empirical studies with theoretical research. Another clear sign of the central role that inter- and intra-speaker variability play in current formal analyses, is the appeal to statistical learning (Lidz & Gagliardi, 2015) and other algorithms, such as the *Tolerance Principle* (Yang, 2016), when researching L1 acquisition from a minimalist perspective. The third methodological assumption of GL, *All languages are equal*, is often underestimated, while we feel that it should be made some justice.

2.2.3 What's missing

In the previous parts of this section we have tried to highlight the positive ethical aspects of adopting a GL methodology. However, some strikingly negative aspects continue affecting generative literature. We examine some of them here.

Language names In most generative papers, as well as in many typological papers, no information is provided next to the examples other than the name of the language. First, it is worth remembering that the name that linguists attribute to languages is not always the same that the linguistic community chooses for itself. The dilemma on which version of the language name to choose is not of immediate solution (see a recent debate between Haspelmath 2017 and Dryer 2019 on who should decide the names of languages), but it needs to be put on the chart for GL as well. Then, in addition to being disrespectful of the opinion of speakers about their own language, this attitude is also a sign of inaccuracy: very often no other information other than the name is provided about the language, where it is spoken, by how many speakers, in which conditions. This information does not cost too much trouble but improves the quality of the article tremendously, and could be easily implemented.

Absence of maps One particularly striking subcase of what we just discussed, namely the lack of information about the language, is the systematic absence of maps in GL articles. It is not immediately obvious to everyone where languages, especially of minorities, are spoken. Aside from ethical concerns, indicating the exact point of data collection is a sign of accuracy, as well as respect for those who produced the data.

The attitude of the speakers and their background We discussed the role of the speaker (or lack thereof) in theoretical, generative research. One piece of information which has not been mentioned but is extremely relevant when trying to ensure the quality of the data collection is regarding the attitude of the speaker, as well as their fluency, and their relationship with the other languages. This information is rarely provided, but it should: on the one hand, because the data must be as accurate as possible; on the other, though, because speakers are not inanimate non-sentient robots with no impact on the data they are collecting.

In what follows, we present three very diverse case studies with some insights and tips on how to carry out theoretical research in different parts on the world, highlighting the relevance for what's been said so far: (i) languages of minorities and indigenous communities

in India, (ii) languages of heritage Italian communities in Latin America as well as in Italy, (iii) heritage Germanic languages spoken in the US. Although each of the scenarios presents unique and pressing challenges, they throw light on how to reconcile GG goals with new data-collection methods, and the ethical practices that should be followed when we work with languages of underrepresented communities.

3 The application of ethical linguistic research across three continents

After discussing the methodologies and standards of practice followed in the GG tradition, in some detail in the last section, along with some current trends towards data collection among minorities and heritage language communities, we now turn to three case studies that exemplify the importance and need for the application of these standards, moving forward. In Section 3.1, we take a closer look at the nature of fieldwork in India. Here we consider how linguistic research should address, and circumvent the differences imminent in state-offered standard/non-standard language tags as well as those arising from the social divisions that exist in the Indian Society. Next we transition to Romance heritage varieties in the Americas as well as in Italy in Section 3.2. In this section we speak of the challenges researchers face when dealing with elderly and (often) illiterate speakers of a given language. Finally, we examine situations often encountered in research on varieties of heritage Germanic in Section 3.4. Although there is a certain degree of overlap between the former group and this one with respect to shared concerns regarding the inclusion of the elderly and illiterate in this research program, a unique issue that emerges in heritage German centers on the use of technology when interviewing conservative Anabaptist Christians, e.g., the Amish, Hutterites, and Old Order Mennonites.

3.1 The Language Situation in India

After its independence from British rule in 1947, the identity of India as a nation was deeply informed by debates on language. The Constituent Assembly Debates of 1947-49 addressed three important topics regarding language that were to profoundly impact the structure of the emerging nation state. The first was the recognition of Hindi/Hindustani as the official language of the country; with English acting as the second official language. The

second concerned the federal structure of the nation, based mostly on linguistic identities (e.g., Punjab as a Punjabi-speaking state, Kerala as a Malayalam-speaking state). The third was the inclusion of standard/mainstream languages in the Eight Schedule of the Indian Constitution, guarantying state support for their sustenance and growth. The total number of scheduled languages currently stands as twenty-two, and there are suggestions for including thirty-nine more languages in the future.

In India, language plays a dual role. On one hand, it is used to create the concept of a homogeneous nation-state, through state-controlled propagation of an single Official Language 'Hindi'. Simultaneously, language-triggered identities underscore the heterogeneous nature of the Indian society, which is already marked by differences based on class, caste, religion and gender. The Eight Schedule adds to these differences by co-opting only a handful of languages in its purview, leaving out hundreds of non-standard languages to struggle and survive on their own.

The binary opposition between standard/non-standard or scheduled/unscheduled languages is however too simplistic for a nation like India, where on the one hand, scheduled, standard languages are sometimes marginalized and on the other, unscheduled languages assume different names and identities in the struggle for state recognition. A researcher interested in studying Indian languages should be aware of the ever shifting linguistic identities and landscapes, and accordingly select the right methods for data-collection in the region, without disturbing the ethos of these societies (for discussion on fieldwork preparation, behavior and dressing codes, see Abby 2001, Krishnamurti 2003, and Chelliah and Reuse 2011). When publishing research on these languages, one must also be careful in truly representing them as scientific objects on par with other languages. Below, we first present the case of Urdu, which is a scheduled language but remains marginalized because of its association with the Muslim minority in India, and then proceed to discuss Rajbanshi, a language of an indigenous community in the eastern state of West Bengal, which, forced by external circumstances, has taken multiple identities over the last many decades.

3.1.1 Scheduled, but still marginalized languages

The Hindi-Urdu debate serves as an instrument to divide the people in the northern states along religious and linguistic lines. It is a sensitive topic that often ignites communal tensions in the country, because of the alleged links between Hindi and Hindus, and between

Urdu and Muslims. From a linguistic standpoint, however, the two languages are very similar in their morphology and syntax, though they are written in different scripts: Hindi uses the Devanagari script while Urdu uses the Nastaliq script.

Urdu was one of the original fourteen languages to be included in the Eight Schedule. Interestingly however, despite being a widely used language in the state of Uttar Pradesh (UP), the first Ordinance to recognize it as a second official language of the state, alongside the official language Hindi, was proposed as late as April 1982. The Ordinance triggered violent protests throughout the state, and multiple petitions were filed against it in the Supreme Court of India. It was in 2014, that Urdu was finally given its constitutional rights as a second official language of UP; though in reality, it is still to get its rightful place in education, media and other official spaces (Chandra 2014, 2019).

Despite its scheduled language status, Urdu thus remains a largely marginalized language because of the lack of incentives available to its speakers. Children of affluent Muslim families are sent to English-medium, private schools where Urdu is not included in the curriculum. Children from poorer backgrounds, on the other hand, attend local, religious schools called *madrasas* where they learn Arabic and Persian classics and take Urdu lessons, along with other traditional disciplines. The use of Urdu is thus largely confined to certain economic groups and a small set of religious-educational institutions.

The case of Urdu testifies that standard, official languages of minorities can be marginalized. In addition, it also informs us that speakers of a standard language do not necessarily constitute a homogeneous group and their social standings and sensibilities can be very different. An instance of this heterogeneity is recorded by Ahmad (2010), where he recounts that despite being a Muslim researcher himself, he received very different treatment from different Urdu-speaking Muslims of Old Delhi, where some welcomed him warmly, while others found him suspicious and not trustworthy.

It is therefore doubtful that popular perceptions about minorities, including those about the Urdu speaking Muslim minority, are helpful when one wants to investigate these languages. In fact, the socio-cultural tags ascribed to them may present more challenges to linguists than not, as they force the internal dynamics of these communities and their complex relationships with 'outsiders' to come to the fore. For a scientific study of these languages and their varieties, it is imperative that researchers discard typical conceptions about minorities and approach their speakers from a completely neutral standpoint - simply, as native speakers of Urdu varieties. It is also important to acknowledge that these

communities constitute smaller groups based on region, education and professions, and therefore it is natural to expect intra-community linguistic variation as well. To gain a fuller understanding of the linguistic diversity within these communities, we must record data from speakers from multiple groups from all walks of life, and analyze the data, using the same theoretical tools and unbiased scientific outlook that are used for analysing the languages of non-minorities.

3.1.2 Non-scheduled, indigenous languages

India has a notable number of indigenous languages that have not been accorded an official status. Most of these linguistic communities are economically deprived, with some voluntarily staying away from the glare of human civilization. From time to time, a tribal group may protest for better allocation of resources, more opportunities, and for an official recognition of its language, forcing the government to take notice and do the necessary. This happened to the Austro-Asiatic language Santhali of the indigenous Santhal community that was co-opted into the Eight Schedule in 2003 after several efforts from the community members. Others such as some aboriginal languages spoken in the Andaman and Nicobar islands may be on the verge of extinction, as their speakers continue with their primitive forms of existence (Abby, 2020).

Research on indigenous languages often focuses on some exotic properties of these languages, and the (apparently) outlandish customs and behaviour of the speakers. By emphasizing on these 'sensational' features, we undermine the natural growth of these languages and communities, and on their assimilation, adaptation and change over time. Field researchers also tend to disregard speakers' agency and their changing attitudes about their languages and the expectations they have from society and government.

We illustrate this with Rajbanshi, which is a language spoken in the state of West Bengal (and in adjoining Assam and Bangladesh). In West Bengal, its speakers are mainly found in the districts of Dinajpur, Malda, Jalpaiguri and Cooch Behar, with the regional linguistic variation expected of any language spoken over such a large area. The language also goes by the names of Matribhasha, Kamtapuri, Surjapuri and Koch. Rajbanshi is an interesting instance of a language where its speakers reinterpreted and rewrote their past to carve out new identities for themselves (cf. Debnath 2007; Nandi 2014). The language itself gained a new identity in recent decades because of social regrouping within the community.

Rajbanshi identity first became a serious issue in 1882, when the Indian Census Report categorized Rajbanshi and Koch as a single tribe. The Rajbanshis rejected this clubbing, with the claim that Koch is an inferior (*Dalit*) caste, speaking an inferior language and living a lowly life, while the Rajbanshis belong to the higher (*Kshatriya*/warrior) caste, coming just behind the *Brahmin*s. In 1926-1927, however, the community had a new narrative, when the caste-based organizations put a requisition to the government for get the Rajbanshis recognized as *Dalits*, a demand that was finally granted to them in post-independent India. The shift was a necessary compromise for the community to get access to the government funds and resources reserved for upliftment of the downtrodden castes and tribes. The Rajbanshis who were once against the idea of being clubbed together with the Kochs are now perfectly content with the new identity 'Koch-Rajbanshi' or 'Kamata'. The Rajbanshi language is considered now as a conglomerate of multiple languages including the languages spoken by the Koch community members.

Rajbanshi is a representative example of indigenous languages in India that have very flexible identities, as their speakers keep changing their identities, while they continue to negotiate with the state for official recognition and better resource allocation. Linguists working on these languages must appreciate this essential nature of indigenous languages as it is unethical to impose a fixed identity on a group, which is actively seeking out multiple ones. One must also take into account that some members of the community have also moved up in the social hierarchy, gained higher education and taken up jobs as educators and administrators. They are highly respected members of the community, and have a strong voice in how their cultures and languages are projected to the outside world and media. In many instances, language researchers have had to first approach these members, to get access to the linguistic data of the other community members. While this emerging practice of gate-keeping by some members has some disadvantages (including the potential risk of highlighting one variety while undermining others), it is always useful to have the support of some 'insiders', in order to gain the confidence of the others.

To summarize, we have discussed two cases from India: the first, where the language is recognized in the Eight Schedule but is marginalized because it is spoken mostly by a minority group, and the second, where the language is not yet recognized by the Eight Schedule, and its members are content forming new identities. Both studies present India as a country which has very complex language relations and linguistic identities, some of which also shift with time. It is thus imperative that researchers contextualize the languages that

they want to examine. They should not be influenced by popular ideas about minorities and indigenous people since these are, in most cases, misleading and wrong. They should also avoid social and cultural biases, and try to keep the study as objective as possible. It should not be assumed that minority and indigenous communities are homogeneous groups with a fixed set of beliefs and knowledge systems. Individual members may have different narratives about their histories and cultures. The researchers should be mindful of these intra-group differences, but should always remain as neutral observers. Since other factors, such as the minority and caste status of the community, also influence how languages are treated, it is imperative that researchers do due diligence and undertake their own in-depth inquiry of the social and political contexts of the languages they want to examine. The information given in government census about languages should be double-checked against literature on the sociology and politics of the languages and their communities. Since the same language may have different names in different areas, it is also advised to adopt the names given by the speakers of specific regions, so as to not hurt the sensibilities of the users. The researcher must also be in contact with people, who are considered to be the advocates for the language and the community. Trust-building is extremely important in such situations.

Along with these issues, there are some specific methodological concerns that GL researchers have to address, especially when collecting minority and indigenous language data in small towns and rural areas in India. With the GL inquiry still focused on I-languages of individual speakers, the ideal situation is where the researchers meet a single consultant at any given point of time. However in most contexts, consultants come in groups - sometimes, it is a group of family members who assemble in the visitors' area of their family home (this is mostly true for young female homemakers who are accompanied by other females of a joint family), or a group of men who engage in collective work such as carpentry or farming, or who have simply gathered together at a local eatery. These people are ideal speakers, as their language remains mostly untainted by Indian English and other standard languages spoken in the neighbourhood. In such situations, the researchers have the formidable task of extrapolating linguistic patterns from individual speakers, and simultaneously, record the 'communicative integrity' (Cowart, 1997: ix) of the members, i.e. an array of sentences that are similarly experienced and accepted by all community members. This task requires the researcher to transit seamlessly between an individual speaker and her intuitions, and the collective linguistic judgments given by the entire group. Collective judgments are also very

useful when investigating discourse-level phenomena such as honorificity. Since elicitation of such forms (sometimes absent, and sometimes rare and still emerging in the language) requires the researchers to provide a context, it is always useful to have some members who, with minimal prompting, provide the context as well the actual utterance in response to the context. Group settings are also great in bringing out variable patterns, with communal acceptability ratings, as the members engage in serious debates on each others' intuitions. Thus, while communal settings may have their own problems, from the viewpoint of a theoretical pursuit like GL, there are many advantages to it, most specifically, its participatory nature, which turns the consultants into researchers of their own languages.

3.2 Italian diaspora in the Americas

3.3 The status of Italo-Romance diaspora in America

The situation of heritage Italo-Romance languages in the Americas and Italy has been largely documented in the *Microcontact* project, and reported in D'Alessandro et al. (2021), Andriani, Casalicchio, et al. (2022); Andriani, D'Alessandro, et al. (2022). The large inquiry targeted Italian Romance minority languages, both in Italy and in America, namely in Argentina, Brazil, United States, and Canada. One of the aims of the project was finding heritage speakers with a specific profile, namely first-generation immigrants, or elderly speakers. The language situation of these minorities in the various American countries is varied, and it is itself different from that which is found in Italy. However, certain features are common, and these will be considered here.

To begin with, minority Italo-Romance languages are rather stigmatized both in Italy and in America. In Italy, the legacy of the fascist regime with respect to these minorities is still heavily ingrained in their speakers: Italo-Romance minority languages are considered inferior to Italian, they are still called *dialects*, they are to be avoided in formal contexts (with some regional exceptions). In general, speaking these languages is considered a sign of illiteracy rather than a welcome exhibition of bilingual knowledge; no education is provided in the languages (again, with some exceptions limited to Friuli Venezia Giulia and Sardinia), and they are slowly dying out.

A somewhat similar situation is found in the Americas, where the speakers of these languages, who left Italy before the spreading of standard Italian, carry with them an ambivalent feeling of speaking both a socially inferior language but yet the language linked

to their identity. In North America, many minority speakers send their children to Italian schools, so that these Italo-Romance varieties are practically erased from their language repertoire. The situation is different in Latin America, especially where largish communities of speakers persist, as in Brazil, where the Venetan-based variety called *Talian* has been officially recognized in 2009 as a heritage language in the states of Rio Grande and Santa Catarina (Estado do Rio Grande do Sul, 2009), and finally as Brasilian Cultural reference language by the Instituto do Patrimônio Histórico e Artístico Nacional of Brazil in 2014 (*Inventário Nacional da Diversidade Linguística (INDL), Decreto 7387/2010*, 2014).

The language policies of the host countries have also played a very important role with respect to the survival of these varieties. In Brazil, the government started so-called *assimilation* policy, culminating in the constitutional act of 1934 which established that "The concentration of immigrants anywhere in the country is prohibited, the law should govern the selection, location and assimilation of the alien"; this resulted in an almost complete wipe-out of the Italian minority languages in cities; they survived instead in rural areas, in particular in southern states like Espirito Santo and Santa Catarina, where less governmental control took place.

The Argentinean situation is slightly different, also due to the very close proximity of Spanish to the Italo-Romance varieties: Italian speakers assimilated spontaneously to the local population and adopted Spanish already in the first generation. Italo-Romance varieties were still spoken in the household, but got somehow quickly lost.

The first-generation speakers targeted by the research had a very specific profile: they were mostly monolingual speakers of one of the Italo-Romance varieties, had a very basic knowledge of Italian, and were illiterate. This profile is more difficult to find in Italy, where it is however also found. The main issues with collecting data from these speakers are common to all elderly and illiterate speakers; we list some of them here.

Difficulty in interacting with computers/recorders When a speaker has never used a computer in their life, the first interaction can be challenging. Obviously, the researcher should present the stimuli directly; there are however very strict constraints against language priming; if the researcher speaks a slightly different dialect of the same language, they will trigger priming. Italian can't be used as it would interfere in the judgments, so ideally one needs to use recorded stimuli from the same variety. Reconciling methodological integrity with physical and educational barriers is a big issue for GG studies. Data accuracy needs to be granted, as the difference between the presence or absence of an *a* can amount

to the difference between two totally separate constructions. While data accuracy needs to be granted, the question is: how do we ensure that the speakers are perfectly at ease with the inquiry?

Difficulty with abstraction Illiterate speakers often have difficulty with understanding tasks, even repetition tasks which seem to be so straightforward. One of the main issues is making sure that these speakers understand the task, both because of the results of the study themselves and because there are strict requirements about informed consent and the clarity of the information, which is not always reached.

Informed consent and social acceptability Informed consent requires full understanding of the purposes for which the research is taking place. As just mentioned, this is not always possible with informants with a low level of literacy. It is crucial that the researcher tries their best to explain the use of the data, though, together with the property: the data belong to the speaker (which is not a controversial statement, given that according to most internal regulations the data are the property of the host institute). The most common form of informed consent is written. While this is easier to store and control, it is not always appropriate. People who are not fluent readers, or elderly people, might find it really difficult or even threatening to have to sign something. In the case of *Microcontact*, oral informed consent was also accepted, given that this means is more in conformity with what is socially correct for the speakers.

Conflicting rules The rules that are established by data protection regulations such as the GDPR (European Union, 2018) are often in conflict with what the speakers would like or expect. One of the most evident areas of conflict is anonymity: speakers very often do not want to remain anonymous, they perceive the request (or the insurance) of anonymity as an undesirable appropriation of their identity, or of a sort of cancellation of their persona. The conflict between regulations and wishes of the speakers is not only relevant for Heritage Italo-Romance speakers, obviously, but it is something that we perceived particularly strikingly.

Social/cultural appropriateness. GL investigates structures, as repeated above. Sometimes these structures, or the items used, are uncomfortable for speakers. As an example, consider transitivity: prototypically transitive/agentive verbs like *kiss* or *kill* can put the speakers in a situation of discomfort. Sentences like "X kissed Y" may feel like inappropriate, especially if one of the names corresponds to someone the speaker knows.

While there seems to be an obvious way out, for instance by changing the names, the

solution is not so straightforward. Foreign names might sound incomprehensible, and the stimuli might consequently be perceived as less natural. Changing the stimulus on the spot is not possible, as all speakers have to have access to the same stimuli for accuracy; furthermore, the fieldworker might speak a different variety than that of the rest of the stimuli, stimulating priming of some sort. The only way out is to avoid words and sentences that might create uneasiness. To do that, the assistance of people from the community is crucial, as will be argued in the next subsection. [4].

3.3.1 Citizen science: involving the speaker in the study of their language

The speaker's contribution is fundamental in scientific research; however, linguists tend to forget the importance of hearing what those who "possess" the language need to say. This aspect was discussed above in section 2.2 on the importance of the speaker's intuitions, but it also branches out to other aspects.

The speakers can and should be involved in the documentation and study of their own languages, and not just be considered "sources of data". An example of this involvement, which resulted in the 2nd place in the PERA prize (ERC, 2022) for the *Microcontact* project, is involving the elderly speakers in the recording of the stimuli as well as in establishing which stimuli are socially and culturally acceptable. Just like in the case of India, the ideas of the fieldworker should be put aside as much as possible. No matter what the ideas and needs of the researchers might be, the speaker will always no better what makes them feel at ease and what they would not like. The researcher might be tempted to think that even if the stimuli can trigger an uncomfortable feeling for the informant they can still be used because science is more important. The accuracy of the judgment and of the data produced could be however heavily impacted by a reluctant response. This needs to be avoided, and the only way to do it is involve the community members to co-design the data.

Involving trained community members is a well-known strategy in generative research dealing with minority languages (Cornips & Poletto, 2005). The technique of the youth documenting the old, largely exploited in the realization of the *Microcontact* atlas, has not only scientific but also very strong social repercussions, as it serves as a reminder for the youth that a language is shared and passed from a generation to the next, that they are

[4]for more tips and issues on fieldwork on heritage Italian varieties in America please refer to Andriani, D'Alessandro, et al. (2022))

the people in charge of passing on their identity. This realization was very strong in the *Microcontact* project, and is something that more generative linguists should consider.

3.4 Heritage Germanic in the US

Finally, we turn to field work focusing on Heritage Germanic languages in the US. The documentation and study of this group of languages has been the focus of linguistic research since the mid-19th century, exemplified by research on Danish (Grøngaard Jeppesen, 2005), Dutch (Marte & Smits, 1989; Smits, 1993; Webber, 2011), German (Gilbert, 1971), Icelandic (Arnbjörnsdóttir, 2006), Norwegian (Haugen, 1969), and Swedish (Andreen, 1900; Bergin, 1909; Nelson, 1943). The interested reader is directed to the following sources for more a more detailed overview of the history and current status of research on these varieties: Johannessen and Putnam (2020), Johannessen and Salmons (2021), and Putnam (to appear). For our current purposes, we will home in on the challenges of working with these communities in connection with notions of ethical research.

Putnam (to appear) makes an important distinction between two existing groups of Heritage Germanic speakers. He classifies one group as **Moribund Heritage Language speakers (MHLs)** and the other as **Diglossic Heritage Language speakers (DHLs)**:

- **Moribund Heritage Language speakers (MHLs):** Speakers of a heritage language of which the eldest or penultimate generation of speakers represent the final highly proficient speakers of the language. The heritage language is endangered, and in most cases moribund.

- **Diglossic Heritage Language speakers (DHLs):** Speakers of a heritage language that continues to be transmitted as the L1 to children within the community. Most speakers are diglossic bi/multilinguals, possessing high proficiency in the sociolinguitically-dominant L2 of mainstream society, especially in particular domains.

The vast majority of research that has been carried out in Heritage Germanic languages-speaking communities has to date focused primarily on MHLs. Unfortunately, the number of remaining proficient and healthy speakers in these communities has significantly dwindled over the past two decades to the extent that most of these varieties will be extinct by 2050 (if not sooner). In order to conduct valuable, and ethical, research in these communities, linguists must be keenly aware of the unique and fragile state of these moribund languages.

Finding ways to gain trust, either through the inclusion of proficient speakers of the language or through trusted individuals (such as close family (e.g., children or grandchildren) or assistants (e.g., workers and directors at a health care facility)) is a critical first step. Second, when conducting fieldwork, whenever possible, seek oral rather than written consent to participate in any oral interviews or on-line tasks. Written forms are often the source of confusion and can lead to distrust between the informant and the linguist, and, in the worst case scenario, can result in a rejection to participate in interviews and tasks. As noted in the previous section (§3.2), similar care and precautions must be carried out to ensure that any sorts of experimental tasks or conversation prompts should be age appropriate and accompanied by enough trials to eliminate potential task effects or uncomfortable situations.

Although the research on MHLs and their speakers has primarily shifted towards language documentation and generative/formal studies based predominantly on corpora and language production, the second group of speakers, namely, the DHLs, present unique and exciting opportunities, and equally daunting ethical challenges with respect to linguistic research. The vast majority of DHLs are affiliated with conservative Anabaptist (Christian) groups, such as the Amish, Hutterites, and Old Order Mennonites. Although there is a healthy research tradition of language documentation, lexicography, and in this communities, many of them do not allow research to record them, based primarily on their religious beliefs in connection with the sin of recordings making a 'graven image' of them and their identity. The relationship that these groups have with modern technology is very fluid (Ems, 2022; Kraybill, Johnson-Weiner, & Nolt, 2013), hence it is unwise to purport rigid expectations for 'all' Amish or 'all' Mennonites. Some will allow themselves to be recorded, and others will also participate in on-line experiments.[5] Experimental tasks should be 'ecologically valid' for these groups in order to address concerns of ethical practice and to encourage participation. For example, in their research on complex *wh*-questions, i.e., those in which the filler (*wh*-item) and the associated gap span the distance of more than one clause, Hopp, Putnam, and Vosburg (2019) designed and implemented story-telling tasks that are commonly used in L1-acquisition research rather than an experiment that would have been strictly computer-based.[6] To be clear, longstanding successful experimental

[5] Anecdotally, most conservative Anabaptists are more prone to participate in experiments if they can be ensured that their personal information (such as their names) are in no way associated with results.

[6] Another pitfall avoided by story-telling tasks is the problem of *yes*-bias associated with acceptability judgment tasks in heritage populations (Polinsky, 2018b: 97).

methods such as the Wug-test can be used in these groups, so long as (1) informed consent is sought and approved, (2) the design items are culturally relevant and appropriate, and (3) the administration of the experiment is delivered auditorily. The theoretical analysis of plural allomorphy in Pennsylvania Dutch carried out by Fisher, Natvig, Pretorius, Putnam, and Schuhmann (2022) signals that important theoretical work can be successfully carried out in DHL-speaking communities when ethical boundaries are factored into the design and implementation of the study.

4 Conclusions

Generative linguistic research on non-standard and heritage languages continues to make lasting contributions to multiple (sub)fields of linguistic inquiry. It is paramount to emphasize the bidirectional influence and benefits that this research has taken, because the unique conditions under which these languages are acquired and maintained throughout the lifespan have the chance to shape existing theoretical models and beyond. For example, the algorithms and restrictive nature of generative models pose serious challenges to tradition notions such as *language death* as this term pertains to individual grammars (Bousquette & Putnam, 2020).

As recently discussed by D'Alessandro et al. (2021), there are indeed a number of challenges, including ones that center on the ethics of experimental design and implementation among these groups, that we collectively face when carrying out (generative) linguistic research with these speakers. In the pages above, we expanded up their initial discussion, highlighting areas that require more care and caution moving forward. In summary, by keeping the ethical concerns of research design and implementation at the forefront, we can ensure that research on non-standard, heritage, and endangered languages can move forward. This is certainly excellent news for theoretical linguists like us, who view these findings and the contributions that may make to theory-shaping efforts as a true treasure.

References

Abby, A. (2001). *A manual of linguistic field work and indian language structures.* Lincom Europa.

Abby, A. (2020). A sixth language family of india: Great andamanese, its historical status and salient present-day features. In *The dynamics of language. plenary and focus lectures from the 20th interanational congress of linguists.* UCT Press.

Ahmad, R. (2010). How do i know you are not a cbi agent?: Examining the identity of researcher in sociolinguistic fieldwork. In I. Hasnain & S. Chaudhary (Eds.), *Problematizing language studies: Cultural, theoretical, and applied perspectives* (p. 426-435). Aakar Books.

Andreen, G. (1900). *Det svenska språket i Amerika.* Stockholm: Bonnier.

Andriani, L., Casalicchio, J., Ciconte, F. M., D'Alessandro, R., Frasson, A., van Osch, B., ... Terenghi, S. (2022). Documenting italo-romance minority languages in the americas. problems and tentative solutions. In A. Nevins & M. Cole (Eds.), *Contemporary research in minority and diaspora languages of europe.* Language Science Press.

Andriani, L., D'Alessandro, R., Frasson, A., van Osch, B., Sorgini, L., & Terenghi, S. (2022). Adding the microdimension to the study of language change in contact. Three case studies. *Glossa: a journal of general linguistics*, *7*(1), 1–36. doi: 10.16995/glossa.5748

Arnbjörnsdóttir, B. (2006). *North American Icelandic: The life of a language.* Winnipeg: University of Manitoba Press.

Barbiers, S., Koeneman, O., Lekakou, M., & van der Ham, M. (Eds.). (2008). *Microvariation in syntactic doubling.* Leiden: Brill.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3-4), 129–181.

Bergin, A. (1909). *Lindsborg. En svensk-amerikansk kulturbild från mellerstra Kansas.* Rock Island, IL: Augustanda Book Concern.

Berwick, R. C., & Stabler, E. P. (Eds.). (2019). *Minimalist parsing.* Oxford: Oxford University Press.

Boecks, C., & Hornstein, N. (2010). The varying aims of linguistic theory. In J. Franck & J. Bricmont (Eds.), *Chomsky notebook* (p. 115-141). Columbia University Press. doi:

https://doi.org/10.7312/bric14474-006

Borsley, R., & Börjars, K. (Eds.). (2011). *Non-transformational syntax: Formal and explicit models of grammar*. London: Wiley-Blackwell.

Bousquette, J., & Putnam, M. T. (2020). Reassessing language death: Evidence from moribund grammars. *Language Learning*, *70*(S1), 188–225.

Chandra, P. (2014). English language teaching in indian madrasas: Accommodation or complicity? In *Proceedings of the 7th free linguistics conference.* Hong Kong Polytechnic University.

Chandra, P. (2019). Urdu: A minority language. In R. Kumar & O. Prakash (Eds.), *Language identity and contemporary society.* Cambridge Scholars Publishing.

Chelliah, S. L., & Reuse, W. J. D. (2011). *Handout of descriptive linguistic fieldwork*. New York: Springer.

Chomsky, N. (1954). Review of eliezer rieger, modern hebrew. *Language*, *1*(30), 180-181.

Chomsky, N. (1957). *Syntactic structures*. Mouton: The Hague.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: Mass: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N., & Schützenberger, M. (1963). The algebraic theory of context-free languages*. In P. Braffort & D. Hirschberg (Eds.), *Computer programming and formal systems* (Vol. 35, p. 118-161). Elsevier. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0049237X08720238` doi: https://doi.org/10.1016/S0049-237X(08)72023-8

Cornips, L., & Poletto, C. (2005). On standardising syntactic elicitation techniques. part i. *Lingua*, *115*(7), 939–957. (Reporting year: 2005 Metis note: Engels, gedrukt MI$_S$AND) doi: 10.1016/j.lingua.2003.11.004

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. London/Delhi: Sage Publications.

Cummins, J. (2005). A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*, *89*(4), 585–592.

D'Alessandro, R. (2019). The achievements of generative grammar. *Catalan Journal of Linguistics*, 7-26.

D'Alessandro, R., Natvig, D., & Putnam, M. T. (2021). Addressing challenges in formal

research on moribund heritage languages: A path forward. *Frontiers in Psychology*, *12*(700126), 1-7.

Debnath, S. (Ed.). (2007). *Social and political tensions in north bengal (since 1947*. Siligur: N.L. Publishers.

Dryer, M. (2019). Language names and nonlinguists: A response to haspelmath. *Language Documentation Conservation*(13), 580-585. Retrieved from `http://hdl.handle.net/10125/24900`

Ems, L. (2022). *Virtually Amish: Preserving community at the Internet's margins*. Cambridge, MA: MIT Press.

ERC. (2022). *Erc public engagement research award 2022.*

Estado do Rio Grande do Sul. (2009). *Lei nº 13.178, de 10 de junho de 2009.*

European Union. (2018). *The general data protection regulation*. European Parliament. doi: doi/10.2811/538633

Fisher, R., Natvig, D., Pretorius, E., Putnam, M. T., & Schuhmann, K. (2022). Why is inflectional morphology difficult to borrow? – Distributing and lexicalizing plural allomorphy in Pennsylvania Dutch. *Languages*, *7*(86), 1–25.

Fukui, N. (1986). *A theory of category projections and its applications* (Unpublished doctoral dissertation). MIT.

Gilbert, G. G. (Ed.). (1971). *The German language in America: A symposium*. Austin: University of Texas Press.

Grøngaard Jeppesen, T. (2005). *Danske i USA 1850-2000. En demografisk, social og kulturgeografisk undersøgelse af de danske immigranter og deres efterkommere*. Odense: Odense Bys Museer.

Haspelmath, M. (2017). Some principles for language names. *Language Documentation Conservation*(11), 81-93. Retrieved from `http://hdl.handle.net/10125/24725`

Haugen, E. (1969). *The Norwegian language in America: A study in bilingual behavior*. Bloomington, IN: Indiana University Press.

Haugen, E. (1997). Language standardization. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook* (p. 341-352). Macmillan.

Hopp, H., Putnam, M. T., & Vosburg, N. (2019). Derivational complexity vs. transfer effects. long-distance *wh*-movement in heritage and L2 grammars. *Linguistic Approaches to Bilingualism*, *9*(3), 341–375.

Huang, C.-T. J. (1982). *Logical relations in chinese and the theory of grammar* (Unpub-

lished doctoral dissertation). MIT.

Inventário Nacional da Diversidade Linguística (INDL), Decreto 7387/2010. (2014).

Jespersen, O. (1925). *Mankind, nation and individual from a linguistic point of view.* Institutet for Sammenlignende Kulturforskning.

Johannessen, J. B., & Putnam, M. T. (2020). Heritage Germanic Languages in North America. In M. T. Putnam & B. R. Page (Eds.), *The Cambridge handbook of Germanic Linguistics* (pp. 783–806). Cambridge: Cambridge University Press.

Johannessen, J. B., & Salmons, J. C. (2021). Germanic heritage varieties in the Americans: Social and linguistic perspectives. In S. Montrul & M. Polinsky (Eds.), *The Cambridge handbook of Heritage languages and linguistics* (pp. 252–276). Cambridge: Cambridge University Press.

Kraybill, D. B., Johnson-Weiner, K. M., & Nolt, S. M. (2013). Technology. In D. B. Kraybill, K. M. Johnson-Weiner, & S. M. Nolt (Eds.), *The Amish* (pp. 312–355). Baltimore: Johns Hopkins.

Krishnamurti, B. (2003). *The dravidian languages.* Cambridge: Cambridge University Press.

Leivada, E., D'Alessandro, R., & Grohmann, K. (2019). Eliciting big data from small, young, or non-standard languages: 10 experimental challenges. *Front. Psychol*, *10:313*. doi: 10.3389/fpsyg.2019.00313

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal Grammar and statistical learning. *Annual Review of Linguistics*, *1*(1), 333–353.

Marte, J., & Smits, C. (1989). Morphological erosion in American Dutch. In N. Boretzky, W. Enninger, & T. Stolz (Eds.), *Vielfalt der Kontakte* (pp. 37–65). Bochum: Universitätsverlag Brockmeyer.

Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard english.* London: Routledge.

Nandi, R. (2014). Spectacles of ethnographic and historical imaginations: Kamatapur movement and the rajbanshi quest to rediscover their past and selves. *History and Anthropology*, *25*(5), 571-591.

Nelson, H. (1943). *The Swedes and the Swedish settlements in North America.* Lund: C.W.K. Gleerup.

Nevins, A. (2022). *When minoritized languages change linguistic theory.* Cambridge: Cambridge University Press. doi: 10.1017/9781009029889

Polinsky, M. (2018a). *Heritage languages and their speakers.* Cambridge: Cambridge University Press.

Polinsky, M. (2018b). *Heritage languages and their speakers.* Cambridge: Cambridge University Press.

Polinsky, M., & Putnam, M. T. (Eds.). (2023). *Formal approaches to complexity in heritage language grammars.* Berlin: Language Science Press.

Postal, P. M. (1966). On so-called pronouns in english. In F. Dinneen (Ed.), *Nineteenth monograph on language and linguistics.* Georgetown University Press.

Putnam, M. T. (to appear). Germanic languages in contact in North America. In S. Kürschner, A. Dammel, J. Hoeckstra, S. Höder, O. Schallert, & F. Van de Velde (Eds.), *The Oxford Encyclopedia of Germanic Linguistics.* Oxford: Oxford University Press.

Reiger, E. (1953). *Modern hebrew.* New York: Philosophical Library.

Smits, C. (1993). Resistance to erosion in American Dutch inflection. In G. Booij & J. Van Marle (Eds.), *Yearbook of Morphology 1993* (pp. 155–184). Dordrecht: Springer.

Webber, P. E. (2011). *Pella Dutch: Portait of a language in an Iowa community.* Iowa City: University of Iowa Press.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language.* Cambridge, MA: MIT Press.