Fieldwork Challenges in Diaspora Communities: Arabic Speakers in Germany and Pennsylvania German Speakers in Canada*

Miriam Neuhausen
University of Heidelberg

Muhadj Adnan
University of Bayreuth

This paper explores the complexities of conducting fieldwork and processing the collected data in the following two diaspora communities: Arabic speakers in Germany and Pennsylvania German speakers in Canada. Pointing out differences and similarities in working with the two socially and linguistically heterogeneous groups, we highlight the need for culturally sensitive research methodologies that acknowledge and adapt to the respective sociocultural and linguistic settings. If we do not acquire and implement community-specific sociocultural knowledge before, during and after the data collection process, important findings may not come to the surface. Providing examples from our fieldwork, we show how the researcher's identity and interaction dynamics with the respective community have an immediate impact on the data collection and analysis processes. We attempt to offer practical recommendations and encourage linguists and social scientists to conduct fieldwork in similarly complex contexts to advance sociolinguistic theory in multilingual minoritised communities.

1   Introduction

Fieldwork is the attempt to collect data in its natural setting (Bowern 2015: 2). Traditionally, it is defined as a "long-term, often uninterrupted, involvement

---

with the people, community or practices being studied" (Heller, Pietikäinen & Pujolar 2018: 73). In sociolinguistics, ethnolinguistic fieldwork, i.e. including ethnographic observations, helps researchers gain a better understanding of the prevailing social dynamics and observed linguistic variation. It involves many jobs on behalf of the fieldworker: collecting the data, communicating with the community, taking care of ethic formalities, making anthropological observations and reflecting one's own behaviour (Bowern 2015: 3–4). Scholars have long been aware that fieldwork observation cannot be objective; rather, the fieldworker has an inevitable impact on the data collection process and can rarely (or never) become an objective and detached observer (Rosaldo 1989: 169).

To accompany fieldworkers from the field to the analysis, linguists have developed a range of highly valuable and relevant research guides, such as Schilling (2013), Bowern (2015), Meyerhoff, Schleef & MacKenzie (2015) and Meakins, Green & Turpin (2018). Fieldwork in lesser-studied communities is often culturally biased by previous findings based on Western, Educated, Industrialised, Rich and Democratic (WEIRD) communities, which is addressed by an increasing number of researchers, such as Mansfield & Stanford (2017), Heller, Pietikäinen & Pujolar (2018) and Werner & Klimiuk (2019).

The present paper illustrates that methodological approaches are often shaped by researcher biases, which is also replicated in the development and advancement of language technology and research tools for diasporic languages. We hope that this article, which is based on our fieldwork experience and insights in two diasporic communities, provides ideas and strategies that encourage more linguists to explore the diasporic field. We highlight some practical challenges we faced when dealing with diaspora communities located in WEIRD societies and offer a comparative perspective on the complexities of sociolinguistic research within highly diverse sociocultural and linguistic settings. We address methodological challenges that we encountered while doing fieldwork with two diaspora communities, namely Arabic speakers in Germany (Adnan) and Pennsylvania German speakers in Canada (Neuhausen).

For groups to be traditionally defined as diasporas, they need to share the same national, ethnic and religious background and have migrated "en masse at a certain historical moment" in response to a crisis (Barontini & Wagner 2020: 246), which is often traumatic (Cohen 2008: 180). The term, originally referring to Jewish dispersion, does not only include the geographic condition but also the community's "spiritual longing for the land from which they have been exiled" (MacCabe & Yanacek 2018: 95). In this respect, diasporic communities differ from other lesser-studied communities.

The diasporic movements of both the Arabic and the Pennsylvania German community have been prompted by such crises. The political turmoil in

Iraq and Syria led Arabic speakers to leave their homelands and many refugees have undertaken perilous journeys to Europe (Sirin & Rogers-Sirin 2015). In the summer of 2015, Germany recorded the highest number of asylum applicants ever (Crul et al. 2017). In contrast, oppression during the 17th-century religious persecution in Europe caused the dispersion of the Pennsylvania Germans from Europe to North America, including many other religious groups (Raith 1982: 9–10; Lepore 2018: 50). While the Arabic speakers have recently arrived in Germany, the Pennsylvania Germans have maintained their diasporic status in North America since the late 18th century (Fretz 1989: 5; Burridge 2002: 204). Both communities express longing towards the "home country"; while most of the Arabic speakers in Germany have experienced life there themselves, most Mennonites in Canada have never visited Germany or Switzerland, as becomes evident in examples (1) and (2).[1]

(1)   *yaʕni ṭabʕan min aððakkar iʃ-ʃaɣḷaat yaʕni naʃaaṭaatna b-baɣdaad, yaʕni tinzil id-damʕa, ma tigdar, manṭaqa yaʕni mawluud biiha uw mitʕawwid ʕaleeha*
        'I mean of course, when I remember the things, I mean our activities in Baghdad, I mean, tears fall, you can't, it's an area you're born in and used to.' (Mahir)

(2)   I'm too old now but I've often wish I would get a chance to go to Germany, just see how far I can get [with my German]. (Cleon)

As a result, in both cases, the communities' identities are strongly religious and the reason for their dispersion is war or religious persecution.

In diaspora communities, group identity is often strongly maintained over generations and shaped and determined by the collective memory and myths about their homelands. This is manifested in cultural practices that are distinct from the surrounding communities. According to Hall (1990: 235), such perceived cultural differences are needed for a community to be considered diasporic; the notion *diaspora* is defined "by the recognition of a necessary heterogeneity and diversity; by a conception of 'identity' which lives with and through, not despite, difference; by *hybridity*" (Hall 1990: 235). In other words, social difference from the surrounding mainstream society is a key aspect in

---

[1] All names have been changed. To increase legibility, we inserted commas and full stops for pauses. Words indicated in square brackets may provide additional information. Extracts may be shortened and restricted to relevant context. Arabic was transcribed based on the International Phonetic Alphabet. Double vowels represent long vowels. All translations are our own.

diasporic cultures. This also applies to the two communities; Arabic speakers in Germany organise their community around religious events that differ from German mainstream religious customs, such as Ramadan and big family celebrations; and the Pennsylvania Germans maintain their cultural distinctiveness by objecting to modern technology, e.g., cars, mobile phones and the internet. In addition to cultural difference, both communities also display linguistic difference by maintaining their heritage languages.

In the following, we first define the notion of fieldwork and describe the two settings, before we assess the effect of researcher identity on interaction dynamics with the community. We then outline challenges before, during and after fieldwork. Finally, we identify common ground and differences in working with the two communities in the hope that future researchers working with multilingual diasporic communities benefit from our experiences.

2   The fields

2.1   *Arabic speakers in Germany*

In this section, we describe the two fields in which we conducted fieldwork. The first fieldwork to be discussed was conducted in two Arabic-speaking communities located in Bavaria, southern Germany, in 2020–21. Five years prior to that, Germany faced a notable influx of asylum seekers, with Syrians and Iraqis among the top nationalities seeking asylum, particularly in Bavaria and North Rhine Westphalia (Bundesamt für Migration und Flüchtlinge [BAMF]). The Syrian community in Germany arrived during several phases of forced migration motivated by political turmoil, particularly following the civil war in 2011.[2] This violent conflict led to a second wave of displacement, establishing Germany as a major destination for Syrians seeking refuge in Europe (Ragab, Rahmeier & Siegel 2017: 26).[3] Similarly, Iraqi migration to Germany did not start in 2016 but goes back to the mid-1960s. Initially involving labour workers, Iraqi migration to Germany reached its peak during the 2003 Iraq War and the rise of the Islamic State (Candan 2017: 8–9). Ongoing unrest and conflict in Iraq have led to continued waves of displacement.

Linguistically, Iraqi and Syrian migrants bring a diversity of dialects to Germany, namely the linguistic varieties of Iraqi and Syrian Arabic, displaying variation both within and between the groups. The vast majority of speakers of

---

[2] The European Commission (n.d.) defines the term *forced migration* as a "migratory movement in which an element of coercion exists, including threats to life and livelihood, whether arising from natural or man-made causes".

[3] Even at the time of writing, Germany was still the preferred destination for Syrians seeking refuge (Deutschlandfunk 2024).

both groups also understands Modern Standard Arabic which is used in writing and in most formal speech throughout the Arab world, but is not acquired as a first language (Ferguson 1959; Boussofara-Omar 2006). Although Modern Standard Arabic is widely understood, it is not typically used in everyday conversations. Local dialects vary significantly within countries and regions and are not always easily understood by those who speak a different dialect. The dialects spoken in the capitals Damascus (Syria) and Baghdad (Iraq) are widely understood, used in the media and associated with prestige (Procházka 2020: 84).[4]

Concerning the status of Arabic in Germany, the community appears to follow in other Arabic communities' footsteps: Feeling uncertain in Arabic, some young speakers already begin to substitute Arabic expressions with constructions similar to German. As a result, they may increasingly become dominant in German. The language situation draws parallels to, for example, Boumans' (2006) and Boumans & de Ruiter's (2012) studies, who found that among themselves, Moroccan young people in the Netherlands mostly speak Dutch, the dominant societal language. Similarly, in Adnan's data, young Iraqis and Syrians display an overproportionally high use of analytic genitive forms, such as Iraqi *maal*, as in example (3), and Syrian *tabaʕ*, as in example (4), instead of synthetic genitive forms. In this function, the items are comparable to a preposition, facilitating the periphrastic genitive, similar to English 'of' – or German 'von' in this case (see also Rosenbach 2002):

(3)   *iħiṭṭa bi-l-firin <u>maal</u> il-<u>bakeraay</u>*
        'He puts it in the oven <u>of</u> the bakery.' (Bilal)

(4)   *iṭṭarit hiyye tsaafir ʕa-l-urlaaub <u>tabaʕ</u> iṭaalia*
        lit. 'She wanted to go on the vacation <u>of</u> Italy.' (Ashraf)

Arabic dialects typically include a genitive exponent that can replace the synthetic genitive construction (also called *Iḍaafa*). This construction can manifest as the "classical" *Iḍaafa*, where the possessed and possessor are juxtaposed, or as the analytic genitive, where the two are connected via an

---

[4] Baghdad is by no means a homogenous dialect region. The dialects spoken by different religious and ethnic groups in Baghdad, such as Christians, Jews (when they were still present) and Muslims, show notable differences (see Abu-Haidar 2006 for more information). Given this diversity, the initial research plan was to focus on speakers of Muslim Baghdadi Arabic for reasons of consistency and comparability. Including multiple sub-dialects from the same city would have complicated the analysis and made it challenging to achieve a balanced and representative sample.

independent word. This preference highlights a broader trend towards favouring analytic structures, especially when incorporating lexical items from other languages, such as *bakeraay* 'bakery' and *urlaaub* 'holiday', which is consistent with previous studies by Owens (2005) and Boumans (2006).

In her doctoral thesis, Adnan investigates contact-induced language change within Syrian and Iraqi diaspora communities in Germany. The research focuses on intergenerational linguistic differences and the potential emergence of a common koine between Syrian and Iraqi groups, more specifically Syrian and Iraqi speakers who had lived in Bayreuth and Nuremberg since 2014–15. Based on the analysis of sociolinguistic interviews, the findings indicate a high adherence to established dialectal norms in individual interviews, while group conversations reveal pre-koineisation phenomena. Whilst Iraqi speakers tend to adapt Syrian features, Syrians display both converging and diverging patterns. While Iraqi speakers display multiple features to accommodate to Syrian Arabic, Syrian speakers only use one feature to accommodate to Iraqi speakers and additionally show diverging behaviour. Iraqi speakers reduce typical Iraqi features like /č/ to palatalised /k/, as in /čam/ 'how much' being realised as /kam/. Other Iraqi features that are reduced are *da-* (pragmatic imperfect marker) and *wiyya* 'with' (*wiyya uṃṃi* vs. *maʕa uṃṃi* 'with my mother'). By contrast, Syrians increase the use of [h] of pronominal suffixes as in Iraqi, e.g. *beet-a* becomes *beet-ha* 'her house'. For the purpose of linguistic divergence, they use, among others, the marker *ʕam-* 'immediate present', as in *ʕam byiʕməl Ausbildung* 'he is doing a vocational training'.

2.2   *Pennsylvania German speakers in Canada*

The second fieldwork to be discussed took place in a community of Pennsylvania German-speaking Mennonites in southern Ontario, Canada.[5] In 2018–19, Neuhausen stayed there for five months and conducted sociolinguistic interviews with Mennonite communities ranging from the very conservative bilingual *horse and buggy* Old Order Mennonites to the monolingual mainstream group of the Conference Mennonites who only speak English.[6]

---

[5] In the Mennonite community, the terms (*Pennsylvania*) *Dutch*, *Deitsh* and *German* are used interchangeably with *Pennsylvania German*. For the sake of consistency, we use only the latter.

[6] Despite their shared nomenclature, Mennonite communities are themselves heterogenous. In Ontario, the Mennonite community constitutes four branches: the very conservative Old Order Mennonites, who wear traditional clothing and reject modern technology, cars and the internet; the Markham Mennonites, who have restricted access to modern technology and drive only black cars; the Conservative Mennonites, who use modern technology, drive any-

Unlike the vast majority of immigrants, the diaspora community of the Old Order Mennonites have maintained their first language Pennsylvania German since they left Europe in the early 18th century (Draper 2010: 216). At the time, they left Europe for freedom of religion which was promised in Pennsylvania (Frantz 2017: 131–2).

The trauma of their early ancestors is still very tangible in the community. This becomes apparent, for instance, in the school curriculum, which strongly focuses on the religious persecution experienced by the Mennonites' ancestors in Switzerland and Germany, or the fact that nearly every household owns a copy of the *Martyrs Mirror* (Van Bragt [1660] 1998), documenting the sufferings of the early Mennonites during the religious persecution in Europe. Holding on to such books serves to maintain the collective memory of the early Mennonites' suffering (Assmann & Assmann 1994), and simultaneously emphasises the Mennonites' desire to remain separate from "the world".

To this day, the traditional Mennonite communities still consciously separate themselves from the government and mainstream society, for example by wearing traditional clothes, driving horse and buggy and rejecting modern technology (cars, internet, computer, cell phones). The degree of bilingualism in the different communities further indexes this social continuum; while the Old Order Mennonites and Markham Mennonites still speak Pennsylvania German as a first language, the Conservative and the Conference Mennonites have shifted to English.

The local area in southern Ontario constitutes a German pocket, with German place names like *Baden*, *Mannheim*, *Schindelsteddle*, *Bamberg*, *Heidelberg* and *New Hamburg*. At the time of the fieldwork, German was the most common immigrant language in the area of Waterloo, Kitchener and Cambridge and accounted for 9.5% of all local immigrant languages (Statistics Canada 2017).[7] While different varieties of European German are spoken by recent immigrants from Europe, Pennsylvania German is spoken only by conservative Mennonite and Amish groups.

Pennsylvania German is usually acquired as the first language and spoken at home, in church and in the community. English is acquired in school as a second language in this community (Burridge 1998: 85–86); outside the

---

coloured cars, but still wear plain clothes; and the Conference Mennonites, who socioculturally fully blend in with mainstream Canadian society.

[7] Interestingly, in the updated 2021 census (Statistics Canada 2022), which focuses on Waterloo only (instead of the wider region including Kitchener and Cambridge), the number of German speakers has been surpassed by Arabic speakers. This is likely a result of many members of the Arabic community fleeing from war and political turmoil in 2015 and 2016, many of whom first arrived in Europe and then made their way from there to Canada.

community, it serves as the language of communication and within the community for all written communication (as Pennsylvania German is not written). Today, members of the traditional communities are usually bilingual and acquire the two languages successively. Members who leave the conservative Mennonite and Amish communities usually stop using it and may eventually lose the language as soon as they fully integrate into a more modern group. For instance, during the conversation with Neuhausen, Timothy realises that his father Lloyd, who left the Old Order Mennonites and who used to speak Pennsylvania German all of his life, has now, in his 90s, trouble producing an utterance in Pennsylvania German. Timothy expresses his surprise in the following example:

(5)   I'm a little surprised at how little he, how hard he has to work at it to, to speak it 'cause, oh my goodness! Like most of my life it just rattled off his tongue. […] Oh, he would never have dreamed about speaking English to them, like that would've been unheard of, it woulda been like really *vershnapt* or stuck up. (Timothy)

In her doctoral thesis, Neuhausen investigated what role identity plays on the sociophonetic, lexicogrammatical and pragmatic level of Mennonite English. On the sociophonetic level, she found a statistically significant correlation between the adoption of marked linguistic variants, i.e. those of Pennsylvania German origin or older stages of English, and the speakers' sociocultural relationship to the traditional group of the Old Order Mennonites. Speakers who grew up in the Old Order group and left the community employ language to index their different sociocultural standings. They produce sociophonetic realisations that are drastically different from all other groups. Moreover, she found that Mennonite English constitutes two linguistic repertoires (Benor 2010), one based on religion, to which speakers of all Mennonite affiliations have access, and the other one based on ethnicity, to which only speakers have access who grew up in the Old Order community.

      These short descriptions of the two communities illustrate that despite obvious cultural and linguistic differences, the communities also share some common ground and can be classified as diasporic communities with strong religious group identities. Both communities exhibit languages and cultures that are distinct from mainstream society – and at the same time are linguistically and socially inherently heterogeneous with a high degree of intra-group variation.

2.3   Similarities

Both languages, Arabic and Pennsylvania German, are clearly shaped by the diasporic contact settings and display a high amount of code-switching. Particularly younger Iraqi and Syrian speakers increasingly incorporate German loanwords, as can be seen in examples (6) and (7), where speakers use the German nouns *tsaayt* ('Zeit', 'time'), *ʃuula* ('Schule', 'school') and the connector *sonst* ('otherwise') within Arabic utterances. Example (8) displays the item *testaat* for German *Test* 'test' with the Arabic plural ending *-aat*. Code-switching is by no means restricted to Arabic, but also happens in German.

(6)   *lamma ṣaar il-<u>koroona tsaayt</u> lamma iʃtiɣaḷna, arbaʕ asʃhur aani ʃtiɣaḷit […] ma činit aruuḥ li-ʃ-<u>ʃuula</u> wa la ʃii.*
'When the <u>corona time</u> happened, when we worked, for four months I worked, […] I didn't go to <u>school</u> or anything else.'

(7)   *waʔət il aḥiss innu biddi ana laḥ iži uw isʔal, <u>sonst</u> ana tamaam.*
'Whenever I feel like it, I will come and ask; <u>otherwise</u>, I'm fine.'

(8)   *hassa ʕindi tlaθ <u>testaat</u> laazim axaḷḷiṣha.*
'Now I have three <u>tests</u> that I have to finish.'

Also Pennsylvania German displays contact features from the mainstream variety, i.e. English, as illustrated in examples (9)–(10).[8] The examples show that Pennsylvania German has its own grammatical system distinct from English but also displays influences from English. While some constructions are borrowed from English in their entire forms, such as *how come* and *refused*, others are integrated into Pennsylvania German morphosyntax, as the phrasal verb *turn out*, where the preposition is translated to Pennsylvania German *aus* 'out' and attached to the English verb *turn*. This innovation is integrated into Pennsylvania German morphosyntax, signalled by the third person marker *-t*: <u>*austurnt*</u> ['turn<u>s</u> out'].

---

[8] The Mennonites' ancestors likely originally spoke Upper German and Swiss German (Raith 1996: 317), but shifted to Palatinate German within one generation when they fled Zurich and sought refuge in the Palatinate (Gratz & Geiser 1973 ctd. in Raith 1996: 317), a region in the southwest of Germany. The shift to Palatinate German took place before the migration to Pennsylvania and provided the basis for what would later be known as Pennsylvania German (Raith 1996: 316). Thus, despite its Swiss German roots, Pennsylvania German displays many similarities to Palatinate German.

(9)   *Un sie het gsagt, "how come het sie refused?"* (Rebekah)
        'And she said, "how come she refused?"'

(10)   *Well, we'll find out wie sell alles austurnt.* (Selema)
        'Well, we'll find out how this all turns out.'

In English, code-switches to Pennsylvania German are restricted to culturally salient lexical items that cannot be translated to English, as shown in example (11). By contrast, the code-switching of other words, i.e. words for which an English concept exists, may be negatively evaluated, as illustrated in example (12). Here, Louise explains a little boy's use of such a lexical item in Pennsylvania German in an otherwise English utterance with confusion on behalf of the boy:

(11)   I'm *glustish*, for a piece of chocolate cake; "I want a piece of chocolate cake" but it's much much more than that when you're *glustish*. Then you have a real desire for it and […] people that have grown up with the Pennsylvania Dutch, sometimes when they're talking about a word like *glustish*, they'll say that yet because it just describes it so much better. (Lou)

(12)   *De glee buva het gmeent* ['the little boy said'], "I have to go home and get my *handshing*." See, "have to go home and get my mittens" is English but *handshing* is German but it's … well, he was a little *verhoodelt* ['confused'], I guess. (Louise)

As discussed in this section, the minority languages spoken in both diasporic settings display a wide range of code-switching. While this applies also to the dominant societal language in the Arabic-German setting, this is not the case in the Pennsylvania German-Canadian context. While code-switching is common and widely accepted in Pennsylvania German, Mennonite English is strongly influenced by prevailing prescriptive language ideologies and thus features only a limited set of culturally salient code-switches.

3   Researcher positionality and identity

We now turn to the researcher's impact on the community and data collection processes. Data are not "'out there' waiting to be sampled" in an objective manner but represent the "product of a consensus on what particular experiences or materials can be used for analysis" (Heller, Pietikäinen & Pujolar 2018: 74;

see also Bowern 2015: 2). Researchers thus actively influence the data collection and generation at all times; they constantly make decisions on what counts as data, how to approach them and how to interpret them. After all, it is the fieldworker's responsibility to not only observe the "things going on out there", but also to select interesting aspects and to choose an angle from which to approach these (Heller, Pietikäinen & Pujolar 2018: 81).

But fieldwork is also closely related to how researchers present themselves in the field. For instance, Henry (2003: 231) defines fieldwork as a "process of representing oneself" and compares it to everyday processes outside the field. In this sense, fieldwork does not only happen in interaction with the community of interest, but also in fieldworkers' private lives, e.g. how they represent themselves and interact with others in their own homes.

In sociolinguistics, it is considered common knowledge that informants can change their way of speaking when performing different identities (as for example shown by Podesva 2007). What is often not mentioned is the fact that the fieldworker, too, displays different identities in different social settings. All fieldworkers display what Rosaldo (1989: 194) calls "multiple identities", with "[m]ore a busy intersection through which multiple identities crisscross than a unified coherent self". As a result, in different contexts, fieldworkers may be forced to accept different facets of their identity (Narayan 1993: 676).

How a fieldworker is perceived also has an impact on the produced data. The following examples illustrate how the speakers' perception of Neuhausen prompted lexical and stylistic variation as well as code-mixing. By some, Neuhausen was perceived as an expert of (Pennsylvania) German in the position to make decisions on the authenticity of some words, as illustrated in example (13). In this example, Mark explains that his teacher of High German used the verb *benootse* 'to use', while he himself would say *use* 'to use'. Despite the teacher being in an officially assigned position of authority, Colleen questions this authority and asks Neuhausen whether this is a word. Note that she does not ask whether this is a word Neuhausen *knows* but she asks whether it *is* a word, putting Neuhausen in a position of higher authority than the quoted teacher of High German. Moreover, in Neuhausen's presence, some speakers felt the urge to speak Pennsylvania German that is "good enough", as illustrated in example (14). Others were conscious of their use of English words in Pennsylvanian German, as shown in example (15), where Glen notices his use of English numbers, immediately translates them into Pennsylvania German and approvingly adds that he knows the words in Pennsylvania German. Neuhausen did not ask Glen to speak in Pennsylvania German; by contrast, knowing that she was interested in the language was sufficient for him to produce as much Pennsylvania German as possible and even commented on that.

(13)   Mark: *Mir soge* <u>use</u> *and un mein teacher, uh, er used* <u>benoots</u>.
         'We say <u>*use*</u> and my teacher, uh, he used <u>*benoots*</u>."

       Colleen [to MN]: Is sell a vart?
                        'Is that a word?'

(14)   *Is des dort Deitsh gut genug fur dich?* (Elon)
       'Is this Deitsh good enough for you?'

(15)   *Mir hen wo ich schon, oh,* <u>nineteen</u> *oder* <u>twenty</u> *jahr alt war –* <u>neinzeh</u>
       oder <u>zwanzig</u> *jahr alt war,* I do know the words! (Glen)
       'We did already when I was, oh, <u>nineteen</u> or <u>twenty</u> years old –
       <u>*neinzeh*</u> 'nineteen' or <u>*zwanzig*</u> 'twenty' years old, I do know the
       words!'

These examples show that certain types of linguistic data may be influenced by
the presence of the researcher, such as word choice, style and code-switches. If
the informant knows that the researcher speaks the minority language, they may
be more likely to switch to that language; if they usually converse, as in the case
of the Mennonites, with non-Mennonites in English, switching into
Pennsylvania German may, however, feel unnatural to them.

       Challenging the concept of researcher objectivity and neutrality, these
examples visualise that the researcher's *multiplex subjectivities* turn the field
into a "site of complex power relations" (Henry 2003: 239). The collected data
are evidently heavily influenced by the fieldworker's self-presentation and
different identity facets, such as the perceived authority in a certain domain, e.g.
speaking German. We argue that this phenomenon goes beyond the observer's
paradox as it interacts with the different identities of the researcher; Neuhausen
was perceived as a German-speaking cultural outsider, a researcher interested in
the Mennonites' languages and an expert of Pennsylvania German.[9] With other
cultural outsiders who may not speak German as a first language, the
Pennsylvania Germans might have been less conscious of their use of
Pennsylvania German. Strikingly, unlike the more modern Mennonites, Old
Order Mennonites did not show any such awareness or consciousness of the
impact of English on their Pennsylvania German; these speakers never
commented on how "English" their Pennsylvania German has become. A reason
may be that they do not evaluate the use of English words negatively. Arguably,
Pennsylvania Germans outside the Old Order community face a greater risk of

---

[9] The *observer's paradox* describes the contradiction that linguists observe the most relevant
speech when speakers are not being observed (Labov 1972: 113).

not being considered authentic speakers of Pennsylvania German and are hence more conscious of their language use – particularly in the presence of a speaker of so-called "real" German. Therefore, Neuhausen, a cultural outsider travelling to Canada from Germany, represented different identities to the Mennonites that prompted different reactions, impacting individual speech productions.

Diaspora communities often feel a strong sense of solidarity with co-ethnic speakers. In these communities, the groups' ancestral homes are often idealised, whereas the relationships with the host society may be troubled. This creates (imagined) closeness to the geographically distant homeland and emotional distance from the local society. We can confirm this for our fieldwork; Adnan as an Arabic speaker of colour and Neuhausen as a white German speaker apparently shared common ground with the respective communities on the linguistic and the ethnic level. This can be observed, for example, in metalinguistic comments. The Pennsylvania German informants often asked Neuhausen whether a Pennsylvania German word was part of her (European) German repertoire, as shown in example (16). Furthermore, she was even asked for her opinion concerning language policies in the public space, as illustrated in example (17).

(16)   *Gallesa, is sell familiar?* (Elon)
         'Suspenders, are you familiar with this [word]?'

(17)   *Wenn noch eppes shons se shopping dus un ich shvatz Deitsh, is sell polite?* (Ada)
         'When somebody else is in the store and I talk Deitsh, is that polite?'

Nonetheless, the line between cultural outsiders and in-group members is not always clear-cut. While "insider" linguists speak the language, have a relationship to the community prior to fieldwork, and easily detect "potential ethical dilemmas", "outsider" linguists are often members of the dominant community that may be "responsible for the marginalisation" of the community of interest (Meakins, Green & Turpin 2018: 4–5). A third group involves the "insider-outsider" linguists, who are also members of a marginalised community but maybe not of this particular one (Meakins, Green & Turpin 2018: 5). Henry (2003) discusses the complexity of being such an "insider-outsider" fieldworker, somewhat in-between a cultural outsider and a member of the community. She grew up in Canada with her parents having moved there from Pakistan. When she did fieldwork as a "diasporic" in India, she expected it to feel more like home. But not having grown up in India, during fieldwork she was not considered a member of the researched community either. Henry emphasises that the researcher's identity and how they are perceived by the community

strongly influences the data collection process and prompts the question how a fieldworker who "exists both on the 'inside' and 'outside'" can appropriately label and name themselves (Henry 2003: 239).

Similar to Henry, Adnan found herself both inside and outside the Arabic community in Germany with a "hybrid identity".[10] Born in Iraq and raised in Germany, she is a member of the Arabic community. Yet, she was perceived as a cultural outsider in some cases. While community members frequently commented on the fact that she looked like them, they sometimes clarified that she did not sound like them and even came up with explanations as to why that was the case. Others admired Adnan's multicultural background, having social and linguistic insights into both cultures, and were even surprised by her Iraqi language skills. Through her cultural background, Adnan had an awareness of both German and Iraqi cultures. Additionally, she had had prior contact with Syrian speakers and was able to share personal experiences concerning the challenges faced by individuals with migration backgrounds living in Germany. At the same time, she also had some insights into the perspectives of Germans on these communities.

We have found it important that fieldworkers acknowledge both their multiplex identities and what relation they have to the community. As a woman pursuing higher education and using modern technology (the recording device), Neuhausen clearly represented a cultural outsider, embodying multiple aspects objected to by the Old Order Mennonite community. However, she also found common ground with them based on her German background, which allowed her to easily acquire Pennsylvania German and converse bilingually with the participants. Adnan, despite being a member of the Arabic community herself, was not fully accepted by the speakers as one of their own. She represented an "insider-outsider" researcher, who shares some similarities with the group of interest, but is still perceived as an outsider. This draws parallels to Liebow's (2003: xxviii) metaphor of the chain-link fence, where you can see each other, walk alongside each other and talk with speakers from another community, but the (sociocultural) barrier remains. This social difference should be acknowledged during both data collection and the analysis, as it might not only affect the collected data but also the interpretation and subsequent presentation of the results.

In this section, we have illustrated the immediate effect the researcher has on the data collection and outlined both our relations to the communities of

---

[10] "Hybrid identities" are considered intercultural, transcultural and multicultural. People with a hybrid identity are often described as having dual or multiple national affiliations, being bicultural or trinational. They are represented as either being caught between two stools or sitting on a third stool (Badawia 2002; Mecheril 2004).

interest. We have found it essential to acknowledge and report potential biases and impacts on the findings and analysis.

4   Before fieldwork

Before the data are collected, researchers can prepare for the fieldwork in different ways. If they are already familiar with the community, they may benefit from previously built relationships. Adnan's relationships with Iraqi and Syrian speakers already began five years prior to her fieldwork when she started working within a project at the service point for refugees and migration at the International Office of the University and built close contact to many Iraqi and Syrian students living in Bayreuth.[11] Having grown up in an Iraqi family, she was also already familiar with the Iraqi culture and language. Similarly, Neuhausen's contact with the community did not only begin with her fieldwork but six years prior, when she first visited and befriended local Mennonites. That way, she was also already somewhat familiar with the Pennsylvania German culture, history and language when she began her fieldwork.

In some communities, fieldworkers cannot rely on previous reports or documentations of the community, their culture and/or their language. In preparation for fieldwork, both authors familiarised themselves with pre-existing studies on the respective communities and sociolinguistic settings. In both cases, at the time, previous studies were not necessarily up to date. Adnan consulted publications dating back to the 1960s, with the most recent ones being already ten years old at the time of fieldwork preparation (Grotzfeld 1965; Ambros 1977; Arnold 1998; Behnstedt 1997; Cowell 1964; Gralla 2006; Yoseph 2012). Studies of Iraqi-Arabic at the time focused to a large extend on the varieties in Baghdad and Mosul (Malaika 1963; Blanc 1964; Van Ess 1978; Abu Haidar 1991; Erwin 2004; Jastrow 2006). Similarly, the majority of previous linguistic research in the Mennonite context addressed Pennsylvania German syntax and was mostly at least thirty years old when Neuhausen began her fieldwork in 2018 (Richter 1969; Costello 1978; Enninger 1984; Huffines 1984; Burridge 1989; 1992; 2002; Van Ness 1990; Keiser 2015; Louden 2016). Research on Mennonite English was even more scarce (Wilson 1948; Huffines 1986; Burridge 1998; Kopp 1999). Moreover, these publications do not describe how initial contact was established and a relationship was built with the community – which may also be due to the community's special interest in not receiving too much attention. As a result, both authors prepared for fieldwork relying on

---

[11] This was in the context of the DAAD project "Integra" to help prepare refugees to study in Germany and the DAAD project "Welcome" to support refugees at universities.

guides for sociolinguistic fieldwork, such as the above-mentioned fieldwork guides.

Additionally, Neuhausen got affiliated with the University of Toronto as a visiting scholar, conducting fieldwork within the frame of Sali Tagliamonte's Dialects of Ontario research project (http://ontariodialects.chass.utoronto.ca). Being affiliated with a local university facilitated access to local libraries and archives, including those outside of Toronto. This was helpful even before fieldwork began, as Neuhausen did not have access to locally published Mennonite books and articles in Germany. Before and during fieldwork, affiliation with the University of Toronto also enabled her to connect with the Variationist Sociolinguistics Laboratory, where she was trained in conducting fieldwork and later also received technical support and had highly valuable exchange with other (more experienced) fieldworkers. Furthermore, being affiliated with a local university meant that Neuhausen had a contact person there for follow-up questions on behalf of the participants, which helped locals to contextualise the research project.

Initial contact with the Mennonite community was established through the "friend of a friend" approach (Milroy 1980: 53) and through cultural "brokers" (Schilling-Estes 2007: 178) – or people who Walt Wolfram informally refers to as "professional stranger handlers" (quoted in Schreier 2013: 25). These people represent the bridge to outside communities; not only are they used to dealing with foreigners on a regular basis, but they are also widely connected and may help the fieldworker to establish contact with a range of diverse speakers.

Importantly, when working with diasporic communities, linguists should reflect on naming practices of contact languages. Some varieties may not have names yet, such as Gurindji Kriol, spoken in northern Australia (Meakins, Green & Turpin 2018: 250); others may have names that were assigned to them by cultural outsiders. Mufwene (2000: 67) criticises some linguists' "self-licence to go around the world baptising some vernaculars 'creoles'" – despite the fact that these "creole" speakers do not even know the word *creole*. Similarly, the English spoken by the Mennonites is often referred to as Pennsylvania German English – as it is shaped by their first language Pennsylvania German (Huffines 1986; Kopp 1999). However, this is not a term used by Mennonites themselves; instead, Neuhausen opts for the notion *Mennonite English*, which is closer to how they would describe it ("Mennonite tongue" or "Mennonite accent").

Data storage and processing is another concern linguists have to reflect on. In our globalised world, there are a range of means to store data online and use automated speech detection tools, helping to speed up the data processing. However, many of these tools make temporary backups in clouds, where linguists are no longer in control of who might get access to the data. It is thus necessary to always safe word-protect data files or entire hard-drives and use

locally run programmes and tools, such as *CorpusCompass* (Adnan & Brandizzi 2023; see the following section).

5   During fieldwork

5.1   Finding speakers

In small and kinship-oriented communities, it may be difficult to find speakers; speakers may have socio-political reasons to distrust cultural outsiders conducting research on their language and culture (Mansfield & Stanford 2017: 117). As a member of secular mainstream society, Neuhausen encountered rejection and prejudices against the secular world (see also Neuhausen 2023: 226), amplified by the increasing general public interest in Amish communities and the release of films and documentaries. It is not in the interest of the Mennonite community to be at the centre of attention, which makes it a fine line between carrying out research on the culture and language and not overstepping the communities' boundaries.

   Signing ethics forms that guarantee the informants' protected anonymity and enable them to withdraw their consent at any time is a requirement at many universities, including the University of Bayreuth and the University of Toronto. Signing such forms, however, may pose a problem in fieldwork in vulnerable communities as it may raise suspicions as to the motives of the respective fieldworker. For instance, Arabic speakers in Germany learn the importance of bureaucratic processes in Germany and that signing forms always has immediate consequences. As a result, even if speakers agree to participate, they might not agree to signing official-looking forms and want to be well-informed about what they sign. Similarly, Pennsylvania German speakers, some of whom did not have regular interactions with the secular world, were reluctant to sign official-looking forms. Without these forms, however, fieldworkers cannot use the collected data for analysis.[12] Ladefoged (2003: 16) notes that speakers of a vulnerable community in the Amazon rain forest were open to sitting down and talking to the researcher but "reluctant to do anything more"; similarly, in the geographically isolated island of Gozo in Malta, speakers consented to being recorded but did not seem comfortable or "consenting of […] [their] own volition" (Klimiuk & Lipnicka 2019: 26). Likewise, speakers may not be literate and may not be able to read the consent form (Neuhausen & Kinsey 2019). As a result, we planned extra time to discuss the consent form with the speakers.

---

[12] This issue has also been reported for some other communities, for example, in the Amazon rain forest (Ladefoged 2003: 15–16) and in East Africa (Copland & Creese 2015: 185–6).

It may also pose a problem for female fieldworkers to interview men in heteronormative and gender-segregated communities, such as the Arabic and Pennsylvania German communities. In some communities, speakers of different genders only intermingle when they are close kin or romantically intertwined, as anything else could result in embarrassment or jealousy (Mansfield & Stanford 2017: 121; Meakins, Green & Turpin 2018: 248–9). As many fieldworkers work alone, the question arises how they can establish contact with, in this case, participants of a different gender. As a woman, Adnan (or her assistant) would not have been able to interview men by herself. However, having their wives present circumvented this issue and allowed her to also include men in the data collection process.

Once speakers are found, the next issue concerns the inclusion of speakers. In reality, speakers are often included in data collection for all kinds of reasons, including availability, compatibility, age, gender and language skills (England 1992: 31), but fieldworkers should reflect on the following questions: Which speaker(s) do they want to be represented in the analysis? Who counts as representative of the community? In some communities, it is not easy to answer the question of who belongs to a community (Heller, Pietikäinen & Pujolar 2018: 47) – what about speakers who leave the community? Adnan met some Arabic speakers who explicitly distanced themselves from the community, leaving behind their traditional Arabic lifestyle and increasingly blending in with German mainstream society. The Arabic speakers were linguistically quite heterogenous, making it challenging to define "representative" speakers. In her doctoral thesis, Neuhausen found that speakers who left the traditional Old Order group diverged linguistically the most from all other groups. This suggests that extending the data set to speakers who left may yield fruitful results – and prompts the question what other types of speakers should be included. What about "semispeakers", who are no longer fluent in the language, "rememberers", who have passive knowledge of the language (Meakins, Green & Turpin 2018: 11), or "near-passive bilinguals" (Dorian 1982: 26)?

Moreover, the question of "language purity" should be scrutinised particularly in diasporic language contact contexts – how "pure" (or non-mixed) can a language be that is or has been in direct contact with the dominant language for a considerable amount of time, i.e. Pennsylvania German with English for the past two hundred years? To put it differently, it is most unlikely that a Pennsylvania German in Canada displays a form of Pennsylvania German in its "purest" form without a trace of English. The same is true for Arabic, which had been in direct contact with German for the past six to seven years at the time of fieldwork. Unlike the *horse and buggy* Mennonites, many Arabic speakers are in intensive contact with German speakers and the German language on a day-to-day basis, creating an even more intense contact setting.

Particularly when working with diasporic communities whose languages are scarcely documented, researchers may not have an initial hypothesis regarding social or linguistic categories. Neuhausen followed a bottom-up approach and explored and familiarised herself with the language and culture before she developed a working hypothesis. Asking in-group members who else could be interviewed may yield valuable insights and further establish contact with new speakers. In-group members are often able to point researchers towards "good speakers"; in the Mennonite community, Neuhausen was frequently pointed towards some speakers who were perceived as being representative of the community, as shown in example (18):

(18)   I almost wish you'd talk more with – well, you talked with Phoebe. I have – don't you think I have a little more of an English accent? Okay, you should hear a *real* Pennsylvania German! (Elon)

Continuing this line of thought, we wonder whether speakers should ever be excluded and if so, on what grounds. It is often advised to exclude speakers of official status, such as priests, teachers and community leaders, because their speech is relatively close to the standard (Tagliamonte 2006: 22). Including such speakers in fieldwork in vulnerable diaspora communities may, however, be the only way to extend a fieldworker's network and facilitate the speakers' decision to participate in the study. As being recorded for the purpose of linguistic research represents a grey area in the Old Order community (it is not absolutely objected to, as it would be for the purpose of entertainment), many speakers were uncertain as to the regulations of the Old Order Church. Speaking to an important member of the community, e.g. of the Old Order Church, helped Neuhausen shed light on potential areas of conflict and confirmed the community's consent to participating in this study.

Community-specific constraints may have a significant impact on the data collection process. When collecting data in the Pennsylvania German community, Neuhausen's goal was never to create a balanced sample but to collect data from anyone who was willing to talk to her. The secluded group has successfully maintained a linguistic and social barrier from mainstream society, which makes it difficult for outsider researchers to make initial contact with them. Many Old Order Mennonites did not want to be interviewed since the recording device embodied one of the core aspects that the Church rejects (modern technology). As a result, the data set Neuhausen collected is by no means representative of the Old Order Mennonite community. Enninger (1987: 149–150), who carried out research on an Amish community, rightly notes: "In

this culture, the choice the field worker has is to work on the basis of the obtainable data, or to gain no insights at all."

In order to detect "new" variables, fieldworkers need to carefully observe and understand the dynamics of the community during fieldwork (ideally before data collection begins): What is socially and culturally important in the community? What is (dis)approved of by the in-group? What are the social and linguistic expectations of particular speaker groups? During her ethnographic fieldwork, Neuhausen gained the impression that the speakers' individual relationships to the Old Order Mennonites determined their language use. But how can this be classified as a social variable? Hazen (2000: 151) developed the concept of cultural identity, distinguishing between speakers who are socially more oriented towards the in-group (what he terms "local identity speakers") and those who are oriented more towards external groups (what he terms "expanded identity speakers"). These social orientations are also reflected in the way these speakers produce speech. Similarly, Reed (2016, 2020) documents that the degree of rootedness in a particular place can show in a speaker's prosody. Additionally, Stuart-Smith, Timmins & Tweedie (2007: 255) point out the importance of acknowledging a community's socio-spatial history; just because a group is not very well connected with other communities does not mean that they are not affected by other more open and mobile communities, such as in-migration in the neighbourhood and increasing tourism. Combining both the concept of cultural identity and socio-spatial history, in her doctoral thesis, Neuhausen developed the social variable of socio-spatial distance, classifying how socioculturally close a speaker is to the traditional Old Order Mennonite community, i.e. whether a speaker is still affiliated with them, has left the community, or whether their parents or grandparents left the community. This variable was statistically significant for all three sociophonetic case studies (/ʍ ~ w/, /l/ velarisation, Canadian Raising). This finding would not have emerged if Neuhausen had solely investigated community affiliation – which did not significantly correlate with the linguistic variables. In fieldwork on multilingual diasporic communities that may have different cultural expectations, norms and values from the fieldworker's native culture, fieldworkers need to keep on their toes and take notes of even minor ethnographic observations they make. During later stages of the fieldwork and analysis, such notes may help them to make sense of the bigger picture and accurately interpret the findings.

As we have seen, finding participants may pose particular challenges in diaspora communities and raises questions concerning the representativeness and inclusion of speakers. Speakers may have a distrust towards cultural outsiders and external or community-specific constraints may further aggravate fieldwork, such as the Old Order Mennonites' objection to modern technology.

Last but not least, if fieldworkers carefully observe the social dynamics of a group, they may also detect socially important variables accounting for individual linguistic variation.

5.2   *The sociolinguistic interview*

The sociolinguistic interview constitutes a widely used method for data collection in variationist sociolinguistics. The goal is to obtain a sample of natural and informal speech (Schilling 2013: 7). This method typically involves a one-on-one conversation between the researcher and the participant, where the fieldworker mostly listens and gives prompts, and the informant mostly talks. Finding an ideal environment for conducting such interviews, specifically "alone in a quiet place with a single interviewee" (Mansfield & Stanford 2017: 121), might not align with the realities of lively rural communities or households with many family members, which is particularly true for tight-knit diaspora communities, including Arabic speakers in Germany and Pennsylvania German speakers in Canada. Moreover, approaching the participants' vernacular, as suggested in traditional fieldwork guides (Tagliamonte 2006: 46), may neither be feasible nor appropriate for researchers who are cultural outsiders (Mansfield & Stanford 2017: 121) – or who find themselves both at the cultural "inside" and "outside". Heller, Pietikäinen & Pujolar (2018: 87) suggest that it is vital to "look not only at the narratives themselves, but also at the conditions in which they are provided, for whom and over the course of what kind of activity". After all, from a sociolinguistic point of view, it may be highly interesting to observe what linguistic variants are performed for cultural outsiders.

Evidently, outsider researchers as interviewers are not able to remove the observer's paradox (Labov 1972: 113). Al-Wer et al. (2022: 13) argue that it may be impossible to eliminate the observer's paradox completely, i.e. to avoid that speakers change their speech because a linguist is present. Instead, they propose that the focus should be on developing and refining methods that aim to reduce this effect (Al-Wer et al. 2022: 13). Even some of the Iraqis with whom Adnan shares a native dialect tended to standardise their speech in conversation with her, likely also prompted by the rather formal recording situation. These speakers had already begun to standardise their language or increasingly integrate German borrowings into their speech during the initial meetings before the recordings; by contrast, during conversations with the Iraqi and Syrian assistants, speakers did not standardise their speech as they likely considered them as in-group members. As an attempt to reduce the effect of the observer's paradox, Adnan recruited Iraqi and Syrian interviewers to interview some speakers.

Unable to ask in-group members to conduct the interviews (as they do not use modern technology), Neuhausen took a different route and included group

interviews, which may be a fruitful alternative circumventing this issue (Copland & Creese 2015: 30). In some cases, Neuhausen observed that some speakers did not feel as if they were in the right setting to speak Pennsylvania German; Mennonites usually do not use Pennsylvania German with cultural outsiders and some quickly switched back to English. In group conversations, however, speakers felt more natural conversing in Pennsylvania German.

During interviews, researchers should remain alert to any linguistic forms they were previously not aware of. They might stumble upon unexpected grammatical forms or phonological features. For example, prior to fieldwork on the Mennonite community, Neuhausen was not aware of the lexeme-specific pronunciation of *December* with a voiced /z/.[13] After noticing this in the speech of multiple speakers – including some who otherwise display mainstream Canadian English – she included a question in each interview that aimed to elicit the lexical item, e.g. "In what months would you usually get the most snow?" Examples like these demand the researcher's attentiveness and adaptability to explore and document previously unnoticed (by the fieldworker) linguistic aspects. Here, our backgrounds as cultural outsiders that were partially also insiders allowed us to ask questions that in-group members usually cannot ask.

Fieldworkers need to be mindful of what questions to ask, particularly in politically marginalised communities. Fieldworkers in diasporic communities often mention modifying the questions based on the local cultural and social setting (Al-Wer et al. 2022: 13). What questions are asked, how they are asked and to whom plays a major role in sociolinguistic fieldwork. The act of asking questions varies across cultures, both in form and meaning. This led us to reflect on asking different speakers the "right" questions. The appropriateness of direct questioning varies across cultural settings. Direct questions may be perceived as impolite (Heller, Pietikäinen & Pujolar 2018: 58) or be met with silence (Kate Burridge, p.c. June 2020, on interviewing David Martin Mennonites, another traditional community in southern Ontario).

Fieldworkers should attempt to get a feeling for what topics are particularly liked in the community. Individual speakers might feel vulnerable when being questioned or when being asked about certain topics – particularly so by cultural outsiders. We suggest to watch out for what topics may be fruitful subjects to prompt elaborate narratives. When recording Arabic speakers, Adnan realised that all participants enjoyed talking about the pandemic. Interestingly, she also observed some age-related differences in preferred topics. While

---

[13] This feature has been documented for varieties of Scottish English (Scots Online 2024) and likely represents a remnant of contact with Scottish settlers in the mid-19th century, when Scottish settlers accounted for the majority of immigrants in Ontario (Boberg 2010: 82).

younger participants liked to discuss topics related to their employment, leisure activities and social media use, older participants tended to show a preference for conversations on the importance of their children maintaining Arabic and on reflections on how their lives have changed since moving to Germany.

One of the most famous and controversial questions in sociolinguistic interviews is Labov's (1972: 113) "Danger of Death" question which is supposed to prompt vivid narratives. Yet, requesting individuals to share experiences of life-threatening situations can lead to high discomfort and may trigger traumatising memories, particularly in the case of diaspora communities. It goes without saying that in the context of forced migration, the "Danger of Death" question is highly inappropriate and may retraumatise individuals. In these groups, given the potential for causing harm, emotional distress or embarrassment, researchers are strongly cautioned against addressing such highly sensitive topics (Al-Wer et al. 2022: 14). In any case, this approach may not yield fruitful outcomes. Milroy (1980) observed that in Belfast, where life-threatening situations happened on an everyday basis during the Northern Ireland conflict, the question failed to elicit emotional narratives. Similarly, in the *horse and buggy* Mennonite community, where many families live on farms, speakers are used to fatal accidents, either on the road or on the farm, and often neutrally described such instances.

Questions should always be reassessed in their individual contexts. As an alternative to potentially highly distressing questions, we opted for more sensitive questions aiming to trigger different emotions, such as "How did the pandemic change your lives?", "What challenges do you encounter as a migrant in Germany?" and "What did your parents say when you left the church?" These types of questions, as we hoped, do not cause harm and do not confirm potential suspicions towards the fieldworkers either.

But how much prompted emotion is appropriate when engaging with vulnerable communities? We suggest that fieldworkers familiarise themselves with the sociocultural norms and carefully navigate the fine line between questions covering emotional and inappropriate topics. As Heller and colleagues (2018) state, flexibility is key throughout the interview process. Interviewers must remain adaptable, "think on their feet" (Heller, Pietikäinen & Pujolar 2018: 91) and carefully observe both verbal and non-verbal signals. Paying attention to social interactions and dynamics is central for prompting authentic responses and enabling a respectful exchange.

Last but not least, the emotional impact of the discussed topics on the interviewer should not be overlooked. Dealing with topics that are heavy or offensive to the fieldworker – with the potential to trigger or traumatise the fieldworker – can lead to uncertainty and mental overload (see also Neuhausen 2023: 236). This is particularly true because interviewers are restricted from

discussing the contents outside the sociolinguistic interviews. Importantly, interviewers need to take care of their own well-being while conducting research.

6   After fieldwork

6.1   Transcribing the data

After the data collection process, researchers should be mindful of their powerful positions and consider the wider political implications their reports may have (e.g. scientific publications or outreach presentations). For diasporic communities, academics may provide the only impression (the non-local) mainstream society, e.g. other linguists, have of the community and their language(s). Arabic communities, which have been established in Germany due to forced migration, often face marginalisation and discrimination based on their cultural and religious practices and language use, specifically due to their skin colour and accented German. Their languages have a low status and are perceived as less privileged. Similarly, the Pennsylvania German community also represents a community based on forced migration. However, instead of integrating into Canadian mainstream society, the traditional group of the Old Order Mennonites chooses to remain separate from it. Maintaining Pennsylvania German serves as a sociocultural barrier between them and outside communities; if it was not for Pennsylvania German and the maintenance of old customs (horse and buggies, traditional clothes), these speakers – who are all white – would have likely quickly blended in with the white Canadian population (Mufwene 2022: 174, 191). Instead, they consciously keep separate from mainstream society. It is therefore necessary that researchers represent the community's interests in their final reports; otherwise they may jeopardise the social status of already vulnerable groups, e.g. by reproducing or creating stereotypes, and/or barring the way for future fieldworkers.

Especially in tight-knit communities like the Old Order Mennonites where everybody knows everybody it may be particularly challenging to protect the informants' anonymity. In such cases, descriptions like "a person who works in a restaurant" or "who has a particular pet/object" might already reveal a speaker's identity. Careful attention is required on behalf of the researchers in how informants are portrayed and described. It may be helpful to ask the following questions: What information is necessary for the line of argumentation to make sense? Can this type of information be generalised in any way? Tagliamonte (2006: 51) suggests assigning pseudonyms to the participants based on their initials or on common local surnames. In a tight-knit community, however, the first option may reveal the informants' identities. Both Adnan and

Neuhausen selected names for pseudonyms that were common in the Mennonite and Arabic communities but not related to the informants' actual names.

Transcriptions are not simply written and neutral representations of speech but carry wider social meaning. All linguistic examples are embedded in particular sociocultural contexts and represent broader social meanings (Bucholtz 2000). The act of transcription itself is not neutral, given its (interpreted) transformation of the spoken word (and non-verbal gestures) into written form (Heller, Pietikäinen & Pujolar 2018: 84). In communities with a (non-written) language lacking an orthographic norm, this requires particularly careful consideration of orthographic representations (but also applies to other non-standard varieties, e.g. Honkanen 2023). When basing Pennsylvania German orthography on standard German, its speakers are likely excluded: They are not able to read standard German because they only read English (and some Lutheran German in the Bible with great difficulty). By contrast, when basing Pennsylvania German orthography on English phonology, they are included in the process and can make sense of the written form. This, however, may be challenging for transcribers who then need to establish a consistent writing system.[14] Neuhausen (2023: 237) suggests involving participants in the transcription process and asking them for feedback concerning how they would write their language. This approach is feasible when researchers have established good access to and relations with the speakers.

For a range of standardised languages, such as English and German, a number of technological tools, such as automated speech recognition, help researchers to speed up the segmentation, transcription and annotation processes. Automatic speech recognition programmes like CLOx (Wassink et al. 2020) are available for a growing number of linguistic varieties, including different varieties of English, German and Arabic. While some dialects are recognised as different enough so that they receive their own automatic speech recognition categories, such as Iraqi, Syrian and other varieties of Arabic in CLOx, contact varieties like Mennonite English and Pennsylvania German are spoken by too few people to be recognised as separate entities on such platforms. This is in part due to the scarce documentation of these languages, but also to their speakers' rejection of technological devices, barely allowing the language to enter the digital space. Evidently, automatic speech recognition programmes can hardly be trained for such languages if there are not many data points publicly available.

In order to improve automated processes like speech recognition, comprehensive Arabic speech corpora are required. The considerable amount of

---

[14] First attempts to standardise Pennsylvania German have already been made, e.g. by Miller (2013); see also the discussion in Hans-Bianchi (2014).

linguistic variation and regional differences evidently pose challenges for automated transcription software, often not resulting in near-accurate results. However, for Arabic, such informative corpora are not available, reusable, informative and/or of sufficient quality (Abushariah et al. 2010). These aspects make it difficult to capture the full range of linguistic features in Arabic.

Complicating the matter of automatic speech recognition processes even more, such programmes are usually not trained on contact varieties, e.g. Arabic/German or Syrian Arabic/Iraqi Arabic, and most will not be able to accurately capture code-switching or hybrid forms in such contact varieties (see section 2.3). Arabic speakers do not only interact with Germans and the German language but also with different varieties of Arabic which they would normally not be exposed to in their home countries (in real-life interactions). Some of the Iraqi speakers in the study had spent time in Egypt or Syria, a scenario not uncommon among migrants from Arab countries, who often reside in other countries before moving to Germany. In contact situations with Syrians, for instance, Iraqi speakers use non-Iraqi features that they probably acquired in the respective countries, as shown in examples (19) and (20), where the speakers use Syrian *ktiir* instead of Iraqi *hwaaya* ('much') and *kamaan* instead of *hamm*/*hammeen* ('also'). These speakers tend to reduce lexical Iraqi features that may not be understood by their interlocutors and instead choose more widely understood terms. This strategy enhances mutual comprehension in conversations between speakers of the different Arabic dialects – but makes automatic speech recognition, which is often trained on one variety of Arabic only, more challenging for such contact varieties.

(19)   *kulliʃ ktiir!* (Amira)
       'Very much!'

(20)   *ṣaḥḥ, hassa kamaan rijʕat ḥatta flaawanzat il-xanaaziir qabil fatra bi-l-axbaar* (Munir)
       'That's right, now also the swine flu has been back in the news recently.'

Furthermore, Arabic has its own alphabet, which does, however, not include all sounds that are present in informal spoken varieties of the language, such as /č/, /g/ and /p/.[15] Consequently, the question arises what alphabet should be used for the transcription.[16] Adnan chose to transcribe using the International Phonetic

---

[15] Unlike Modern Standard Arabic, Arabic dialects have no standard orthography.

[16] Although Arabic speakers commonly write their dialects using the Arabic alphabet without standardised rules, another popular method is *Arabizi* or *Franco Arabe*. This system uses the

Alphabet (IPA), a uniform representation of all possible Arabic sounds.[17] This, however, introduces an additional layer of complexity to the transcription process, which requires even more time-intensive labour and in-depth linguistic knowledge. Al-Wer et al. (2022: 35) suggest that the level of transcription should be restricted to the project's specific needs. For projects requiring detailed phonetic analysis, a narrow transcription using the complete set of IPA offers maximum precision; for studies where phonetic details are less critical but phonemic distinctions are essential, a broader and less precise transcription approach is sufficient; and for studies interested in language attitudes, orthographic transcription will suffice.

Lack of resources may entail that only one annotator processes the data. This limitation raises concerns linked to subjectivity, potential errors and inconsistencies in the annotations, which ultimately affect the analysis. For the case study, Adnan made multi-layered annotations on the phonological, morphological and lexical level, with items originating in four different linguistic varieties (Iraqi Arabic, Syrian Arabic, German, English) and mixed forms.

Despite the complexity of the linguistic forms, Adnan aimed for a comprehensive and precise extraction of the annotations in spreadsheets that could be used later for statistical analysis. Given the diverse languages within the corpora and the diverse types of annotations, an imminent need for a tool capable of handling it emerged. The tasks of processing such highly complex data are not only time-consuming, but prone to error and require a high level of technical expertise. Regarding this matter, Gries (2009) points out that acquiring programming skills and developing custom analytical tools can be an effective solution to overcome the limitations posed by existing corpus tools.

Not all linguists are, however, computer scientists. The field of corpus linguistics offers a variety of tools that require no knowledge of programming, such as *AntConc* (Anthony 2023), providing a wide range of functions and allowing for keyword analyses, word frequency counts and collocate searches. Despite the wide availability of these tools, their application often falls short for under-resourced languages and, in particular, does not account for contact

---

Latin alphabet combined with numbers to represent Arabic sounds that do not have corresponding letters in the Latin script. Developed by Arabic-speaking youths, *Arabizi* allows for the transliteration of Arabic with Latin letters and numerals, making it an effective solution for digital communication. For more information on *Arabizi*, see Yaghan (2008).

[17] While it is important for the community members to be able to read the transcriptions of their own data, Adnan chose to transcribe using the IPA system because it covers all Arabic sounds – unlike the Arabic writing system. As a result, when making the transcriptions readable to the participants, an additional step needs to be taken, i.e. transforming the transcriptions into the Latin or Arabic writing system.

varieties emerging in diasporic contexts. As Adler et al. (2024) show, almost every work step is carried out manually in sociolinguistic research on Arabic varieties, ranging from error-prone manual text analysis to inconsistent data management approaches. This is particularly true for research focusing on multiple variables, factors, and speakers. Importantly, this even applies if researchers have advanced programming skills. To address these challenges, Adnan and Brandizzi created *CorpusCompass* as an open-source tool for annotation extraction and dataset creation (Adnan & Brandizzi 2023; see also Adler et al. 2024). Adnan used the programme to generate a structured spreadsheet (including dependent variables, independent variables, and metadata), to select the variables, and to compile frequency counts.[18] The programme helped her preselect variables and prepare the data for the quantitative analysis (e.g. by treating it as binary categories or count data), which was then carried out in a different programme, i.e. in the programming language R (R Core Team 2021). It is specifically intended for under-resourced languages and non-standard annotations in mono- and multilingual corpora. The tool is flexible enough to accommodate diverse workflows and data types, all the while overcoming shortcomings found in existing tools.

Meeting individual research needs, *CorpusCompass* enables researchers to define dependent and independent variables of interest, numerous speakers within one conversation as well as customised tags. It is designed for data that has already been transcribed and annotated. It allows for an in-depth analysis on the phonological, morphological and lexical level and simultaneously ensure error-free and consistent processing of annotations. The output is a structured dataset in a CSV format that incorporates pre-defined linguistic variables and potentially relevant metadata, thereby facilitating corpus exploration and statistical analysis. Targeting researchers with less technical proficiency, the tool transforms the input data into a structured spreadsheet that can be used for subsequent quantitative and qualitative analyses. It helps to decolonise linguistic data processing and analysis, enabling students, early-career researchers and linguists without programming skills to use it.

As a result, research processes involving diasporic languages can be even more time-intensive and long-winded than projects based on fieldwork on well-

---

[18] *CorpusCompass* was developed in collaboration with computer scientist Nicolo' Brandizzi who focused on its programming aspects. The project also received support from Jelke Bloem and the University of Amsterdam, particularly through the *Research Engineering Support Grant* from the UvA Data Science Centre (DSC). Additionally, financial support for the further development of the tool was provided by the Arabic Studies Department at the University of Bayreuth. Currently, *CorpusCompass* is a work-in-progress and developed by Jonas Adler and Carsten Scholle, who significantly contribute to making the tool accessible to users with no programming background.

researched linguistic varieties, like American English. As a consequence, particularly early-career researchers in less privileged academic positions with little financial stability may be deterred from investing much time and energy in lesser documented languages; this often involves carrying out fieldwork in lesser-known communities in far-away countries where they may need to fly to and maybe even more than once, to do follow-up research and maintain long-term relationships with the communities (as some groups cannot be reached via the internet or other global means of communication).

These examples show that the development of language technology and research tools is biased towards monolingual mainstream varieties. Automatic speech recognition systems accommodate only a fraction of the global linguistic diversity. This is particularly true for linguistic varieties in diasporic communities with contact languages and for varieties that are not or barely present in the digital space. As a consequence, researchers of contact languages need to rely more on manual processes. Manual transcription slows down the entire research process immensely and increases the time spent on the transcription process. This may result in more (early-career) researchers in precarious work situations focusing on better-studied communities, which enables them to produce results at a greater speed.

6.2 *Interpreting the data*

Not only data collection methods, but also data interpretation methods are often culturally biased (Dimmendaal 2001: 69). As a result, in (multilingual) diasporic communities, social variables unknown to the linguist may play a role in language variation and change, such as the above-mentioned variable of the socio-spatial distance from the Old Orders (see section 5). In the traditional Old Order Mennonite community, social variables that are relevant in many other traditional sociolinguistic settings, such as socioeconomic class and education, do not play a role as they are consistent across the entire community; there is no socioeconomic class system, and Old Order Mennonites do not pursue education beyond grade eight (i.e. they attend school until the age of 14).

Asking the community for feedback is highly valuable. In the Arabic case study, an elderly female participant, a journalist from Baghdad, provided valuable insights into the speech of Iraqi communities both in Iraq and in Germany. She explained the linguistic differences in Baghdad, specifically between the two districts *al-Adhamiyya* and *al-Kadhimiyya*, two of nine administrative districts of Baghdad. She mentioned that these areas are generally perceived by Baghdadis as "manaaṭiq baɣdaad il-aṣliyya" ('the true/original districts of Baghdad') and that outside these two linguistically similar areas, linguistic features differ, which she believes have developed due to language contact from earlier migrations from the south of Iraq to Baghdad. Mithun

(2001: 48) emphasises the importance of recognising speakers as experts, particularly during the initial stages of collaboration. Insights provided by speakers can help linguists understand language attitudes and uses, enrich the analyses, validate emerging hypotheses and reveal fine layers of linguistic variation which might otherwise remain hidden. This is particularly the case in smaller communities where the local knowledge is taken for granted and not necessarily shared with the fieldworker.

Community feedback was also necessary in understanding how specific features, such as "Allah expressions", varied among different age groups. Examples of such expressions include إن شاء الله /inshallah/ 'if God wills' and الحمد لله /alḥamdulillaah/ 'thank God'. These expressions were more frequently used by elderly speakers compared to younger speakers. Asking the community about this proved to be particularly useful. They explained that in Iraq and Syria, unlike in Germany, these religious expressions are frequently used across all age groups. They associated these expressions with characteristics of older Arab speakers, whom young people in Germany may not want to be associated with. The involvement of speakers in shaping the linguistic record not only enriches the documentation of the language but also expands our understanding of linguistic variability (Mithun 2001: 51).

Similarly, asking the Mennonite community for feedback also shed light on linguistic variation and social meaning attached to it. Without the feedback, Neuhausen would have interpreted these differently. During the transcription of Pennsylvania German, she observed that some tokens of *mom* were realised with the long Canadian vowel /ɑ:/, while others featured the German short vowel /ɐ/. What Neuhausen initially took for individual language transfer from Pennsylvania German turned out to carry community-relevant information. An informant pointed out to her that /mɑ:m/, with the standard Canadian vowel, is used by the Markham group, while /mɐm/ is used by the traditional Old Order Mennonites. In other words, socially meaningful variation may only emerge in sociolinguistic analyses when community members are included in the analyses.

We also found that it is important to pay attention to metalinguistic comments. They may shed light on the speakers' perceptions and intuitions of the factors causing linguistic variation, the changes speakers perceive and comment on and the distinct roles each language plays in relation to their attitudes towards these languages.

The use (or avoidance) of linguistic variants imbued with social meaning may also reveal insights into the speakers' language attitudes. For instance, in the context of the Arabic diaspora, Iraqi participants showed a positive attitude towards Syrian Arabic, a view not reciprocated by the Syrian participants. This became apparent in the levelling of marked and dialect-specific linguistic features in Iraqi Arabic, such as stigmatised /č/ for /k/ (e.g., /čam/ vs /kam/ 'how

much?') or the verb prefix *da-* (Adnan & Owens forthcoming).[19] Iraqi speakers maintained these features in intra-group settings but employed them less in conversations with Syrians, accommodating their speech in inter-group contexts.

For the interpretation of the data, involving the community will yield the most genuine and accurate observations that also represent the speakers' voices. We have therefore found it helpful to include community feedback and metalinguistic comments in conceptualising prevailing social dynamics and linguistic variation.

7   Conclusion

Fieldwork in diasporic communities, such as the Arabic community in Germany and the Mennonite community in Canada, presents both unique community-specific and shared challenges. Although the vast majority of the Arabic-speaking community actively seeks to integrate into German society and puts effort into doing so, they remain marginalised and (structurally) discriminated against by mainstream society. The traditional Mennonite community, however, chooses to remain separate from mainstream society and represents a secluded group that has consciously maintained a sociocultural barrier from mainstream society, by ways of dress, language and mode of transport. While the Mennonite community has maintained their heritage language since the 17th century, the Arabic community in Germany follows in other Arabic communities' footsteps: Younger speakers increasingly appear to experience language uncertainty in Arabic and may become stronger in the dominant societal language.

Interestingly, we also identified common ground shared by these communities, such as the multilingual environments of the speakers, resulting in hybrid linguistic forms and the absence of an orthographic norm of the first language. All these factors influence not only data production, but also how we as fieldworkers collected and interpreted the linguistic data.

We have shown that the fieldworker's identity and self-presentation impact the data collection process. The approach to the two communities was similar on behalf of the fieldworkers, as Adnan was an "insider-outsider" researcher and Neuhausen more on the "outside" with some "inside" knowledge, such as her apparent expert status of the German language. This facilitated social interactions with community members and simultaneously allowed us to approach the interview situation from an outsider perspective and ask questions

---

[19] The verb prefix *da-* is used to emphasise or spotlight a specific event, often indicating urgency or relevance to the topic of discourse (Adnan & Owens forthcoming). For instance, in the sentence *il-ḥanafiyya da-t-naqqiṭ* ('The faucet is dripping') (Erwin 2004: 139), *da-* implies an urgent need for action, rather than a habitual action.

an in-group member cannot ask. At the same time, the above-mentioned chain-link fence, i.e. the sociocultural barrier between the fieldworker and community, may put fieldworkers in an ambiguous position where they are always positioned outside the community (and within the scientific setting).

We want to emphasise the necessity for fieldworkers and linguists to remain critically aware of their own impact on the community and their methodological approaches. The overwhelming focus of research on WEIRD communities is also replicated in the development of language technology and research tools. As a result, the need to adjust and continue developing fieldwork methods in diaspora communities is critical, and adaptable culturally sensitive methodologies are needed; for that purpose, we have presented the open-access tool *CorpusCompass* (Adnan & Brandizzi 2023). This tool, which helped Adnan create a structured spreadsheet with multiple variables and factors, aims to help linguists working on lesser-studied and diasporic languages process complex linguistic data and produce an output that can be used for subsequent statistical analysis – without requiring advanced programming skills.

We hope to have shown that by discussing fieldwork processes and cultural biases openly, researchers can enrich their studies and provide valuable insights that can lead to the development of new sociolinguistic methodologies. This approach has the potential to "pioneer new forms of sociolinguistic methodology" (Mansfield & Stanford 2017: 130). Only when researchers critically evaluate the cultural biases they bring to the field, they are able to identify how their observations are coloured by their own assumptions – even outside the field. In order to authentically represent the voices and complexities of the speakers, the speakers should be involved in the data analysis process if possible. We hope that this paper encourages more linguists to conduct exciting fieldwork in diasporic communities and produce linguistic findings that genuinely reflect the lived experiences of the communities under study as well as advance sociolinguists' understanding of diaspora communities and their languages.

8   References

All hyperlinks were last accessed on 05 August 2024.

Abu-Haidar, Farida. 1991. *Christian Arabic of Baghdad*. Wiesbaden: Harrassowitz.

Abushariah, Mohammad A. M., Raja N. Ainon, Roziati Zainuddin, Bassam A. Al-Qatab & Assal A. M. Alqudah. 2010. Impact of a Newly Developed Modern Standard Arabic Speech Corpus on Implementing and Evaluating Automatic Continuous Speech Recognition Systems. In Gary Geunbae Lee, Joseph Mariani, Wolfgang Minker & Satoshi Nakamura (eds.), *Spoken Dialogue Systems for Ambient Environments*

(Lecture Notes in Computer Science), vol. 6392. Berlin & Heidelberg: Springer Berlin Heidelberg, 1–12. https://doi.org/10.1007/978-3-642-16202-2_1.

Adler, Jonas, Carsten Scholle, Daniel Buschek, Nicolo' Brandizzi & Muhadj Adnan. 2024. User-Centered Design of Digital Tools for Sociolinguistic Studies in Under-Resourced Languages. In *Proceedings of the 3rd Workshop on NLP Applications to Field Linguistics (Field Matters 2024)*, 12–27, Bangkok, Thailand. Association for Computational Linguistics. DOI: 10.18653/v1/2024.fieldmatters-1.3 .

Adnan, Muhadj & Nicolo' Brandizzi. 2023. CorpusCompass: A Tool for Data Extraction and Dataset Generation in Corpus Linguistics. In Federico Boschetti et al. (eds.), *Proceedings of the 9th Italian Conference on Computational Linguistics, Venice, Italy, November 30 – December 2, 2023. CEUR Workshop Proceedings,* vol. 3596. https://ceur-ws.org/Vol-3596/.

Adnan, Muhadj & Jonathan Owens. Forthcoming. Imperfect Verbal Prefixes as Discourse Markers. In *Perspectives on Arabic Linguistics XXXV*. Amsterdam: John Benjamins.

Al-Wer, Enam, Uri Horesh, Bruno Herin & Rudolf de Jong. 2022. *Arabic Sociolinguistics*. Cambridge: Cambridge University Press.https://doi.org/10.1017/9781316863060.

Ambros, Arne. 1977. *Damascus Arabic*. Malibu: Undena.

Anthony, Laurence. 2023. AntConc. Tokyo: Waseda University. https://www.laurenceanthony.net/software.

Arnold, Werner. 1998. *Die arabischen Dialekte Antiochiens*. Wiesbaden: Harrassowitz.

Assmann, Aleida & Jan Assmann. 1994. Das Gestern im Heute. Medien und soziales Gedächtnis. In Klaus Merten, Siegfried J. Schmidt & Siegfried Weischenberg (eds.), *Die Wirklichkeit der Medien. Einführung in die Kommunikationswissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 114–140.

Badawia, Tarik. 2002. *"Der Dritte Stuhl" – Eine Grounded Theory-Studie zum kreativen Umgang bildungserfolgreicher Immigrantenjugendlicher mit kultureller Differenz*. Frankfurt am Main: IKO.

Barontini, Alexandrine & Lauren B. Wagner. 2020. Diasporic Arabic(s): Speakers, Usages and Contacts. In Reem Bassiouney & Keith Walters (eds.), *The Routledge Handbook of Arabic and Identity*. New York: Routledge, 245–258.

Behnstedt, Peter. 1997–2000. *Sprachatlas von Syrien*. Wiesbaden: Harrassowitz.

Benor, Sarah Bunin. 2010. Ethnolinguistic Repertoire: Shifting the Analytic Focus in Language and Ethnicity. *Journal of Sociolinguistics* 14(2): 159–183.

Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge, MA: Harvard University Press.

Boumans, Louis. 2006. The Attributive Possessive in Moroccan Arabic Spoken by Young Bilinguals in the Netherlands and their Peers in Morocco. *Bilingualism* 9(3), 213–231.

Boumans, Louis & Jan Jaap de Ruiter. 2012. Moroccan Arabic in the European Diaspora. In Aleya Rouchdy (ed.), *Language Contact and Language Conflict in Arabic*. *Variations on a Sociolinguistic Theme*. London & New York: Routledge, 259–285.

Boberg, Charles. 2010. *The English Language in Canada: Status, History and Comparative Analysis.* (Studies in English Language.) Cambridge: Cambridge University Press.

Boussofara-Omar, Naima. 2006. Diglossia. In Kees Versteegh (ed.), *Encyclopedia of Arabic Language and Linguistics.* Vol. 1. Leiden: Brill, 629–637.

Bowern, Claire. 2015. *Linguistic Fieldwork. A Practical Guide.* London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137340801.

Bucholtz, Mary. 2000. The Politics of Transcription. *Journal of Pragmatics* 32(10): 1439–1465. https://doi.org/10.1016/S0378-2166(99)00094-6.

Bundesamt für Migration und Flüchtlinge (BAMF). Aktuelle Zahlen zu Asyl. Berichte von 2015–2023. https://www.bamf.de/DE/Themen/Statistik/Asylzahlen/BundesamtInZahlen/bundesamtinzahlen-node.html

Burridge, Kate. 1989. *Canadian-German Folklore: A Localized Study of P.G. Dialect in Waterloo County, Ontario.* Vol. 6. London, ON: The Pennsylvania German Folklore Society of Ontario.

Burridge, Kate. 1992. Creating Grammar: Examples from Pennsylvania German, Ontario. In Kate Burridge & Werner Enninger (eds.), *Diachronic Studies on the Languages of the Anabaptists.* Bochum: Brockmeyer, 199–241.

Burridge, Kate. 1998. Throw the Baby from the Window a Cookie: English and Pennsylvania German in Contact. In Anna Siewierska & Jae Jung Song (eds.), *Case, Typology and Grammar.* Amsterdam & Philadelphia: John Benjamins, 71–93. https://doi.org/10.1075/tsl.38.08bur.

Burridge, Kate. 2002. Steel Tyres or Rubber Tyres – Maintenance or Loss: Pennsylvania German in the "Horse and Buggy" Communities of Ontario. In David Bradley & Maya Bradley (eds.), *Language Endangerment and Language Maintenance: An Active Approach.* London: Routledge, 203–229.

Candan, Menderes. 2017. *Die irakische Diaspora in Deutschland. Struktur, Organisation, Aktivitäten und entwicklungspolitisches Engagment* [sic]. Bonn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). https://diaspora2030.de/fileadmin/files/Service/Publikationen/Studien_zu_Diaspora-Aktivitaeten_in_Deutschland/giz-2018-de-diasporastudie-irak.pdf

Cohen, Robin. 2008. *Global Diasporas.* 2nd ed. London: Routledge. https://doi.org/10.4324/9780203928943.

Copland, Fiona & Angela Creese. 2015. *Linguistic Ethnography: Collecting, Analysing and Presenting Data.* London: Sage.

Costello, John R. 1978. Syntactic Change and Second Language Acquisition. The Case for Pennsylvania German. *Linguistics* 16(213): 29–50. https://doi.org/10.1515/ling.1978.16.213.29.

Cowell, Mark. 1964. *A Reference Grammar of Syrian Arabic (based on the Dialect of Damascus).* Washington D.C.: Georgetown University Press.

Crul, Maurice, Elif Keskiner, Jens Schneider, Frans Lelie & Safoura Ghaeminia. 2017. No Lost Generation? Education for Refugee Children: A Comparison between Sweden, Germany, the Netherlands and Turkey. In R. Bauböck & M. Tripkovic (eds.), *The Integration of Migrants and Refugees: An EUI Forum on Migration, Citizenship and Demography*. Florence: European University Institute & Robert Schuman Centre for Advanced Studies, 62–80.

Deutschlandfunk. 2024. Syrische, afghanische und iranische Flüchtlinge kommen bevorzugt nach Deutschland. *Deutschlandradio*, 4 August 2024. https://www.deutschlandfunk.de/syrische-afghanische-und-iranische-fluechtlinge-kommen-bevorzugt-nach-deutschland-102.html.

Dimmendaal, Gerrit J. 2001. Places and People: Field Sites and Informants. In Paul Newman & Martha Ratliff (eds.), *Linguistic Fieldwork*. Cambridge: Cambridge University Press, 55–75. https://doi.org/10.1017/CBO9780511810206.004.

Dorian, Nancy C. 1982. Defining the Speech Community to Include its Working Margins. In Suzanne Romaine (ed.), *Sociolinguistic Variation in Speech Communities*. London: Edward Arnold, 25–33.

Draper, Barb. 2010. *The Mennonites of St. Jacobs and Elmira: Understanding the Variety*. Kitchener: Pandora.

England, Nora C. 1992. Doing Mayan Linguistics in Guatemala. *Language* 68(1): 29–35. https://doi.org/10.1353/lan.1992.0073.

Enninger, Werner. 1984. Funktion, Struktur und Erwerb der Varietäten Pennsylvaniadeutsch, Amisch Hochdeutsch und Amerikanisches Englisch bei den Altamischen. In Els Oksaar (ed.), *Spracherwerb – Sprachkontakt – Sprachkonflikt*. Berlin & New York: Walter de Gruyter, 220–242. https://doi.org/10.1515/9783110859751.220.

Enninger, Werner. 1987. On the Organization of Sign-Processes in an Old Order Amish (OOA) Parochial School. *Research on Language & Social Interaction* 21(1–4): 143–170. https://doi.org/10.1080/08351818709389288.

Erwin, Wallace M. 2004. *A Short Reference Grammar of Iraqi Arabic*. Washington, DC: Georgetown University Press.

European Commission. n.d. *Forced Migration*. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/forced-migration_en#.

Ferguson, Charles A. 1959. Diglossia. *WORD* 15(2): 325–340. https://doi.org/10.1080/00437956.1959.11659702.

Frantz, John B. 2017. Religion. In Simon J. Bronner & Joshua R. Brown (eds.), *Pennsylvania Germans. An Interpretative Encyclopedia*. Baltimore: Johns Hopkins University Press, 131–147.

Fretz, Joseph Winfried. 1989. *The Waterloo Mennonites. A Community in Paradox*. Waterloo, ON: Wilfried Laurier University Press.

Gralla, Sabine. 2006. *Der arabische Dialekt von Nabk (Syrien)*. Wiesbaden: Harrassowitz.

Gries, Stefan Th. 2009. What is Corpus Linguistics? *Language and Linguistics Compass* 3(5): 1225–1241. https://doi.org/10.1111/j.1749-818X.2009.00149.x.

Grotzfeld, Heinz. 1965. *Syrisch-arabische Grammatik: Dialekt von Damaskus*. Wiesbaden: Harrassowitz.

Hall, Stuart. 1990. Cultural Identity and Diaspora. In Jonathan Rutherford (ed.), *Identity: Community, Culture, Difference*. London: Lawrence and Wishart.

Hans-Bianchi, Barbara. 2014. Pennsylvaniadeutsch: Wege der Verschriftung einer Minderheitensprache. *Bollettino dell'Associazione Italiana di Germanistica* 7: 113–131. https://www.associazioneitalianagermanistica.it/docs/rivista_aig/baig7/9_Hans-Bianchi_113-131_DEF.pdf.

Hazen, Kirk. 2000. *Identity and Ethnicity in the Rural South: A Sociolinguistic View through Past and Present Be* (Publication of the American Dialect Society 83). Durham, NC: American Dialect Society & Duke University Press.

Heller, Monica, Sari Pietikäinen & Joan Pujolar. 2018. *Critical Sociolinguistic Research Methods: Studying Language Issues that Matter*. New York: Routledge.

Henry, Marsha Giselle. 2003. "Where are You really from?": Representation, Identity and Power in the Fieldwork Experiences of a South Asian Diasporic. *Qualitative Research* 3(2): 229–242.

Honkanen, Mirka. 2023. Orthography and the Sociolinguistics of Writing. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.998.

Huffines, Marion Lois. 1984. The English of the Pennsylvania Germans: A Reflection of Ethnic Affiliation. *The German Quarterly* 57(2): 173–182.

Huffines, Marion Lois. 1986. Intonation in Language Contact: Pennsylvania German English. In Werner Enninger (ed.), *Studies on the Languages and the Verbal Behavior of the Pennsylvania Germans I*. Stuttgart: Franz Steiner, 25–36.

Jastrow, Otto. 2006. Iraq. In K. Versteegh (ed.) *Encyclopedia of Arabic Language and Linguistics*. Vol. 2. Leiden & Boston: Brill, 414–424.

Keiser, Steven Hartman. 2015. Religious Identity and the Perception of Linguistic Difference: The Case of Pennsylvania German. *Language and Communication* 42: 125–134.

Klimiuk, Maciej & Maria Lipnicka. 2019. Dialectology in Practice: Notes from Fieldwork in Gozo. In Werner Arnold & Maciej Klimiuk (eds.), *Arabic Dialectology: Methodology and Field Research*. Wiesbaden: Harrassowitz, 23–32. https://doi.org/10.11588/HEIDOK.00029350.

Kopp, Achim. 1999. *The Phonology of Pennsylvania German English as Evidence of Language Maintenance and Shift*. Selinsgrove: Susquehanna University Press.

Labov, William. 1972. Some Principles of Linguistic Methodology. *Language in Society* 1(1): 97–120. https://doi.org/10.1017/S0047404500006576.

Ladefoged, Peter. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA: Blackwell.

Lepore, Jill. 2018. *These Truths. A History of the United States*. New York: W. W. Norton & Company.

Liebow, Elliott. 2003. *Tally's Corner. A Study of Negro Streetcorner Men*. New ed. New York: Rowman & Littlefield.

Louden, Mark L. 2016. *Pennsylvania Dutch: The Story of an American Language*. Baltimore: Johns Hopkins University Press.

MacCabe, Colin & Holly Yanacek (eds.). 2018. *Keywords for Today. A 21st Century Vocabulary. The Keywords Project*. Oxford: Oxford University Press.

Malaika, Nisar. 1963. *Grundzüge der Grammatik des arabischen Dialektes von Bagdad*. Wiesbaden: Harrassowitz.

Mansfield, John & James N. Stanford. 2017. Documenting Sociolinguistic Variation in Lesser-Studied Indigenous Communities: Challenges and Practical Solutions. *Language Documentation & Conservation Special Publication* 13: 116–136.

Meakins, Felicity, Jennifer Green & Myfany Turpin. 2018. *Understanding Linguistic Fieldwork* (Understanding Language Series). London & New York: Routledge.

Mecheril, Paul. 2004. Andere Deutsche gibt es nicht. Zusammenhänge zwischen subalterner Erfahrung und diskursiver Praxis. In AntiDiskriminierungsBüro Köln, Öffentlichkeit gegen Gewalt e. V & cyberNomads (eds.), *TheBlackBook – Deutschlands Häutungen*. Frankfurt am Main: IKO, 82–90.

Meyerhoff, Miriam, Erik Schleef & Laurel MacKenzie. 2015. *Doing Sociolinguistics: A Practical Guide to Data Collection and Analysis*. London: Routledge, Taylor & Francis Group.

Miller, David. 2013. *Pennsylvania German Dictionary*. Deitsh Books. https://www.deitshbooks.com.

Milroy, Lesley. 1980. *Language and Social Networks*. Oxford: Blackwell.

Mithun, Marianne. 2001. Who Shapes the Record: The Speaker and the Linguist. In Paul Newman & Martha Ratliff (eds.), *Linguistic Fieldwork*. Cambridge: Cambridge University Press, 34–54. https://doi.org/10.1017/CBO9780511810206.003.

Mufwene, Salikoko S. 2000. Creolization is a Social, not a Structural, Process. In Ingrid Neumann-Holzschuh & Edgar Werner Schneider (eds.), *Degrees of Restructuring in Creole Languages* (Creole Language Library 22). Amsterdam & Philadelphia: John Benjamins, 65–83.

Mufwene, Salikoko S. 2022. The Verticalization Model of Language Shift from a Population-Structure Perspective: A Commentary. In Joshua R. Brown (ed.), *The Verticalization Model of Language Shift: The Great Change in American Communities*. Oxford: Oxford University Press, 166–194. https://doi.org/10.1093/oso/9780198864639.003.0007.

Narayan, Kirin. 1993. How Native Is a "Native" Anthropologist? *American Anthropologist* 95(3): 671–686. https://doi.org/10.1525/aa.1993.95.3.02a00070.

Neuhausen, Miriam. 2023. Understanding, Collecting and Presenting Data in New Englishes Research: Insights from Ethnographic Fieldwork in an Old Order Mennonite Community. In Guyanne Wilson & Michael Westphal (eds.), *New Englishes, New Methods* (Varieties of English around the World). Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/veaw.g68.11neu.

Neuhausen, Miriam & Sonya Kinsey. 2019. Out in the Field: Old Order Mennonites & the Wit'suwit'en First Nation. Paper presented at the workshop *New Englishes, New Methods*, University of Freiburg, December 2019.

Owens, Jonathan. 2005. Bare Forms and Lexical Insertions in Code-switching: A Processing-based Account. *Bilingualism* 8(1): 23–38.

Podesva, Robert J. 2007. Phonation Type as a Stylistic Variable: The Use of Falsetto in Constructing a Persona. *Journal of Sociolinguistics* 11(4): 478–504. https://doi.org/10.1111/j.1467-9841.2007.00334.x.

Procházka, Stephan. 2020. Arabic in Iraq, Syria and Southern Turkey. In Christopher Lucas & Stefano Manfredi (eds.), *Arabic and Contact-Induced Change*. Berlin: Language Science Press, 83–114. https://doi.org/10.5281/ZENODO.3744507.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ragab, Nora Jasmin, Laura Rahmeier & Melissa Siegel. 2017. *Mapping the Syrian Diaspora in Germany. Contributions to Peace, Reconstruction and Potentials for Collaboration with German Development Cooperation*. Maastricht Graduate School of Governance (MGSoG): Commissioned by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). https://www.merit.unu.edu/publications/uploads/1487758705.pdf.

Raith, Joachim. 1982. *Sprachgemeinschaft, Sprachkontakt, Sprachgebrauch: Eine Untersuchung des Bilinguismus der anabaptischen Gruppen deutscher Abstammung in Lancaster County, Pennsylvania* (Zeitschrift für Dialektologie und Linguistik, Beihefte, 36). Wiesbaden: Franz Steiner.

Raith, Joachim. 1996. Religiöse Migration. In Hans Goebl, Peter H. Nelde, Starý Zdeněk & Wolfgang Wölck (eds.), *Kontaktlinguistik: Ein internationales Handbuch zeitgenössischer Forschung*, vol. 1. (Handbücher zur Sprach- und Kommunikationswissenschaft 12.1). Berlin: Walter de Gruyter, 311–320.

Reed, Paul E. 2016. *Sounding Appalachian: /ai/ Monophthongization, Rising Pitch Accents, and Rootedness*. University of South Carolina Doctoral Thesis.

Reed, Paul E. 2020. Prosodic Variation and Rootedness in Appalachian English. *University of Pennsylvania Working Papers in Linguistics* 25(2): 106–114.

Richter, Manfred. 1969. *The Phonemic System of the Pennsylvania German Dialect in Waterloo County, Ontario*. Toronto: University of Toronto PhD Thesis.

Rosaldo, Renato. 1989. *Culture and Truth. The Remaking of Social Analysis*. Boston: Beacon Press.

Rosenbach, Anette. 2002. *Genitive Variation in English. Conceptual Factors in Synchronic and Diachronic Studies* (Topics in English linguistics, 42). Berlin & New York: Mouton de Gruyter. https://doi.org/10.1515/9783110899818.

Schilling, Natalie. 2013. *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511980541.

Schilling-Estes, Natalie. 2007. Sociolinguistic Fieldwork. In Robert Bayley & Ceil Lucas (eds.), *Sociolinguistic Variation: Theories, Methods, and Applications*. Cambridge: Cambridge University Press, 165–189. https://doi.org/10.1017/CBO9780511619496.010.

Schreier, Daniel. 2013. Collecting Ethnographic and Sociolinguistic Data. In Manfred Krug & Julia Schlüter (eds.), *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press, 17–35. https://doi.org/10.1017/CBO9780511792519.004.

Scots Online. 2024. Pronunciation. *Wir Ain Leed – Scottish Standard English*. https://www.scots-online.org/grammar/sse.php.

Sirin, Selcuk R. & Lauren Rogers-Sirin. 2015. *The Educational and Mental Health Needs of Syrian Refugee Children*. Washington, DC: Migration Policy Institute.

Statistics Canada. 2017. Proportion of Mother Tongue Responses for Various Regions in Canada, 2016 Census. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dv-vd/lang/index-eng.cfm.

Statistics Canada. 2022. Mother Tongue by Geography, 2021 Census. https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/dv-vd/language-langue/index-en.html.

Stuart-Smith, Jane, Claire Timmins & Fiona Tweedie. 2007. "Talkin' Jockney"? Variation and Change in Glaswegian Accent. *Journal of Sociolinguistics* 11(2): 221–260.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801624.

Van Bragt, Thieleman. [1660] 1998. *Martyrs Mirror: The Story of Seventeen Centuries of Christian Martyrdom from the Time of Christ to A.D. 1660*. Rev. ed.; 1st Dutch ed. 1660. Scottdale, PA: Herald Press.

Van Ess, John. 1978. *The Spoken Arabic of Iraq*. 2nd ed. Oxford: Oxford University Press.

Van Ness, Silke. 1990. *Changes in an Obsolescing Language. Pennsylvania German in West Virginia*. Tübingen: Narr.

Wassink, Alicia, Rob Squizzero, Campion Fellin & David Nichols. 2020. Client Libraries Oxford (CLOx). University of Washington. https://clox.ling.washington.edu/.

Werner, Arnold & Maciej Klimiuk (eds.). 2019. *Arabic Dialectology. Methodology and Field Research*. Wiesbaden: Harrassowitz.

Wilson, Arthur Herman. 1948. English Spoken by Pennsylvania Germans in Snyder County, Pennsylvania. *American Speech* 23(3/4): 236–8.

Yaghan, Mohammad Ali. 2008. "Arabizi": A Contemporary Style of Arabic Slang. *Design Issues* 24(2): 39–52. https://doi.org/10.1162/desi.2008.24.2.39.

Yoseph, Jean. 2012. *Der arabische Dialekt von Mharde (Zentralsyrien)*. Wiesbaden: Harrassowitz.

Contact information:

Miriam Neuhausen (she/her)
University of Heidelberg
miriam.neuhausen[at]as.uni-heidelberg.de

Muhadj Adnan (she/her)
University of Bayreuth
muhadj.adnan[at]uni-bayreuth.de