

2 DOCUMENTING ENDANGERED ORAL HISTORIES OF THE ARCTIC: A PROPOSED SYMBIOSIS FOR DOCUMENTARY LINGUISTICS AND ORAL HISTORY RESEARCH, ILLUSTRATED BY SAAMI AND KOMI EXAMPLES

*Michael Rießler and Joshua Wilbur*¹

Introduction

In this chapter, we argue that documentary linguistics, particularly as we practice it in our own projects, can provide valuable resources for social science research. Especially in our experience as documenters and researchers of endangered Uralic languages spoken in the Arctic, our projects and the data we collect can be considered an additional source for future oral history studies.

We first briefly discuss the history of the relationship between oral history studies and linguistics (specifically documentary linguistics), or more appropriately the lack of any significant relationship. We then go on to explore current social science projects in the Arctic and how these have the potential to relate to language documentation as a background to the further discussion. The next section deals with documentary linguistics in general, as a field within digital humanities, and how it strives to collect multifunctional, quality data for long-term preservation of linguistic and cultural knowledge. In this, it has significant potential to inform the field of oral history and other social sciences. Specifically, even if the sizes of our text collections pale in comparison with corpora for major European languages, our heavily annotated (“thick”, see § *Oral History, Language Technology and “Thick” Metadata Descriptions* below) metadata² make our texts qualitatively rich, and more accessible, even to oral historians, anthropologists and other non-linguists, in part because they include translations into the respective majority languages as well as English. Indeed, our extensive metadata provide relevant background information on speakers, the recording session

-
- 1 The order of the authors’ names is alphabetical. We would like to express our gratefulness to Rogier Blokland, Stephan Dudeck, Erich Kasten, Sonya Kinsey, Niko Partanen and Katja Roller for their insightful comments on our paper.
 - 2 As is common in documentary linguistics, we use the term “metadata” to refer to data *about* the content of recordings or other primary data, for instance dates, names, or descriptions of recording events. This can be compared to “annotations”, which typically mean textualizations of primary recordings, such as transcriptions and translations. In the strict sense, however, annotations are of course a special subtype of metadata *about* primary data (the text itself).

itself, as well as project details; this then allows non-linguists not only to contextualize recordings, but also to filter them for potentially relevant categories such as location, recording date, and, most importantly, topical keywords. We follow this up with several examples from our own documentation projects that illustrate how transcribed recordings from such projects often include narratives that are of potential interest for oral historians. Specifically, we provide excerpts from narratives by speakers of the Uralic languages Izhva Komi, Kildin Saami, Skolt Saami and Pite Saami.

Finally, we present our vision for creating extensively linked documentations of oral histories for future research in both linguistics and the social sciences. In order to ensure the creation of multifunctional and sustainable databases for multiple fields of research, we highlight some best practices for digital data archiving and publication. Furthermore, we indicate how tagging the metadata and annotations for relational linking and keywords on relevant subjects in individual recordings allows searches both within a single archive, as well as in connection with external archives and/or search engines, for both linguistics and other disciplines, thus increasing discoverability.

Oral History and Linguistics – a Missing Link?

To begin with, we want to point out that when we use the term “oral history” we are primarily referring to an oral speech genre that consists of an individual’s narrative about a historical event that he/she personally observed or participated in (this is a countable noun, as in “a collection of oral histories”). These are the oral accounts that (re)present vernacular historical knowledge or perspectives of the individuals that tell them. The same phrase is also used by the entire social science discipline that utilizes such texts to come to conclusions about such events (this is uncountable, as in “oral history as a discipline has become established relatively recently”). Note, however, that we do not use the term to refer exclusively to any abstracted analyses that are based on the study of such narratives.

A seeming lack of a relationship between social scientists and documentary linguists concerning the collection and analysis of oral histories appears to have been around for as long as both disciplines have co-existed. For instance, in discussions on oral history, linguists or the like are clearly missing, even in recent works such as in Freund et al.’s extensive list of oral history practitioners. This includes “archivists, historians, geographers, ethnographers, ethnomusicologists, folklorists, educators, museum curators, journalists, broadcasters, and authors” (Freund et al. 2015: 7), but does not mention anyone from language sciences.³ This seems to be the case despite

3 This is perhaps even more surprising considering the obvious use of oral history materials for descriptive linguistics, as evidenced by Katja Roller in chapter 7 in this volume.

a goal which is clearly common to both disciplines: again, Freund et al. mention that oral historians “attempt to stimulate the narration of stories ... rather than simply recording short responses as in a questionnaire or survey” (Freund et al. 2015: 7). While the goals of the research done on the resulting material differ significantly, the raw materials themselves are strikingly similar, although the breadth of topics that documentary linguists collect is typically broader than that of oral historians whose focus is on historical events. Indeed, at the very least concerning the observer’s paradox, documentary linguistics can certainly stand to learn from oral historians, who, again according to Freund et al., “found that in their relationship with interviewees they could not claim to be detached and objective” as far back as the 1950s (Freund et al. 2015: 7). Furthermore, the ethical obligations of researchers (typically outsiders from the majority group, although not exclusively so) towards the members of minority groups they work with have obvious parallels (cf. Freund et al. 2015: 8). Finally, a number of shared practical challenges concerning the preservation and presentation are shared by both disciplines, again emphasizing how a mutual dialog could benefit everyone involved (cf. e.g. the chapters in part 3 “Preservation and Presentation” in Llewellyn et al. 2015).

In linguistics, this missing link is not quite as obvious, at least in the discourse around documentary linguistics.⁴ It is most obvious in Himmelmann’s seminal paper “Language Documentation: What is it and what is it good for?”, in which the author lists three fields of research as examples for disciplines with a potential interest in language documentations: “linguistics, anthropology, oral history, etc.” (Himmelmann 2006: 2). The author further mentions that interdisciplinary projects working on language documentation aiming for truly comprehensive results would include “anthropology, ethnomusicology, oral history and literature” (Himmelmann 2006: 15). Himmelmann even concludes that “[d]ocumentary work that aims at a truly comprehensive record of a language also has to engage with ethnobotany, musicology, human geography, oral history, and so on” (Himmelmann 2006: 28). Aside from Himmelmann, oral history studies are occasionally mentioned in name, although never in much detail (cf. e.g. Bower 2011: 464, 480; Woodbury 2011: 162). In fact, Woodbury even mentions that one of his first language documentation projects arose out of a local oral history project (Woodbury 2003: 9). Otherwise, oral history can only be inferred in formulations concerning the potential audience of language documentation projects such as creating a “multi-purpose record” (Himmelmann 2006: 1).

4 See § *Oral Histories in Endangered Language Documentation* (below) for a more detailed description of what exactly this sub-discipline entails.

Oral Histories in the Arctic – Examples of Ongoing Projects in the Social Sciences

The title of this section is inspired by the project “Oral History of Empires by Elders in the Arctic (ORHELIA)”⁵ which was carried out by a team of anthropologists led by Florian Stammer at the University of Lapland between 2011 and 2015. This project can be taken as an example for recent anthropological approaches to oral history research in the Arctic during the last century. The arctic cultures investigated by ORHELIA include different Saami and Komi groups and therefore overlap considerably with our own work. Last but not least, it was the collaboration with Florian Stammer and the ORHELIA group which initially attracted our own special interest in oral history research.

Our observation that ORHELIA recorded oral histories almost exclusively in the respective majority languages (Russian and Finnish, cf. the community DVD published by the project Dudeck et al. 2015),⁶ is particularly relevant for the discussion in the present paper. This is despite the fact that the project descriptions stress “indigeneity” as well as the common language roots of the different Saami, Nenets and Komi groups (which all speak Uralic languages) investigated by ORHELIA and linked to this “indigeneity”. A similar approach using non-native languages predominantly (or exclusively) seems characteristic in other projects as well,⁷ although there are obvious exceptions such as those thoroughly discussed by Erich Kasten in chapter 1, or the studies presented in the book series *Languages and Cultures of the Russian Far East* (by the same publisher as the current volume).⁸

Two further examples of projects working in the same region and on similar topics are the recent works by Lukas Allemann⁹ (2010, translated into English in 2013) and Anna Afanasyeva (2013). Whereas the forced relocations of Kola Saami during the Soviet Union was the specific topic of Afanasyeva’s study, relocations were only one of several topics in Allemann’s study on the history of Saami people during the Soviet Union. The data analyzed in both studies consists mostly of oral history recordings done with ethnic Kola Saami informants. Note that all informants mentioned by Afanasyeva and Allemann are fully bilingual in Saami and Russian (they are mentioned by name in Allemann 2010: 127–128, and personally known to us; Afanasyeva anonymizes her informants, but their proficiency in Saami can be inferred with near certainty from the given birth years between 1930 and 1940).

5 <http://www.arcticcentre.org/EN/research/anthropology/ORHELIA> [20.02.2017]

6 The only exception being the sub-project on Tundra Nenets (Florian Stammer, pc, 20.02.2017).

7 Cf. e.g. the seeming lack of discussion on the role that information provided in native languages as opposed to lingua franca in the contributions in Ziker and Stammer (2011), with the exceptions of chapter 5 and chapter 8, which are notably also the only contributions by linguists.

8 http://www.siberian-studies.org/publications/lc_E.html [20.02.2017]

9 Lukas Allemann joined the ORHELIA team after the project.

Like ORHELIA, Afanasyeva's and Allemann's studies cover topics related to Saami ethnic identity under Russian assimilation pressure. Generally speaking these and similar projects are typically influenced by contemporary frameworks in so-called indigenous studies and often include the specific aim to record local (or "indigenous") memories and knowledge and preserve these for later generations. The ORHELIA project description states this specifically:

[...] the project also contributed to preserve [sic] incorporeal cultural heritage among Uralic speaking northern minorities of Europe and study [sic] the transmission of historical heritage between different generations.¹⁰

As linguists we are particularly interested in the different *languages* spoken by Arctic people and how *communication through language(s)* functions in Arctic societies, and we were therefore surprised to observe that the majority of oral history recordings by ORHELIA (and similar projects) were done in the corresponding majority languages. This was especially unexpected for us because methodological discussions of the oral history approach (e.g. by Dudeck and Allemann 2016) specifically mention the importance of interpreting the *context* in which the life stories are told by the informants and stress the *polyphone character* of oral histories as subjective accounts of one's personal life history:

Ein solch polyphones Verständnis von Oral History setzt jedoch auch voraus, dass sich die Forscher und Forscherinnen mit dem Kontext beschäftigen, in dem die Geschichten erzählt werden und in den sie sozial und kulturell eingebettet sind. Das Zelt oder die Wohnung, in der wir eine Geschichte zu hören bekommen, steht nicht in einem Vakuum. Die Art der Kommunikation wird von Konventionen beeinflusst, wie sie beispielsweise in einer von Rentierzucht geprägten Gemeinschaft generell vorherrschen. (Dudeck and Allemann 2016: 85)¹¹

There is of course an obvious practical reason for the predominant use of majority languages for oral history interviews, namely when a researcher team does not have any members sufficiently fluent in the relevant native language or languages. In many other cases the informants themselves may also choose the majority language, perhaps because they simply prefer to do so, because they are not fluent speakers of the native language either, or for other reasons.

¹⁰ <http://www.arcticcentre.org/EN/research/anthropology/ORHELIA> [20.02.2017]

¹¹ "Nonetheless, such a polyphonic understanding of oral history also requires researchers to deal with the context in which the stories are told and in which they are embedded socially and culturally. The tent or the apartment in which we get to hear a story is not located in a vacuum. The type of communication is influenced by conventions such as for instance those that generally prevail in a community shaped by reindeer husbandry." (our translation)

What is striking to us, however, is that the multilingual context in which all Arctic indigenous societies exist today has scarcely been considered relevant by oral history researchers in social sciences. We are not aware of any ongoing methodological discussion about the potential role of language choice by the interviewer and the interviewee in oral history research. We believe, however, that the following questions are methodologically valid and should to be taken up in future research:

1. Does the choice of language influence the results of the transmission of historical heritage and knowledge between different generations, or between informant and researcher?
2. If the answer is yes, what should be best practice in oral history research, i.e. in what language(s) should data be collected, especially when informant and researcher are not native speakers of the same language?

One aim of the present chapter is to try to stimulate such a discussion in the future.

Oral Histories in Endangered Language Documentation

Language documentation (also referred to as documentary linguistics) is an emerging sub-field of applied linguistics. Research in language documentation aims at the provision of long lasting, comprehensive, multi-faceted and multi-purpose records of linguistic practices characteristic of a given speech community, often in conditions where these languages are under threat of disappearing (cf. Himmelmann 2006; Woodbury 2011; Austin 2014). Although it evolved out of traditional fieldwork methodology used primarily by descriptive linguists and language anthropologists, language documentation is no longer merely a method, as it has its own primary aims and methodologies. One of the most important purposes of language documentation is ensuring that data are available for further research on and for endangered languages, for both further theoretical and applied research, as well as for direct use by the respective language communities.

Ideally, the data pool provided by the language documenter includes a comprehensive, deeply annotated and easily accessible corpus¹² of primary language recordings, representing a wide variety of texts in terms of chronology (e.g. age of recorded speakers), geography (e.g. dialects), and other sociolinguistic variables (e.g. gender and educational background of speakers, speech registers, text genres, etc.). In addition to annotations, cataloging metadata are crucial in ensuring the intellectual accessibility of the documented data and concern both the *content* of the recorded speech sample (typically represented as orthographic or phonological transcriptions, morphosyntactic tagging, and free translations into other languages) as well as the *context*

12 The term “corpus” is typically used in linguistics to mean a database which is systematically annotated for specific data features in order to investigate a research question using a data-driven and quantitative approach.

(such as actors, places, and speech events, but even meta-documentation about the project itself, cf. Austin 2013).

Along with methodologies and best practices related to fieldwork and archiving (including questions of research ethics, protection of copyrights, resource discoverability, data standards and long term data preservation), the usefulness of the actual product of language documentation for linguistic research hinges on the quality and quantity of digitally accessible *annotations* as the basis for further analyses and data derivations. With the awareness of such collections and the increase in the quality and quantity of a number of such collections, the use of language documentations for scientifically significant corpus-based investigations on endangered and lesser-known languages as well as the role of computational methods in this have frequently been a driving topic in recent years. (For a specific discussion of these questions on our own projects, cf. Gerstenberger et al. 2016.)

While the data typically gathered in endangered language documentation projects correspond to a wide variety of genres, a common type of recording can clearly be considered oral history. Our main motivation as linguists collecting such recordings is to secure non-elicited, unplanned examples of the target language in a spoken modality on topics that speakers can relate to in a comfortable, relaxed and natural way. Precisely this approach aligns well with expectations of oral history. It is just such recordings that can prove to be valuable multi-functional sources for other disciplines as well, particularly historically oriented social sciences.

The following section provides a few concrete examples from our language documentation projects on Pite Saami, Kildin Saami, Skolt Saami and Izhva Komi for recorded and annotated texts in these native languages which could potentially serve as useful sources for oral history or other social science research looking at Arctic peoples.

Case Studies

The languages which our own documentation projects are concerned with are all spoken in the Barents Sea region and belong to the Saamic and Permic branches of the Uralic language family. The data, which we discuss below, stem from our own projects: the Pite Saami Documentation Project, which has been carried out by Joshua Wilbur since 2008 and which works with the Pite Saami language spoken in northern Sweden around Arjeplog (cf. Wilbur 2008–2017); the Kola Saami Documentation Project, which has been carried out by Michael Rießler (and collaborators) since 2005 and which works with all four Saami languages spoken (or formerly spoken) on the Kola Peninsula in Russia (cf. Rießler 2005–2017); and the Izhva Komi Documentation Project by Michael Rießler together with Rogier Blokland, Marina Fedina, and Niko Partanen (and other collaborators) which works with speakers of the Izhva

Komi dialect diaspora both within and outside the Komi Republic (cf. Blokland et al. 2009–2017).

All texts collected by our projects in the field are available at least in audio, and many also include video. In addition to our own field work data, we include available legacy data in our archives whenever possible. By “legacy data” we mean for instance fieldwork data collected by other projects (annotated or not) and stored in various language archives, as well as spoken texts which were transcribed, translated and published in books and are available with or without original audio/video recordings.

Further processing of legacy data basically follows the same processing as with our own fieldwork data, and thus includes segmentation into utterances in the ELAN program,¹³ followed by orthographic transcription and translation into at least one (inter)national language. The majority of our data are transcribed and translated into either English or Russian.

As indicated above, oral histories have been collected by researchers in diverse fields, although no special framework of oral history research seems to exist in which methodology or theory would overlap between these different fields. On the one hand, in our own projects, oral history is but one of many categories found in our collected texts (more detail on this below in § *Oral History, Language Technology and “Thick” Metadata Descriptions*) since we as linguists do not have a particular focus on oral history. On the other hand, anthropologists and historians use oral history as a framework, but without any specific intention to “document” when the collected oral histories are simply a tool used to provide an empirical foundation for further theoretical study. Still other projects collecting data on the cultures and languages mentioned here neither “document” nor carry out “oral history” research, but nonetheless collect oral history without mentioning this specifically, for instance the work by the Norwegian political scientists Overland and Mikkel-Berg (2012), which is also partly based on fieldwork interviews. Nonetheless, one of the most typical goals of documentary linguistics is the recording of “natural” language, and precisely this goal provides significant common ground with oral history and fieldwork-based research in social sciences in general.

With this in mind, the following sections present some examples from our text collections that show how oral history stands to benefit from language documentation, even if such texts were not initially collected with oral history in mind. Topics covered include for instance reindeer husbandry, life in Soviet times, or the introduction of modern technologies. These examples are also presented here to illustrate the following discussion about how not only the texts alone, but our extensive “thick” (see below § *Oral History, Language Technology and “Thick” Metadata Descriptions*) metadata can and should be utilized by non-linguistics researchers not only as an

13 ELAN (EUDICO Linguistic Annotator) is free software for annotating and presenting multi-modal language data, and developed by the Technical Group of the Max Planck Institute for Psycholinguistics; <https://tla.mpi.nl/tools/tla-tools/elan/> [20.02.2017]

access point, but as a source of contextualization. After presenting some best practices for both archiving and publishing such texts, we share our vision for a unified digital infrastructure allowing access to a large number of documentations from various archives and with various target users in mind.

Pite Saami

Pite Saami is an indigenous language spoken in and around Arjeplog municipality in Swedish Lapland, and historically in adjacent territory in Norway. While there may be more than a thousand ethnic Pite Saami individuals (as stated, for instance, by Krauss 1997: 24), the language has suffered significantly under dominant Swedish/Norwegian social, cultural and political pressure over the course of the last half century, such that practically an entire generation of parents ceased to pass the language on to their children, resulting in a likely irreparable break in the transmission of the language (Wilbur 2014: 6–7; Valijärvi and Wilbur 2011). Currently, Pite Saami has around forty speakers.

The Swedish anthropologist Ernst Manker compiled a significant amount of research on various Saami groups, a small part of which concerns Pite Saami, particularly reindeer herders (cf. eg. Manker 1947). Otherwise, the only studies in the social sciences specifically concerning Pite Saami are very recent and can be found in Evjen and Myrvoll (2015). However, linguists have been studying Pite Saami for more than a century, and a number of texts are available which are potentially very interesting for oral history researchers. The earliest texts are short narratives with Hungarian translations in Halász 1893 and with German translations in Lagercrantz 1957 (but originally transcribed in 1921). The Swedish state agency *Institutet för språk och folkminnen* (ISOF)¹⁴ has an archive which includes a significant collection of Pite Saami legacy materials. While many of these texts are, strictly speaking, written (even if presenting a spoken modality of language), the archive also includes several hundred untranscribed Pite Saami recordings by Israel Ruong, Olavi Korhonen and others, recorded throughout the second half of the 20th century. While we have not yet had the opportunity to examine these written and audio texts in any detail, particularly concerning their potential usefulness for oral history, a trip to the archive in Uppsala is planned for summer 2017, with the intention of eventually ingesting as many of these heritage texts as possible into our data collection.

The Pite Saami recordings we are most familiar with stem from the Pite Saami Documentation Project, which has been carried out by Joshua Wilbur since 2008, and it is these recordings we will focus on here. All recordings are transcribed and translated into either English or Swedish or both. A number of these recordings seem to be exemplary for how language documentation can provide a valuable source for

14 Swedish Institute for Language and Folklore: <http://www.sprakochfolkminnen.se> [01.03.2017]

oral history as these present insights into Saami culture as primary sources portraying Pite Saami life in previous decades, including the changes it has undergone in Sweden, by individuals who experienced these firsthand. One reason these recordings are particularly valuable is because they contain a number of interviews with settled Saami, i.e. those who survived mainly on fishing and small-scale agriculture, in addition to reindeer herding Saami, who tend to receive a disproportionately large amount of focus in studies on Saami culture and history (relative to the actual proportion of the Saami population in the whole of northern Europe).

In example (1), from a recording done in 2015, a reindeer herder (born in 1977) thumbs through his journal and recalls specific details about the highly unusual weather during the winter of 2014–2015:

(1) *Men dā lij gu buhtin jåvlå vuásste, dā älgij nievrut huj spajjta, ... dā lij muohta bāhtam, ja dā huj garra bivval budij. ... årrå jage biejeven dā lij gákkkse plus grader.*

“But then as we got closer to Christmas, it started to get bad really fast, ... the snow came and then really warm weather came along, ... on New Year’s Day it was six degrees above freezing.” (sje20150329b.056-061)¹⁵

While this only documents the weather during one specific winter, it nonetheless bears witness to the effects of climate change on reindeer herding. It will potentially be a useful source for future oral history research on the effects of climate change, especially if adequate metadata on its content are available (as is the case for any archived recording; see § *Oral History, Language Technology and “Thick” Metadata Descriptions*, below).¹⁶

In another recording, an elderly speaker (born in 1927) presents some memories from her childhood, at which point her family still migrated seasonally with their reindeer. In this particular example (2), which was recorded in 2009, she recollects what her family did while in the village of Arjeplog on special occasions such as church holidays or market days, including their relationship with tourists.

(2) *Da gāde ma dāle lä, da lä sjaddam manjel, ja dānne inijme del omassev duogajme, dā slöjdojd ja gajka, dujjid ja málestijme guasmagav ja duogajme aj dajda turistajda ma buhtin.*

“Those huts which are here now, they were put up later, and we had all kinds of things here, we sold handicrafts and such, and we prepared coffee and also sold those things to tourists who came here.” (pito90915.254-256)

15 Here and below, references in parenthesis after examples from our data indicate the session name (here: sje20150329b) and sentence/utterance numbers (here: 56 through 61) in the relevant data collections so that the source data can be located in the archives.

16 A good example for how primary oral history research can be used in scientific studies on climate change and reindeer herding in the larger Arctic can be found in Forbes et al. (2016).

With a similar focus on past lifestyles, in the following example (3), recorded in 2013, another speaker (born in 1949) provides insight not only into life on a subsistence farm by providing a simple inventory of the family's animals:

(3) *Jå, dubben hiejman gålmå gusajd inijme, ja gålbmå gajtsa, ja hiestav.*
 “Yes, at home there we had three cows, and three goats, and a horse.”
 (sje20131025.072-075)

But she also discusses some of the personal frustrations and limitations she experienced while living there in (4):

(4) *Men då sujttijiv fiksav, dajd gajk judosijd, ja biebmov mån lägiv ja bähkkujiv ja men itjiv då dárbahe bassat, då muv äddne dav dágaj. Ja då nagin då turista buhtin diht då almatj inij nagan biednegijd. Itjiv mån ietjá biednegijd ane, men ij lam del nåv nävvre danne árrot, men almatj sidaj ulgus vuállget, ja kan lij nagin suohtas radnav gávdam nagin sájen, men idtjiv. Mån iv diede jus muv äddne itjij sida att mån galgav naginav adnet uddne.*

“So I took care of the barn, all the animals, and I prepared food and baked, but I didn't have to wash up, my mother did that. And when some tourists showed up, then we had some money. I didn't have any other money, but it still wasn't so bad living there, and yet one wanted to get out into the world more, and maybe to have met some fun friend some place, but I didn't. I don't know if my mother didn't want me to have anyone today.” (sje20131025.023-033)

The following example (5), recorded in 2009, illustrates some of the inherent risks involved in depending on reindeer herding as a way of life. Here, after relating the story of how a large portion of her husband's family's reindeer herd was lost because all but the strongest reindeer drowned while swimming across a large lake, the speaker (born in 1927) indicates the significant effects this event had on her husband's family.

(5) *ja då virrtijin häjttet då gu då iello sággoj. Dä idtjin disste dárbahe jáhtet ... ja då dale då genugin då dasa dán Áhkaj*
 “and so they had to stop [herding] after the herd drowned. They didn't have to migrate any more ... and then they settled here at Áhka.” (pito90609b.029-033)

In continuing the themes from the recording in the previous example, in the following example (6), also from 2009, the same speaker's husband (born in 1927) talks about some of the activities his family were dependent on for survival after losing their reindeer herd (as described in example 5).

(6) *Dä inijmä gusajd, gålma nällje gossa, ja så dä gájtsajd ja dä dajna vanj viesojme ja ja dä fiskodijme ja inijme del nan buhtsuv. Men nähköm lidjin vanj, da uvduť ajge lidjin urrum nāv gārrdok, att nuhkin buhtsu, så idťjin del mādde bātse, så dä sladjime danne ja pirunijd sadjijme ja ārojme ja viesojme gulij ja mielkijn ja dä slaktijme nan gusav.*

“Well we had cows, three or four cows, and goats, and we lived with them, and we fished and had some reindeer. But then they [the reindeer] were gone, it was tough in the old days, that the reindeer were gone, there weren’t many left over, so we harvested hay here, and planted potatoes, and stayed and lived on the fish and the milk, and sometimes we slaughtered a cow.” (pito90609a.009-018)

In this final example (7), again from 2009, this speaker (born in 1954) indicates how modern technology has been both a blessing and a curse to life in remote villages such as the one she grew up in before moving to town.

(7) *Mij lip aj adnám telefonav dán dāben, ja dun stuordāben aj. Men dā, jáhkav lā guākt-jage urrum, dā Telia vāldij bārtā dav telefonav, ja mān iv dav ābbā tuhtje. Da āddā telefona, da mobiltelefona, da lij, eller nāv buoraga, halva ij gullu danne, ja gu ij lā akktak eliktrisitehta dānne, dā ij almatj māhte dajd lāddit dānne, så dā lā tjiervas Mān sidav dav gambal telefonav ruopto.*

“We also had a telephone in this house, and in the big house, too. But then, I believe it has been two years since Telia took away the telephone—I didn’t like that very much. These new telephones, these mobile phones, they are quite good, but they almost don’t belong here, and since there isn’t any electricity here, you can’t charge them here, so that is difficult. I want that old telephone back.” (pito90823.151-161)

Kola Saami

Kola Saami is the common denomination for the four Saami languages of northwestern Russia (sometimes including and sometimes excluding the Skolt Saami now living in Norway and Finland). Today, Russian is the dominating language in practically all domains of life for the Kola Saami groups of Russia. Contacts with Russians are centuries old and go back to the establishment of the first orthodox monasteries in the area in the 16th century and subsequent Russian colonization. Russian cultural and linguistic influence culminated after the creation of the Soviet Union in the 20th century when a significant number of Russian speaking people moved to the area as a result of the industrialization and militarization on the Kola Peninsula (cf. Siegl and Rießler 2015).

Like the other languages described in this chapter, the Kola Saami languages have been studied for more than a century and a variety of texts have been collected which

are potentially very interesting for oral history researchers. Unlike Pite Saami, there are published legacy recordings relevant for oral history because portions of these text collections could be described as primary oral history sources. In addition, there are many more unprocessed recordings stored in archives, for instance in Helsinki, Petrozavodsk, and Tallinn.

The four Kola Saami languages are Ter Saami, Kildin Saami, Akkala Saami, and Skolt Saami. Kildin Saami is spoken by a total number of around 500 native speakers,¹⁷ most of whom live in the municipality of Lovozero, where the majority of them were forced to resettle to during the 1950s and 1960s.¹⁸ Originally, Kildin Saami was spoken all over the central inland parts and the north-central coastal parts of the Kola Peninsula, and the language had several significant dialectal variants.

The neighboring Kola Saami dialects in the northwest belong to the Skolt Saami language, which is hardly spoken in Russia any longer. After their families were forced to leave Russia when the Winter War broke out in 1939, most speakers of Skolt Saami moved to Sevettijärvi, Nellim and other places in the Finnish municipality of Inari, where they still live today. The total number of Skolt Saami speakers is roughly similar to Kildin Saami (cf. Siegl and Rießler 2015), but basically all Skolt Saami speakers live in Finland today.

Ter Saami dialects were formerly spoken in the eastern parts of the peninsula, but there are practically no Ter Saami speakers left in these areas today. The last Ter Saami speakers live in various other places such as Lovozero, Murmansk or elsewhere (Scheller 2011a). The fourth Kola Saami language, Akkala Saami, was originally spoken to the southwest of the Kildin Saami dialect area, but is also moribund or perhaps even already extinct (Rantala et al. 2009; Scheller 2011a).

The examples we provide here are from Kildin and Skolt Saami, which are the most vital of the Kola Saami languages. The following short extracts provide not only instances of oral history in our own data, but describe also the approach taken in our projects to include legacy data into our corpora.

Kildin Saami

The first example from Kildin Saami is from an original speech recording stored at the spoken language archive at the Institute for the Languages of Finland in Helsinki.¹⁹ Metadata in the archive's catalogue are sparse, but we found out that the recording was done in Petrozavodsk (Karelian Republic, Russia) by the Finnish linguist Terho Itkonen on 8th June 1965 on analogue tape, and only recently digitized in Helsinki.

17 Cf. Rießler 2013 and the estimates by Scheller (2011), who differentiates between “active” and “passive” speakers.

18 Cf. also the studies by Allemann (2010) and Afanasyeva (2013) mentioned above.

19 <http://www.kotus.fi> [20.02.2017]

The recorded speaker is Pavel Polikarpovich Yuryev (1936–1983) from the town of Lovozero. Additional pieces of metadata were recorded in Finnish at the beginning of the tape and include the exact location (the recording was done at the Karelian branch of the Institute of Language and Literature of the Academy of Sciences of the USSR), the profession of the speaker (geography teacher) and the reason for his stay in Petrozavodsk (he was participating in a course). The recording is 32 minutes long and includes four different topics about the life and work of the Saami: descriptions of fishing in former and contemporary times, the seasonal weather changes, making skis, as well as the oral history described below. The speaker talks almost exclusively in monologues, with only very little interference by the interviewer. All questions asked by the interviewer are in Russian, while the speaker consistently answers in Kildin Saami, although he uses many loanwords and other Russian-language influences are plentiful.

Surprisingly, we found the same recording completely transcribed phonetically and translated into English in an unpublished M.A. thesis from Indiana University (Bjarnson 1976a). We have digitized these annotations, re-aligned them with the digital audio data, and added additional annotations, such as an orthographic transcription in standard written Kildin Saami, a Russian translation, and a few additions and corrections to the original annotations. According to our conventions concerning legacy data, the original annotations are kept unchanged and all new annotations are added as additional layers.

We think that one part of the recording is especially interesting from the perspective of oral history. In this 8 minute extract, the speaker remembers a school outing to the Tundra which he organized for his students from the boarding school in Lovozero in the winter of 1963. The story starts like this:

(8) *Кудтлоагкь куалмант ыгесът январь мәнэсьт мунн учениками авцант клāссэсьт выййлэмь чйррэ, нүдзэгүэйм. Тэнн райя мыйй сōāгэмь соанэть запрāватъ кўнт харянь. Колхозс энтьэ мийенъ вийем нүдзэть, мыйй вāльтэмь нёрк мйльэтэ, мāльцетъ, ниматъ, туберкэть, савехетъ, рōāлхэть, сāллас, кōмпас, тетрāдэть, фатанарāт оаккиэ э выййлэмь чйррэ. Биштэмь соан эл кутэ-колмэ олмэ.*

“In the year of 1963, in January, me and my ninth grade students went to the tundra by reindeer. For this purpose we got reindeer sleighs, with equipment and two wooden reindeer driving sticks. The kolkhoz²⁰ gave us draught reindeer, we took along food, fur parkas, fur stockings, fur boots, skis, poles, tarpaulins, compasses, textbooks, cameras, an ax and we left for the tundra. We sat on sleighs in twos and threes.” (sjd19650608kotus5493-1az.200-279)

20 A kolkhoz was a form of collective farm in the Soviet Union (and other communist countries).

At the time it was recorded, this story was still quite recent for the speaker, and it includes interesting personal accounts on life and work in Lovozero in the 1960s. Note that Pavel Yurev is a local celebrity²¹ in Lovozero and his biography is well-known to most local people (who did not know, however, about the existence of this old recording of him and were excited about it when we played it to them). He is especially famous for the after school club for “Young Reindeer Herders”, which he established and for which he also led excursions like in the one described here. Yurev was also a local historian and the founder of a small museum at the boarding school, which later developed into the Museum for the History and Culture of the Kola Saami People.²²

The 1965 interview described here is an oral history account of Pavel Yurev’s life and work. It also provides some details about the zeitgeist in the local Soviet society of Lovozero, where for instance schooling, spare time activities, and work at the kolkhoz were intertwined. For instance, Yurev describes how the kolkhoz, as the most important entrepreneur and potential future employer of the school pupils, supported the outing with material and personnel.

Finally, this oral history also provides us with a typical example of discourse *about* the Soviet Union.

(9) Тэсьт ... унчтэль ... географья, мунн раз- мойнсэ парнатъ пугк чър баяс ... кōххт адть изменился чърр, тэяя ыгкя ... Кōххт ёлешкуэдтэшь пэря ... И так далее.

“Here, (I as) the teacher of geography, I told the kids all about the tundra, how the tundra has changed now, up to now ... How we’ve started living better ... And so on.” (sjd19650608kotus5493-1az.255-259)

The changes in the tundra and the lives of the local people mentioned here refer of course to the alleged improvements in the economy and society in the Soviet Union. Such standard propaganda phrases were obligatory in official speeches. As a teacher and communist representing the town of Lovozero in a recorded interview (even given to a foreigner), the speaker presumably felt obliged to build such clichés into his speech.

Another example of an oral history recorded in Kildin Saami in the 1960s and included as legacy data into our corpora is about the personal experiences of Lazar Dmitrievich Yakovlev (1916–1993) from the village of Kildin. He tells about his own experiences as a primary school pupil in his village, his later studies at the pedagogical college in Murmansk, and the boarding school in Lovozero (the same as in the oral history above), where he was working as a teacher when the interview was recorded by the Russian linguist Georgi Kert in Lovozero in 1960.

21 Note also the entry about him in the online encyclopedia *Kolskij Sever*, http://lexicon.dobrohot.org/index.php/ЮРЬЕВ_Павел_Поликарпович [20.02.2017]

22 Cf. https://ru.wikipedia.org/wiki/Музей_истории_кольских_саамов [20.02.2017]

(10) *Мйн сыйтэсът, Кйллт сыйтэсът лййй шкōла уже мауна түййиэнч. Тэnn шкōласът ðһnnуввэнь сāmь парна, кӯһт(эмн)лоагкь и выдт(эмн) лоагкь пāррьиэнче. Сйнегуэйм муnn тэсът-иэ ðһnnуве.*

“In our village, in the Kildin village, the school only was built later. Saami children learned in this school, twelve (girls) and fifteen boys. I learned with them, too.” (sjd1960000oldjkert1961a-1.04-06)

In this text, the interviewee also mentions a historically significant event: the arrival of the first native Saami teacher Ivan Andreevich Osipov after his studies at the Institute of the People of the North at Herzen Institute in Leningrad. Ivan Osipov started working in Lazar Yakovlev’s village school in the 1930s.

(11) *Мауна нуэдтэль мйн сыййта Осипов Эвван Вуэnnьтре алльк, тэдта пёрвэ сāmь вуэnnьсэй.*

“Later, Ivan Andreevich Osipov, the first Saami teacher, came to us.” (sjd1960000oldjkert1961a-1.21)

Last but not least, typical communist propaganda phrases are also included in this interview.

(12) *Адтб сāmь парна, ыжэм парна я рӯши, кōгк тэсът Луйвьрэсът лёв, аллькэв рōбуушиэ ёмьне альн, аллькэ куаййвэ ёмьне, кōххт сёмьтб сёйе и ðһnnуввэ, кōххт пэрямп пынне нӯдзэтб, штобэ ваньса коадхэ, штобэ нӯдтэ нӯдзэ лйһченб пэрямп, штобэ вял колхозс ðллма пэрямп ёлешкудтэв. Тэдта пайнэмүшиэ вёкяһт советскэ правительствэ и коммунистическэ пāртья. Сыйй лыһкэв пай пэрямп, штобэ сāmь ðллма ялченб пэрямп.*

“Now, the Saami, the Izhva Komi, and even the Russian children in Lovozero have started working on the fields, cultivating the soil, sowing, and learning how to herd reindeer better, in order to have fewer losses and have the reindeer grow better, in order for the people to live better in the kolkhoz. These improvements are due to the assistance from the Soviet government and the communist party. They do everything in order to let the Saami people live a better life.” (sjd1960000oldjkert1961a-1.41-43)

The interview with Lazar Yakovlev was printed as a phonemic transcript with a Russian translation in the text collection by Kert (1961). The original transcript was later converted into orthography by our project and included in our corpus.²³ The original audio is probably even available for this recording in digitized form in the Phonogram

23 This specific text was also included in a small Kildin Saami book called “Lazar Dmitrievich’s stories” published by the Kola Saami Documentation Project (Afanaseva and Riefler 2008).

archive of the Karelian Branch of The Russian Academy of Sciences in Petrozavodsk²⁴ and could be re-aligned to the written annotations. Unfortunately, the archive has not made its Saami language recordings available to the public.

These two examples of oral histories collected unintentionally, i.e. collected not for oral history research but as linguistic data, also illustrate both the potentials and the limitations of the approach of reconstructing oral history data from already existing collections. This approach seems to differ considerably from the fieldwork-based participation approach (described for instance by Dudeck and Allemann 2016), which pays careful attention of the (spacial, personal, etc.) specific context of the interview situation itself and the personal interaction between the interviewer and the interviewee. The approach we describe here comes much closer to “source criticism”, i.e. a critical analysis of (written) historical sources, which is a traditional method in historical sciences. However, as we will argue below, the inclusion of the *complete context* in the interpretation of oral history during fieldwork-based participation is also only an ideal and depends on the unplanned outcome of the interview, spontaneous reflexions by the researcher and his/her questions for the interviewee during the recording, as well as on the later interpretations.

Furthermore, using different layers of “thicker” (see below § *Oral History, Language Technology and “Thick” Metadata Descriptions*) and retroactively better contextualized metadata can help enable the reconstruction of context. Note also that the two examples of texts described above are not only linked to each other on different levels (for instance using “Lovozero”, “schooling”, “Saami teachers”, etc.) but are also linked to other collected histories with shared topics. Last but not least, they are linked to various other kinds of historical data found for instance in photographs, written primary or secondary sources, and they provide information on the Kola Saami people or local northern Russian history in general. For instance, the two teachers Pavel Yurev and Lazar Dmitrievich were colleagues at the same boarding school in Lovozero and both oral histories mention events frequently taken up in current social scientific research on Kola Saami or northern Russian societies. Note also that the book mentioned above with transcribed and translated interviews (among them several oral histories about WW II, work in the kolkhoz, schooling, etc.) by Kert (1961) as well as several of the non-printed interviews listed in the catalogues in the archives in Petrozavodsk or Helsinki have obviously never been thought of as historical sources by the numerous social scientists who have been carrying out research on the topics mentioned here.

24 According to the webpage, the Kola Saami audio collection has been digitized and this text is included in the catalogue available there, cf. <http://phonogr.krc.karelia.ru/section.php?id=27> [20.02.2017]

Skolt Saami

Skolt Saami is culturally and linguistically closely related to Kildin Saami, although both languages are not completely mutually intelligible and their writing systems are also standardized differently. The oral history we describe here was recorded with Zoya Mikhailovna Nosova (born in 1937 in the village of Muotka), one of the very few Skolt Saami speakers on the Russian side today. She is also considered by other community members to be the last remaining truly fluent speaker of Skolt Saami in Russia.

We want to describe this recording in detail for methodological reasons. The data originate from an ongoing project in visual arts carried out as a collaboration between a documentary filmmaker,²⁵ a language documenter (i.e. one of the present authors) and a native Skolt Saami speaker with the aim of collecting texts about the personal history of this speaker and her family during Soviet times. The topics touched in the interviews with Zoya Nosova are typical for oral history studies and include for instance: the protagonist's childhood in a kolkhoz where the family was forced to resettle after their village was closed down by the government, how this original collective farm was forcibly subsumed into a larger state-owned farm (which again resulted in forced resettlement to a new place), her life as a Saami teenager and young adult in the small town of Verkhnetulomsk, and the private and societal changes during perestroika. Similar topics have been central in anthropological and historical research on Northern Russian society during recent years. As a result, the interviews collected for this documentary film include valuable information for social science researchers.

Although "indigeneity" (and even exotic ideas about it) are often a driving force behind the interest in the topics of such projects to begin with, all documentaries we are familiar with concerning the Saami of Russia are done completely in a lingua franca (while perhaps including a few symbolic sentences in the native languages). The project described here is unique in having the protagonist speak exclusively in her native Skolt Saami language. This approach does not only seem sensible from the point of view of native representativeness, but it also results in many hours of recorded video useful for language documentation. In fact, the linguist (Michael Rießler) only agreed to participate in the project on the condition that the film is (predominantly) recorded in Skolt Saami and that the resulting materials are properly archived and made available for future research. Zoya Nosova, who would have been ready to be interviewed in Russian, agreed also readily under the condition that the final translations (as subtitles) are rechecked with her before launching the film.²⁶

25 Solvej Dufour Andersen, <http://www.solvejdufourandersen.com> [20.02.2017]

26 Having an agreement is hardly a given; indeed, as evidenced by Erich Kasten's experience with one filmmaker, portrayed in chapter 1 of this volume, neither is keeping such an agreement.

In addition to requiring significantly more working hours for interviews, transcriptions, and translations (compared to making films in a dominant majority language), there are several crucial differences between a filmmaker's approach and that of anthropological and/or linguistic fieldwork. Although the filmmaker has a "documentation" in mind and is also carrying out ongoing research on the biographic-historical topic of the planned film (i.e. by going to archives, doing preliminary interviews before the actual filming starts, and even when reflecting on events and posing questions while filming), the research aim is typically not primarily scientific but artistic. In interviews and filming, the director has to find a compromise between the biographic-historical topic developing through the continuing insight from interviews and the plot he or she has prepared in order to present the narrative of the film to a general public in a visually, artistically appealing way. This also includes the fact that technical questions of audio and video quality and the composition of recorded sound and pictures are more important than in normal scientific fieldwork. Even in documentary filmmaking, scenes are sometimes repeated until the director is satisfied. However, care must be taken when prioritizing filmmaking aspects, as this can cause the value of the data to be diminished significantly from a research point of view, as illustrated by several anecdotes in chapter 1.

Because the informant is aware of these priorities, she dramatizes the situation somewhat on occasion. And it is likely that this specific situation of filming affects her own attitude or willingness to speak about certain sensitive questions or how she answers them. However, we believe that the recorded materials are nevertheless very valuable and usable in scientific research, for instance oral history and language documentation, because there is no clear boundary between recordings done by social scientists or linguists and those done by documentary filmmakers. Indeed, these exist on a continuum between "spontaneous" and "dramatized".²⁷ Indeed, as highlighted by the observer's paradox, the very fact that filming is taking place creates a special situation, and researchers also stage their interviews, even if to a more limited extent (for instance by selecting certain backgrounds or activities, or choosing the location with the best lighting, etc.). In addition, researchers typically also have a bias towards topics that are currently trending. On the other hand, a good filmmaker is very well prepared, after spending months or years becoming familiar with the topic of the film, often more than a typical linguist who focuses on linguistic aspects and has only a fleeting familiarity with the non-linguistic topics involved. With this in mind, a filmmaker's approach, if it includes constant reflection and further questioning as in the described project, is much closer to the approach of anthropologists (Dudeck and Allemann 2016). On top of that, the quality of the images and audio is professional,

27 Interestingly, the example provided in chapter 1 (Erich Kasten) on the life history of the Even elder Nadezhda Barkavtova indicates that different recordings of one and the same story can exist at the different ends of this continuum, which in fact can provide a unique opportunity to compare the "spontaneous" with the "dramatized" version.

and typically much better than that recorded by linguists or anthropologists, who are at best amateur filmmakers. The raw film footage, which is ideally archived and available for research, covers not only the topic at hand, but everything discussed before and after the actual scenes used in the final documentary film, including material on other topics which may also be of interest. In this way, the context of the recordings and potentially even the degree to which the situation was “dramatized” can later be reconstructed. All in all, such recordings undoubtedly have significant potential to be useful sources for research in linguistics and the social sciences. In our specific case concerning the documentary film about the Skolt Saami speaker, we were fortunate because, on the one hand, the filmmaker was well aware of the relevance of her materials for research, and, on the other hand, both the filmmaker and the informant agreed to allow the materials to be archived and made available to researchers.

The following example (13) is an extract from the interview recorded in the town of Kola on the 4th September 2015. The recording took place outdoors, on the street where a former boarding school was situated and where a playground is now there in its place.

(13) *Täšt leäi internat, mij täšt mätt'tōōdin. Jiännai pärnzed le'jje, i sää'm i ruošš. I le'jje i Muörmaš- i Tuállämpäärna i Tuulomapäärna i ru'vddčuâkkaz mie'ldd. Le'jje tâk, Laplandija räija le'jje i Puljââu'rest. I Loparskast, i Kicast, i Tajbolast. I Šongast. Puk le'jje, i vøl le'jje mättam pärna â'lddla puättam. Le'jje mättam ruoššpäärna, säälдатпäärna le'jje.*

“This was the boarding school, here we went to school. There were many children, both Saami and Russian ones. And there were both the kids from Murmashi and from Verkhnetulomsk, and from Tuloma and even those who lived along the railroad. They were, from Laplandiya station and from Pulozero. And from Loparskaya, Kitsa, Taibola. And from Shonguy. From everywhere and there were even kids from far away. There were also a couple of Russian kids, children of soldiers.” (sms20150904Kola-Internat-007-013)²⁸

Zoya Nosova went to this vocational school between 1956 and 1958, so she was 19 when she started there. Upon being asked whether it was hard for her to live at the boarding school, she answers:

(14) *Lossâd leäi, tōndiōtt što meešt jee leämma ... Mij to'lko poorin internatpoormōōžž. Jââdin bännest. Täšt ij leämma puõ'lli čääcc. Nu i... jiiij pōsslōōjjin. Jiijän päikka jââ'ttiin to'lko õhttešt, tälvva. Vot, tâk rosttovkanikului i vot, to'lko*

28 As the film project is only in the works, the processing of these recordings is not finished and they are not archived yet. We therefore provide only a translation of the extracts without a transcript of the Skolt Saami original. No specific identifier for the single sentences/utterances is currently available.

kuõ'htt neä'ttel. A nu't pirree'jj täšt jeälstin. Da, lossâd leäi.

“It was hard, because we didn’t ... We only ate the boarding school food. They went to the sauna [at home]. But there was no hot water here [at the boarding school]. And ... we did the washing ourself. To our families we went only once, in winter. Well, the Christmas holidays, only two weeks. But the whole year we stayed here. Yes, it was hard.” (sms20150904Kola-Internat-044-052)

Later in the recording, Zoya Nosova repeats information which she already gave and continues speaking about additional memories from the boarding school. She does this in Russian, and not in Skolt Saami as before, because she is addressing her granddaughter and not the camera.

(15) *Холодно. Ветер с залива ... есть ... На втором этаже ... было видно, Кольский залив ... Магазин был и залив. Магазин был, вот здесь, вот где [кралечка]. Вот тут магазин был. Насыти не было. Нет, такой большой магазин был; в основном был рыбный, но и продавленный. Потому что привозили с моря рыбу и прямо выгружали. А магазин высоко к берегам был и ... ящиками рыбу. А мы отсюда продукты брали, с этого магазина.*
 “It is cold. There is wind from the bay ... From the second floor you could see ... the Kola Bay ... There was a shop and the bay. The shop was, well, over there where the Kralechka store is. Well, over there was the shop. The levee wasn’t there. No, it was such a large shop, basically a shop for fish, but also for selling things. Because they brought fish from the sea and unloaded the ships here directly. And the shop was a tall building at the shore and ... [they loaded] the fish into crates. And we got food from here, from this shop.”
 (sms20150904Kola-Internat-095-100)

Izhva Komi

The Izhva Komi (in Komi *izhvatas*, in Russian *izhemtsy*) is an ethnic group which came into being in the 18th century as a mixture of a number of Komi ethnic subgroups, primarily consisting of speakers of the Vym and Udora varieties in the western and north-western areas of the present-day Komi Republic. These people intermarried while moving northwards along the river Izhma, but also married Russians (mostly from Novgorod and Arkhangelsk) and later also Nenets, a reindeer-herding people in the very north of European Russia. The language spoken by the Izhva Komi is the northernmost variety of Komi-Zyrian, spoken both in the north of the Komi Republic and in a number of small diaspora settlements in a wide swath of territory from the Kola Peninsula in the northwest of European Russia to northwestern Siberia.

Whereas the language is vibrant (and learned by children as their first language) in the majority of places where it is spoken inside the Komi Republic, it is critically

endangered in most of the diaspora settlements. On the Kola Peninsula, where they have been existing in close interaction with Kola Saami people for more than a century, the number of ethnic Izhva Komi is at most one thousand. However, according to our estimates much less than half of them speak the language (cf. Blokland and Rießler 2011).

The Izhva Komi Documentation Project began in 2014 and has been collecting speech recordings in all areas where Izhva Komi live and has been systematically annotating legacy data from other projects and archives. In 2016, we launched a multimedia database²⁹ including the recordings we have annotated so far. Among them are also multiple examples of oral history.

Our recordings from the Kola Peninsula are often interlinked at several layers with the Kola Saami oral histories since both typically mention one another, as in the following extract from an oral history recorded in Izhva Komi with Marfa Maximovna Andreeva (born 1922) from Lovozero. The recording was done by Valentina Filippova, Paul Fryer and Paula Kokkonen in 2000 (cf. Kokkonen 2004) and transcribed and translated by the Izhva Komi Documentation Project in 2014.

(16) А сэсса тридцать втором году, кор раскулачитисны, дак сыа муні мамыс доре. Мамыскед оліс. А мян, ми бара код кыче. Митрей дяде муні. Э, мыйке, сылэн вöлі бабаыслэн вок да сыа пыыянсэ сетіс, сэтэн олісны семьяыс. А ми вот эта лопарскей керкаас, аддьылін тай, эта лёк, дёлиндик – сэтте петім. Сэтте керкаас петім.

“And when they were dekulakized in 1932, she moved to live at her mother’s place. She lived there together with her mother. But we, [lived] wherever. Grandpa Dmitri passed away. He had a brother-in-law and he [the brother-in-law] gave them his sauna, there they lived with the family. And we, well [we moved] into this Saami house, you have seen it, the small one—over there, we moved to live there. We moved into that house.” (kpv_izva20000320-1AndrejevaMM.173-182)

In the next example (17), Marfa Andreeva recollects the period of political repressions during the early 1930s, the so-called dekulakization. When she was a 10 year old girl, her family’s property—as allegedly better-off peasants—was expropriated.

(17) Ме пöмнита, ми кор аптека керкаас олім, мян зэй ыджыд эстием вöлі этаа джудждатэм сундук. Сэтэн вöлі мамелэн пимыыс, сэсса, лёк, маличаясыс, паркаясыс быдсэн сэн куйлісныс. [Interviewer: Ыхы.] Ставсэ босьтісныс. [Interviewer: Код нэ босьтіс? Код сэттием?] Мун да тöд. Комиссия босьтіс, а кыче карисныс, ме ог тöд. [Interviewer: Аха.

29 <http://videocorpora.ru> [20.02.2017]

А сэсэя тiянтэ кыче, керкасыыс вöтлiсны?] Вöтлiсныс. Ми лопарьяс ордэ сэк петiм. [Interviewer: Аха.] Öтiк изба сэтэн обiшей вöли да сэтте ставнум воим.

“I remember, when we lived in the building with the pharmacy, we had such a very big box of this height. Mom’s fur boots, fur parkas, parkas, everything was stored inside there. [Interviewer: Aha.] They took everything away. [Interviewer: Who took it away? Who would be like this?] Who knows. The commission took it away, but I don’t know who they gave it to. [Interviewer: Aha. And you, you were turned out of the house?] We were turned out. We went to the Saami place. [Interviewer: Aha.] There was a shack we shared, and all of us moved in there.” (кpv_izva20000320-1AndrejevaMM.256-270)

Summary of the case studies

These example case studies from Pite Saami, Kildin Saami, Skolt Saami, and Izhva Komi illustrate both the opportunities and the limitations for oral history research using oral histories which were not intentionally collected as such. Here, we have provided a variety of examples from very recent field recordings, legacy recordings and even language recordings done for visual arts rather than primarily for language documentation. Our case studies also provide insight into how language documentation projects can provide useful oral history sources, even inadvertently, and that such data can be accessed by interested researchers, regardless of background, via the international, digital archives. Finally, these examples also demonstrate how various themes are influenced by the zeitgeist of the time of recording (e.g. using propaganda phrases, or the positive attitude towards boarding school), as well as even the selection of topics and texts which are included.

In the following sections, we will present our vision of how oral history and other social sciences can gain to learn from the practices and experiences linguistically informed documentation projects have. Specifically, this covers the use of language technology and metadata, on the one hand, and some best practice suggestions for both archiving and making data available, on the other.

Oral History, Language Technology and “Thick” Metadata Descriptions

Although being multifunctional in principle and including data on linguistic and cultural knowledge, the virtual research infrastructures created by endangered language documentation projects are still predominantly used as databases for structural linguistic investigations. Furthermore, structural linguistically oriented documentation projects typically pick out only one endangered language, or ethnic culture using this language, merely in relation to its geographical location, as if the corresponding lan-

guage or culture existed separately from its neighbors. But this scarcely reflects social reality. Ethnic and linguistic identity can overlap between people in contact with each other and even between minority and majority groups. Even more so, it is in fact the situation of multilingualism and cultural contacts—combined with a social status asymmetry between different groups—that often leads to one group’s assimilation to another group and its ultimate disappearance.

Sociolinguistic approaches, i.e. the study of linguistic variation and change in language structures determined by social variables, have a long tradition in English and other major languages and are becoming more and more popular even in endangered language documentation. Most typically, however, documentary linguists work separately on their own single languages of interest and construct monolingual corpora. This is despite the fact that their speaker informants are almost always multilingual themselves. This is definitely the case for all Arctic cultures and languages dealt with in our chapter. Sometimes, one and the same speaker is recorded by different projects and included in three different corpora, in our own projects for instance in Russian, Kildin Saami, and Izhva Komi or in Russian, Nenets and Izhva Komi. Furthermore, whereas linguists prefer to collect data exclusively in the respective target language(s) under investigation, cultural and social anthropologists or historians often work in the same area and with the same individuals as informants but use the majority language as lingua franca and create a corpus of data *about* a group’s culture but without including the target language.

In order to make future qualitative and quantitative sociolinguistic and other cross-disciplinary investigations possible on these and other multilingual situations as well as the cultural-historical contexts in which these situations are embedded, we believe it is imperative to better *interlink* our newly collected data with previously archived data on both linguistic and other cultural practices of the different and ethnically overlapping and or culturally interacting communities in the Arctic. Ideally, our collections can be enriched by and interlinked with non-linguistic legacy data from archives and existing publications (e.g. photographs, biographies, written documents and all kinds of secondary sources already based on such data), as well as speech data not recorded in the target languages specifically (e.g. audio and video interviews exclusively in Russian, rather than in the native language).

The way this could be done is by rigorously applying methods from language technology to automatically create metadata and other annotations for large amounts of data. Language technology can be defined as the applied side of computational linguistics, as it aims at analyzing (and eventually also generating) natural language. Whereas a variety of language technology tools are available for larger languages (including Russian), Saami languages and Komi are still under-resourced in this respect. However, the relevant technologies are available in principle, and could be applied to these small languages as well. Note, that our idea goes far beyond cataloging and digitally publishing searchable data collections. Specifically, we mean the autom-

atization of annotating both textual and non-textual data in order to build ontologies of relevant categories which in turn can be automatically linked to each other across different collections, across different time spans and across different indigenous communities. As a result, a larger amount of data can be browsed and analyzed, ultimately resulting in scientific generalizations with more significant empirical support. Note that computerized work is crucial in our work because our own projects already have hundreds of hours of transcribed interviews, and these cannot be processed manually any longer in an efficient way.

As one example of a technology to be utilized, we want to mention “named-entity recognition”, a method applied in linguistic corpus creation. It results in the automatic parsing and tagging of text strings belonging to predefined categories (simple examples are for instance dates, names of persons, places, companies, etc.). Applying named-entity recognition in processing our databases automatically results in a number of searchable cross-database links throughout the corpora. As an example, one can think about the numerous place names mentioned in Zoya Nosova’s oral history about the boarding school in (13–15), which also occur in other texts recorded by our projects or mentioned in other documents. Once extracted and tagged as named-entities, these place names can be linked not only to each other across different recordings, but even to external geodata in order to be visualized on maps. Another example is the name of Ivan Osipov, mentioned in the oral history by Lazar Yakovlev (11), which is also relevant to other recordings or various other documents on Kola Saami language teachers, especially concerning the well researched period of Kola Saami language planning in the 1930s.

In a similar way, keywords can be extracted automatically from our data and inter-linked between recordings and other documents. The resulting relations can be formalized semantically in order to automatically construct relevant ontologies and then use these in digital catalogs of the data.

The description above is still merely an idea towards a new research approach, combining methods from computational sciences and linguistics with oral history research. The ultimate aim we envision could be described as “thicker metadata description”. This is a metaphor which we have borrowed from Clifford Geertz (1973). At first glance, anthropologist Geertz’s qualitative approach to interpreting and describing culture seems completely opposed to our quantitative way of working. However, our quantitative methods for data annotation are a tool to attain a better empirical foundation for qualitative interpretation, rather than the interpretation as such. What we aim for is a method for making more and better data available for qualitative analyses. Our “thick (digital) metadata” should therefore be compared to Geertz’s preliminary analyses in his field diary, rather than his final interpretations and descriptions.

Whereas the change towards data-driven research has already been accepted by most linguists, we are aware that the methods we are describing here will definitely

challenge the field of anthropology, which still relies exclusively on the qualitative-hermeneutic approach and takes a rather sceptical stance in regard to the relevance of “data” and quantitative interpretations. We nevertheless believe that anthropology, other fields in the social sciences as well as linguistics stand to profit significantly from our approach to documentation, archiving and publication (the latter two are treated in more detail below).

Archiving and Publishing Oral History Data

Over the last two decades, documentary linguistics has become more and more established as a subdiscipline of applied linguistics with its own theoretical approaches, methodologies and best practices, as detailed more distinctly in § *Oral Histories in Endangered Language Documentation*, above. Archiving and publication are two aspects of contemporary, digital and data-driven language documentation that are particularly relevant to our discussion. In the following we provide a general outline of these from the perspective of documentary linguistics in hopes of providing not only an example but also an impetus for oral history to consider adopting similar standards, and in doing so improve how oral history data are dealt with. Much of what we discuss here can be found in various scholarly works in language documentation and will sound quite familiar to most documentary linguists, but we particularly want to point out the insightful and thorough discussion and best practice guidelines provided in Bird and Simons’ article “Seven Dimensions of Portability for Language Documentation and Description” (2003), as well as the very recent white paper by Ameka et al. (2017). To be clear, we are talking about the archiving of digital data, an essential part of digital humanities, big-data research and e-sciences in general.

First of all, archiving—as we practice it in our projects—is in itself one way to publish data. However, the main function of digital archiving is twofold:

1. *scientific preservation*: guarding and making multifunctional data discoverable and available beyond one’s own project, and
2. *scientific reproducibility*: ensuring that the entire dataset still exists in an immutable form in order to enable future replication of an analysis.

Both are basic principles in digital humanities. The first principle is also especially important for oral history researchers who want to make their research and the collected data available to the communities they investigate in the long run.³⁰ The second

30 Making data available to communities on temporary data media such as DVDs or Flash drives is certainly a legitimate short-term solution, particularly for communities without unproblematic access to the internet. Indeed, this is even an essential solution for moribund languages whose communities should have access to such data before it is too late, in order to be able to utilize them in revitalization efforts, for instance. However, we would like to emphasize that such solutions are not sustainable in the long-term.

principle is not only generally accepted in natural sciences, but also in quantitative linguistics and quantitative social sciences, which believe that true scientific claims need to be reproducible (and thus verifiable or falsifiable).³¹

Archiving is really only useful when the archive itself is likely to survive long-term. In this, we do not only mean that the archive will exist as an institution, but also that all data in the archive will be migrated into new formats as these become relevant before the data are no longer accessible as older formats become obsolete. In other words, “archiving” data by keeping them on DVDs in a box is hardly sufficient for a number of reasons, such as the short life-expectancy not only of the physical disks themselves, but even of the format (as nowadays even blu-ray disks are less frequent as streaming services and cloud-storage become prevalent). For the same reasons, storing data on university servers cannot be considered proper archiving, although the bitstream (i.e., the actual computer data in raw form) may be safe in such cases. Instead, an actual digital archive which takes its archiving practices seriously in the long-term is clearly preferable. Two examples of such digital archives are the ones that we store our projects’ data in: the Endangered Language Archive (ELAR),³² which is part of the library at SOAS/University of London, and The Language Archive (TLA)³³ at the Max Planck Institute for Psycholinguistics in Nijmegen/Netherlands. The latter even hosts the recent social scientist oral history research found in the deposits by Anna Afanasyeva and Lukas Allemann (cf. § *Oral Histories in the Arctic*).

Metadata (as discussed in more detail above in § *Oral History, Language Technology and “Thick” Metadata Descriptions*) should be as extensive and detailed as possible, covering topics beyond simple cataloguing facts such as who, where and when. Rich, detailed descriptions of the interview situation and even interpersonal relationships, as well as details about participants’ backgrounds should be included.

The metadata should be structured in a clearly understandable way, ideally in an open-source, plain text standard, and ideally in xml format, such as in the IMDI³⁴ or CMDI³⁵ format, two current standards. When metadata are stored in this way, then it is possible for these to be harvested by meta search engines (search engines that search multiple other search engines) such as the Virtual Language Observatory (VLO)³⁶ or the Open Language Archives Community Language Resource Catalog.³⁷ Similarly,

31 Social scientists and many linguists who work in a qualitative way only make “data” available as excerpts and as far as these support the argumentation (i.e. by including snippets from oral histories, or example clauses from a language corpus in a published article), but not the data in their entirety. As a result, the analyses cannot be reproduced, and the scientific community is ultimately left with no choice but to either believe the claims or not believe them.

32 <https://elar.soas.ac.uk> [02.03.2017]

33 <https://tla.mpi.nl> [02.03.2017]

34 <https://tla.mpi.nl/imdi-metadata/> [02.03.2017]

35 <https://www.clarin.eu/content/component-metadata> [02.03.2017]

36 <https://www.clarin.eu/content/virtual-language-observatory-vlo> [02.03.2017]

37 <http://search.language-archives.org/index.html> [02.03.2017]

transcriptions, annotations and similar text-based data should also be structured in a clear and understandable way, using an open-source, plain text standard, ideally in xml format. If possible, recordings are also linked to the relevant annotations (even including time-alignment whenever useful) and to metadata (this can be done for instance using the ELAN tool (cf. § *Case studies*, above).

Any written text should be stored in non-proprietary, plain-text format (note that this does not include Microsoft Word or Microsoft Excel), and using the Unicode standard for encoding text characters. Audio and video files should be recorded in high-quality, open formats (or at the very least standard formats); smaller sized versions can be created for publication (such as streaming or distribution to the native community) if necessary, but higher quality cannot be created if the original data is poor from the very beginning.

Concerning publication, we mean this in the broadest sense, and use the term in reference to very simply making data available to others.³⁸ In many cases, modern digital archives provide a way to access the data they store, and many of the points presented above are equally valid in this respect (such as using open formats, or including as much descriptive detail as possible in metadata). But in addition to that, a major point concerns ensuring that metadata are available to repositories and catalogues that are used as search engines by researchers looking for data. While keywords are certainly a useful tool in ensuring discoverability, making not only all metadata, but in fact all textual data searchable for search engines increases discoverability even more.³⁹

In case some data are sensitive and require restricting access as a result, one should work with archives and publication outlets that have a robust and explicit implementation of access rights. The archives mentioned above employ systems that implement such access restrictions, even on a file-by-file basis.

Conclusion

Language and social sciences, and particularly language documentation and oral history, take significantly different approaches to collecting and analyzing data. One important reason for these differences is the qualitative approach preferred by social sciences compared to quantitative approaches typically used in some sub-disciplines of language sciences, such as language documentation and corpus linguistics. Another reason is the focus on the multi-functionality of collected data (a main goal of language documentation) as opposed to a more focused, research-driven incentive (i.e.

38 While community members should definitely be included when publishing materials, our focus in this chapter is on researchers; for more on communities as a target audience, see chapter 1 [Kasten].

39 Cf. Nathan et al. 2004 for an insightful discussion on what all can count as metadata.

using oral history for a specific research question and without considering later re-use of the collected data). In this chapter, we have presented our vision of a symbiosis of language documentation and oral history research based on the idea that both disciplines could learn from one another's standard practices, and benefit greatly from each other's data collections (even if these were not collected with the other field's research goals in mind). As documentary linguists, we have focussed on presenting aspects of language documentation that may be particularly innovative for oral history studies, and indeed for social sciences in general.

After determining that currently there is *de facto* no significant interdisciplinary collaboration between the two disciplines, we first outlined some oral history projects concerning Arctic peoples that share a similar approach with our language documentation projects. In this, we also discussed the seemingly contradictory lack of focus on the native languages common in such projects. We then discussed how oral histories can in fact be found in the recordings done by many language documentation projects, and included a number of examples for such data in our Pite Saami, Kildin Saami, Skolt Saami and Izhva Komi documentation projects. We then outlined how language technology and "thick" metadata can be utilized to ensure discoverability of data, and how this can be particularly useful for oral historians. Finally, we presented an overview of some basic best-practice recommendations that we think should be implemented on a wider scale in the social sciences, ultimately aimed at increasing the empirical base; even qualitative research stands to benefit from this. With such practices in place, it is possible for archived materials to become oral again.

In Table 1, we provide a summary of relevant criteria, and how the social sciences and language sciences relate to these in general. While there are certainly numerous counter examples, the trends we observe support these categorizations. Here, we have set up a binary opposition between generally being applicable (+) or not being applicable (-), but this is of course hardly black and white, but rather a continuum.

While language documentation and corpus linguistic projects tend to have large amounts of data which are digitally archived and annotated to various extents, oral history and other social sciences tend to have much smaller collections, and these are only occasionally available to anyone outside a specific project. On the other hand, the quality of social science data tends to be much higher concerning the particular topic being studied in a particular research project, while the contents of linguistics collections tend to be fragmentary and random. However, due to the very fragmentary and random nature of linguistics collections, they tend to be more multifunctional, as they provide a wider variety of topics, while social science project data are less useful for other disciplines because they are more focussed on a single topic. Since language is the core topic of linguistics projects, they have high potential for providing unique insights into the actual contents of collected texts, as this is potentially only accessible through the native language. Many social science projects use majority languages when working with informants, and this may exclude such language-specific insights.

While linguists collect metadata, these typically do not include nearly as many details for the contextualization of a recording, while this is standard fare for, and indeed the core of social sciences. Language documentation projects often include legacy data, and thus cover a greater time-span, which then allows for diachronic comparisons; for social science projects this is less common. Finally, due to the standard practice of archiving materials in digital archives, language documentation and other linguistics data are in general more accessible and sustainable, as well as more easily verifiable (and thus accountable) than social science data that is not publically available in its entirety.

All in all, we hope that the future leads to more interdisciplinary co-operation between oral history and language documentation. At the very least this should be in the form of mutual consultations concerning the advantages that one field may have over another (including the points we have presented in our chapter). Ideally, this would consist of carrying out interdisciplinary projects that include both documentary linguists and oral historians working together on the same team.

		Linguistics	Social Sciences
1.	Large quantity of available original data (digitally archived, transcribed, translated, keyword-tagged and catalogued)	+	-
2.	High quality of available original data (e.g. topic-specific, reflective questions in the interview)	-	+
3.	Multifunctionality (providing potentially different topics)	+	-
4.	Potentially unique insights via native language	+	-
5.	Contextualization (available via metadata)	-	+
6.	Time-span represented (availability of legacy data)	+	-
7.	Accessibility, accountability, sustainability (through digital archives)	+	-

Table 1: A summary of the pros and cons of data from documentary linguistics (and other areas of linguistics) as opposed to oral history (and other social sciences); + = more applicable, - = less applicable.

References

- Afanasyeva, Anna 2013. Forced Relocations of the Kola Sámi people. Background and Consequences. M.A. thesis. Tromsø: Universitetet i Tromsø.
<http://munin.uit.no/bitstream/handle/10037/5241/thesis.pdf> [01.03.2017]
- Afanas'eva, Nina E., and Michael Rießler (eds.) 2008. Лăзер кăллса моаййнас. Кыр̄ха лй Ёльцэ Нйна. Berlin: Humboldt-Universität zu Berlin.
- Allemann, Lukas 2010. Die Samen auf der Kola-Halbinsel. Über das Leben einer ethnischen Minderheit in der Sowjetunion. Frankfurt am Main: Peter Lang.
- 2013. The Sámi of the Kola Peninsula. About the Life of an Ethnic Minority in the Soviet Union. Tromsø: University of Tromsø.
<http://dx.doi.org/10.7557/10.2546> [01.03.2017]
- Ameka, Felix, Jonathan Blumtritt, Lissant Bolton, Vera Szöllösi Brenig, Irmgarda Kasinskaite Buddeberg, Brian Carpenter, Hilaria Cruz, Sebastian Drude, Patience L. Epps, Vera Ferreira, Colleen Fitzgerald, Ana Vilacy Galucio, Lauren Gawne, Brigit Hellwig, Oliver Hinte, Gary Holton, Maja Kominko, Manfred Krifka, Susan Kung, Miyuki Monroig, Bhanu Neupane, Ayu'nwi Ngwabe Neba, Sebastian Nordhoff, Brigitte Pakendorf, Kilu von Prinz, Felix Rau, Keren Rice, Michael Rießler, Mandana Seyfeddinipur, Nick Thieberger, Paul Trilsbeek, Hein van der Voort, Tony Woodbury 2017. Public Access to Research Data in Language Documentation. White paper. In *Language Documentation & Conservation*. (in print)
- Austin, Peter K. 2013. Language Documentation and Meta-Documentation. In *Keeping Languages Alive. Documentation, Pedagogy and Revitalisation*. M. Jones and S. Ogilvie (eds.), 3–15. Cambridge: Cambridge University Press.
- Austin, Peter K. 2013. Language Documentation in the 21st Century. *JournaLIPP* 3: 57–71.
- Bird, Steven, and Gary Simons 2003. Seven Dimensions of Portability for Language Documentation and Description. *Language* 79(3): 557–582.
- Bjarnson, Donald Einer 1976. A Phonemic Transcription of Lovozero (Kildin) Lapish. M.A. thesis. Bloomington: Indiana University.
- Blokland, Rogier, Marina Fedina, Niko Partanen, and Michael Rießler 2009–2017. “Izhva Kyy”. In *The Language Archive (TLA). Donated Corpora*. In collab. with Vasilij Čuprov, Marija Fedina, Dorit Jackermeier, Elena Karvovskaya, Dmitrij Levčenko, and Kateryna Olyzko. Nijmegen: Max Planck Institute for Psycholinguistics.
<https://hdl.handle.net/1839/00-0000-0000-0005-8A34-E@view> [01.03.2017]
- Blokland, Rogier, and Michael Rießler 2011. Saami-Russian-Komi Contacts on the Kola Peninsula. In *Language Contact in Times of Globalization*. C. Hasselblatt, P. Houtzagers, and R. van Pareren (eds.), 5–26. Amsterdam: Rodopi.
- Bowern, Claire 2011. Planning a Language-Documentation Project. In *The Cambridge Handbook of Endangered Languages*. P. K. Austin and J. Sallabank (eds.), 459–482. Cambridge: Cambridge University Press.

- Dudeck, Stephan Johannes, Florian Stammeler, Nina Messhtyby, Lukas Allemann, Nuccio Mazzullo, and Roza Laptander 2015. *Nomadic Memories. People and Oral History in the 20th Century along the Shores of the Arctic Ocean*. Rovaniemi: University of Lapland. DVD/USB flash drive.
- Dudeck, Stephan, and Lukas Allemann 2016. Indigene Oral History entlang des Eismeereres von Lappland bis zum Lena-Delta. In *Arktis und Subarktis. Geschichte, Kultur und Gesellschaft*. G. Saxinger, P. Schweitzer, and S. Donecker (eds.), 83–102. Wien: New Academic Press.
- Evjen, Bjørg, and Marit Myrvoll (eds.) 2015. *Från kust till kust – åhpegáttest åhpegáddáj. Møter, miljø och migrasjon i pitesamisk område*. Stamsund: Orkana Akademisk.
- Forbes, Bruce C., Timo Kumpulainen, Nina Meschtyb, Roza Laptander, Marc Macias-Fauria, Pentti Zetterberg, Mariana Verdonen, Anna Skarin, Kwang-Yul Kim, Linette N. Boisvert, Julianne C. Stroeve, and Annett Bartsch 2016. Sea Ice, Rain-on-Snow and Tundra Reindeer Nomadism in Arctic Russia. In *Biology Letters* 12.20160466.
- Freund, Alexander, Kristina R. Llewellyn, and Nolan Reilly 2015. Introduction. In *The Canadian Oral History Reader*. K. R. Llewellyn, A. Freund, and N. Reilly (eds.), 3–21. Montreal: McGill-Queen's University Press.
- Geertz, Clifford 1973. Thick Description. Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures. Selected Essays*. C. Geertz (ed.), 3–30. New York: Basic Books.
- Gerstenberger, Ciprian, Niko Partanen, Michael Rießler, and Joshua Wilbur 2016. Utilizing Language Technology in the Documentation of Endangered Uralic Languages. In *Northern European Journal of Language Technology* (4): 29–47.
- Halász, Ignác 1893. *Népköltési gyűjtemény. A Pite Lappmark Arjeppluogi egyházkerületéből*. Vol. 5 of Svéd-Lapp Nyelv. Budapest: Magyar tudományos akadémia.
- Himmelfmann, Nikolaus 2006. Language Documentation. What is it and what is it good for? In *Essentials of Language Documentation*. J. Gippert, U. Mosel, and N. Himmelfmann (eds.), 1–30. Berlin: De Gruyter.
- Kert, Georgii M. 1961. Obraztsy saamskoi rechi. Materialy po yazyku i fol'kloru saamov kol'skogo poluostrova (kil'dinskii i iokan'gskii dialekty). Moskva: Nauka.
- Kokkonen, Paula 2004. Kuolon niemimaan komilaiset. *Suomalais-ugrilaisen Seuran Aikakauskirja* 90: 371–384.
- Krauss, Michael 1997. The Indigenous Languages of the North. A Report on their Present State. In *Northern Minority Languages. Problems of Survival*. H. Shoji and J. Janhunen (eds.), 1–34. Osaka: National Museum of Ethnology.
- Lagercrantz, Eliel 1957. West- und südlappische Texte. Gesammelt und herausgegeben von Eliel Lagercrantz. In *Lappische Volksdichtung*. Vol. 1. Helsinki: Suomalais-Ugrilainen Seura.
- Llewellyn, Kristina R., Alexander Freund, and Nolan Reilly (eds.) 2015. *The Canadian Oral History Reader*. Montreal: McGill-Queen's University Press.

- Manker, Ernst 1947. De svenska Fjällapparna. (STF:s handböcker om det svenska fjället 4). Stockholm: Svenska Turistföreningens Förlag.
- Nathan, David, and Peter K. Austin 2004. Reconciving Metadata. Language Documentation through thick and thin. In *Language Documentation and Description*. Vol. 2. P.K. Austin (ed.), 179–187. London: SOAS, University of London.
- Øverland, Indra, and Mikkel Berg-Nordlie 2012. *Bridging the Divides. Ethno-Political Leadership among the Russian Sámi*. Oxford: Berghahn Books.
- Rantala, Leif, and Alefina Sergina 2009. *Ahkkila sápmelaččat*. Oanehis muitalus sámejoavkku birra, man manimuš sámegeielalaš olmmoš jámii 29.12.2003. Rovaniemi: Lapin yliopisto.
- Rießler, Michael 2005–2017. Kola Saami Documentation Project. Linguistic and Ethnographic Documentation of the Endangered Kola Saami languages. In *The Language Archive (TLA). DoBeS archive. Digital language archive*. In collab. with Anna Afanas'eva, Anja Behnke, Svetlana Danilova, Andrej Dubovcev, Aleksandra Erštadt, Dorit Jackermeier, Elena Karvovskaya, Kristina Kotcheva, Jurij Kusmenko, Maryna Litvak, Sergej Nikolaev, Kateryna Olyzko, Niko Partanen, Elisabeth Scheller, Nina Šarshina, Ganna Vinogradova, Joshua Wilbur, Evgenia Zhivotova, and Nadežda Zolotuchina. Nijmegen: Max Planck Institute for Psycholinguistics. <https://hdl.handle.net/1839/00-0000-0000-0005-8A34-E@view> [01.03.2017]
- Rießler, Michael 2013. Towards a Digital Infrastructure for Kildin Saami. In *Sustaining Indigenous Knowledge: Learning Tools and Community Initiatives for Preserving Endangered Languages and Local Cultural Heritage*. E. Kasten and Tjeerd de Graaf (eds.), 195–218. Fürstenberg/Havel: Kulturstiftung Sibirien.
- Scheller, Elisabeth 2011. The Sámi Language Situation in Russia. Finno-Ugric Minorities. In *Ethnic and Linguistic Context of Identity. Finno-Ugric Minorities*. R. Grünthal and M. Kovács (eds.), 79–96. Helsinki: Suomalais-Ugrilainen Seura.
- Siegl, Florian, and Michael Rießler 2015. Uneven Steps to Literacy. The History of Dolgan, Forest Enets and Kola Saami Literary Languages. In *Cultural and Linguistic Minorities in the Russian Federation and the European Union. Comparative Studies on Equality and Diversity*. H. F. Marten, M. Rießler, J. Saarikivi, and R. Toivanen (eds.), 189–229. Cham: Springer.
- Valijärvi, Riitta-Liisa, and Joshua Wilbur 2011. The Past, Present and Future of the Pite Saami Language. Sociological Factors and Revitalization Efforts. *Nordic Journal of Linguistics* 34: 295–329.
- Wilbur, Joshua 2008–2017. Pite Saami. Documenting the Language and Culture. In *Endangered Languages Archive (ELAR). Digital Language Archive*. In collab. with Iris Perkmann, Elsy Rankvist, and Peter Steggo. London: SOAS University of London.
- Wilbur, Joshua 2014. *A Grammar of Pite Saami*. Berlin: Language Science Press.
- Woodbury, Anthony C. 2003. Defining Documentary Linguistics. In *Language Documentation and Description*. Vol. 1. P.K. Austin (ed.), 35–51. London: SOAS, University of London.

- Woodbury, Anthony C. 2011. Language Documentation. In *The Cambridge Handbook of Endangered Languages*. P. K. Austin and J. Sallabank (eds.), 159–186. Cambridge: Cambridge University Press.
- Ziker, John, and Florian Stammler (eds.) 2011. *Histories from the North. Environments, Movements, and Narratives*. Proceedings of the Final BOREAS Conference, Rovaniemi, Finland, October 29–31, 2009. (Faculty Authored Books 279). Boise: Boise State University.