Language Documentation
and Description

ISSN 2756-1224

This article appears in: *Language Documentation and Description, vol 21: Special Issue on the Social Lives of Linguistic Legacy Materials*.
Editors: Lise M. Dobrin & Saul Schwartz

Philology in the folklore archive: Interpreting past documentation of the Kraasna dialect of Estonian

TOBIAS WEBER

Cite this article: Weber, Tobias. 2021. Philology in the folklore archive: Interpreting past documentation of the Kraasna dialect of Estonian. Language Documentation and Description 21, 70-100.

Link to this article: http://www.elpublishing.org/PID/246

This electronic version first published: December 2021

This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See http://creativecommons.org/licenses/by-nc/4.0/

EL Publishing

For more EL Publishing articles and services:

Website:	http://www.elpublishing.org
Submissions:	http://www.elpublishing.org/submissions

Philology in the folklore archive: Interpreting past documentation of the Kraasna dialect of Estonian

Tobias Weber
Ludwig-Maximilians-Universität München

Abstract

In this paper I share my experience of studying past documentation of the extinct Kraasna variety of South Estonian without access to native speakers, their descendants, or the researchers who conducted fieldwork in the community. After an overview of the historical and sociolinguistic contexts for early 20th-century Kraasna documentation projects, I illustrate the work I had to do to reconstruct information about Kraasna legacy materials in order to use them in my own research. The main obstacle lay in incomplete information about the original researchers and consultants, the circumstances of their fieldwork, and the methods utilised to produce the documentary artefacts that are available to work with today. The process of reconstructing this information led me to realise that the philological work of restoring the meta-documentation behind a project has as much academic value as the initial recording of material or the subsequent linguistic description. At the same time, the work of editing and curating legacy materials folds secondary researchers into the picture that they reconstruct about previous documentation. Thus, users of legacy materials need to adopt a reflexive stance on their own role in the research process, one that is best served by a philological approach that recognises the human involvement in every step of the creation, reception, and replication of documentary linguistic materials.

1. Introduction

In recording, analysing, and disseminating people's languages and stories, field linguists share those people's knowledge and perspectives with the world. At the same time, fieldworkers create traces of themselves and their research – another layer of narrative created by *their* project. But with legacy materials from more distant past projects, the researchers and their consultants are often just faded names on a piece of paper, on the front or back of a book,

Weber, Tobias. 2021. Philology in the folklore archive: Interpreting past documentation of the Kraasna dialect of Estonian. *Language Documentation and Description* 21, 70-100.

or in a box in an archive. Their traces are not clearly visible anymore. In this paper I share my experience of studying such artefacts of the extinct Kraasna variety of South Estonian, something I did without access to native speakers, their descendants, or the researchers who conducted fieldwork in the community. After an overview of the historical and sociolinguistic context of Kraasna documentation in the early 20th century, I present the detective work I had to do in order to use the materials for my own research. The main obstacle I found lay in incomplete accounts of the participants, circumstances, and research methods underlying the materials' production. This experience convinces me that philologically restoring a previous project's "meta-documentation", as I have done, has as much academic value as the initial recording of material or the subsequent linguistic analysis and description. Since reconstructing meta-documentation for legacy materials can be as time consuming as starting a new documentation project, there is a need for academic merit to be awarded not only for language description and documentation but also for the curation of legacy materials, which may require extensive archival research, studying secondary literature, and conducting interviews with people who are knowledgeable about the materials. Moreover, the work of editing and curating legacy materials folds secondary researchers into the picture that they reconstruct about previous documentation. As a result, users of legacy materials need to take a reflexive stance on their own role in the research process, one that is best served by a philological approach that recognises the human involvement in every step of the creation, reception, and replication of documentary linguistic materials.

2. Meta-documentation of Kraasna

I discovered Kraasna as a research topic through reading an edited collection of dialect texts (Mets et al. 2014) as I explored my interest in South Estonian and variationist sociolinguistics. Noticing that the Kraasna section was the shortest in the collection, I became curious about why there were so few sources available on Kraasna. As I studied the texts, I grew interested in the personal stories of the much-lauded last speakers, "guardians of language" (Coulmas 2016; see also Heller-Roazen 2005; Dobrin & Berson 2011) who enable future audiences to learn about the history of their community and keep the memory of their dialect alive. My initial plan was to describe Kraasna through the idiolect of "the last speaker", inspired by works on Kamas (Klooster 2015), a South Samoyedic language with a prominent last speaker (see Künnap 1991).[1] In the Kraasna case, however, the exact circumstances surrounding the last speaker's recordings were not as clear. Using the legacy materials and accompanying texts as an entry point, I took the preliminary steps of identifying and transcribing the sources, in some cases after digitising them first. This turned out to be a major project, which taught me that there is value in the curation of legacy materials that is independent from their subsequent use in language description or publications. When curating texts and textual artefacts, the most useful methods derive not from language documentation or description but rather from philology. In the following sections, I draw on my experience of working with Kraasna documentation to illustrate the philological approach to curating linguistic legacy materials. Scholars in language documentation and description have much to learn from revisiting legacy materials and trying to make sense of them as an activity that is valuable in its own right. In the Kraasna case, we can even consider the linguistic analysis of these materials which eventually followed (Weber 2021a) to be a secondary product of the curation process, rather than the curation being merely a preparatory step for linguistic description.

---

[1] Like Kraasna, Kamas underwent language shift to Russian around the time of the First World War. No fieldwork was done on the language between the 1920s and the

When I started working with Kraasna legacy materials, I faced a network of artefacts which were obviously related, but there was little information on the precise links between them. The archival sources seemed to overlap in their contents, with additional or missing content and at times misleading metadata. This required me to take additional steps not only to better understand the materials but also to understand the relationships among them in order to provide a holistic picture of different documentary efforts as interconnected events and not just as singular occurrences. I reconstructed the network or "ecology" (Good 2007) of the Kraasna materials by piecing together information about the relevant agents (including researchers, consultants, funding bodies, and archives), their activities, and the resulting artefacts of raw, primary, and secondary data (Lehmann 2004), including all research I could find that was based on Kraasna documentation (Weber 2016). Uncovering the story behind these artefacts led me to investigate the human factor of fieldwork through the traces of human agency and decision-making the legacy materials contain. This task is well-labelled by the term "meta-documentation" (Austin 2013; see also Nathan 2010), or the "documentation of the documentation research itself" (Austin 2010: 29). Meta-documentation involves reconstructing the context of linguistic fieldwork and the factors impinging on the research that led to the production of the documentary

---

1960s, although phonograph recordings were made in 1912–1914 prior to the shift (see Klumpp 2013). But then, forty years later, researchers found two last speakers who were previously unknown to the scholarly community, one of whom worked with linguists throughout the 1960s and 1970s, providing about another fifteen hours of recorded interviews in the language.

artefacts we now want to learn from. This task foregrounds the human rather than technical side of metadata.

In sharing my experience of restoring meta-documentation for the Kraasna materials, I have two goals. First, for researchers interested in Estonian or Uralic linguistics, my work presents an overview of the sources on Kraasna and serves as a guide for others who want to consult these materials for their own purposes.[2] Second, for a more general audience, this case illustrates the value of a philological approach to legacy materials using a focused example with a manageable amount of primary and secondary sources. There have been only a few publications on Kraasna in this century. Twenty-two Kraasna texts with a translation and grammatical sketch were included in a collection published in a series on Estonian dialects (Mets et al. 2014; see also Faster 2014; Iva 2015). A two-page chapter was devoted to Kraasna in a compendium on Estonian dialects (Pajusalu et al. 2018), and there is a special issue of the *Journal of Estonian and Finno-Ugric Linguistics* on Kraasna and other linguistic enclaves in 2021. Other than that, there have been only a handful of journal articles about Kraasna over the last twenty years (e.g., Pajusalu 2005; Ernits 2012, 2018). While some of the details discussed below are of course peculiar to the Kraasna case, other researchers in Uralic linguistics have reported encountering similar challenges in their own work with legacy materials (Winkler 1994, 1997; Helimski 1997; Katz 1979). Lessons from the Kraasna case may therefore be useful to other researchers, since metadata and meta-documentation will always be (at best) partial, and it is often necessary to reconstruct missing contextual information for legacy materials in order to understand what they are and use them appropriately.

3. The Kraasna dialect of Estonian and its documentation

The Kraasna *maarahvas* (lit. ‘rural population, country folk’) was a South Estonian-speaking community which existed as a linguistic enclave in a Russian-majority region around the city now called Krasnogorodsk in the Pskov *oblast* or region (once a ‘Governorate’) of the Russian Federation between sometime in the 17th century and the first half of the 20th century (see Figure 1).

The earliest information we have about Estonians in this area dates from 1701 (Grünthal 1912), although it is unclear whether records from that time refer to Kraasna or the Lutsi, a neighbouring Estonian-speaking community

---

[2] More information on the archival sources, including brief descriptions of the contents of each source, is included at the end of this paper (see Weber 2016, 2021a).

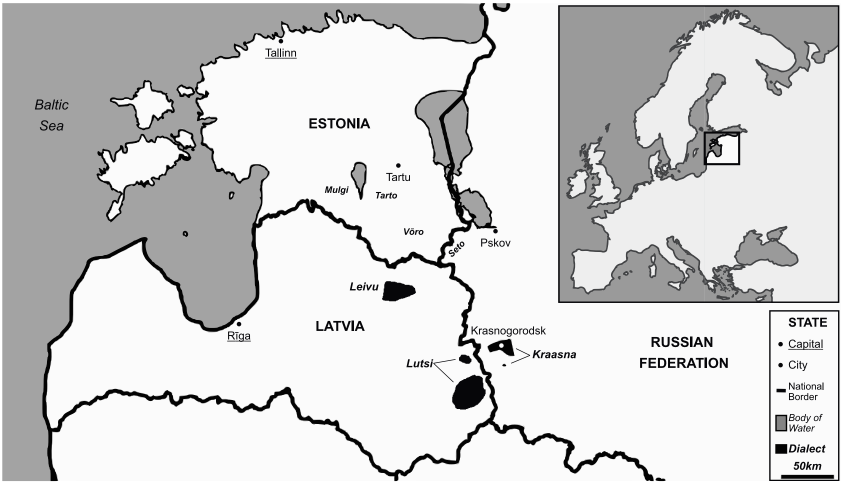
Tallinn
Baltic Sea
ESTONIA
Tartu
Mulgi
Tarto
Võro
Seto
Pskov
Leivu
LATVIA
Rīga
Krasnogorodsk
Lutsi
Kraasna
RUSSIAN FEDERATION
STATE
● Capital
● City
National Border
Body of Water
Dialect
50km

Figure 1. The distribution of the South Estonian linguistic enclaves in the late 19th *century. The data on Kraasna is based on population figures given by Kallas (1903). The Leivu and Lutsi labels are meant to indicate their approximate geographic spread based on Mets et al. 2014. Modern political boundaries are included for reference.*

less than 60 km away in Latvia. The Lutsi and Kraasna *maarahvas* were not linguistically and culturally identical, yet both of these South Estonian language communities likely have their origins in relocated peasant serfs or refugees fleeing from southeastern Estonia due to religious persecution during the period of the Polish-Swedish Wars in the early 17th century.[3] There were no clear accounts of these Estonian speakers until the first scholarly enquiries by Friedrich Reinhold Kreutzwald and his colleague Adolph Johann Brandt in the mid-19th century (see Ernits 2012, 2018). As a result of changes in administrative borders, culminating in the formation of independent nation states in the early 20th century, Kraasna developed independently from the other South Estonian varieties in Latvia or Estonia in its own enclave for two hundred years.[4] Kraasna remained more archaic than other varieties and is thus especially important for research on the history of the South Estonian language.[5] The Kraasna population was already assimilating linguistically and culturally to the Russian majority when the first Estonian researchers learned about the existence of the Kraasna community towards the end of the 19th century and directed their attention towards documenting its language and culture.

---

[3] The Polish-Swedish Wars were a series of skirmishes between the Polish-Lithuanian Commonwealth and Sweden which were fought in the Baltics. This was a period of major political reshuffling after the collapse of medieval German rule in the 16th century and before the ultimate inclusion of the Estonian-speaking regions into the Russian Empire in 1721. One of the Kraasna manuscript sources mentions a ‘Lithuanian’ War, during which the speaker’s ancestors fled from the Seto regions to the area around Krasnogorodsk (AES 202).

3.1 The South Estonian linguistic enclaves

Alongside the Lutsi and the Leivu varieties formerly spoken in Latvia, the Kraasna dialect forms part of the *keelesaared* 'language islands', a set of South Estonian linguistic enclaves which have conventionally been treated as separate "sub-dialects" in Estonian dialectology (Pajusalu 2007). Simplifying somewhat, I will refer to Kraasna, Lutsi, and Leivu as "dialects" here. The suggested position of Kraasna in Estonian dialectology can be seen in the lower part of Figure 2.[6]

As can be seen from Figure 2, South Estonian forms a separate branch in the Finnic group of Uralic. Descending from the historical Ugala language, South Estonian is believed by several scholars to be the first language to separate from the Finnic proto-language (Sammallahti 1977; Kallio 2014). As a result, the South Estonian varieties serve as important reference points for reconstructing Proto-Finnic (see, e.g., Pajusalu 2009, 2012). Since their split from Proto-Finnic, the South Estonian varieties developed independently of Northern Estonian until their much later convergence due to the creation of an Estonian Literary Standard in 1908. Consequently, knowledge of South Estonian is useful for making sense of historical sources and investigating the history of the southern Estonian regions.

---

[4] Information in Kallas's monograph (1903) and Voolaine's diaries hint at familial ties through arranged marriages between members of the Kraasna community and the Seto-speaking communities around the city of Pskov, as well as pilgrimages from Krasnogorodsk to Pskov. Despite these interactions, the Kraasna dialect maintained some archaic features (see Pajusalu et al. 2018). It is also linguistically different from the Lutsi dialect (see Figure 2), which implies that cross-border exchanges did not leave traces on Kraasna. Language contact between the linguistic enclaves calls for further investigation.

[5] I refer to South Estonian as a distinct language partly based on its historical development and partly based on its socio-political status within the modern Baltics. In certain parts of the Estonian and general Uralic research literature, it is referred to as one of the main dialects of the Estonian language, itself a 20th century political construct (see Figure 2).

[6] In Estonian dialectology, this lowest level of regional varieties (usually tied to a parish) is called a "sub-dialect", while the term "dialect" covers more abstract language forms found across several parishes. The use of parishes, i.e., the area of a church's authority, as the basis for these categories is rooted in the religious past of the Baltics, which was governed by Christian Orders and Bishoprics until the Northern Wars. As the church was an important meeting point and later played a part in schooling, local language use centered roughly around the administrative boundaries of the parishes until increasing mobility in the early 20th century led to a levelling of parish-bound sub-dialects.

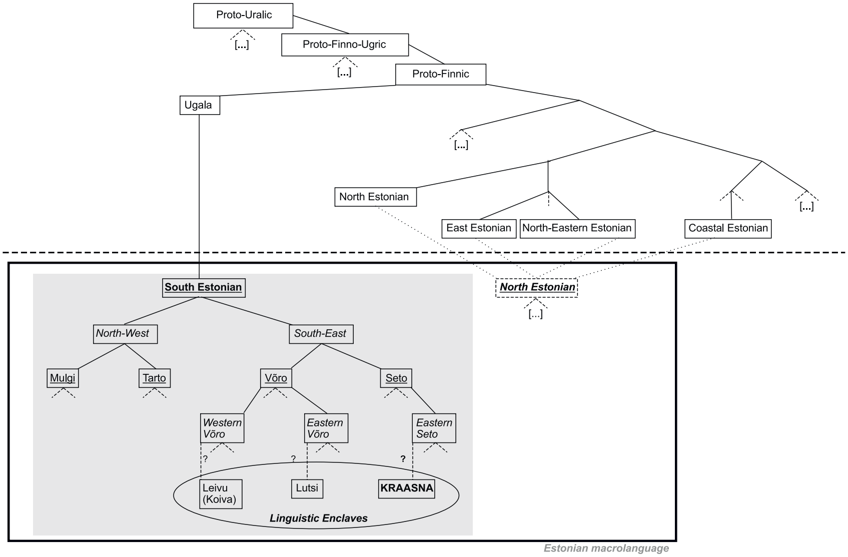

Figure 2. An overview of the position of the Kraasna sub-dialect within Estonian dialectology and the Finnic dialect continuum. The diagram is non-exhaustive and selective: only branches which contribute to the Estonian macrolanguage are shown in the upper part, based on Viitso (1978, 1985, 2007, 2008), Laanest (1982), and Kallio (2014). The lower section illustrates the classification used in Estonian dialectology for the modern varieties of South Estonian, based on Pajusalu (2007) and Pajusalu et al. (2018). The tentative positions of the linguistic enclaves shown are based on suggestions by Kask (1956) and Mets et al. (2014)

Like the Leivu and Lutsi, the Kraasna community was geographically separated from the other languages in the South Estonian dialect continuum. Each of these South Estonian enclaves had its own religious and cultural customs, neighbouring languages and settings for multilingual exchange, and political and administrative systems. These differences in their contact histories set the language enclaves apart from each other and from all other Estonian communities. Yet, after the Estonian national awakening in the late

19th century, Estonian philologists aiming for a uniform Estonian national and linguistic identity united all three (pre-)historical Estonian languages – North, South, and Coastal – into one national language. This standardisation happened between 1908 and 1911, before Estonian independence, in a step toward the goal of an Estonian-speaking nation state. Residents of the linguistic enclaves were not party to deliberations about the standard, as they lived in predominantly Russian- or Latvian-speaking regions that were negotiating their own issues of nationhood and linguistic identity around this same time, often to the detriment of the Estonian enclaves. The political decision to subsume South Estonian under a unified Estonian umbrella disenfranchised South Estonian-speaking communities within the territory of the Republic of Estonia following independence in 1918 (Koreinik 2011). In the 2011 Language Act (KeeleS 2020[2011]: Section 3), the Estonian government declared that the relationship between South Estonian and the Literary Standard is one of mere dialect difference.[7] Because it was considered a dialect of Estonian, South Estonian was long left to Estonian philologists to study and did not receive the same attention from Uralicists, who created thorough and internationally accessible research and teaching materials on the other Uralic languages in the past century; in contrast, only a few items are available for South Estonian (see also Stipa 1990). South Estonian has thus been under-researched, in large part because it has been treated as a dialect of Estonian for political reasons since the early 20th century.

3.2 The legacy of past research on Kraasna

Before delving into the Kraasna legacy materials themselves, it is worth taking a moment to describe the broader context in which they were collected. The European traditions of philology, folklore, dialectology, and linguistics have their roots in the 18th-century Enlightenment and subsequent questions about nationality and state formation following the reshuffling of political boundaries after the Napoleonic era. Humanists of this time had an eclectic approach to the humanistic sciences that supported work at the interfaces between philosophy, culture, language, arts, and politics. Their work might appear interdisciplinary to modern scholars in these fields, which have since established themselves as individual disciplines in the academic landscape. But for scholars of the time, the study of language, text, and culture were inseparably linked as vehicles for appreciating the classics, engaging in philosophical debates, and understanding folklore. For example, the Grimms' gathering of folklore from all parts of the German-speaking world led them to engage not only with literature but also with linguistics, leading to the development of the theory of sound changes which we still know as Grimm's law. Other collectors of folk songs and tales were trying to unearth the national spirit and culture of variously defined and imagined groups and communities, ideas that in the 19th century culminated in revolutions and nationalist movements.

---

[7] This is the central regulation of language policy for the Republic of Estonia, first passed by the Estonian parliament on 2011-02-23. Such legislation makes South Estonian an interesting case for language policy research (Gibson 2017).

In the Uralic sphere after 1770 – when a linguistic relationship among Uralic languages was first demonstrated – there was a bipartite focus on gathering linguistic data and folk stories from distant relatives in Siberia while also investigating the Finno-Ugric languages in Europe (see Stipa 1990). The earliest collections of Uralic folk songs and tales also originate from that time and were motivated by a desire to compile material that exemplified national cultures. In Hungary, the poet, philosopher, and ethnographer János Erdélyi collected vast amounts of folklore in the years prior to the 1848 revolution, embodying the national romantic spirit of the time. During the same period, Finnish folklorist Elias Lönnrot collected the stories that came to be known as the Finnish national epic, the *Kalevala*. Since the seminal work of Finnish folklorist Julius Krohn, the texts in these folklore collections were analysed in terms of their geographic distribution (Frog 2013). Folklore was thus closely tied to early dialectology, as both disciplines aimed to understand the geographic spread of – and connections between – national narrative or linguistic traditions. Estonia at the time was no different, with many of the 19th-century humanists (e.g., Friedrich Reinhold Kreutzwald, Jakob Hurt) working simultaneously on folklore, literature, and linguistics, all in support of a national awakening. Researchers not only published ethnographic studies with linguistic data on the linguistic enclaves, but also cataloged Estonian folk songs (Kallas 1901). Thus, for the researchers whose Kraasna materials we consult today, folklore, philology, and dialectology were intertwined fields of study.

The few existing records of the Kraasna dialect were compiled by three researchers. The first was Estonian folklorist Oskar Kallas, who conducted fieldwork in the Kraasna community in 1901. Kallas (1903) reported fewer than 100 speakers, with the youngest at the time belonging to the parent generation. Kallas's reports are one of the two major sources that exist on the Kraasna Estonians. A second documentary researcher, and the last who was able to record the language in active use, was a Finnish linguist named Heikki Ojansuu. Ojansuu visited Kraasna twice, once in 1911–12 and then again in 1914. Ojansuu made the only existing audio recordings of spoken Kraasna from the most fluent speakers he could find out of the 26 mentioned in his 1911–12 manuscript. His return visit in 1914 was impacted by the First World War, when an atmosphere of distrust led the local population to be suspicious of a foreigner collecting information (see ES MT 224). In the interwar period, the Kraasna Estonians found themselves on the Soviet side of the border, which separated them from their linguistic and cultural brethren in the new nation states of Estonia and Latvia. Only after the dust of the Second World War had begun to settle and the Baltic states had come more fully into the Soviet sphere of influence did the third expedition take place. It was conducted in 1952 by Estonian folklorist Paulopriit Voolaine. Voolaine continued to visit Kraasna several times in the 1960s to salvage the last bits of information that he could from rememberers of the language and culture. The last expeditions by philologists in 2004 (see Harju 2004) and archaeologists in 2014 did not yield any new linguistic material.

Working with the primary and secondary sources on Kraasna created by these three researchers requires digging deep into the folklore archives, where the original materials were deposited by their authors for posterity. While modern archives may make old data available in digitised formats, some important pieces of metadata may have become obscured over the years, the result of changing archival customs and the creation of copies, new versions, or reprints shared between archives. As a result, any work with the available documentation requires the user to do the detective work of establishing links between individual segments of the data in light of their archival and artefactual histories, restoring obscured information, identifying errors in the metadata, and reconstructing the stories behind the original researchers' projects.

4. Philological analysis of linguistic legacy materials

While the term "philology" is still used in Europe, often in the names of institutions researching and teaching language, literature, culture, and linguistics, the term might not be as familiar to American readers (Turner 2014; Gurd 2015). Philology aims "to preserve, monitor, investigate, and augment our cultural inheritance" (McGann 2013: 334) and "to provide an academic commentary or explanation of a text, and possibly also a parallel translation of the text, through critical edition and varied interpretation of the text" (Palola 2020: 159). These activities are linked by a focus on textual artefacts and the contextual knowledge necessary to interpret them, including culture, history, and arts. Philology as a method thus draws primarily upon textual criticism, textual curatorship, and interpretation. These are skills that are needed in work with legacy materials and go beyond linguistic knowledge and the ability to understand the language the text is written in.

My point of access to Kraasna was the textual artefacts produced by researchers in the past. In order to unlock the knowledge contained in these texts, it was necessary to make sense of them by learning about their relationship to other texts and the contexts of their creation. The task is not unlike that of epigraphers aiming to interpret ancient inscriptions in possibly unknown languages, papyrologists seeking to link and better understand different texts through collocation in edited collections, or codicologists comparing versions of medieval manuscripts to find out more about the scribes who produced them. The investigation of textual artefacts as carriers of language data is a comparative project that requires “an attitude of respect for the datum” (Wenzel 1990: 18) in which each textual artefact is treated as an object of inquiry in its own right. At the same time, there is the possibility of human error, making a critical and self-critical stance necessary (see Gurd 2015). Older versions of data can generate as much insight as more recent versions, as they provide contextual information about the data’s creation, such as the editors’ motivations and the technologies they employed (Shils 1981). Philology seeks to learn as much as possible about the humans involved in the creation of artefacts, since the work of editors, translators, and transcribers shapes and leaves traces in the artefacts themselves. Their actions are affected by subjective judgements and individual abilities (Weber & Klee 2020), which makes philology a “fundamental science of human memory” (McGann 2013: 345). The philological investigation of texts thus illuminates the human factor in an artefact’s creation.

As outlined in the previous section, folklore collections are products of philological research. Critically editing them, adding commentary, and contextualising them is a continuation of the philological tradition in which they were created, yet it is not necessarily the end, since our own work can itself serve as the basis for future philological enquiry (see Weber 2020). In the philological approach to documentary linguistics for which Seidel (2016) advocates, interpretive contexts are inseparable from the texts linguists collect, even when the information may be stored separately in what we are now used to calling “the metadata”. Rethinking the entire trajectory of language documentation, archiving, language description, and subsequent scientific discourse in light of the philological process helps bring the researcher into focus for our discipline alongside the textual artefacts they create and use. Philology is thus more than a canon of methods and tools; it is a distinctive way of thinking about our work and the artefacts that result from it. In the following section, I will illustrate how making sense of the Kraasna legacy materials requires a philological orientation.

5. Making sense of the sources

The materials most easily accessible to me were the published transcriptions of Ojansuu’s manuscripts (Mets et al. 2014) and the publications by Oskar Kallas (1903, 1904) based on his own fieldwork.[8] I naively believed that these materials would be exhaustive and suffice for linguistic description. Kallas’s

---

[8] There are different types of *artefacts*. *Texts* are coherent sources, such as transcribed narratives. Standardised representations of spoken language are called *transcriptions* here. *Manuscripts* are recorded in handwriting, as opposed to typewritten or typeset versions found in publications.

(1903) *Kraasna maarahvas* was the first monograph on this community, providing an ethnographic and folkloristic overview while also commenting on language use and history. The linguistic data consists mostly of single words and simple phrases which were often inserted in the running ethnographic text and, as I found out later, morphologically and syntactically altered and recomposed from his fieldnotes – sometimes dramatically – to fit more smoothly into the Estonian sentences in the monograph. For example, Kallas's notepad contains *morol* 'in the yard' in the adessive case (marked with the suffix *-l*) which is changed in the published version to *morole* 'into the yard' in the allative case marked by *-le*, while ignoring suffixal vowel harmony which does not exist in Standard Estonian but would be expected in this South Estonian variety (Kallas 1903: 93). Kallas also changed word order, so that, e.g., what was written in the notepad version as (1) was rendered in the publication as (2) in a description of Easter customs (Kallas 1903: 91). This represents a creative recomposition of the field materials, significantly altering the word order and adding components that were not recorded but may have been discussed with informants in the fieldwork interview.[9]

(1) *Lihavõde* *aegu* *kannedi* *värähtõ* *pääle*
Easter.GEN time.PART carry.PASS.PST gate.GEN onto
'Easter time, it was carried onto [outside] the gate'

(2) *Värähtõ* *pääle* *kannedi* *kõik* *süüüki*
gate.GEN onto carry.PASS.PST all food.PL.PART
'Onto the gate, all foods were carried'

The phrases were elicited by Kallas from several different consultants who are identified in his monograph. But while full texts are attributed to specific consultants, the individual phrases and words are not, so their provenance is unknown. Kallas's monograph is an insightful anthropological source, and his editing of the Kraasna language data makes the Estonian text easily readable. However, the discrepancies between the Kraasna examples in the monograph and the notepad versions may represent something other than manipulation for simple stylistic reasons. The changes may reflect words or phrases which Kallas remembered at the time of writing that he did not note down during the interview. He might have decided to "fill in" as much information as possible for his audience, who were not Kraasna people but readers from scholarly circles in Estonia. Or Kallas may have been prompted to recompose the words and phrases he elicited out of an ideological commitment to what he believed was the “correct” way of speaking the language. Kallas’s monograph also includes a song adapted from an 1849 manuscript by Brandt (see Ernits 2012) which was published in a collection of folk songs (Neus 1850), but Kallas admits that he edited the version he used by choosing one spelling over another. In any event, given the uncertain basis for his revisions, a researcher today must be careful in using the monograph as a source for linguistic or dialectological enquiry.

---

[9] Abbreviations in glosses are: ALL – allative; DIM – diminutive; GEN – genitive; ILL – illative; IPS – impersonal; NOM – nominative; PART – partitive; PASS – passive; PL – plural; PST – past; SG – singular.

Ojansuu’s material had not been published prior to its inclusion in the 2014 collection of dialect texts, and he himself only published a short phonological essay based on his data in 1912. The Ojansuu texts underwent minor editing as they were transcribed for the 2014 collection, as the selection process required the editors to identify (or attribute) a start and end point for each text. The collection names one “Matrëna Razivonova” (who I later found to be Matrëna Rodionovna Kuznecova) as the source for all of the Kraasna texts, whereas I strongly believe that they originate from at least two different speakers (or potentially even more) because person or place names are present as headings at the top of every page of the corresponding manuscripts, such as *Feodosia*, a female name, *Ivatsova P’etò* ‘Pëtr [from/in] Ivatsova’, or *Panki*, which could be a hypocoristic form of *Stepan* or a misspelt reference to *Tańka*, an Estonian village mentioned by Kallas. One of these names is *Ulla*, the likely second consultant for the materials. But the editors interpreted this as the title of the text (i.e., “205. Kraasna Ulla”) rather than the name of the consultant who produced it (Mets et al. 2014: 291).

While surveying possible sources on Kraasna, I consulted with Karl Pajusalu, Professor of History and Dialects of Estonian Language at the University of Tartu, one of the editors of the 2014 text collection and an expert on South Estonian dialectology. He provided more information on the original sources and suggested that I visit the archives to check the manuscripts myself. He also disclosed that the manuscripts were richer than the published accounts and that diaries by a third scholar, Paulopriit Voolaine, might also contain valuable data. As I began shifting my focus towards the philological work of recovering and collating all of the Kraasna sources, I discovered they were scattered across five archives in three cities in two countries.

5.1 Assessing Kallas’ seminal work

During his 1901 fieldwork, Kallas took scratch notes, notes taken “contemporaneous with or soon after the events observed or words heard” (Sanjek 1990: 96). He took these in three notepads that incorporate monologic texts and phrases in Kraasna; comments and translations in a literary style of Estonian (in modern handwriting), Russian (in Cyrillic script), or German (in Kurrent, a German cursive script); and other points of enquiry. Dialect material was rendered in an Estonian-based orthography with some idiosyncratic innovations, such as rendering Estonian's famed over-long vowel length with a trigraph as in (2), and using adapted Finno-Ugric transcription conventions, like marking palatalisation with an acute accent, that were codified by Finnish linguist Setälä in 1901, the year of Kallas's fieldwork. The Finno-Ugric transcription system, due to its early codification, has been used instead of the IPA for Uralic studies and is the standard system used for transcribing Estonian dialects. Along with Kallas's notepads in the Estonian Folklore Archives of the Eesti Kirjandusmuuseum (Estonian Literary Museum) in Tartu, Estonia, are some letters and official documents that help us understand the motives behind Kallas's expedition and the perspective he took in documenting Kraasna.

Kallas had funding from the Finnish Literature Society to travel to Krasnogorodsk and salvage any material he could get from the "kindred" Estonian-speaking community recently rediscovered there. As Kallas (1903: 10) expressed it, "often, a worthless keepsake from a dying relative can be dear" (here and elsewhere, quotes from Kallas in English are my translation). His aim was to make the Kraasna people and their culture known to the scholarly community as well as to an Estonian society striving for nationhood. Kallas's ethnographic description emphasises the threat to the Kraasna people's Estonian identity and their potential reconciliation with the Estonian nation. The Kraasna community is likened to the other Estonian "lost tribe", the Lutsi across the border in Latvia, whom Kallas (1894) visited some years prior. He presents a fatalist account of the chances for Estonian culture's survival in Kraasna, and he implicitly belittles the community, which was assimilating linguistically under Russian influence, as "dissolving physically and psychically, changing in language and mind" (Kallas 1903: 10). Contradicting Kallas's narrative of language attrition, Ojansuu's consultants from a decade later were described as speaking "very good and fluent Eastern Seto" (Iva 2015: 516 [my translation]), which suggests that Kallas may have been inflating the threat of assimilation in order to galvanise Estonians to aid their imperiled kindred in Russia. Although speakers' own ideologies about their language are not clearly articulated, they are presented by Kallas (1903: 30) as not speaking either Estonian or Russian well, making them "even weaker than a Russian". Yet, they could tell stories and keep their customs in South Estonian, while participating in social and religious life in Russian. It seems they found ways to create a somewhat diglossic environment in which even some of their Russian neighbours understood Estonian (Ernits 2012: 48). But for Kallas, a national romanticist dreaming of a nation for the Estonian people, this "dying kindred tribe" remained entangled in a conflict between the Estonian “mother tongue” and the “foreign language” Russian.[10] Kallas’s description has coloured the entire discourse surrounding the Kraasna dialect and its speakers to date. The discussions of assimilation and language death which can be found in modern descriptions of the community (e.g., Mets et al. 2014; Iva 2015) bear witness to Kallas’s influence. Critically assessing and reviewing his work and its impact is an important element of the philological process linking various research activities to the texts that resulted.

5.2 Sorting out intertextual links between artefacts through Voolaine’s diaries

The same archive which holds Kallas’s collection also holds the diaries of Paulopriit Voolaine, an Estonian folklorist who first visited Kraasna in 1952. While there, he wrote down a few words from the grandchild of Timofej Rodionov, a speaker who likely also worked with Ojansuu (AES 202: 5 [notes with “Timoški”]). Although the diaries do not yield any relevant new Kraasna language data, Voolaine’s ethnographic accounts provide information useful in reconstructing consultants’ backgrounds. Ideally, consultants’ genealogies could be reconstructed using official accounts from Krasnogorodsk, but I do not know whether these public registers still exist in Russia. Through the diaries, we get a glimpse of what happened in the thirty-odd years between Ojansuu’s and Voolaine’s fieldwork. We learn that the last native speakers died in the 1930s, while the last rememberers disappeared from accounts before 1968. These details link the different Kraasna documentation projects through their consultants, who were members of the same families. As a result, family stories and family language use could possibly be followed through the documentary record. Voolaine’s details about consultants help to establish the social (as well as artefactual and textual) networks which bear on the interpretation of the language data.

The identity (and indeed even gender) of Timofej Rodionov’s grandchild remains opaque, as Voolaine’s half-page of notes on Kraasna words from 1952 bears several names. On Voolaine’s 1966 field trip, Ernits (2018) identifies Voolaine’s sources as Egor Vassilev and Vassili Davydyč as well as Ivan Mihailovič Kuznecov (born 1892), the son of the speaker Matrëna Rodionovna [Kuznecova], who (as mentioned above) worked with Ojansuu.

---

[10] In describing the Lutsi, in contrast, Kallas (1894: 17) admires that the speakers have retained their Estonian language skills despite the community’s multilingualism. He gives the example of a man using four languages on a daily basis: Estonian for talking to his father, Latvian when speaking to his Latvian wife, and Russian for communicating with their common children who learn Russian in school, while his Catholic catechism, a book containing prayers and religious texts, was printed in Polish.

These names are technically part of the materials' contents, but they function like metadata: they provide contextual information that allows us to better understand the data and interpret related references in other writings about the community. For example, through anthroponyms we learn about consultants' lives, find family names and occupations (the name Kuznecov means 'blacksmith'), and gain insight into those families in the community who best preserved the language (see Weber 2021b). Russian patronyms link children to their fathers. Timofej Rodionov[ič] may be the brother of Matrëna Rodionovna, whose husband's name likely was Mihail, as can be seen in the patronym of Ivan Mihailovič Kuznecov. Ideally, these relationships would be confirmed by public records from Krasnogorodsk, but even without access to those we can still reconstruct the information.

Voolaine's diaries contributed to my understanding of other Kraasna texts as well. For example, interviews in Russian with Voolaine's 1966 consultants were also tape recorded for the University of Tartu Archives of Estonian Dialects and Kindred Languages in 1968 (Lindström et al. 2019). Ivan Kuznecov and Z. Stepanov are named as the consultants in the metadata for these recordings. From Ivan Kuznecov we learn more about Matrëna, who apparently lived to the age of 96 (ca. 1840–1936). She was possibly the last native speaker, as Timofej died in 1915, and Kallas's consultants were all over the age of 70 in 1901. Placing the death of the last native speakers in the 1930s fits with several sources, among others the report of an interviewee (born ca. 1924) for a 2004 newspaper article who confirms that he heard his grandfather speak a foreign language when he was young (Harju 2004). Kraasna is said in the literature to have disappeared sometime early in the 20th century, or in the first half of the 20th century, or before the Second World War. But this evidence from Voolaine's manuscripts allows us to narrow down the death of the last native speakers specifically to the mid-1930s.

As suggested by the preceding discussion, the knowledge embedded in the linguistic artefacts is more than just their "contents". Only when different projects and their associated sources are linked and combined does a holistic image emerge of the documentary activities and work trajectories that resulted in the materials. The interviewers in 1968 knew which words to elicit from their consultants, as these words had been documented for the same families some years prior. So the recorded interviews add a new layer of context to the elicited words, a new node in the network. As Clifford (1990: 57) reminds us, "noting of an event presupposes prior inscription": observations are made because they are linked in some way to information outside of the discourse such as a theoretical framework, other speech events, or examples in the literature. Thus, understanding a researcher's decisions "requires knowing something about what [context] was taken for granted when the notes were written" (Van Maanen 1988: 124). Voolaine knew Ojansuu's and Kallas's works and had previous fieldwork experience that shaped his records (e.g., through the places he went, the families he sought out, and the kinds of data he tried to elicit). On the level of content, it is thus possible that one textual artefact is influenced by two or more previous artefacts in ways that are not always obvious and transparent. In the Kraasna materials, this helps to account for the similarity of the songs recorded by Brandt, Kallas and Ojansuu, as well as the presence of a story about a fox tricking a wolf in both Kallas's and Ojansuu's manuscripts.[11] So for the user of legacy artefacts, intertextual links are part of the context that needs to be charted and recovered.

5.3 Rediscovering Ojansuu's recordings

The most interesting example of productive philological sleuthing comes from my study of Ojansuu's materials. After I had compared Kallas's manuscript and published sources and retrieved information from Voolaine's diaries, I knew enough to approach Ojansuu's artefacts with care. My work with Ojansuu's recordings shows how I could not begin to use the materials as data without first doing the detective work of restoring the links between different artefacts and versions of the texts and the underlying network of agents that were involved in their production.

Ojansuu's Kraasna materials had been copied and exchanged between various archives. The 1911–12 manuscripts form the fifth part of the *Estonica*, a 2,000-page collection of linguistic examples on 27 different Estonian varieties which is stored in the Literary Archives of the Finnish Literature Society, Suomalaisen Kirjallisuuden Seura. Yet, excerpts were also hand-copied for the Estonian Mother Tongue Society, Emakeele Selts, an academic society focusing on Estonian linguistics and philology. The hand-copied excerpts are stored in the Archive of Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language (Ermus et al. 2019). They consist of texts accompanied by unattributed commentary in an appendix on the last page. This last page contains an account of Ojansuu's fieldwork as told by his wife in 1938 as well as additional information about the circumstances of Ojansuu's fieldwork and the creation of the recordings. Apparently, he took a phonograph recorder along on both field trips. Some of the recordings he made had to be destroyed during fieldwork upon the consultant's request, likely because the speaker feared repercussions for interacting with Ojansuu, a foreigner whose appearance and questioning were suspicious at the start of the First World War (see also Weber 2021b). The remaining recordings had been given to Kalevalaseura, the Kalevala Society, a Finnish society for folkloristic research. The commentary on the last page of the manuscript concludes by saying that surviving recordings had probably been lost from the possession of the Kalevala Society; it is unclear if this was based on a comment by Mrs. Ojansuu or the unknown author's personal views. If it were true that the recordings had been lost, it would explain why they were unknown to the scholarly community. Out of curiosity, I decided to email the Kalevala Society and asked whether the recordings said to be lost in 1938 had since resurfaced. They forwarded my query to the staff at the sound archives of the Finnish Literature Society, who promptly offered to send me the digitised versions of the phonograph recordings they held on Kraasna so that I could listen to them and assess the value of their contents for my linguistic research.

---

[11] The story about the wolf and the fox is contained as fragments in Kallas 1903: 126 with dialect material inserted into the Estonian story. It was also collected by Ojansuu in 1912 for his manuscripts, possibly elicited from the same consultant. Ojansuu's phonograph recordings, presumably from 1914, contain a part of this story which is difficult to understand due to wear. It differs in wording which may indicate a different speech event or even a different consultant, but it is still the same story: The fox promises the wolf that it could catch fish by hanging its tail into an ice hole. The tail freezes onto the ice. The fox goes snitching and leads the village people to the hole where they batter the wolf because "the wolf poops into the well", a phrasing that is notably used in all three versions.

In the sound archives of the Finnish Literature Society there still exist three wax cylinders of Ojansuu's, one of which bears the title "Kraasna".[12] There are also five cylinders on Kraasna belonging to the collection of Otto Armas Väisänen, an ethnomusicologist who never visited the Kraasna community himself. Väisänen, a compatriot of Ojansuu who was working on South Estonian folk music around the same time, may have borrowed the phonograph recordings either from the archives or directly from Ojansuu. All the recordings had been copied to audiotape by the archive in 1963 and re-copied in the 1980s, possibly from the previous tape recording, with an introductory comment by the archivist in Finnish. When I listened to them I could not believe what I was hearing: it was actually the voice of a Kraasna speaker recorded one hundred years earlier! The narratives were, without a doubt, the texts contained in the manuscripts. Initially, it was difficult to understand the recordings, as wear on the material caused them to skip. Luckily, I managed to understand some phrases after several rounds of listening, which then enabled me to recognise what I was hearing as one of the texts published in the 2014 dialect collection. It was a narrative about a funeral which was transcribed for the text collection from manuscript AES 202. But despite having access to these transcriptions, it was not always easy to link what I heard on the audio recordings with the manuscript or published texts, as the various transcribers introduced innovations, changed word orders, and added or left out forms in the transcription process. Ultimately, I managed to transcribe seven of the eight recordings anew by tacking back and forth between the recordings and the manuscripts. The eighth recording contains songs which I do not recognise and cannot understand in their sung form. As they do not seem to be included in the manuscripts, I did not find an entry point to understanding, interpreting, and transcribing them. Overall, the recordings represent only a small percentage of the manuscript texts. Perhaps it is not surprising that several of the monologic narratives are preserved, since they feel object-like and are easier than other forms of discourse to transcribe. Since we know that Ojansuu’s 1914 fieldwork resulted in the production of more materials, it is possible that there are still more Kraasna recordings, treasures hidden in archives, yet to be rediscovered.

---

[12] For a detailed discussion of Ojansuu's phonograph recordings, see Weber (2021a).

6. My contribution to the research on Kraasna

The creation of new transcriptions was an integral part of my Master’s thesis, in which I aimed to establish links between the recordings and the manuscript texts. But the transcription process was anything but straightforward. As mentioned above, damage to the cylinder caused skips in the recordings, the age of the material created crackles, and the tape recording added another layer of white noise; all of this made the captured speech very difficult to hear. As a result, I opted to use the manuscripts as a basis for the new transcription, using slowed playback speed and rewind to make corrections and pick up small words missing from the manuscripts, as when the consultant appears to have faced away from the phonograph talking to a bystander or the researcher.

Compare the following three transcriptions of the same sentence from the narrative about burial customs. In Ojansuu’s manuscript, example (3), an abbreviated form of the previous sentence’s subject, ‘priest’, is copied into the example sentence, which is in a personal/active voice. The voice has implications for the case of the noun ‘candle’. Examples (3) and (4) both put ‘candle’ in the genitive, likely drawing on the transcribers’ knowledge of Estonian grammar. In the Mets et al. (2014: 280–81) text collection, the abbreviated transcription of ‘priest’ is expanded but is erroneously merged with the verb, resulting in a different verb in (4). To my ear, the recorded sentence without the subject occurs in the impersonal passive voice, which makes the object’s case nominative, as I transcribe it in (5). My transcription also resolves the concatenation of adverbial phrases *ḱät̀te kurra ḱät̀te a hüv̀vä ḱät̀te* ‘into the hand, into the left hand, but into the right hand’ in (3) and (4), which is nonsensical because it contains no object after *kurra ḱät̀te*. This additional object is sadly unintelligible. It sounds to me like *padaroži̯j*, likely a grave good of some kind (see Weber 2021a).

(3) *täl̆lè p. ańd́ ḱät̀te kura ḱät̀te, a hüv̀vä ḱät̀te ańd́ kündlik̆keze*

täl̆lè	***p.***	***ańd́***	*ḱät̀te*	*kura*	*ḱät̀te*	***a***	*hüv̀vä*
3SG.ALL	priest	give.3SG.PST	hand.ILL	left	hand.ILL	but	right.ILL

ḱät̀te	***ańd́***	***kündlik̆kez̦e***
hand.ILL	give.3SG.PST	candle.DIM.GEN

‘The priest gave into his hand, into his left hand, but into his right hand [he] gave a candle’ (Ojansuu 1914 in AES 202: 16)

(4) *täl̆lè p[ap̀] pāńd́ ḱät̀te kura ḱät̀te, a hüv̀vä ḱät̀te ańd́ kündlik̆keze*

täl̆lè	***p[ap̀]***	***pāńd́***	*ḱät̀te*	*kura*	*ḱät̀te*	*a*	*hüv̀vä*
3SG.ALL	priest	put.3SG.PST	hand.ILL	left	hand.ILL	but	right.ILL

ḱät̀te	*ańd́*	*kündlik̆keze*
hand.ILL	give.3SG.PST	candle.DIM.GEN

‘The priest put into his hand, into his left hand, but into his right hand [he] gave a candle’ (Ojansuu 1914, digitized by Mets et al. [2014: 280–281] from AES 202: 16)

(5) *täl̆lè andas kurra ḱät̀te (padaroži̯j?), hüv̀vä ḱät̀te andas kündlek̆kene*

täl̆lè	***andas***	*kurra*	*ḱät̀te*	***[padaroži̯j]***	*hüv̀vä*
3SG.ALL	give.IPS/PASS	left.ILL	hand.ILL	[object]	right.ILL

ḱät̀te	***andas***	***kündlek̆kene***
hand.ILL	give.IPS/PASS	candle.DIM.[NOM]

‘They gave [= It was given] into his left hand an [object], into his right hand they gave a candle’ (Ojansuu 1914, transcribed by Weber from SKSÄ A 530/6, fonokop 32/6)

In other words, the different sources present different descriptions and analyses of recorded Kraasna utterances. As the transcription differs between textual artefacts, the grammatical structure, features, and glosses change. This has consequences for users of these archival materials and may hinder the use of these data sets altogether. At the same time, publications based on these data sets or citing examples from them become new artefacts which need to be studied as well (Weber 2019, 2020).

Without the manuscripts and the published text collection based on the manuscripts, I would not have been able to progress through the audio transcription process nearly as swiftly as I did, as the existing transcriptions offered alternate points of view on the recording.[13] The manuscript sources embody Ojansuu's interpretation and understanding of Kraasna grammar, as he supplements his transcriptions with occasional commentary on grammatical forms. Ojansuu's manuscripts thus contain an abstracted representation of his grammatical knowledge, which I repurpose in my new transcriptions. Although I tried to give primacy to the recording, my version is ultimately a new interpretation of the fieldwork event that draws from the recording, the manuscript, and my own understanding of how to represent Kraasna in written form. This means that my transcriptions convey not only the narrative content and language use on the recordings, but are also filtered through Ojansuu's and ultimately my own understanding and interpretation, making me a part of the artefacts I produce.

While my transcriptions may be closer to the spoken form than the archived manuscripts, I cannot claim that my transcriptions are the "authoritative version" (Seidel 2016: 31) or the "correct" interpretation. Linguistic transcriptions, as value-adding procedures, are intrinsically agent-driven and not predetermined by or a part of the data; nor are they replicable using rules and conventions. Anyone may create new and divergent transcriptions based on the recordings by applying their own knowledge and craftsmanship, and in the process they leave traces of their interaction with the artefacts (see Weber & Klee 2020). Consequently, the recordings never become dispensable just because we have (several) transcriptions of them; nor can we discount the older transcriptions in favour of a newly transcribed version.

Minor differences in the data, e.g., an additional or missing word or an omitted diacritic, may appear negligible from a broad perspective, as we accept that to err is human. Yet, as linguists, we are also interested in the rare and rarest phenomena, providing support for or against our hypotheses; a singular occurrence may give important insights into the inner workings of a language or highlight a current trend in language development. For the Kraasna data, the inconsistencies between the recordings and the transcriptions raise questions about language use in ways that make a difference for linguistic analysis or even disprove theories. Some of the changes introduced by the transcriber affect phenomena used for determining isoglosses, and thereby the classification of the Kraasna dialect, e.g., the use of *um* for third person singular 'be' with a raised vowel instead of *om* as in other South Estonian dialects. Some of the word order patterns found in the Kraasna manuscripts also have typological implications, as the high frequency with which the subject occurs in sentence-final position is atypical for Finnic. Lastly, the Kraasna dialect is said in the literature to iotate (i.e., insert *j* before) the vowels *i* and *e* in word-initial position, likely a contact phenomenon under Russian influence. The manuscripts provide several instances without iotation, yet in the recordings, one can clearly hear that these instances contain iotated vowels (Weber 2021a). So, without careful analysis of *why we find what we do* in all the sources, we may draw unwarranted conclusions about the language.

---

[13] I should qualify my use of the term "swift". I listened to each recording (2–4 minutes each) for about six to eight hours, a total of sixty hours for all of the recordings. And there are still some parts I cannot understand.

I am proposing this perspective on language research because it helps us to have a more realistic image of our work. Despite my own initial views, I have to conclude that Ojansuu's transcriptions are not "wrong", even if the recordings suggest additional words or different word order. Instead they are reflections of his interpretations, influenced by the decisions he made. For example, Ojansuu opted to leave out certain repetitions or self-corrections, as they may not have seemed useful to his research objective, whereas modern research may be very interested in investigating these phenomena as they pertain to cognitive processes in self-correction or stylistics. But my transcription is not simply a matter of "correcting" Ojansuu's mistakes, because in making my own transcript, I also brought my own decisions into the process.

For example, in my working drafts I used the modern Võro notation of <q> for the glottal stop instead of the standardised transcription with the symbol <ˀ>. (The name *Võro* refers to the largest and most standardised variety of South Estonian and is often used interchangeably with 'South Estonian' despite the existence of the other South Estonian varieties *Seto*, *Mulgi*, and *Tarto*.) This decision facilitated my use of a standard keyboard layout, and the symbol could be altered quickly with a simple search-and-replace. While this choice was initially made spontaneously without a particular theoretical framework in mind in order to enable fast typing on my computer, it is an example of the researcher using their judgement to make a decision that affects the representation of the data. This decision, as with any of the others I made, is certainly open to discussion and evaluation by colleagues present and future.[14] For example, those who see my working drafts may comment on and criticise the foreign orthographic element I introduced. Although I do not consider myself a player in South Estonian language politics, the fact that I chose to use the modern Võro <q> in my drafts in part reflects my stance on Kraasna as a South Estonian variety and my ideas about the orthography of South Estonian in general. I restored this symbol to <ˀ>, the symbol of the Uralic Phonetic Alphabet used in Estonian dialectology, in the published materials, but I could have decided to use an entirely different approach, e.g., Cyrillic-based characters, to make the material more accessible for the modern population of Krasnogorodsk. My decision contrasts with the one made by Balodis (2015, 2019), who created a Latvian-based practical orthography for Lutsi that makes the language more easily accessible to the Latvian public. Even though members of the Võro-speaking public have criticised <q> for its foreignness, Balodis and I both used this grapheme at various points in our transcriptions and publications (see also Koreinik 2013: 8). This is something future users may wonder about when they look at my unpublished records. And it shows how as researchers we leave our traces on a project, even when we are trying to avoid involving ourselves in the meta-linguistic discourses, language policies, and ideologies at play in the linguistic communities where we work. My brief and not entirely reflexive decision about transcription and orthography ties me personally to the artefacts of my research.

---

[14] As Gurd (2015) puts it: "One cannot describe past philologies as erroneous without acknowledging the likelihood that one's own certainties will one day fall under a similarly critical eye".

The personal or political element in language documentation brings us to the final human factor in working with legacy materials – the editor and author functions of the researcher. This is one of the reasons why researchers may be reluctant to share their fieldnotes or working drafts of materials and datasets: so much of themselves is in them (see Jackson 1990). And if we as researchers are already part of the research, it only seems sensible to address our role in the process openly by telling our own narratives, rather than feigning objectivity. My own decisions as well as my interpretations of Ojansuu's decisions become part of the artefacts I produce. My version is a new instance in the genealogy of Kraasna legacy materials, whether I justify and record my decisions in the meta-documentation or in my methodology section or not (though, of course, it is preferable if I do). In fact, acting as if I somehow stood outside the materials and did not leave traces on them would *reduce* the transparency of the research. When it comes to legacy data, it is not just an absence of metadata that makes reconstruction necessary. It always takes work to (re)discover the narratives behind our linguistic research.

7. Conclusion: A philological approach to linguistic legacy materials

The philological approach to legacy materials that I adopt here follows Seidel (2016) in drawing attention to the human factors in the artefacts created through fieldwork and the broader research process, with the goal of making such factors visible, if not manageable. Taking this critical look into our work might be frightening at first, as researchers may fear being judged on their inadequacies (Jackson 1990; Burrell 2016). Yet, putting the researcher and their artefacts at the centre of the research provides the only honest basis there is for scientific and meta-scientific discourse. It also links linguistic fieldwork and the curation of legacy materials to discourses about the role of the researcher in ethnography.

Ideally, the researcher's own accounts would be sufficient to enable future researchers to revisit the original research situation. This is why so much attention has been given to metadata and transparency in documentary linguistics. But even the thickest metadata cannot replace philological care in working with and carefully comparing language data, the textual artefacts that contain the data, and secondary sources. Metadata also cannot replace the need for researchers to adopt a reflexive stance, because we ourselves form part of the artefacts we leave behind, through the traces left by decisions we make in creating, adapting, or interacting with our data. Philology respects historical researchers and their decisions by treating each version of an artefact in its own right, while supporting future researchers by making our traces in our own present-day research as transparent as we can through metadata and meta-documentation. But even these efforts will likely provide only a few breadcrumbs for those in the future to use as a starting point.

I would like to conclude with some reflections on what we, the current generation of researchers, may learn from the study of legacy materials, whether or not we ourselves are making direct use of any data from past documentation. A crucial question is how we should manage our own subjectivity in our research. It is certainly tempting to aspire to objectivity by imagining a research agenda that can be independent of the researchers, allow for straightforward reproducibility, or rely only on random sampling. But these goals are not in fact applicable to current documentary linguistic research, just as they create some of the challenges we encounter when working with legacy data, such as authoritative versions of sources without a clear editor, only partial information behind the sources, missing accounts of past decision-making processes, and orphaned data sets which are not clearly linked to other artefacts. Certainly we do our best to provide as much as metadata we can, but we do not in fact know how much will be enough, nor do we have time to lay out every last thing that could be said. So as a scholarly community, we need to find ways to address the irreducible role of the researcher in our endeavour and recognise the academic value of working with language data. Instead of treating the curation of artefacts as an ancillary activity along the lines of a literature review, we should recognise that philological work with legacy materials generates new knowledge that can help reconcile current and past research, preserve the efforts of our predecessors, and support a reflexive stance towards our own work.

Archival sources

AES 202 = Akadeemilise Emakeele Seltsi Ülevaated 202. Häälikuloolisi andmeid ja tekste Kraasna murdest (Archive number AES0202) [Overviews of the Academic Mother Tongue Society 202. Texts and data on historical phonology from the Kraasna dialect.]. Authored by Heikki Ojansuu. 1938. The Archive of the Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language (Eesti Keele Instituudi Eesti Murrete ja

Soome-Ugri Keelte Arhiiv). PID: 11297/3-00-0000-0000-0000-02D36L. http://emsuka.eki.ee/view/book/175/0 (accessed 2021-12-30).

Comment: “1938” is the year given in the archival record and indicated on the front page of the materials. But it cannot have been written by Ojansuu in 1938, as he passed away in 1923. My best inference is that 1938 was the date when the materials were compiled or donated to the archive. The texts are derived from Ojansuu’s 1914 field research, as indicated on top of each page.

ES MT 224 = Emakeele Seltsi Murdetekstid. Kraasna murdetekste ja muid märkmeid (Archive number ESMT0224) [Dialect texts of the Mother Tongue Society. Kraasna dialect texts and other notes.]. Kopeeritud H. Ojansuu 1911. või 1912. a. reisu ülestähendustest [Copied from Heikki Ojansuu’s 1911/12 fieldwork records.] The Archive of the Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language (Eesti Keele Instituudi Eesti Murrete ja Soome-Ugri Keelte Arhiiv). PID: 11297/3-00-0000-0000-0000-031BEL. http://emsuka.eki.ee/view/book/78/0 (accessed 2021-12-30).

Comment: I include information about the date the records were copied because there is no author or date mentioned in the title of the materials. The final page, titled “appendix”, contains information on personal communication with Mrs. Ojansuu in January 1938, making this the earliest possible date of the manuscript’s compilation.

Estonica I–V = Heikki Ojansuu 1910–1911. Archive of Heikki Ojansuu. Literary Archive at the Finnish Literature Society. Contains a typewritten copy of Ojansuu’s 1914 materials from AES 202, prepared on 05.04.1939.

Comment: The Estonica contains a typewritten copy of Ojansuu’s 1914 fieldwork materials as volume one which was copied from a manuscript in Tartu, AES 202. The copies were created in 1939; the identity of the typist is unknown. Volume five contains the remaining Kraasna materials as manuscripts. It is unknown when they were written but they may originate from Ojansuu himself, while the associated field trip is identified as 1911/12 by ES MT 224. Page 1954 contains the name of “Matrëna Razivonova” next to the text of a song.

EFAM Kallas M 4 = Materials of Estonian Folkloristic History. Authored by Oskar Kallas 1901. Estonian Folklore Archives at the Estonian Literary Museum.

Comment: Kallas’s collections consist of several binders of documents and personal diaries. M4 contains three notepads marked with red Roman numerals. They were cut up from the calligraphy book of a second grader named Boris Vaxtin issued in 1884. The collection also contains a pamphlet titled 'Krassnojer Esten' (German for Krasnoj Estonians) with excerpts from Neus' 1950 collection of folk songs.

EFAM Voolaine M 1 = Materials of Estonian Folkloristic History. Authored by Paulopriit Voolaine. Estonian Folklore Archives at the Estonian Literary Museum.

Comment: Voolaine's collection consists of his field diaries, in which he records his travel logs and takes notes during interviews. The notebooks are not exclusively related to his Kraasna research and do not always bear a date.

SKSÄ fonokop 32/4–8 & SKSÄ A 530/4–7, 9 = Perinteen ja nykykulttuurin kokoelman äänitteet [Recordings of the collection of traditional and modern culture]. Collected by Armas Otto Väisänen [Heikki Ojansuu] 1914. Digitised tape recording from 1963 and the 1980s. Sound archives of the Finnish Literature Society.

SKSÄ fonokop 136/7–9 & SKSÄ A 502/15–17 = Perinteen ja nykykulttuurin kokoelman äänitteet [Recordings of the collection of traditional and modern culture]. Heikki Ojansuu [1914]. Digitised tape recording from 1963 and the 1980s. Sound archives of the Finnish Literature Society.

Comment: For more information on the contents of the individual recordings, see also Weber (2021a).

Acknowledgments

I would like to express my deepest gratitude to Peter Austin and Lise Dobrin, who both encouraged me to look at the hidden stories within the Kraasna legacy materials and to explore their implications for documentary linguistics. I am also grateful to all colleagues who supported and commented on my work on the Kraasna legacy materials, including four anonymous reviewers who provided diverse and insightful comments on this paper. Special thanks to the societies, archives, and their staff who enabled me to conduct my archival research: the Estonian Folklore Archives of the Estonian Literary Museum in Tartu, the University of Tartu Archives of Estonian Dialects and Kindred Languages, the Archive of the Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language, the literary archive at the Finnish Literary Society, the sound archive of the Finnish Literary Society as well as the Kalevala Society and the KOTUS archives.

References

Austin, Peter K. 2010. Current issues in language documentation. *Language Documentation and Description* 7. 12–33.

Austin, Peter K. 2013. Language documentation and meta-documentation. In Mari Jones & Sarah Ogilvie (eds.), *Keeping languages alive: Documentation, pedagogy, and revitalisation*, 3–15. Cambridge: Cambridge University Press.

Balodis, Uldis. 2015. Writing down Lutsi: Creating an orthography for a South Estonian variety of Latgale. Rakstu formas izveide Ludzas apkārtnes dienvidigauņu valodas izloksnei. In Andra Kalnača (ed.), *Valoda: Nozīme un forma 6. Valodas sistēma un lietojums*, 55–67. Riga: University of Latvia.

Balodis, Uldis. 2019. Expeditions among the Lutsi Estonians and the design of language learning materials. In Sofia Björklöf & Santra Jantunen (eds.), *Multilingual Finnic: Language contact and change*, 439–478 (Uralica Helsingiensia 14). Helsinki: Suomalais-Ugrilainen Seura.

Burrell, Jenna. 2016. "Through a screen darkly": On remote, collaborative fieldwork in the digital age. In Roger Sanjek & Susan W. Tratner (eds.), *eFieldnotes: The makings of anthropology in the digital world*, 132–152. Philadelphia: University of Pennsylvania Press.

Clifford, James. 1990. Notes on (field)notes. In Roger Sanjek (ed.), *Fieldnotes: The makings of anthropology*, 47–70. Ithaca: Cornell University Press.

Coulmas, Florian. 2016. *Guardians of language: Twenty voices through history*. Oxford: Oxford University Press.

Dobrin, Lise M. & Josh Berson. 2011. Speakers and language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 187–211. Cambridge: Cambridge University Press.

Ermus, Liis, Mari-Liis Kalvik, & Tiina Laansalu. 2019. The Archive of Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language. In Sofia Björklöf & Santra Jantunen (eds.), *Multilingual Finnic: Language contact and change* (Uralica Helsingiensia 14), 351–366. Helsinki: Suomalais-Ugrilainen Seura.

Ernits, Enn. 2012. Fr. R. Kreutzwald lõunaeestlaste piire kompimas. In Jüvä Sullõv (ed.), *Õdagumeresoomõ piiriq*, 30–65 (Võro Instituudi Toimõndusõq 26). Võru: Võro Instituut.

Ernits, Enn. 2018. Kraasna rahvalaulude esimestest üleskirjutustest. In Jüvä Sullõv (ed.), *Valitsõmisjaotusõst keeleaoluuni*, 157–201 (Võro Instituudi Toimõndusõq 33). Võru: Võro Instituut.

Faster, Mariko. 2014. Jumalõluuk nõssi taivadõ ehk keelesaared jälle pildil. *Emakeele Seltsi Aastaraamat* 60. 263–272.

Frog. 2013. Revisiting the historical-geographic method(s). In Karina Lukin, Frog, & Sakari Katajamäki (eds.), *Limited sources, boundless possibilities: Textual scholarship and the challenges of oral and written texts*, 18–34 (The Retrospective Methods Network Newsletter 7). Helsinki: University of Helsinki.

Gibson, Catherine. 2017. From literary languages to dialectal varieties to micro languages? Historical perspectives on language policy towards South Estonian and Latgalian. In Marja Siiner, Kadri Koreinik, & Kara Brown (eds), *Language policy beyond the state*, 163–182 (Language Policy 14). Cham: Springer.

Good, Jeff. 2007. The ecology of documentary and descriptive linguistics. *Language Documentation and Description* 4. 38–57.

Grünthal, V[illem]. 1912. Eesti asundused lätlaste seas. *Eesti Kirjandus* 9. 369–373.

Gurd, Sean. 2015. Philology and Greek literature. In *Oxford Handbooks Online*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199935390.013.65 (accessed 2021-12-30).

Harju, Ülle. 2004. Eesti jalajälg’ Pihkvamaal. *Uma Leht* 53. https://vana.umaleht.ee/vana/index.php?leht=73&keel=3 (accessed 2021-12-30).

Helimski, Eugen 1997. *Die Matorische Sprache* (Studia Uralo-altaica 41). Szeged: University of Szeged.

Heller-Roazen, Daniel. 2005. *Echolalias*. New York: Zone Books.

Iva, Sulev. 2015. Liivi, Ludzi ja Kraasna maarahva kiil. *Keel ja Kirjandus* 7. 515–517.

Jackson, Jean E. 1990. “I am a Fieldnote”: Fieldnotes as a symbol of professional identity. In Roger Sanjek (ed.), *Fieldnotes: The makings of anthropology*, 3–33. Ithaca: Cornell University Press.

Kallas, Oskar. 1894. *Lutsi maarahvas*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kallas, Oskar. 1901. *Die Wiederholungslieder der estnischen Volkspoesie. I. Folkloristische Untersuchung* (Suomalais-Ugrilaisen Seuran Toimituksia 16). Helsingfors [Helsinki]: Suomalais-Ugrilainen Seura.

Kallas, Oskar. 1903. Kraasna maarahvas. *Suomi* 4(10). Helsinki: Suomalaisen Kirjallisuuden Seura.

Kallas, Oskar. 1904. *Die Krasnyjer Esten. Volkskundliche Beschreibung* (Verhandlungen der Gelehrten Estnischen Gesellschaft 21). Jurjew [Tartu]: Gelehrte Estnische Gesellschaft.

Kallio, Petri. 2014. The diversification of Proto-Finnic. In Joonas Ahola & Frog (eds.), *Fibula, fabula, fact: The Viking Age in Finland*, 155–168 (Studia Fennica Historica 18). Helsinki: Suomalaisen Kirjallisuuden Seura.

Kask, Arnold. 1956. *Eesti murded ja kirjakeel*. Tallinn: Valgus.

Katz, Hartmut. 1979. *Selkupische Quellen. Ein Lesebuch* (Studia Uralica 2). Vienna: Verband der Wissenschaftlichen Gesellschaft.

KeeleS. 2020 [2011]. Keeleseadus [Language Act]. https://www.riigiteataja.ee/en/eli/ee/521032019016/consolide/current (accessed 2021-06-30).

Klooster, Tiina. 2015. Individual language change: a case study of Klavdiya Plotnikova’s Kamas. MA thesis. University of Tartu.

Klumpp, Gerson. 2013. On Kai Donner’s phonograph records of Kamas. *Finnisch-ugrische Mitteilungen* 37. 45−59.

Koreinik, Kadri. 2011. Agency lost in the discourse of language endangerment: Nominalisation in discourse about South Estonian. In Helle Metslang, Margit Langemets, & Maria-Maren Sepper (eds.), *Eesti Rakenduslingvistika Ühingu aastaraamat 7. Estonian Papers in Applied Linguistics 7*, 77–94. Tallinn: Eesti Rakenduslingvistika Ühing.

Koreinik, Kadri. 2013. *The Võro language in Estonia: ELDIA Case-Specific Report* (Studies in European Language Diversity 23). Mainz: ELDIA.

Künnap, Ago. 1991. Zur Erinnerung an die letzte Sprecherin des Kamassischen. *Linguistica Uralica* 27(2). 150–151.

Laanest, Arvo. 1982. *Einführung in die ostseefinnischen Sprachen*. Hamburg: Helmut Buske.

Lehmann, Christian. 2004. Data in linguistics. *The Linguistic Review* 21(3–4). 275–310.

Lindström, Liina, Pärtel Lippus, & Tuuli Tuisk. 2019. The online database of the University of Tartu Archives of Estonian Dialects and Kindred Languages and the Corpus of Estonian Dialects. In Sofia Björklöf & Santra Jantunen (eds.), *Multilingual Finnic: Language contact and change*, 327–350 (Uralica Helsingiensia 14). Helsinki: Suomalais-Ugrilainen Seura.

McGann, Jerome. 2013. Philology in a new key. *Critical Inquiry* 39(2). 327–346.

Mets, Mari, Anu Haak, Triin Iva, Grethe Juhkason, Mervi Kalmus, Miina Norvik, Karl Pajusalu, Pire Teras, Tuuli Tuisk, & Lembit Vaba. 2014. *Eesti murded IX. Lõunaeesti keelesaarte tekstid*. Tallinn: Eesti Keele Instituut.

Nathan, David. 2010. Archives and language documentation: From disk space to MySpace. *Language Documentation and Description* 7. 172–208.

Neus, Heinrich. 1850. *Ehstnische Volkslieder*. [Reval] Tallinn: Ehstländische Literärische Gesellschaft.

Ojansuu, Heikki. 1912. Ein südestnischer beitrag zur stufenwechseltheorie. *Finnisch-ugrische Forschungen* 12. 147–149.

Pajusalu, Karl. 2005. Keelesaarte *se*-minevik. In *Tartu Ülikooli Lõuna-Eesti keele- ja kultuuriuuringute keskuse aastaraamat*, vol. 4, 145–150. Tartu: Tartu Ülikool.

Pajusalu, Karl. 2007. Estonian dialects. In Mati Erelt (ed.), *Estonian language*, 231–272 (Linguistica Uralica Supplementary Series 1). Tallinn: Eesti Teaduste Akadeemia.

Pajusalu, Karl. 2009. The reforming of the Southern Finnic language area. In Jussi Ylikoski (ed.), *The quasquicentennial of the Finno-Ugrian society*, 95–107 (Suomalais-Ugrilaisen Seuran Toimituksia 258). Helsinki: Suomalais-Ugrilainen Seura.

Pajusalu, Karl. 2012. Phonological innovations of the Southern Finnic languages. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric northern Europe*, 201–224 (Suomalais-Ugrilaisen Seuran Toimituksia 266). Helsinki: Suomalais-Ugrilainen Seura.

Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll, & Jüri Viikberg. 2018. *Eesti murded ja kohanimed.* Tallinn: Eesti Keele Sihtasutus.

Palola, Elina. 2020. What is philology? Philology and its utilisation in the study of linguistic variation and change. In Margit Langemets, Maria-Maren Linkgreim, Helle Metslang, & David Ogren (eds.), *Eesti Rakenduslingvistika Ühingu aastaraamat 16. Estonian Papers in Applied Linguistics 16*, 157–175. Tallinn: Eesti Rakenduslingvistika Ühing.

Sammallahti, Pekka. 1977. Suomalaisten esihistorian kysymyksiä. *Virittäjä* 81. 119–136.

Sanjek, Roger. 1990. A vocabulary for fieldnotes. In Roger Sanjek (ed.), *Fieldnotes: The makings of anthropology*, 92–121. Ithaca: Cornell University Press.

Seidel, Frank. 2016. Documentary linguistics: A language philology of the 21st century. *Language Documentation and Description* 13. 23–63.

Setälä, Emil Nestor. 1901. Über die transskription der finnisch-ugrischen sprachen. *Finnisch-ugrische Forschungen* 1. 15–52.

Shils, Edward. 1981. *Tradition*. Chicago: The University of Chicago Press.

Stipa, Günter Johannes. 1990. *Finnisch-ugrische Sprachforschung* (Suomalais-Ugrilaisen Seuran Toimituksia 206). Helsinki: Suomalais-Ugrilainen Seura.

Turner, James. 2014. *Philology: The forgotten origins of the modern humanities*. Princeton: Princeton University Press.

Van Maanen, John. 1988. *Tales of the field: On writing ethnography.* Chicago: University of Chicago Press.

Viitso, Tiit-Rein. 1978. Läänemeresoome esimese silbi õ ajalugu. *Keel ja struktuur* 10. 81–117.

Viitso, Tiit-Rein. 1985. Läänemeresoome murdeliigenduse põhijooned. *Keel ja Kirjandus* 28(7). 399–404.

Viitso, Tiit-Rein. 2007. Rise and development of the Estonian language. In Mati Erelt (ed.), *Estonian language*, 130–230 (Linguistica Uralica Supplementary Series 1). Tallinn: Eesti Teaduste Akadeemia.

Viitso, Tiit-Rein. 2008. *Liivi keel ja läänemeresoome keelemaastikud*. Tartu - Tallinn: Eesti Keele Sihtasutus.

Weber, Tobias. 2016. *Kraasna. The state of documentation and description of an extinct South Estonian dialect*. BA thesis. Ludwig-Maximilians-Universität München. https://kraasna.wordpress.com/ (accessed 2021-12-30).

Weber, Tobias. 2019. Can computational meta-documentary linguistics provide for accountability and offer an alternative to “reproducibility” in linguistics? In Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, & Milan Dojchinovski (eds.), *2nd Conference on Language, Data and Knowledge (LDK 2019)*, 26:1–26:8 (OpenAccess Series in Informatics 70). Dagstuhl: Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Weber, Tobias. 2020. A philological perspective on meta-scientific knowledge graphs. In Ladjel Bellatreche, Mária Bieliková, Omar Boussaïd, Barbara Catania, Jérôme Darmont, Elena Demidova, Fabien Duchateau, Mark Hall, Tanja Merčun, Boris Novikov, Christos Papatheodorou, Thomas Risse, Oscar Romero, Lucile Sautot, Guilaine Talens, Robert Wrembel, & Maja Žumer (eds.), *ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium*, 226–233. Cham: Springer.

Weber, Tobias. 2021a. A linguistic analysis of Heikki Ojansuu's phonograph recordings of Kraasna. In Uldis Balodis & Karl Pajusalu (eds.), *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 12(2), 343–390. Tartu: University of Tartu.

Weber, Tobias. 2021b. Consultant identity in historical language data: Anthroponyms as a tool or as an obstacle? In Anna Choleva-Dimitrova, Maya Vlahova-Angelova, & Nadezhda Dancheva (eds.), *Слѣдовати достоитъ. Proceedings of the International Onomastic Conference "Anthroponyms and Anthroponymic Researches in the Beginning of 21st Century"*, 165–175. Sofia: Bulgarian Academy of Sciences.

Weber, Tobias & Mia Klee. 2020. Agency in scientific discourse. *Bulletin of the Transilvania University of Braşov Series IV: Philology and Cultural Studies* 13(1). 71–86.

Wenzel, Siegfried. 1990. Reflections on (new) philology. *Speculum* 65. 11–18.

Winkler, Eberhard (ed.). 1994. *Salis-Livische Sprachmaterialien* (Veröffentlichungen des Finnisch-Ugrischen Seminars an der Universität München C 21). Munich: Ludwig-Maximilians-Universität München.

Winkler, Eberhard. 1997. *Krewinisch. Zur Erschließung einer ausgestorbenen ostseefinnischen Sprache* (Veröffentlichungen der Societas Uralo-Altaica 49). Wiesbaden: Harrassowitz.