Language Documentation & Conservation

Volume
19 (2025)

Pages
328 - 347

Received
22 April 2025

Handle
https://hdl.handle.net/10125/74838

Online
http://nflrc.hawaii.edu/ldc

Citation
Drayton, Holly. 2025. Translating Language Documentation Texts: A survey of macro-level translation methods used in the translation of endangered language texts in a digital language archive. *Language Documentation & Conservation* 19: 328-347

Translating Language Documentation Texts: A survey of macro-level translation methods used in the translation of endangered language texts in a digital language archive

Holly Drayton

Portsmouth University, Newcastle University

A central aim of Language Documentation is the creation and storage of lasting, multi-purpose collections of endangered-language texts, with accompanying annotation. Despite many important developments in the field over the last two decades, there has been no systematic attempt to research or theorise the translation processes or resulting translated texts involved in language documentation work. In this paper I argue that translation practices in Language Documentation can be linked to the field's social justice goals, and decisions to translate (or not) determine whether a collection can be useful beyond the initial documentation project. I present a study of macro-level translation methods based on a corpus of 1088 texts from 10 deposits in the Endangered Languages Archive (ELAR). I survey the decisions made by researchers of which texts to translate, which genres to translate, which languages to translate into and choice of translator. This research demonstrates how a key development from the field of Translation Studies, the notion of Translator Agency, could be integrated with Language Documentation to facilitate the production of more multi-purpose, useful translated texts that better support the field's social justice aims.

E-ISSN 1934-5275

1. Introduction[1]

Over recent years, the field of Language Documentation has increasingly focused on critically analysing the accessibility and useability of digital language archives which are central to preserving and disseminating language documentation texts (Kung et al. 2020; Austin 2021; Burke et al. 2022). This strand of research aims to better understand and improve the potential for language documentation text collections to be accessed and used by end-users beyond the initial documentation project (Thieberger 2016; Burke et al. 2020; 2022; Thieberger & Harris 2022; Babinski et al. 2022; Yi et al. 2022). Discussion about the accessibility of language archives feeds into ongoing conversations around reproducibility (Berez-Kroeker et al. 2018), open science, and the push for increased recognition of documentation and archiving work (Austin 2016). Most importantly for this paper, data accessibility is also central to ensuring that documentation benefits language communities and supports linguistic and social justice goals.

Many of the languages represented in digital archives are endangered,[2] minoritised, and/or not widely understood. Therefore, they need to be accompanied by metadata and annotation, including transcription, glossing and translation, to render them accessible to a wide range of end-users who do not understand the original texts (Austin 2019; Sullivant 2020). Transcription and glossing tend to be reasonably standardised across documentation projects, and these practices have been explored and theorised (e.g., the Leipzig Glossing Rules[3]). However, as this study demonstrates, translation output varies dramatically between individual documentary collections, and there has been very little discussion, guidance or theorisation on translation methods in the literature. For example, a thorough search of the literature from the past two and a half decades reveals just two published articles that focus on the act of translating documentation texts, both of which, while valuable contributions, were published almost twenty years ago (Evans & Sasse 2007; Woodbury 2007). In fact, translation is most commonly mentioned in the literature when it is noted that translated documentary texts are unsuitable for use in revitalisation initiatives within the communities that they are intended to serve (Evans & Sasse 2007; Hinton 2011; Penfield & Tucker 2011; Rice 2011; Burke et al. 2022). Therefore, there is an urgent need for updated research into translation practices, with the aim of improving the suitability of translated texts for end-users.

In the Language Documentation literature there is "astonishingly little" published guidance on translating language documentation texts (Beier 2015). Although it is acknowledged that professional translation is a skill, art, or science that requires training (Schultze-Berndt 2006: 232), the general impression from the literature is that translation is work that researchers are expected to carry out, almost as an afterthought to the principal activity of data collection, with "no, or minimal, theory and methodological guidance" (Beier 2015). It has been suggested that researchers should not even attempt to "meet the standards of professional literary or scientific translation", and that translations should be considered as providing no more than a "clue" to the meaning of the original text (Schultze-Berndt 2006: 232). This viewpoint appears to be incompatible with producing translated texts that are multipurpose and able to meet the requirements of their target end-users, many of whom cannot understand the original text.

Since the Language Documentation literature is lacking in translation guidance, we can turn to the field of Translation Studies, where researchers have developed both practical and theoretical knowledge in several transferable areas. One key development was the 'cultural turn' in the late 1980s and early 1990s (e.g., Snell-Hornby 1988; Bassnett & Lefevre 1990) which argued for situating language within the context of the culture(s) of its speakers. Since then, Translation Studies researchers have developed theories and practical skills for translating between languages spoken by people from vastly different cultures (Cronin 2003; Pedersen 2011; Katan & Taibi 2021). More recently, the 'outward turn' (Tymoczko 2007; Marais & Delgado Luchner 2018; Bassnett & Johnston 2019; Cronin 2022) has highlighted the Eurocentric traditions of Translation Studies and has called for a broadening of the role, function, and definition of translation, going beyond Western data (Cronin 2017; 2020; Bassnett & Johnston 2019).

[1] I am grateful to Begoña Rodríguez de Céspedes for her encouragement and detailed comments at earlier stages of this research. I am also thankful to thoughtful comments and suggestions from Vasiliki Vita. I also thank the audience of LD&A 2022 for raising thought-provoking questions on translation in language documentation. Finally, I would like to thank two anonymous reviewers whose comments helped to substantially improve the paper.

[2] I note that the term *endangered languages* has, itself, been associated with objectification and extraction (Roche 2022). I use the term in this paper since I refer to the Endangered Languages Archive.

[3] https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

At the centre of these movements is the role, power, and agency of the Translator (Venuti 1998b; 1998a; 2018; Cronin 2006; Tymoczko 2007; Pym 2012; Baker 2013). Acknowledging Translator agency frames translation as a social, human practice. This is in opposition to the creeping 'McDonaldisation' of translation, which is emblematic of the move towards automated, invisible and instantaneous machine translation (Cronin 2017: 30), where the labour of translation is hidden.

A translator's agency applies throughout the translation process, starting from the decision of what to translate or not translate (Pym 2012), then continues with decisions over what to "change, add or omit" (Chesterman 1997: 182). A translator working across cultural differences holds particular power and agency (Tymoczko 2007), which can be realised and manipulated as a means to fulfil the translator's mission. In the most positive and engaged contexts, the translator's power can extend to social agency by contributing to education, human rights, and empowerment of marginalised communities (Inghilleri 2008; Taibi & Ozolins 2016; Taibi 2017; Katan & Taibi 2021). On the other hand, a translator's power and agency can also contribute to negative social outcomes such as the theft of cultural knowledge, objects, or land; imposition of a dominant culture; and forced conversion to religion (Tymoczko 2007).

These examples demonstrate how translators' engagement with and control of their agency has a major impact on translated texts: If a translator acknowledges the power and agency that they hold over their translation choices and practice, they can produce texts that are better suited to their intended audiences, purposes, and goals. However, the lack of training and attention paid to translation in the field of Language Documentation suggests that many translators working with documentation texts may be unaware of the agency they hold, and how their decisions can impact the translated texts they produce at political, ideological, and ethical levels (Beier 2015).

In this paper I draw on developments from Translation Studies to examine the current state of translation in Language Documentation at the macro-level to investigate:

1. The proportion of language documentation texts that are translated and glossed[4]
2. The genres that are translated
3. The languages used for producing translated texts ('Target Language')
4. The identities of translators of language documentation texts

§2 provides the background to the study, firstly exploring the needs of end-users of translated texts and secondly, considering the translator's agency to make choices during the translation process. §3 outlines the study design, while §4 presents an overview of the quantitative findings. §5 discusses the findings from the perspectives of useability of texts, translator agency, and Language Documentation's social justice goals. Finally, §6 presents suggestions for improving translation practices and developing this area of research.

[4] In the absence of previous theorisation or research in this area, it is important to note that this study is not interested in providing a definition of "translation of language documentation texts". Rather, I take a broad definition of "translation" that encompasses any case of rendering meaning from one language into another. This includes summaries, morpheme-level glossing, word-level glossing, faithful word-by-word translation, functional translation, and anything in between.

2. Background to the study

2.1 Potential end-users[5]

This section considers the potential end-users and their requirements for language documentation translated texts.

Many potential end-users of documentary texts, including members of the language community and their descendants, may not know or understand the documented language (Czaykowska-Higgins 2009; Schwartz & Dobrin 2016; O'Shannessy et al. 2019). Therefore, in order for texts to be accessible and useful, annotation – that is, transcription, glossing and translation – is required (Austin & Nathan 2004; Austin 2019: 6). In fact, language documentation material that is not accompanied by adequate annotation has been described as "as helpful as tablets in the Indus or other undeciphered script", in that users may recognise an example of linguistic use, but cannot gain further information from the text alone (Evans & Sasse 2007: 60). In other words, endangered language texts may be carefully recorded and archived, but without sufficient annotation they are unlikely to be accessible to potential end-users outside of the initial documentation project (Penfield & Tucker 2011: 292). This lack of accessibility has two major ethical implications: Firstly, data that cannot be understood cannot be verified, replicated, or reproduced (Berez-Kroeker et al. 2018). Secondly, the lack of accessibility, and therefore lack of usefulness, does not serve language communities and is therefore incompatible with the social justice aims of language documentation.

2.1.1 Requirements for translated texts

A survey of the literature reveals two key features that have been identified as important in facilitating or hindering the use of translated documentation texts beyond the initial documentation project: existence of appropriate texts and quality of annotation.

Language learning is a primary activity for those wanting to revitalise and increase the number of speakers and domains of language use (Taylor-Adams 2019: 428). For this purpose, texts are needed that demonstrate "a rich language environment" (Underriner et al. 2021: 237) and that provide examples of functional language such as, "how conversations are organised in the particular community, how to use language to get people to do things, what is appropriate to say or not say in what situation, how to agree, disagree, or argue with someone, and how to be a functioning speaker of the language" (Austin 2021: 202).

However, this kind of language is often lacking from documentary corpora (Sallabank & King 2021: 35), and where corpora do contain a range of genres, translation tends to be prioritised for texts that are relevant to the researcher's academic interests (Austin 2016: 161; Babinski et al. 2022). In other words, the texts that are selected for translation are unlikely to be those that are most appropriate for community use.

The quality of annotation accompanying documentation texts also affects how texts can be used. Therefore, even where translations exist, low-quality or inappropriate translation choices may render the texts inaccessible or inappropriate for use. Firstly, texts may be glossed and translated into languages that are not relevant to the local linguistic ecology, and therefore not accessible to the local language community (Austin 2016; Taylor-Adams 2019). Furthermore, even when they are provided in an accessible language, translations and glossing may be cryptic, unclear, incomplete, or inaccurate (Evans & Sasse 2007; Austin 2017). It has also been noted that where translated texts omit or change parts of the original source text, translator notes are very rarely included to explain how or why such changes have occurred (Beier 2015).

This brief overview of the literature demonstrates that translated texts are required to support language revitalisation activities such as language learning. To meet this requirement, texts should contain naturalistic and useful language, while translations should be of a high quality and translated into languages that are relevant to the local language community. Currently, these requirements are often reported to be unmet.

[5] Note that there are other potential end-users, including anthropologists, other linguists and academics from other disciplines. I do not discuss the needs of these groups in detail. However, I assume that they share many of the translation needs of linguists and community members.

2.2 Macro-level translation choices

This section identifies the macro-level choices that are made during the translation process, thereby highlighting several areas of translator agency that shape the translated text. Key macro-level translation choices include: choice of text to translate, style of translation (glossing and/or free translation), choice of target language, and choice of translator. These variables reflect research from the field of Translation Studies exploring the interrelated choices and conditions that affect the production of a translated text (e.g., House's (2016) 'complexity of translation').

2.2.1 Choice of text

The choice to translate (or not) allows texts to be accessed (or not) by a range of end-users, illustrating the asymmetric power of researchers and translators in amplifying the voices and knowledge contained within endangered or Indigenous language texts (Cronin 2017: 146). Language documenters, therefore, hold a gatekeeping role whereby certain texts and genres of texts that are selected for translation are more likely to be accessible and useable beyond the initial project; while texts that are not translated are far less accessible and useful to a range of end-users.

2.2.2 Choice of translation style

Style of translation is a complex theoretical area and a full exploration is beyond the scope of this paper. For the purposes of this study, translation style is identified as the choice between *glossing* and *free translation*. Here, glossing is used to describe a morpheme-by-morpheme or word-by-word rendering of the original text, which may also include additional grammatical annotation. Free translation is taken to describe any rendering from an original source text into a translated text that focusses on phrases that are larger than words, including line-by-line translations, summaries etc.

The choice of translation style is important in how the texts can be used, since the information that the translator provides determines how the reader of the translated text interacts with the source text. For example, where a free translation is given, an end-user of the translated text can memorise or learn phrases, but without some form of glossing they cannot understand the individual constituents of the phrases. This has a large impact on how suitable texts are for language learning and whether the texts can be used and verified by other researchers.

To illustrate, the examples below show extracts from ELAN files included in this survey. In both examples, the transcribed source text is printed on line 1. In Example 1, a word-by-word gloss is printed on line 2, showing a 1:1 correspondence to glossed terms and the source text terms. On line 3, a free translation is printed, which conveys the meaning in English and does not have a 1:1 correspondence with the source text. Example 2 demonstrates a different approach to glossing, with the inclusion of abbreviations expressing grammatical information. Importantly, in both examples, the glossing shows a 1:1 equivalence with the transcribed source text, while the free translation does not.

(1)	Source Text	alu sabzi raŋ hɔt ta bã mã nimi a:ts	
	Gloss	potato sabzi with chilte is very tasty happen/become	
	Free Translation	Chilta is very tasty with alu sabzi	(Martinez et al. 2020)

(2)	Source Text	Cei55 nai53 lo55 de31 jou55 nen55	
	Gloss	Ten two MW.pai one CL.string need	
	Free Translation	we need hair which is 12 pai-long	(Li et al. 2017)

2.2.3 Choice of target language

The choice of language of a translated text (Target Language) also determines how and by whom the text can be used. Texts are accessible to the local community, and beyond, when translations are rendered in a language that is part of the local linguistic context. Conversely, English is the most commonly used language of global academia (Bennett 2007; Bassnett & Johnston 2019) and therefore is likely to be the most useful language for academic purposes, as well as for introducing the text into broader contexts.

2.2.4 Choice of translator[6]

A translated text is also impacted by the identity of the translator. Research from Translation Studies has demonstrated how a translator's culture influences translation output (Tymoczko 2007). For example, whether the translator identifies with the culture of the original source text, the target language, or a third culture has an impact on their understanding of the cultural frames involved in translation. Therefore, the choice of translator is an important element of macro-level decision making in translating documentation texts, which is currently not well understood. To illustrate the complexities, an anecdote published on the personal blog of a documentary linguist (Collins 2019) outlines the translation process in the documentation of Sasi, a language of Botswana: the speakers of the endangered language are also fluent in the local dominant language, while the translator speaks English and the local dominant language, but not the endangered language. Therefore, the researcher plays a recording, the source language speakers then translate their own words into the local dominant language, which are clarified and written down by the translator. The translator then translates and writes the phrase in English. In this context, the role of translator is not clearly delimited. Rather, the cultures of each of the participants in the process have an effect on the translated texts. However, due to a general lack of acknowledgement and research within the field, the impact of each participant on the final translated text is not currently known.

3. Study design

3.1 Data

The study surveys a corpus of 1088 endangered language audio and/or audio-visual files plus accompanying annotation extracted from ten deposits from the Endangered Languages Archive (ELAR), which are accessible through the ELAR website.[7] ELAR deposits are collections of audio and audio-visual texts of spoken endangered languages, accompanied by bundles of related text files and/or images. The text files may contain metadata, annotations of the sociolinguistic conditions of the recordings and their participants, transcription, translation, and glossing.

The deposits surveyed were created between 2015–2021 (see Table 1) and therefore, provide a picture of current practices. Only deposits that have been promoted by ELAR as 'closed' were included, as this aimed to ensure that the depositors were not planning to amend their work. However, since ELAR collections are dynamic, it is still possible that the deposits will be updated or modified in the future.

The collections surveyed are anonymised (numbered 1–10) since the aim is not to comment on individual researchers or collectors, rather to provide a picture of current practices in general.

The ten deposits include texts recorded in ten languages spoken in countries representing areas with high levels of linguistic diversity (Table 2). They have been deposited by researchers affiliated with academic institutions in a range of

[6] Vasiliki Vita points out that the identity of the translator may not necessarily be a free choice, depending on constraints within a particular context. However, I use the notion of 'choice' to highlight the agency of the language documenter in making decisions regarding translation within the documentation project.

[7] https://www.elararchive.org.

countries (Table 3), skewed towards the USA, Europe, and Australia/New Zealand, which are areas where funding and expertise on Language Documentation are centred.

Table 1. Year of deposits

Year	Number of deposits
2015	1
2017	1
2018	3
2019	1
2020	3
2021	1
Total	10

Table 2. Distribution of languages

Country (a-z)	Number of deposits
China	1
Colombia	1
India	1
Indonesia	1
Nepal	1
Paraguay	1
Papua New Guinea	2
Vanuatu	2
Total	10

Table 3. Distributions of affiliated institutions

Country (a-z)	Number of deposits
Australia	1
France	1
Germany	2
New Zealand	1
UK	1
USA	4
Total	10

3.2 Variables

The 1088 texts are coded for six variables relating to macro-level translation decisions: genre of text, transcription presence, translation presence, glossing presence, language(s) of translated texts and translator acknowledgement.

These variables are based on a pilot test with an initial subset of the data and are constrained by the information that is available from the archive.

3.2.1 Genre of text

This variable tracks the genres represented in the corpus, and provides a necessary first step in order to determine which genres have been transcribed, translated, and/or glossed. Each text was assigned a label denoting its genre type, simplified based on the information provided by the depositor. Genres were simplified to account for the wide range of variation found in the metadata provided with each deposit: some deposits use more specific descriptions, while others are more general. Therefore, to allow for the data to be collated, the most specific descriptions were subsumed within the least specific descriptions. This process necessarily follows the metadata conventions used by the least specific deposits in the corpus sample.[8]

For example, the genres *conversation*, *dialogue* and *interview* were simplified to *interactive discourse*. *Consent* texts, while also examples of interactive discourse, were kept separate as there are no instances of translated consent texts in the corpus, suggesting that researchers treat these texts differently. Therefore, separating consent texts from other forms of interactive discourse avoids skewing the data. Furthermore, *folktale* and *mythology* are simplified to *traditional tales. Oral history* and *personal narrative* are included in the simplified category *narrative*. Procedural texts were kept separate since they typically display different discourse features to other monologues. Formulaic texts were simplified to a single category: *song, joke, riddle, prayer*, primarily due to the low frequency of these texts in the corpus. *Elicitation* was maintained as a separate category since this genre may be relevant to previous findings that documenters tend to prioritise translation of texts that are directly relevant to their own research (Austin 2016: 161; Babinski et al. 2022). Finally, *observational recording* was maintained as a separate category since this type of text is generally found to be untranslated.

The simplified genres are shown in Table 4, with the number of texts and number of deposits in which they appear. There is an uneven distribution between the numbers of texts included in each genre, which is reflective of the reality and content of the deposits.

Table 4. Simplified Genres

	Number of texts	**Number of deposits represented**
Narrative	467	10
Procedural	47	8
Elicitation	100	7
Consent	77	4
Song, Joke, Riddle, Prayer	37	7
Interactive Discourse	249	10
Observational Recording	39	4
Traditional tales	72	3
Total	**1088**	

[8] Further research with a larger sample would allow for a finer grained analysis of genres.

3.2.2 Presence of transcription, translation, and glossing

These variables were coded on binary scales (YES/NO).

- **Transcription Present YES/NO**

  This variable was coded as YES when the audio(-visual) text was accompanied by a transcription text file, whether using phonetic transcription or another orthography.

- **Free Translation Present YES/NO**

  This variable was coded as YES when the audio(-visual) text was accompanied by a translation text file.

- **Glossing Present YES/NO**

  This variable was coded as YES when the audio(-visual) text was accompanied by a glossing text file.[9]

3.2.3 Language(s) of translated texts

Texts that were coded as YES for the presence of translation, were further coded to indicate the language(s) of the translated texts. Some texts were coded with more than one language, where multiple translations were present.

3.2.4 Translator acknowledgement YES/NO

This variable was coded as YES when there was explicit reference to the role of translator, and a name was given.

3.3 Limitations to methodology

It is important to emphasise that this research is not intended to diminish the important work of researchers depositing material with endangered language archives, who are often performing a high volume of work – much of which is time-intensive and not recognised by academic evaluation metrics. It is clear that depositors are working in a wide range of contexts and there are many reasons why decisions regarding the translation process may be made. Therefore, I agree with Babinski et al. (2022) that "'incomplete but archived" is preferable to "not archived at all"'. Furthermore, this survey represents a first look at some of the translation decisions taken when archiving endangered language materials, but it is necessarily incomplete since the sample size of deposits is relatively small, and just one archive is surveyed. However, anecdotal evidence suggests that the findings are likely to be happening on a larger scale across language archives. In light of the irreplaceable nature and importance of much of the material, it is vital to begin to question and theorise the translation that is included in Language Documentation texts in the hope that the findings presented can provide a call to further study in this area. As such, the aim is to raise these issues for depositors of future documentation materials, so that the field can move forward in providing translated material in a more accessible and useful way for potential end-users.

[9] Note that I manually identified the presence of transcription, translation, and glossing as I collected the data by examining each text or ELAN file associated with the deposits.

4. Findings

4.1 Transcription, translation, and glossing

Of the 1088 endangered language source texts, 526 (48.3%) are transcribed, 528 (48.5%) are translated,[10] and 287 (26.3%) are glossed. The proportions of annotations vary dramatically between deposits (see Figure 1), ranging from the lowest level of annotation: no transcription, translation or glossing (Deposit 10) to the highest level of annotation: transcriptions and translations for all texts, and glossing for over half of texts (Deposit 1). Six deposits include more transcribed and translated texts than glossed texts. Just two deposits (Deposit 1 and Deposit 3) include transcription, translation and glossing for more than half of the texts.

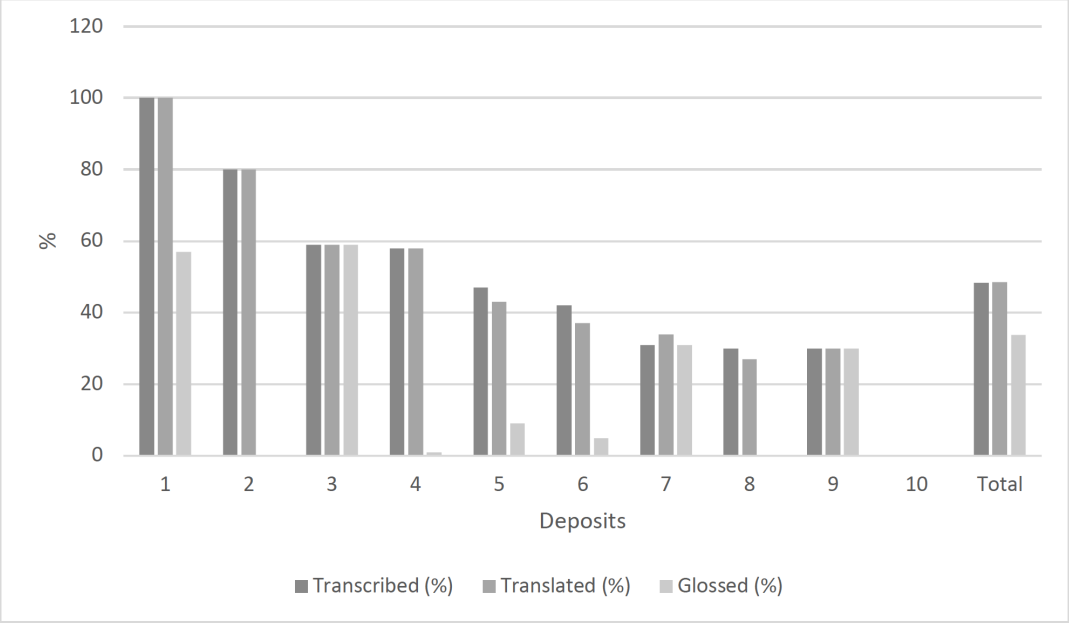

Figure 1. Percentage of Transcription, Translation and Glossing by deposit.

4.2 Genre

There are more texts accompanied by a single level of translation than texts that are both translated and glossed. Narratives are most frequently translated (63%) and glossed (38%), followed by procedural texts (55% translated and 32% glossed). Notably, less than half of interactive texts (48%) are translated, and only 10 percent are glossed. Less than half of traditional tales are translated (47%), with only around 20 percent glossed. Elicitation displays a higher proportion of glossing than other genres, with 46 percent of texts translated, and 36 percent glossed – the second highest rate of glossing. Songs, jokes, riddles and prayers are translated 40 percent of the time, and glossed in 32 percent of cases. The mainly non-linguistic observational texts are translated in 13 percent of cases and glossed in 8 percent of cases. Consent texts are never translated or glossed (See Figure 2).

[10] The disparity between transcription and translation can be accounted for by two texts that are not transcribed, but include a written summary in English. Therefore, the text is translated directly from audio source language to written target language.

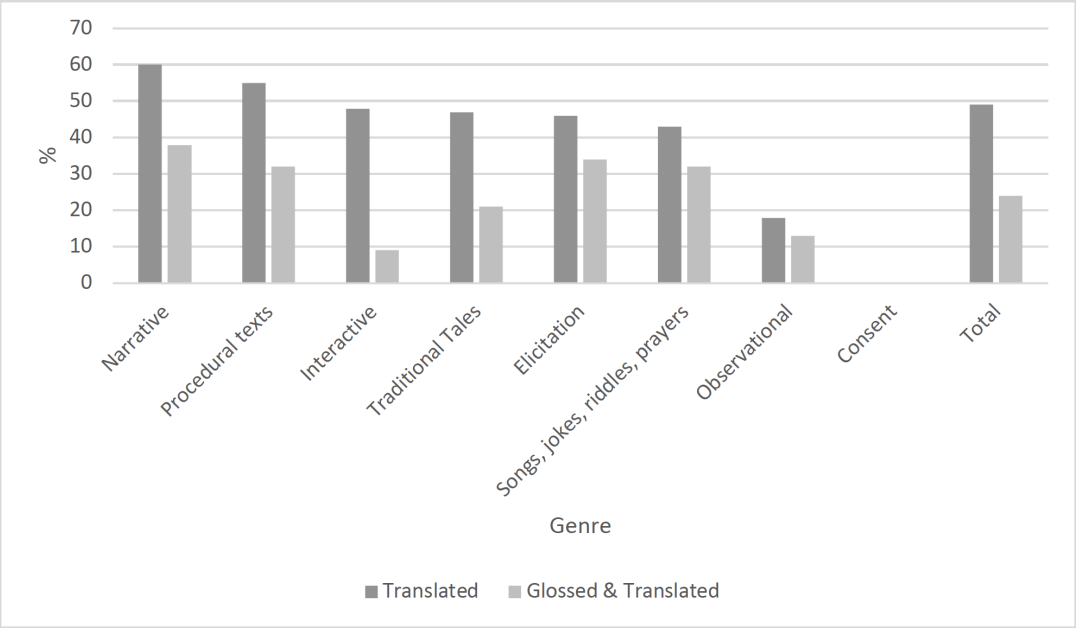

Figure 2. Translation and Glossing by genre.

4.3 Languages of translated texts

Out of 529 translated texts, 370 (70%) are provided in English, while 435 (82%) are provided in another locally relevant language (See Figure 3 for a visual representation and Table 7 for the figures for each deposit). The languages found in the corpus are: Bislama, Spanish, Tok Pisin, Mandarin, Hindi and Indonesian. Three deposits provide translations only in English, while one deposit provides translations only in a locally relevant language.

Table 5. Percentage of Translated Texts by language

Deposit	Total Translated Texts	English		Local Language	
		Number	%	Number	%
1	72	72	100	72	100
2	84	84	100	84	100
3	128	7	5.4	128	100
4	57	57	100	0	0
5	32	0	0	32	100
6	7	7	100	7	100
7	112	106	94.6	112	100
8	30	30	100	0	0
9	7	7	100	0	0
10	0	0	0	0	0
TOTAL	529	370	69.9%	435	82.2%

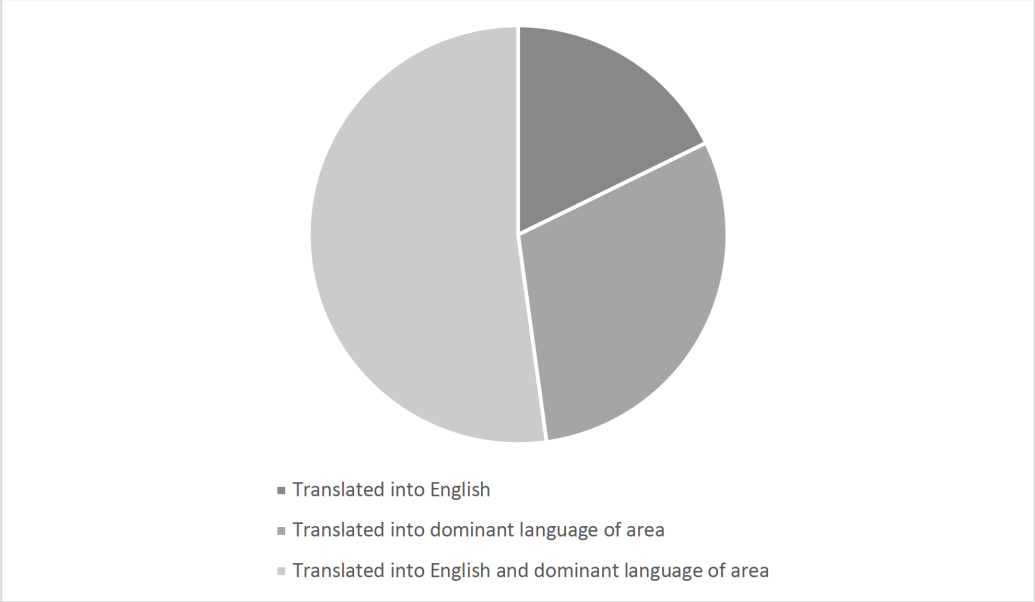

Figure 3: Languages of translated texts (%)

4.4 Translator acknowledgement

The survey reveals a widespread lack of translator acknowledgement. Explicit acknowledgement is given for only 146 (28%) translated texts. Strikingly, as shown in Table 8, these translator acknowledgements are provided in just three deposits, and within these three deposits, the translator is not consistently named (acknowledgement is found in 71%, 52% and 21% of texts, respectively). No information about translators or translation work was found in translated texts in the remaining seven deposits.

Table 6. Translator acknowledgement

Deposit	Number of translated texts with explicit translator acknowledgement	%
1	0	0
2	12	21%
3	0	0
4	55	52%
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	79	71%
TOTAL	146	27.6%

5. Discussion

In this section I discuss what the findings reveal about how useful language documentation texts are beyond the initial documentation project.

5.1 Choice of text

Less than half of the texts surveyed in this study are translated (49%), meaning that just over half of the texts are inaccessible to end-users who cannot understand the endangered source language. These untranslated texts are unlikely to be suitable for revitalisation purposes such as the development of pedagogical materials, since language teachers – who in many contexts are not L1 speakers of the endangered language (Czaykowska-Higgins 2009; Schwartz & Dobrin 2016; O'Shannessy et al. 2019) - are unable to search for and identify useful language (Underriner et al. 2021). In addition, the untranslated texts are unlikely to be suitable for use in wider academic contexts.

It also seems likely that texts that are accompanied by translation, but not glossing have limited use beyond the initial project. For example, locating and isolating corresponding lexical pairs of meaning from an endangered language source text and accompanying translated text is often impossible or inaccurate without the presence of glossing, which provides a link between the two texts. To illustrate, in Example 3, below, without the presence of the gloss in line 2, someone who doesn't understand the source language is able to access a general meaning of the sentence. However, this meaning is vague due to the retention of the source language term *nalza*, and the difficulty in determining which elements of the source text correspond with which elements of the translated text. This seems to support previous observations made in the literature that noted the existence of poor-quality translations (e.g. lacking glossing, translation notes, and translation of cultural terms) that may hinder the usefulness of texts.

(3)	Source Text	dangbo kalo ri ken yoli nalza	
	Gloss	long.ago era LOC food stir pot.small	
	Free Translation	those pots were used to make as nalza	(Martinez et al. 2020)

For revitalisation initiatives such as language learning, someone working with a text that is accompanied by free translation could memorise whole phrases, but they would not be able understand individual elements separately. When glossing is provided, users can determine the correspondences between the source text and the target text, meaning that they can gain a deeper understanding of the phrase, thus helping them to learn how to produce the language.

Similarly, glossing is a crucial element to enable wider academic use of documentary material, as it allows for linguistic analysis and ensures accountability. It also ensures that research findings, as well as translations, are 'transparent' (Woodbury 2003; 2011), verifiable, replicable, and/or reproducible. This is particularly important in the context of endangered language texts, since it is likely that the data cannot be collected again (Gawne et al. 2017: 158).

Only around a quarter (26%) of the texts surveyed provide both glossing and free translation, meaning that only this small proportion of texts could truly be described as 'multi-purpose' for revitalisation use such as developing language learning materials, or for academic and wider use beyond the initial documentation project.[11]

[11] Note that this research does not comment on the usability or quality of translations. See forthcoming work by this author for a detailed discussion °n micro-level translation strategies used in language documentation texts.. As pointed out by Vasiliki Vita, untranslated or unannotated texts can be used for Revitalization purposes if there are resources available to add annotation to the data at a later time. However, this process would add significant time to any project, and would only be possible where a dictionary, linguist and/or other resources are available.

5.2 Genres of translation

This research reveals that only 11 percent of interactive texts, the most useful, 'everyday' genre for language learning, are accompanied by both glossing and free translation. This means that the vast majority of interactive texts are not glossed, and therefore not 'multi-purpose' since they are unsuitable for language learning, or wider use without significant additional processing.

Narrative and elicitation are the genres that are most likely to be both glossed and translated (38% and 36%, respectively). This could be for several reasons: Firstly, priority may be given to processing texts for the research interests of the initial project, accounting for the relatively higher rate of glossing for elicitation. Secondly, it has been suggested that single-speaker texts with a narrow range of vocabulary are less challenging and time-consuming to translate and gloss, since there is less overlapping speech or unknown vocabulary items (Collins 2019). However, the focus on elicitation and narrative does not represent natural, communicative language use, and is therefore less useful to revitalisation or wider use.

5.3 Choice of language of translation

This research reveals that more than half (52%) of the translated texts are provided in both English and a local dominant language, meaning that these texts are potentially accessible to a wide range of end-users – from both within the language communities and more widely. This is a positive finding. However, when we consider that, of the 1088 texts in the corpus, only 286 (26%) provide sufficient annotation (glossing and free translation) to be described as 'multi-purpose' for a range of potential end-users and then that only 52% of those texts are translated into both English and a locally relevant language, we are left with 149 texts (14%) that can be said to be truly 'multi-purpose' in terms of annotation provision (glossing and free translation) and language(s) of translation.

5.4 Choice of translator

This research reveals that translator acknowledgement is given for just 28 percent of translated texts, meaning that the majority of translations surveyed make no reference to the translator.[12] Therefore, it is not possible to determine any generalisable information about the translators. Instead, this research supports previous observations in the literature regarding the lack of prestige given to the translation process within language documentation, that is evident from the lack of research on translation within the field (Beier 2015), thus minimising the labour of translation.

The reality of the labour involved in annotating documentary texts by adding transcription, glossing and translation has been estimated to take between 40 - 150 minutes for every minute of audio (Himmelmann 2018: 34; Collins 2019). Anecdotal evidence indicates that the process often involves multiple actors. Translations that lack acknowledgements are less accountable, verifiable, and reproducible since neither the translators' sociolinguistic backgrounds, nor their personal translation styles, can be taken into consideration, thereby concealing "the social and cultural positioning of the translator" (Baker 2000: 261).

Furthermore, much of the translation work associated with language documentation projects is carried out in collaboration with local community members. Rendering this labour invisible minimises the agency of the translator(s) and may have negative consequences in terms of fostering empowerment and revitalisation in the community. Lack of acknowledgement also raises questions over how much control speakers of the source language have regarding what, when, and how texts are translated into or out of their languages (Cronin 2003: 165–172).

[12] Some deposits include metadata at the collection-level about participants who have worked on translations for the collection, but do not provide acknowledgements with individual texts. These acknowledgements have not been included in this research.

6. Moving forward: Suggestions

The findings produced by this study prompt questions over how the field of language documentation can move forward to improve the quantity and quality of appropriate genres of translated texts for use beyond initial documentation projects. In addition, the lack of acknowledgement of the labour of translation within documentation projects raises concerns. I propose two areas for development:

1. Translator Agency
2. Translation Training

6.1 Building an awareness of translator agency

A first step towards improved translation in the field is to integrate the Translation Studies concept of 'Translator Agency' (e.g. Venuti 1998b; 1998a; 2018; Cronin 2006; Tymoczko 2007; Baker 2013). If we, as translators, or actors responsible for commissioning translation work, understand and acknowledges the power and agency that we hold over our translation choices and practice, we can produce texts that are better suited to their intended audiences and purposes.

I suggest increasing transparency over our translation practices should be the first step, since this can be achieved relatively easily without adding significant burden to the process of documentation. At the level of individual documentary projects, we should think carefully about the translation and glossing process, and document how/why decisions are made. It would be useful to provide 'translation notes' as part of the metadata associated with a text collection, and at the level of individual texts within the collection. Crucially, documenters should ensure that the translator is acknowledged in the metadata associated with each file.

Although translating and glossing increased quantities of text will increase the workload of a project significantly, it seems unavoidable that we need to ensure that more texts are glossed and translated in genres that have been identified as most useful for wider use. Specifically, documenters should aim to translate and gloss those texts that will be of most use to revitalisation initiatives: naturalistic and conversational text, as well as culturally important texts. Furthermore, documenters should aim to provide translations in language(s) that are relevant to the language community.

6.2 Providing translation training, guidance and resources

Translator Agency could be increased through translator training. Language Documenters do not routinely receive translation training in university programs, Field Methods courses or training provided by grant awarding organisations such as ELDP.[13]

More widely, beyond the field of Language Documentation, translator training initiatives rarely include minority or endangered languages due to financial unviability, low demand and lack of qualified teachers (Taibi & Ozolins 2016). The resulting implications are that translations involving minority or endangered languages have been reported to be "poor, inappropriate or less effective" in a wide range of contexts extending beyond Language Documentation (Taibi & Ozolins 2016: 22). Furthermore, the lack of investment in translation skills in this area also means that endangered language contexts are absent from translation theories; and thus the cycle perpetuates. It seems that there is potential for interdisciplinary collaboration between Translation scholars working to broaden the field of Translation Studies beyond its Eurocentric traditions, and Language Documentation researchers. There is productive current research and discussion in the field of Translation Studies on the topics of translation training in crisis and emergency contexts (e.g.

[13] A full discussion of the role of funding agencies on translation and annotation practices is beyond the scope of this paper. However, in order to improve the usefulness of translation within Language Documentation, we also need to see support from funding bodies, as well as revised expectations on output.

Cadwell & Bollig & Ried 2019) Community translation (Taibi 2017) and minority language translation (Cronin 2017; 2020; 2022) which could provide a model for inspiring and developing translation training within the field of Language Documentation.

Finally, as proposed by Beier (2015), "good basic training in relevant aspects of translation theory" should be provided to build up the skills required to produce high quality translations of the rich endangered language material that linguists document[14]. To meet this suggestion, I propose holding workshops or panels focussing on translation, perhaps integrated with other practices involved in annotating and archiving documentary material. I would also advocate integrating translation practices into field methods and documentation courses.

In the first instance, we need to raise awareness of translation within the field so that we can begin to construct our own theoretical underpinnings of the process and how it relates to the goals of language documentation.

7. Conclusions

This research reveals several trends and key observations on how macro-level translation decisions are realised in the translation of language documentation texts.

Bearing in mind the limitations, this research has demonstrated that decisions regarding which texts to leave untranslated, which texts to gloss and translate, which languages to translate into, and who should translate them have an important impact on how, and by whom, translated texts can be used. At the macro-level, there appears to be a lack of language documentation text translations that are suitable for use in community, academic, or wider world contexts, beyond the initial research project. Furthermore, the lack of acknowledgement of translation labour suggests that translation is seen to be a low-prestige activity, and that language documenters are broadly unaware of the agency that they possess as translators or facilitators of translation. Translated texts could be rendered more multi-purpose and useful to language communities as well as the wider group of end-users by increasing transparency over translation practices and acknowledging translation work. Furthermore, we should aim to provide a greater proportion of translated and glossed texts in those genres that are most suitable for end-users' purposes – for example, communicative everyday language for revitalisation, in appropriate target languages.

[14] A comprehensive textbook developed in the field of Translation Studies such as Mona Baker's *In Other Words* (Baker 2018) is a good starting point.

References

Austin, Peter K. 2016. Language documentation 20 years on. In Filipović, Luna & Martin Pütz (eds.), *IMPACT: Studies in Language, Culture and Society*, vol 42. 147–170. Amsterdam: John Benjamins Publishing Company. doi:10.1075/impact.42.07aus

Austin, Peter K. 2017. Language documentation and legacy text materials. *Asian and African Languages and Linguistics* 11. 23–44. (http://hdl.handle.net/10108/89205)

Austin, Peter K. 2019. Language documentation and meta-documentation. In Jones, Mari C. & Sarah Ogilvie (eds.), *Keeping languages alive: documentation, pedagogy and revitalization*, 1st paperback edn. 3–15. Cambridge New York: Cambridge University Press.

Austin, Peter K. 2021. Language Documentation and Language Revitalization. In Olko, Justyna & Julia Sallabank (eds.), *Revitalizing Endangered Languages*, 1st ed. 199–219. Cambridge University Press. doi:10.1017/9781108641142.014

Austin, Peter K. & David Nathan. 2004. Reconceiving metadata: Language documentation through thick and thin. *Language Documentation & Description* 2. 179–188. (http://www.elpublishing.org/PID/029)

Babinski, Sarah, Jeremiah Jewell, Kassandra Haakman, Juhyae Kim, Amelia Lake, Irene Yi, & Claire Bowern. 2022. How usable are digital collections for endangered languages? A review (How to do language archival deposits better). Presented at the *LSA 2022 Annual Meeting,* January. (https://www.youtube.com/watch?v=60Nf7JS8GT4) (Accessed 2025-01-23.)

Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. *Target. International Journal of Translation Studies* 12(2). 241–266. doi:10.1075/target.12.2.04bak

Baker, Mona. 2013. Translation as an alternative space for political action. *Social Movement Studies* 12(1). 23–47. doi:10.1080/14742837.2012.685624

Baker, Mona. 2018. *In other words: A coursebook on translation* 3rd edn. Abingdon, Oxon: Routledge. doi:10.4324/9781315619187

Bassnett, Susan & David Johnston. 2019. The outward turn in translation studies. *The Translator* 25(3). 181–188. doi:10.1080/13556509.2019.1701228

Bassnett, Susan & André Lefevre. 1990. *Translation, history & culture*. Printer Publishers.

Beier, Christine. 2015. *Text translation in the context of endangered language documentation: The case of Iquito*. Presented at *Fieldwork Forum, Department of Linguistics, University of California, Berkeley*, Berkeley, April. (http://www.cabeceras.org/beier_translation_fforum_20150423.pdf) (Accessed 2025-01-23.)

Bennett, Karen. 2007. Epistemicide!: The tale of a predatory discourse. *The Translator* 13(2). 151–169. doi:10.1080/13556509.2007.10799236

Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David I. Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meier, Nick Thieberger, Keren Rice, & Anthony Woodbury. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. Linguistics 56(1). 1–18. doi:10.1515/ling-2017-0032

Burke, Mary, Oksana L. Zavalina, Shobhana L. Chelliah, & Mark E. Phillips. 2022. User needs in language archives: Findings from interviews with language archive managers, depositors, and end-users. *Language Documentation & Conservation* 16. 1–24. (http://hdl.handle.net/10125/74669)

Burke, Mary, Oksana L. Zavalina, Mark Edward Phillips, & Shobhana Chelliah. 2020. Organization of knowledge and information in digital archives of language materials. *Journal of Library Metadata* 20(4). 185–217. doi:10.1080/19386389.2020.1908651

Cadwell, Patrick, Claudia Bollig, & Juliane Ried. 2019. Management and training of linguistic volunteers. In Federici, Federico M., & Sharon O'Brien (eds.), *Translation in cascading crises*, 1st edn. 152–173. Routledge. doi:10.4324/9780429341052-8

Chesterman, Andrew. 1997. *Memes of Translation: The spread of ideas in translation theory*, vol. 22. Amsterdam: John Benjamins Publishing Company. doi:10.1075/btl.22

Collins, Chris. 2019. Rate of transcription in syntactic fieldwork. *Ordinary Working Grammarian,* November 16. (http://ordinaryworkinggrammarian.blogspot.com/2019/11/rate-of-transcription-in-syntactic.html#more) (Accessed 2025-01-23.)

Cronin, Michael. 2003. *Translation and globalization.* Transferred to digital print. London: Routledge.

Cronin, Michael. 2006. *Translation and identity*. London: Routledge.

Cronin, Michael. 2017. *Eco-translation: Translation and ecology in the Age of the Anthropocene*. London: Routledge.

Cronin, Michael. 2020. Minority. In Baker, Mona & Gabriela Saldanha (eds.), *Routledge encyclopedia of translation studies*, 334–338. Routledge.

Cronin, Michael. 2022. From translation zone to sacrifice zone: Minor perspectives on the tradosphere. *Translation in Society* 1(1). 105–124. doi:10.1075/tris.21014.cro

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3(1). 15–50. (http://hdl.handle.net/10125/4423)

Evans, Nicholas & Hans-Jürgen Sasse. 2007. Searching for meaning in the Library of Babel: Field semantics and problems of digital archiving. *Language Documentation & Description* 4. 58–99.

Gawne, Lauren, Barbara F. Kelly, Andrea Berez-Kroeker, & Tyler Heston. 2017. Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation* 11. 157–189. (http://hdl.handle.net/10125/24731)

Himmelmann, Nikolaus P. 2018. Meeting the transcription challenge. In McDonnell, Bradley, Andrea L. Berez-Kroeker, & Gary Holton (eds.), *Reflections on language documentation 20 years after Himmelmann 1998.*, 33–40. Honolulu, HI: University of Hawai'i Press. (http://hdl.handle.net/10125/24806)

Hinton, Leanne. 2011. Language revitalization and language pedagogy: New teaching and learning strategies. *Language and Education* 25(4). 307–318. doi:10.1080/09500782.2011.577220

House, Juliane. 2016. *Translation as communication across languages and cultures*. London: Routledge

Inghilleri, Moira. 2008. The ethical task of the translator in the geo-political arena: From Iraq to Guantánamo Bay. *Translation Studies* 1(2). 212–223. doi:10.1080/14781700802113556

Katan, David & Mustapha Taibi. 2021. *Translating cultures: An introduction for translators, interpreters and mediators*, 3rd edn. Abingdon, Oxon: Routledge.

Kung, Susan Smythe, Ryan Sullivant, Elena Pojman, & Alicia Niwagaba. 2020. Archiving for the future: Simple steps for archiving language documentation collections [OER]. (https://archivingforthefuture.teachable.com/) (Accessed 2025-01-23.)

Li, Y., Li, L. & Li, L-M. (2017). Red wine: Cooking01. In Documentation of Zauzou, an endangered language in China. Endangered Languages Archive. https://www.elararchive.org/uncategorized/SO_0adf3dfc-7cfe-4e05-8544-f457d14ebc14/

Marais, Kobus & Carmen Delgado Luchner. 2018. Motivating the translation-development nexus: Exploring cases from the African Continent. *The Translator* 24(4). 380–394. doi:10.1080/13556509.2019.1594573

Martinez, P. A., Devi, G., Negi, N-P. & Singh, A. (2020). A local dish called hot: TRD_cik04GD20181126. Documentary Corpus of Chhitkul-Rakchham, an endangered Tibeto-Burman language of Northern India – Himachal Pradesh, Kinnaur District. Endangered Languages Archive. https://www.elararchive.org/uncategorized/so_c7c4271c-85bf-4fc7-98ca-33b67d2d0b5d/

O'Shannessy, Carmel, Samantha Disbray, Barbara Martin, Gretel Macdonald, Linda Barwick, Jennifer Green, & Petronella Vaarzon-Morel. 2019. (Re)turning research into pedagogical practice: A case study of translational language research in Warlpiri. In *Archival returns: Central Australia and beyond*, 139–151. Honolulu: University of Hawai'i Press. (http://hdl.handle.net/10125/24881/)

Pedersen, Jan. 2011. *Subtitling norms for television: An exploration focussing on extralinguistic cultural references,* vol. 98. Amsterdam: John Benjamins Publishing Company. doi:10.1075/btl.98

Penfield, Susan D. & Benjamin V. Tucker. 2011. From documenting to revitalizing an endangered language: Where do applied linguists fit? *Language and Education* 25(4). 291–305. doi:10.1080/09500782.2011.577219

Pym, Anthony. 2012. *On translator ethics: Principles for mediation between cultures*. Amsterdam: John Benjamins Publishing Company.

Rice, Sally. 2011. Applied field linguistics: Delivering linguistic training to speakers of endangered languages. *Language and Education* 25(4). 319–338. doi:10.1080/09500782.2011.577216

Roche, Gerald. 2022. The necropolitics of language oppression. *Annual Review of Anthropology* 51(1). 31–47. doi:10.1146/annurev-anthro-041420-102158

Sallabank, Julia & Jeanette King. 2021. What do we revitalise? In Olko, Justyna & Julia Sallabank (eds.), *Revitalizing endangered languages*, 1st edn., 33–48. Cambridge University Press. doi:10.1017/9781108641142.003

Schultze-Berndt, Eva. 2006. Linguistic annotation. In Gippert, Jost, Nikolaus P. Himmelmann, & Ulrike Mosel (eds.), *Essentials of language documentation*, 213–251. Berlin: De Gruyter.

Schwartz, Saul & Lise M. Dobrin. 2016. The cultures of Native North American language documentation and revitalization. *Reviews in Anthropology* 45(2). 88–123. doi:10.1080/00938157.2016.1179522

Snell-Hornby, Mary. 1988. *Translation Studies: An integrated approach*. Amsterdam: John Benjamins Publishing Company. doi:10.1075/z.38

Sullivant, Ryan. 2020. Archival description for language documentation collections. *Language Documentation & Conservation* 14. 520–578. (http://hdl.handle.net/10125/24949)

Taibi, Mustapha (ed.). 2017. *Translating for the community*. Bristol, Blue Ridge Summit: Multilingual Matters. doi:10.21832/9781783099146

Taibi, Mustapha & Uldis Ozolins. 2016. *Community translation*. London: Bloomsbury Academic, an imprint of Bloomsbury Publishing Plc.

Taylor-Adams, Allison. 2019. Recording to revitalize: Language teachers and documentation design. *Language Documentation & Conservation* 13. 426–445. (http://hdl.handle.net/10125/24873)

Thieberger, Nick. 2016. What remains to be done—Exposing invisible collections in the other 7,000 languages and why it is a DH enterprise. *Digital Scholarship in the Humanities* fqw006. doi:10.1093/llc/fqw006

Thieberger, Nick & Amanda Harris. 2022. When your data is my grandparents singing. Digitisation and access for cultural records, the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). *Data Science Journal* 21. 9. doi:10.5334/dsj-2022-009

Tymoczko, Maria. 2007. *Enlarging translation, empowering translators*. Manchester, UK: St. Jerome Pub.

Underriner, Janne, Lindsay Marean, Pigga Keskitalo, Zalmai Zahir, Pyuwa Bommelyn, & Ruby Tuttle. 2021. Teaching strategies for language revitalization and maintenance. In Olko, Justyna & Julia Sallabank (eds.), *Revitalizing endangered languages*, 1st edn., 235–272. Cambridge University Press. doi:10.1017/9781108641142.016

Venuti, Lawrence. 1998a. Introduction. *The Translator* 4(2). 135–144. doi:10.1080/13556509.1998.10799016

Venuti, Lawrence. 1998b. *The scandals of translation: Towards an ethics of difference*. London ; New York: Routledge.

Venuti, Lawrence. 2018. *The translator's invisibility: A history of translation*, 3rd edn. New York: Routledge.

Woodbury, Anthony C. 2003. Defining documentary linguistics. *Language Documentation & Description* 1. 35–51.

Woodbury, Anthony C. 2007. On thick translation in linguistic documentation. *Language Documentation & Description* 4. 120–135.

Woodbury, Anthony C. 2011. Language documentation. In Austin, Peter K. & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 159–186. Cambridge: CUP.

Yi, Irene, Amelia Lake, Juhyae Kim, Kassandra Haakman, Jeremiah Jewell, Sarah Babinski, & Claire Bowern. 2022. *Accessibility, discoverability, and functionality*. Presented at the LSA 2022 Annual Meeting, LSA 2022, January. (https://www.youtube.com/watch?v=SN9Y6uRZ5Rw) (Accessed 2025-01-23.)

Holly Drayton
h.drayton2@newcastle.ac.uk
orcid.org/0000-0001-8392-0731