



Glossing in the Linguistic Survey of India: some insights into early 20th century glossing practices

Aimée Lahaussais

► To cite this version:

Aimée Lahaussais. Glossing in the Linguistic Survey of India: some insights into early 20th century glossing practices. *Historiographia linguistica*, 2021, 48 (1), pp.25-59. halshs-03206953

HAL Id: halshs-03206953

<https://shs.hal.science/halshs-03206953>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Glossing in the *Linguistic Survey of India*: some insights into early 20th century glossing practices¹

Aimée Lahaussais

C.N.R.S., Université de Paris

1. Introduction

Glossing is a widespread practice in cultures with a written tradition and is found over a wide range of historical periods and linguistic traditions. Good descriptions of practices in Medieval Europe (Cinato 2015; Teeuwen & van Renswoude 2017, among many others), the Far East (Whitman et al. 2010; Whitman 2014; Kosukegawa 2014; Alberizzi 2014), modern-day descriptive linguistics (Lehmann 1982; Lehmann 2004) exist, but no conclusive general historiography of the practice has been produced as of yet (see, however, Cinato, Lahaussais and Whitman, forthcoming, which is a first attempt at looking at glossing from a comparative point of view).²

There are a number of reasons that the history of glossing has not yet been written: while cross-cultural and ubiquitous (Whitman 2011), practices in different traditions (both areal and temporal) are nonetheless quite distinct. It appears that the similarities across different traditions, which are at least partially captured by typologies of glosses, such as Wieland (1983) and Cinato (2015), result from a human need to explicate texts rather than from a linear transmission of practices. Despite solid preliminary work, until we have good descriptions of practices for a wide range of time periods and geographical areas, we will not have access to the whole picture and be able to reconstruct which elements have arisen spontaneously and which aspects may be inherited or borrowed from a practice used in another time or place.

The study presented herein attempts to provide a piece of the puzzle, in this instance glossing practices of languages of the Indian subcontinent at the very beginning of the 20th century. The focus is on one specific work, the great descriptive enterprise known as the *Linguistic Survey of India* (henceforth LSI), published over

¹ I am very grateful to *HL* editors and two reviewers (Sebastian Nordhoff and an anonymous reviewer) for helpful feedback on this article. I also thank Matthew Zisk (Tohoku University) for inviting me to present a preliminary version of this work at the *First International Conference on Linguistic Terminology, Glossing and Phonemicization* in Yamagata, Japan, in February 2020.

² The creation of the Network for the Study of Glossing (<http://www.glossing.org>) is a promising initiative in this direction.

the years 1894-1928, under the stewardship of Grierson (1851–1941), and the practices found therein are placed in the context of the period preceding the LSI and of the period following.

Glossing as currently practiced within the subfields of descriptive linguistics and typology is considerably standardized, with most linguistics journals in these subfields recommending that authors present their data according to the Leipzig Glossing Rules³ which evolved out of work by Lehmann (1982).⁴ Establishing a timeline for the stages that lead from earlier glossing to interlinear glossing practices as currently practiced is challenging, for some of the reasons outlined below. I make the claim that the LSI does not suffer from the biases discussed below (as will be shown in Section 2.2) and constitutes a record of a significant step in the evolution of glossing towards current practices.

Firstly, within a grammatical description the data presented as examples and as text specimens may look quite different from the point of view of their annotation. One example of this is the linear alignment of examples found in some grammars, while in the same grammars, longer texts are glossed interlinearly. This suggests that the analysis of glossing practices of a given work or period must specify what it takes as its object of study in order to be able to maintain consistency.

Secondly, the intended audience for the glossed material can affect the output. This is seen for example in the use of exclusively lexical glosses as opposed to a combination of lexical and grammatical glosses. Lehmann, who identifies Finck (1909) as one of the first publications to make use of interlinear morphemic glossing and translation, points out that the latter's use of exclusively lexical glosses in Turkish is "a tribute to the non-specialist readership that the booklet aims at, but necessarily *falsifies the working* [italics mine] of the language by attributing lexical meanings to its grammatical morphemes" (Lehmann 2004: 1837). This suggests that we must consider materials produced for the same (or similar) audiences if we are to achieve meaningful results.

Thirdly, analyzing glossing practices, and indeed data annotation in general, relies greatly on the physical presentation of the data: issues of transcription, of glossing, of translation, and of the relationships (including physical) between these types of analysis. The results of such a study are thus affected by typesetting decisions by publishers, which may not reflect the practices of the linguist producing the material. This raises the question of the respective influence of data compilers, typesetters, journal editors on the presentation of data and on our interpretation of glossing practices and their evolution.⁵ The following notice, inserted between the Table of contents and page 1 of the 1857 issue of the *Journal of the Asiatic Society of Bengal*,

³ <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

⁴ Note that Lehmann continues to refine his interlinear morphemic glossing scheme, the latest version of which is available on his website:
https://www.christianlehmann.eu/ling/ling_meth/ling_description/representations/gloss/index.php

⁵ The recent ability, from approximately the 1990's on, to submit digital versions of manuscripts has of course limited the introduction of these types of errors.

where B.H. Hodgson published the first grammatical sketches of the Kiranti languages Vayu and Bahing, indicates that publishers were well aware of these difficulties.

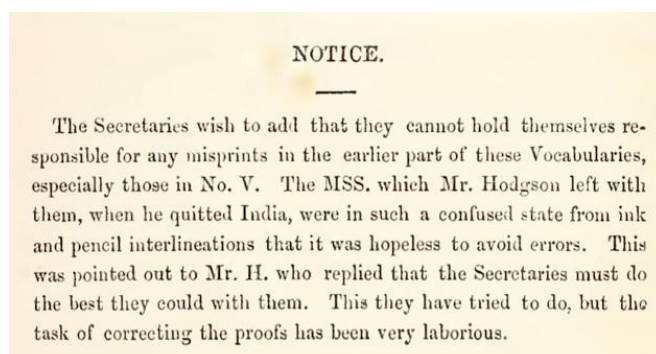


Fig. 1. Notice in the 1857 JASB

The notice in Figure 1 reveals the challenges of converting hand-written manuscripts into printed publications and capturing the at times idiosyncratic nature of annotations accompanying the material, suggesting the care with which we must handle such printed materials in using them to study the evolution of glossing practices. The three points listed above concerning the difficulty of establishing with any definiteness trends in glossing practices over time suggest that one approach is to find a source of materials which limits these variables as much as possible. As we shall see below, the LSI is one such source, and therefore serves as a precious resource in establishing a benchmark for glossing practices at a given place and time.

The structure of this article is as follows: In Section 2, I will first present background information on the LSI and discuss why it is a valuable resource for establishing the state of glossing practices at the beginning of the 20th century. In Section 3, I will lay out some aspects of glossing which are relevant to this study and interesting to explore within a historical framework. In Section 4, I will present general glossing practices as found in the text specimens of the LSI. In Section 5, I place the LSI practices within the context of practices in works from the period preceding and following the LSI, on the basis of materials from Himalayan languages (which are also contained in the LSI). Section 6 will offer a conclusion on the role of the LSI in shaping glossing practices for its time.

2. *The LSI: background and advantages as corpus*

This section presents information about the LSI, such as background information on the resulting publication, advantages as a corpus of study for glossing practices, and information, from the LSI's introduction, that sheds lights on the methodology of the compilation work, specifically as concerns collection and editing of text specimens.

2.1 Background

The *Linguistic Survey of India* is the publication resulting from a vast linguistic enterprise: It is made up of 21 volumes⁶ and covers 723 linguistic varieties spoken in British India, with lexical and grammatical information for 268 varieties (Majeed 2019a), and was begun in 1894, with the final volume published in 1928.

For most languages, the entry is made up of a short grammatical sketch (with an introduction to the people, census data, and a list of authorities and provenance of data) and language specimen(s). Comparative word lists, grouping all languages within a given section, are also provided.⁷

The context in which Grierson presided over the LSI project is presented in great detail in two recent books by Majeed (2019a; 2019b). The LSI generated a good deal of research on various languages of India. Ohannessian and Ansre (1975: 66) list the following as influences of the LSI:

"[I]t is still claimed as an authoritative source by political groups; its results still provide material for university courses; its influence can be seen in the flourishing of linguistics in several distinguished centers in the country; some of the basic techniques it used to collect information are still productive, and a host of other studies both by Indian and other scholars have followed the LSI's lead."

Scholarly reviews of individual volumes of the LSI were published as they appeared in print,⁸ but I have found a single review of the overall Survey. It offers a very flattering report, emphasizing the immensity of the accomplishment and its exceptional quality (See Unsigned 1928). As far as current-day documentary linguists are concerned, the LSI is cited in many linguistic descriptions of the South-Asian linguistic area to this day.⁹ Furthermore the potential of the LSI as a resource has recently been tapped with the conversion of the entire work into a large, searchable database on the languages of South Asia (see Borin, Virk & Saxena 2018).¹⁰

2.2 Advantages as a corpus

The LSI presents a number of advantages as a corpus for the study of glossing practices at the cusp of the 20th century, as it considerably limits the variables that often affect and obscure a longitudinal study into this topic. These advantages are outlined below.

a) It was produced over a limited time period (1894–1921), and is thus temporally bounded, allowing us to delve into practices for a given time frame. This happens to

⁶ In this number I count as volumes Vol 1, Suppl. 2 'Addenda and corrigenda minora' and an Index of language names.

⁷ These lists are based on a list which Campbell had collected by colonial officers (Campbell 1874), but the material in the LSI has been collected anew.

⁸ All of the reviews I have been able to track down are by T. Grahame Bailey, an apparent devotee of Grierson's work. One review ends with the following praise: "English scholarship need not fear comparison so long as it can produce books like the one before us or scholars like Sir George Grierson." (Bailey 1921a: 471); another, with: "I have studied all Sir George Grierson's volumes in this series, and I never rise from the study of one of them without a feeling of astonishment at the mastery of detail, the power of extracting living facts out of a bewildering mass of often discordant minutiae, the sureness of touch and quickness of vision." (Bailey 1921b: 475)

⁹ A rapid survey of 21 descriptive grammars of the Himalayas written over the last 30 years shows that 15 of them cite Grierson's LSI in their bibliography.

¹⁰ The searchable database is available at <https://spraakbanken.gu.se/blogg/index.php/2020/09/01/griersons-linguistic-survey-of-india-as-open-access-digital-data-resource-for-studying-languages-of-south-asia/>

be a period of great change in the greater world of glossing practices, with, according to Lehmann (2004) the first interlinear morphemic glossing, as well as what appear to be the beginnings of morphosyntactic glossing of person indexation on verbs (see Lahaussais forthcoming).

b) The languages described in the LSI are all Indospheric,¹¹ in other words affected by being spoken in an area of contact with Indo-Aryan languages, and share characteristics of their language area, even across the boundaries of language families (Matisoff 1990; Bradley et al. 2003). One of these characteristics is predominantly verb-final word order (Masica 1991: 333), relevant when it comes to identifying the word order found within the glosses.

c) The language descriptions follow the same template: grammatical sketch, text specimens (the first of which is always the Parable of the Prodigal Son), word lists. Even though different languages have descriptions of considerably different lengths, the comparability afforded by the inclusion of the same text specimen is invaluable for the study of glossing practices.¹²

d) The LSI was overseen by a single editorial team over the course of the entire publication, guaranteeing that the materials underwent the same editorial process and were prepared for the same intended audience.

For these reasons, I believe that the use of the LSI as a corpus considerably limits the variables that can be impediments to determining the state of glossing practices at a given time period and place.

2.3 Grierson on methodology

An additional advantage of the LSI is that Grierson provides documentation, in his introduction, about the methodology of text selection and annotation. We learn that it is for its linguistic features that the Parable of the Prodigal Son is chosen to accompany each linguistic description in the LSI:

It was then determined that the first specimen should be a version of the Parable of the Prodigal Son, with slight verbal alteration to avoid Indian prejudices, a passage which has been previously used and is admirably suited for such purposes [fn. 'it contains the three personal pronouns, most of the cases found in the declension of nouns, and the present, past, and future tenses of the verb.']

Acknowledging the unnaturalness of using a translation-based text for linguistic analysis,¹³ Grierson explains the decision to include other specimens in addition to the Parable of the Prodigal Son. These are textual ('a second specimen was also to be called for in each case, not a translation, but a piece of folklore or some other passage

¹¹ The terms 'Indosphere' and its counterpart 'Sinosphere' were coined by Matisoff (1990: 113) to describe the two large spheres of influence that Asian languages (in the article in question, South East Asian languages) are subjected to.

¹² Furthermore, the fact that the same specimen is presented for all languages results in a very large parallel corpus of the Parable of the Prodigal Son, which would have been useful to Grierson when glossing the Parable, which was chosen specifically for its linguistic features (see Section 2.3), especially in languages that were not as well documented. (An interesting contemporary equivalent can be found in a situational questionnaire devised and used by Alexandre François for languages of Vanuatu: the parallel nature of his corpus greatly helped his analysis (François 2019).)

¹³ This is also a topic of concern in contemporary linguistics (see, among others, Cysouw & Wälchli 2007).

in narrative prose or verse, selected on the spot and taken down from the mouth of the speaker" (Grierson 1927: 17)) and lexical ("Subsequently a third specimen was added to the scheme—a standard word list and test sentences originally drawn up for the Bengal Asiatic Society in 1866 by Sir George Campbell and already widely used in India." (ibid, 17)).

We also learn about the methodology for the specimen collection, for which "careful instructions were given" (ibid, 19), including the provision of a "word for word interlinear translation" (ibid, 20), in addition to a clear mention of Grierson's having edited the textual materials: "I was entrusted with the task of collecting the specimens and of editing them for the press." (ibid, 18).

3. *Aspects of glossing*

This section lays the groundwork for some basic concepts in glossing that are relevant to the study presented herein.

One essential breakdown in glossing type is into number and word glosses. Examples of number glosses are the system used by Humboldt for Classical Nahuatl (Humboldt 1968 [1836], one example is on p. 165), as well as the inversion glosses of Japanese *kunten* glossing (Whitman et al. 2010; Kosukegawa 2014).¹⁴ In (1), taken from the just mentioned passage of Humboldt, both 'nacatl' and 'Fleisch' are glossed with the number '5', identifying 'Fleisch' as the translation for 'nacatl'.

(1) 1 2 3 4 5 1 3 2 4 5
ni-c-qua in nacatl, ich esse es, das Fleisch.

Number glosses are used as an aid to restore the word order of the glossed source into that of the metalanguage. Examples are found among the materials in the period leading up to the LSI (See Figure 11).

The majority of glosses are of the word type, using words as glosses. One consequence of this is that practical challenges may arise in cases where a lexeme (either the lexeme being glossed or used as a gloss) is made up of more than a single word, creating a discrepancy in the 'word-for-word' principle. This is seen in (2), where multiple glosses (including the grammatical gloss 'ACC', in addition to lexical glosses) are associated with each of the transcribed words. This is problematic insofar as it does not allow us to access the internal make-up of the transcribed lexemes.

(2) lego librum
 I-read a-book-ACC
 'I am reading a book.' (Law 2003: 166)

The reverse situation is also found: a lexeme of the transcribed language¹⁵ may be written as multiple words, yet glossed with a single word in the metalanguage.

¹⁴ The latter are numerical glosses which make it possible to reorder the elements of a Chinese text for a 'vernacular' (i.e. Japanese) reading, accounting for the different word order of the two languages.

¹⁵ The term 'transcribed language' serves here to refer to the language being described; other terms for this are 'object language' (as per Lehmann, see website), 'source language' (as used in translation studies), both of which seem to have a telic sense which does not seem appropriate here. I feel that 'transcribed language' refers rather

As we are focused here on glossing practices, we shall set aside for now practices affecting the presentation of the transcribed language (i.e. transcription practices) and concentrate on features of glosses. Glosses can be *single-word*, when there is a one-to-one correspondence between the lexeme and gloss, or they can be *multi-word*. The latter involves the situation where more than one lexical word in the metalanguage is necessary to translate a transcribed lexeme. When glosses are multi-word, it is necessary to signal that they correspond to a single transcribed word, and this is often done by using hyphens (as is seen in (2)). Note however that the hyphen is also often used, in the same materials, as a segmentation mark (both within transcriptions and within glosses), to mark boundaries between sub-word elements within the same word. This results in the hyphen having two simultaneous functions—segmentation mark and aggregation mark—leading to potential opacity.¹⁶

The reason this question is important is for the determination of the unit of analysis used by the linguist/glossator: word or sub-word. If the focus is on the word, then analysis is presumably reflective of the word order within the sentence, in other words largely syntactic; if the focus is on the sub-word, then analysis is at a level which can lead to a morphemic analysis, which eventually became the basis for glossing as currently practiced. The presence of multi-word glosses, mismatches between transcription and gloss segments, and the order of segments within glosses (which can respect the word order of the transcribed language or of the metalanguage) thus make up important data to help identify the linguist's analytical approach and whether it can be seen as a first step towards morphemic glossing.

Another relevant dichotomy present in glosses is the distinction between word-for-word glosses and literal translation, even though in many instances (for example, in the work of translation theorists; see eg. Barbe 1996) these two are considered equivalent, and stand in opposition to free translation.¹⁷ Nonetheless, for the purposes of analyzing glosses within annotation schemes, word-for-word glosses must be distinguished from literal translation. There are some diagnostics that can be used to characterize them typographically, in addition to their different functions: Literal translations tend to make use of punctuation (such as periods, commas, quotation marks, hyphens) and sentence-initial capitalization; they also usually feature less segmentation than word-for-word glosses (insofar as segmented words are not all that frequent in writing). They are produced for readability, with an emphasis on conveying the narrative content of the text. Word-for-word glosses, on the other hand,

transparently to the language presented in the transcription line, without any judgement value about what that language actually is and how its transcription may have transformed it.

¹⁶ The Leipzig Glossing Rules provide a mechanism for resolving the discrepancy in number of words within a lexeme ("Rule 4: one-to-many correspondences"), recommending the use of a period to separate the metalanguage elements that correspond to a single transcribed word. As an alternative, an underscore is also proposed.

¹⁷ Note however that this is not always the case: Newmark (1988: 45–46) sets up the opposition between the two in the following way: Word-for-word translation "[...] is often demonstrated as interlinear translation, with the T[arget] L[anguage] immediately below the S[ource] L[anguage] words. The SL word-order is preserved and the words translated singly by their most common meanings, out of context. Cultural words are translated literally. The main use of word-for-word translation is either to understand the mechanics of the source language or to construe a difficult text as a pre-translation process." On the other hand, in literal translation "[t]he SL grammatical constructions are converted to their nearest TL equivalents but the lexical words are again translated singly, out of context.

generally do not contain punctuation or sentence-initial capitalization, operating at the word- rather than sentence-level, and often make use of segmentation marks in order to reveal sub-word level analysis. They are produced with the intent of conveying information about the morphosyntax of the data, and to reflect linguistic analysis. Furthermore, it is important to note that 'word-for-word' here is used to reflect the fact that the unit for dividing up the transcription is the word (with spaces as boundary markers), and that the glosses are aligned under each word. The term 'word-for-word' can, with respect to glossing, subsume both glosses that access only the word-level and those that access the sub-word-level, as it refers to the positioning of glosses as opposed to their analytical reach.

While the difference between word-for-word glossing and literal translation 'glossing' is gradient rather than binary, it is useful to have the two types in mind when analyzing glossing practices in the late 19th and early 20th centuries. This is in part because morphemic glossing emerges from word-for-word glossing, and it is therefore important to document when word-for-word glossing becomes the norm, on the path to morphemic glossing.

Literal translation as a term is therefore used for both translation type and glossing type, referring to a distinct level of textual annotation. While contrasts between literal and free translation have been explored in some detail (see Lahaussais 2016), literal translation has rarely been considered as a type of glossing, even though the material herein suggests that it should be.

A final type of glossing, morphemic glossing, is currently widely practiced in documentary linguistics, and is often referred to as interlinear morphemic glossing (Lehmann 1982; Lehmann 2004). It is at the heart of the Leipzig Glossing Rules. Morphemic glosses are characterized by a matching number of segments, ordered according to the transcribed language (in order to reflect its morphosyntax), and featuring both lexical and grammatical glosses,¹⁸ with the aim to gloss the morphemes within a transcription. The corpus explored herein does not feature any morphemic glosses, but it does contain material which at times approaches them, when transcribed words are segmented and glossed and ordered by segment, allowing us a glimpse into the potential forebearers of morphemic glossing.

4. *Glossing practices in the LSI text specimens*

In this section, I look at the characteristics of glossing in Parable of the Prodigal Son specimens in the LSI,¹⁹ through a selection of five arbitrarily chosen language versions. The languages are Kashmiri, Telugu, Lepcha, Kachari, and Singpho, covering the Indo-Aryan, Dravidian and Tibeto-Burman families.

The text extracts provided in the figures are intended to give a general sense of the presentation of glossing in the specimens, and some examples that are discussed are

¹⁸ The emergence of grammatical glosses for person indexation on verbs constitutes a separate, and interesting, topic, explored for an Algonquian/Kiranti grammar corpus in Lahaussais (forthcoming).

¹⁹ Majeed (Majeed 2019a: 96; 2019c) has suggested that this choice is based on Grierson's personal history, a claim that seems inaccurate given the history of using the Parable as a text specimen in linguistic compilations (Hamans 2017; Chelliah & de Reuse 2011: 48).

from sections of the text which are not shown. For more detail, the reader is referred to the full versions of the specimens, which can be consulted online.²⁰

Throughout Section 4 I pay particular attention to words with segmented transcription and/or glosses, and especially to the order in which the glosses and transcription segments match up, as this provides insight into whether the data is presented with a focus on word- or sub-word-level analysis.

4.1 Kashmiri

The section on Kashmiri, in volume 8.2 ('Indo-Aryan Family. North-Western Group. Specimens of the Dardic or Piśāchā Languages (including Kāshmirī)') of the LSI, contains a very substantial sketch grammar (Grierson 1919: 254–316), including two different versions of the Prodigal Son as well as two other text specimens, both local. The glossing characteristics of the Kashmiri Parable, the first lines of which are presented in Figure 2, are the following: the transcription features segmentation; the text is comprehensively glossed, using a word-for-word glossing type, with segmented glosses.

Akis-mahanivis	ös ⁱ	z ^h	nēchiv ⁱ .	Timau-manza	dop ^a
<i>To-one-man</i>	<i>were</i>	<i>two</i>	<i>sons.</i>	<i>Them-from-in</i>	<i>it-was-said</i>
kūs ⁱ -hih ⁱ	mōlis	ki,	‘hē māli,	mě	dih
<i>by-the-younger</i>	<i>to-the-father</i>	<i>that,</i>	<i>‘O father,</i>	<i>to-me</i>	<i>give-thou</i>
his ^a	yus	mě	wāti.’	Tawa-pata	tam ⁱ
<i>the-share</i>	<i>which</i>	<i>to-me</i>	<i>will-arrive.’</i>	<i>From-that-after</i>	<i>by-him</i>
khōt ^a ra	dana	bög ^a rōwun.	Kētsau-dōhau-pata	suh	kūs ^a
<i>for-the-sake</i>	<i>wealth</i>	<i>was-divided-by-him.</i>	<i>From-some-days-after</i>	<i>that</i>	<i>younger</i>
bōy ^a	sōru-y	raṭith	akis-dūr-dishēs	sakharyāv.	Tati
<i>brother</i>	<i>all</i>	<i>having-taken</i>	<i>to-one-far-country</i>	<i>set-forth.</i>	<i>There</i>
					tam ⁱ
					<i>by-him</i>

Fig. 2. Section of the Kashmiri version of the Parable in the LSI (Grierson 1919: 325)

As far as the correspondence between segments in the transcriptions and the glosses is concerned, two patterns are found in this specimen:

When the transcription is unsegmented, the glosses can be either single-word glosses or multi-word glosses (see Section 3). When the latter, they follow the word order of the metalanguage, namely of English. One such example is *paralōkuk^u* ‘of-heaven’ (ibid, 325). This order is unsurprising: as there is no sub-word structure in the transcribed word to gloss, the multi-word gloss serves to convey the meaning of the transcribed word, with the hyphens thus used for aggregation rather than segmentation, and not its internal structure.

On the other hand, where the transcription is segmented, and there is a corresponding number of segments across the transcription and the glossing (i.e. the glosses are all single-word glosses), the order of glosses matches that of the transcription. This suggests that glosses are mapped onto the segments of the

²⁰ The LSI can be found online at <https://dsal.uchicago.edu/books/lsi/>.

transcription, revealing the sub-word structure of the material. An example is *jān-hyuh*^u ‘good-very’.

In some cases, however, a transcribed word is glossed with both multi-word glosses and single-word glosses, resulting in a mismatch in the number of segments across glossing and transcription tiers. This is seen in the following words:

timau-manza ‘them-from-in’

ditsāv-na ‘was-given-not’

khētē-pēth ‘the-field-on’

pananēn-nōkaran-manza ‘thine-own-servants-from-among’

It is helpful in these cases to identify the multi-word glosses and to consider them as units. Using brackets to do so yields, respectively, ‘them-[from-in]’; ‘[was-given]-not’, ‘[the-field]-on’, ‘[thine-own]-servants-[from-among]’.²¹ The single-word and bracketed glosses now match the number of segments in the transcription, and we see that the glosses are ordered according to the sub-word level of the transcribed language. In this sense, they express (at least partly) a morphemic analysis of the transcribed words.

The bracketing reveals that the order of the segments is in fact consistent, despite appearances to the contrary: the word order of the metalanguage is used to aggregate multi-word glosses that are segmented, whereas the word order of the transcribed language is used when glosses feature segmentation into sub-word elements.

4.2 Telugu

The Telugu chapter of the LSI, found in Volume 4 (‘Munda and Dravidian languages’), features nine different versions of the Prodigal Son (in different dialects), following an introduction to Telugu (Grierson 1906: 576–589) of which the ‘Telugu skeleton grammar’ makes up two pages. The image in Figure 3 is from the specimen labeled ‘standard dialect’.

Voka manushyu-ni-ki yiddaru kumāru-lu vuṇḍiri. Vāri-lō chinnavāḍu, ‘ō
A man-to two sons were. Them-among the-younger, ‘O
 taṇḍri āsti-lō nā-ku vachēhē pālu yimm’-ani taṇḍri-tō cheppin-appuḍu
father property-in me-to coming share give’-so the-father-to said-when
 āyana vāri-ki tana āsti-ni pañchi peṭṭenu. Konni dinamul-aina
he them-to his-own property having-divided put. A-few days-having-become
 taruvāta ā chinna kumāruḍu samastamu-nnu kūrṭṣukoni dūra dēśamu-na-ku
after that younger son all-together having-gathered far country-to
 prayāṇam-ai velli tana āsti-ni durvyāpāramu-valla pāḍu-chēsenu. Ad-antā
having-journeyed having-gone his property bad-behaviour-by waste-made. That-all
 vrayamu-chēsina taruvāta ā dēśam-andu pedda karuvu kaligin-anduna
expending-having-made after that country-in mighty famine having-arisen-because
 ataḍu yibbandi paḍa-sāgenū. Appuḍu ataḍu velli ā dēśa-sṭhu-la-lō
he strait to-suffer-began. Then he having-gone that country-dwellers-among

²¹ This type of bracketing into multi-word glosses is challenging with case-marked noun phrases, as the various elements making up the noun phrase are all inflected (such as the numeral ‘one’ *akis* in the first word in Figure 2, in addition to the noun ‘man’), yet rendered by a single English gloss covering the noun phrase. (see Koul & Wali 2006, Chapter 3; Grierson 1919: 274)

Fig. 3. Image from the Telugu version of the Parable in the *LSI* (Grierson 1906: 591)

This text contains a high number of words with more segments in the transcription than in their glosses. A few examples are: *manyushyu-ni-ki* 'man-to'; *dēsa-sthu-la-lō*, 'country-dwellers-among', *āsti-ni* 'property', *ākāsāmu-na-ku* 'heaven-to'. In such cases, accessing the principles underlying segmentation of the transcription is impossible without sufficient knowledge of the language, something which the short accompanying grammatical sketch does not provide. The reason that what might be called 'over-segmentation' of the transcription is not useful is that, unlike a situation in which the glossing tier contains more segments with respect to the transcription, as discussed in Section 4.1, there is no way, through the glosses, to retrieve information that reveals an analysis at the sub-word level of the language, because there is no possibility of equating fewer glosses with a greater number of transcribed segments.

On the other hand, for words in Figure 3 (and the rest of the Telugu text) with matched numbers of segments in the transcription and the glossing, the glosses reproduce the word order of the transcribed language, as in *dēsam-andu* 'country-in'; *na-ku* 'me-to'. As with Kashmiri in Section 4.1, when there is a mismatch in number, with more glosses than transcribed word segments, we find that once the multi-word glosses are identified, the overall order of glosses matches that of the transcribed language, and the order within the multi-word glosses matches the metalanguage, as expected: recall that multi-word glosses use hyphens for aggregation rather than segmentation and that these do not hinder our ability to use the glosses and their ordering to recover a sub-word analysis. This can be seen, for example, in *taṇḍri-tō* 'the-father-to' ([the-father]-to); *paḍa-sāgenu* 'to-suffer-began' ([to-suffer]-began), both from Figure 3. Although the analysis provided by the glosses does not descend to the morphemic level, it is sufficient that one could use the glosses to compile a sketch grammar of the language in question, of a level that would enable language comparison.

4.3 Lepcha

The Lepcha grammar sketch, in Volume 3.1 ('Tibeto-Burman languages of Tibet and North Assam') of the *LSI*, runs to eight pages (Grierson 1909: 233–241). The Lepcha texts in the *LSI* are the Parable Prodigal Son (ibid, 244-246) and one other specimen.

The specimen in Figure 4 features a word-for-word glossing scheme, with segmented transcriptions and glosses. The segmentation differs from that seen in the Kashmiri and Telugu versions of the parable, as they demarcate every syllable within the transcribed material: the writing system for Lepcha being a syllable-based abugida script, the segments correspond to the glyphs in the version of the parable transcribed in Lepcha (ibid, 242-243) which precedes the Romanized version.

Because segmentation is based on the syllable in this Lepcha text specimen, it is not particularly informative in terms of glossing: the glosses (which are sub-word level) cannot be correlated with the transcribed segments (which are syllabic and do not necessarily match the glosses) in any meaningful way. What can be said, however,

is that the order of the glosses is that of the Lepcha, as can be seen throughout the text: *kāt-sa* 'one-of', *tek-nun* 'small-by', *ā-bo-rem* 'father-to' are but a few examples. Even though the transcribed segments and their glosses cannot be used as the basis for a sketch grammar, as there is no way to establish a correspondence without already knowing the language, the order of the glosses for transcribed words gives us information about the morphosyntax of the language.

Ma-ró	kāt-sa	ā-kup	nyet	nyi.	Ha-nyí	nóng-kā	ā-kup	tek-nun
<i>Man</i>	<i>one-of</i>	<i>sons</i>	<i>two</i>	<i>were.</i>	<i>Both</i>	<i>among</i>	<i>son</i>	<i>small-by</i>
ā-bo-rem	shu,	‘e	ā-bo-wa,	gyū-gi-cho-nun	sa-tet	ka-su	ka-kā	
<i>father-to</i>	<i>said,</i>	<i>O</i>	<i>father,</i>	<i>substance-wealth-from</i>	<i>how-much</i>	<i>my</i>	<i>share</i>	
thūp-shyet	nyi-wung-re	ka-sum	nóng-wa.’	O-thā	hu-nun	ha-yūm	ha-do-sa	
<i>getting-for</i>	<i>being-that</i>	<i>me-to</i>	<i>give.’</i>	<i>Then</i>	<i>him-by</i>	<i>them-to</i>	<i>his</i>	
gyū-gi-cho	rit-bi-fāt-te.	Sa’ayāk	ā-gyāp	ma-bām	ā-kup	tek-nun		
<i>property</i>	<i>divide-give-finished.</i>	<i>Days</i>	<i>many</i>	<i>not-going</i>	<i>son</i>	<i>small-by</i>		
gyū-gi-cho	tyáng	gyom-bu-bān	lyāng	ā-rum	kāt-kā	nóng-lung		
<i>property</i>	<i>all</i>	<i>gathered-carried-having</i>	<i>country</i>	<i>far</i>	<i>one-to</i>	<i>going</i>		

Fig. 4. Image from the Lepcha version of the Parable in the LSI (Grierson 1909: 244).

4.4 Kachari

The Kachari materials are found in volume 3.2 ('Bodo, Nāgā, and Kachin groups of the Tibeto-Burman languages') of the LSI. In addition to the grammatical sketch, a number of language specimens from various collectors are provided, covering a variety of dialects and districts of Plains Kachari. Of these, four are versions of the Parable of the Prodigal Son, the one studied here being the first presented (Grierson 1903: 18–20).

This version features comprehensive segmentation and word-for-word glosses, across the entire specimen. Glosses are ordered according to the transcribed language. Transcribed words featuring more segments than their glosses are frequent in this specimen: they are often verb forms, including passives and converbal forms, which are difficult to render through the lexical glosses which are largely used throughout the LSI, and thus often feature glosses which render the semantics of the verb forms rather than their morphology.

Sā-sè	mānsūi-hā	fśā	z’lā	sā-nūi	dang-man.	Bī-ni	f’sā-z’lā
<i>Person-one</i>	<i>man-of</i>	<i>child</i>	<i>male</i>	<i>person-two</i>	<i>were.</i>	<i>Him-of</i>	<i>child-male</i>
mudūi-ā	bi-fā-ni-āu	khithā-nai-sè,	‘he	ā-fā,	nang-ni	zi	basthu-
<i>little-the</i>	<i>his-father-to</i>	<i>said,</i>	<i>O</i>	<i>my-father,</i>	<i>thee-of</i>	<i>what</i>	<i>property-</i>
ni	zi	bhāgū	āng-ni-āu	gaglai-ū,	bī-khō	āng-nū	hū.’
<i>of</i>	<i>what</i>	<i>share</i>	<i>me-to</i>	<i>falls,</i>	<i>that</i>	<i>me-to</i>	<i>give.’</i>
							<i>That-at (thereon)</i>
bi-fai-ā	gāgai-ni	basthu-khō	bī-sūr-nū	rān-nā-noi	hū-nāi-sè.	Bī-ni	
<i>his-father-the</i>	<i>himself-of</i>	<i>property</i>	<i>them-to</i>	<i>having-divided</i>	<i>completed-giving.</i>	<i>That-of</i>	

Fig. 5. Image from the Kachari version of the Parable in the LSI (Grierson 1903: 18).

Examples of such verb forms from Figure 5 are *khithā-nai-sè*, 'said' (line 3); *gaglai-u* 'falls' (line 5), *rān-nā-noi*, 'having divided' (line 7). The skeleton grammar, brief as it

is, makes it possible to identify these forms (*pluperfect*, *present*, and *participle*, respectively), even though no analysis of the different morphemes in the construction is provided, or indeed accessible.

The glosses for the non-verbal morphology are, however, presented so as to make an analysis of the nominal morphology of Kachari accessible: the glosses match the segments of the transcribed words (except in cases where transcribed words feature additional segments with respect to their glosses), and can be used to determine features such as case-, possessive- and number-marking.

4.5 Singpho

Singpho, also called Kachin in the *LSI*, is also presented in volume 3.2 ('Bodo, Nāgā, and Kachin groups of the Tibeto-Burman languages'). The grammatical description is eleven pages long (Grierson 1903: 499–510). The Parable of the Prodigal Son is one of two specimens provided for the language, and features word-for-word glossing with segmentation in both transcription and glossing tiers.

Like in Kachari, there are quite a few words with more transcription segments than glosses: see *ngā-dai* 'were', *nyē-nā* 'my', *gadē-mā-ngā-dī-mang* 'whatever' (with a footnote explaining the sub-word level structure), *kām-mō-kām-rūm-shā-thūm-hā-dai* 'has-wasted'. Some of these can be interpreted using the sketch grammar (*-dai* is the present tense suffix (ibid, 508); *-nā* is a genitive marker (ibid, 507)), while others cannot. This is the same issue we have seen in the specimens for other *LSI* languages: an overly segmented transcription with respect to the glosses results in analytical gaps, as the meaning of the segments is not accessible. Nonetheless, the Singpho specimen's glossing is unusual for the *LSI* in that it provides glosses that translate the word-level much more frequently than the sub-word level. There are exceptions, such as *nāng-goi* 'you-to'; *mūng-sīdai-goi-nā*, 'country-that-on-of', but they are rare. (These are both in a section not shown in Figure 6; see pages 511–513 of volume 3.2 for these and a few more examples). The data for a correct analysis at the sub-word level can often be recovered from the sketch grammar provided, and sometimes from footnotes, but the glossing is considerably less 'successful' than what we have seen in the other language specimens.

Singphō	aimā-nā	keshā	n'khong	ngā-dai.	Keshā	kachī	ga-wā
Man	a-of	sons	two	were. ¹	Son	(the)-younger	(his)-father
fē	ngā-dai,	'nyē-nā	man	gadē-mā-ngā-dī-mang ²	jō-ū.'	Deng	ga-wā
to	says,	'my	share	whatever	give.'	Then	father
jō-dai.	Ga-wā	rai	gam-jō-dī ³	n'thomī	keshā	kachī	sībā
gives.	Father	property	having-divided	after	son	the-younger	many
yā	n'kring-dī	num	chān	goi	dām ⁴	sā-hā-dai.	Khī
days	not-staying	country	a-far	in	to-visit	has-gone.	He
khī-nā	rai-bok ⁵	kām-mō-kām-rūm-shā-thūm-hā-dai.	Rai-bok ⁵	yōng	rūm-shā-thūm		
his	goods	has-wasted.	Goods	all	wasting		

Fig. 6. Image from the Singpho version of the Parable in the *LSI* (Grierson 1903: 511).

An explanation for this difference can be found in the Singpho materials sent to Grierson, which had been glossed by the collector using a number-type glossing scheme as opposed to a word-type, with a corresponding loss of information at the sub-word level. See Figure 11 below for an excerpt of the data that was sent to Grierson for the LSI (see Section 5.1.2 below): while the date on the LSI version is 1896 and that on the *Specimens* version is 1876, they are both by the same collector (J.F. Needham, Esq) and the transcription is very close between the two, apart from some suprasegmentals and the segmentation of the LSI version. One strong indication that the source is the same is that the *Specimens* version includes, in the English text column, parenthesized words; the LSI version features the same words, still in parentheses (see, for example, '(he)' and '(his)' in the first glossing line in Fig. 6). Additionally, the footnotes which resolve morphosyntactic issues in the *Specimens* version are found, identically phrased, in the LSI version. It is therefore possible that the differences we see in the LSI Singpho specimen's glossing, with respect to other LSI specimens, are a result of the glossing scheme of the original data sent to Grierson.

4.6 Analysis of glossing practices in the specimens of the LSI

The LSI texts examined in Section 4 all share a number of features with respect to their glossing, which are summed up here:

- The specimens are annotated with a consistent presentation style: all texts feature transcription and glossing tiers, with glosses appearing under the transcribed words they correspond to. There is no translation tier present because all specimens are of the same text, the Parable of the Prodigal Son, and presumably familiar enough to readers to forego providing an English version;
- Segmentation is found in both transcription and glossing tiers. Note however that the number of segments across the transcription and glossing tiers does not always match. The various possibilities are listed in Figure 7;
- Glosses are of the word-for-word type, with the glosses grouped according to what is considered a word in the transcribed language;
- The basis for glossing is, whenever possible, the sub-word level; in some cases, however, glosses do not represent an analysis of the sub-word level and are only relevant to the word-level;
- Consistent practices govern the order of segments in glosses.

The glossing practices across the LSI specimens are marked by an impressive consistency, especially considering that the sources were provided by a great number of collectors whose own practices varied widely. (Section 5.1.2 discusses some source texts in more detail). In discussing particularities of the specimens in different languages, I focused in particular on segmentation across transcriptions and glosses. This is because segmentation is highly relevant to the development of morphemic glossing: the latter can only be claimed to be present if each segment of the transcribed text is given a gloss, ordered in such a way as to match the transcribed segments for ease of identification. Grierson's segmentation and glossing does, for the

most part, make it possible to identify the sub-word level elements (often morphemes) of the transcription, even though it looks, at first glance, like his practices are inconsistent: In the absence of a diverse system for signaling different types of boundaries with distinct typographical symbols,²² Grierson makes use of spaces and hyphens to convey all the boundary types needed for the purposes of glossing. Even though this results in the hyphen having two simultaneous functions, segmenting and aggregating, the system yields consistent results.

There are various situations to consider with respect to the number and order of segments within a transcribed word and its glosses:

a) Neither the transcribed word nor the gloss features segmentation marks. Examples of this are Figure 3's *yiddaru* 'two', *pālu* 'share', *vachchē* 'coming'. Transcription and gloss are in a 1:1 correspondence and pose no problem with respect to a morphemic analysis, as they could very well be single morphemes. (While this looks to be the case for the first two examples, it seems less likely for the third.)

b) Segmentation is present in the transcription but not in the glosses. Examples of this are Figure 4's *ma-ró* 'man'; *ā-bo-wa* 'father' in Lepcha. In this situation, the data has been segmented according to a principle which is not transparent: it may be interpretable with the help of the accompanying sketch grammar, as was described in Section 4.5 for some cases in Singpho. The segments may reflect elements that make up a complex word in the transcribed language (these are sometimes elucidated through footnotes, as in Singpho), or they may be based on some other principle.

c) Segmentation is present in the glossing but not in the transcription, as in (1) above. Other examples are Figure 6's *kachī* 'the-younger', *chān* 'a-far', *dām* 'to-visit'. We find this situation in the LSI, and it can generally be explained as cases where the hyphen is used as an aggregation symbol, as opposed to being used to segment material into sub-word level units. These are what we have termed multi-word glosses (see Section 3).

d) Segmentation is present in both the transcription and in the glossing. This yields two possible situations: i) one with a matching number of segments in both the transcription and the glossing, as in (3), and ii) one with a mismatched number of segments across the two lines, as in (4). (Both examples (3) and (4) are from the Kashmiri version of the Parable of the Prodigal Son).

(3) *tsöchyau-sūtin*
loaves-with

(4) *tasandi-khöt^ara*
of-him-for-the-sake

As mentioned above, the number of segments across transcription and glossing lines is important because identification of the transcription segments can only be

22

See https://www.christianlehmann.eu/ling/ling_meth/ling_description/representations/gloss/index.php?open=boundary_symbols

unequivocally made when their number matches that of the glosses. Thus while (3) is potentially a case of morphemic glossing, as the two segments in the transcription match up with the two in the glossing line, example (4) has five glossing segments to the two segments in the transcription, making it unclear how to interpret the correspondence between the two lines.

Even when the number of segments in the transcription and glosses features a mismatch (in other words the situation in d ii and in (4)), there are fairly consistently applied principles at work throughout the LSI specimens which make it possible to extract sub-word level analysis from the glosses. These rules affect the order in which the gloss segments occur.

When c) obtains, in other words an unsegmented transcribed word receives multi-word glosses, the glosses are for the most part ordered according to metalanguage. This is because, as seen above, the hyphens are used to aggregate rather than segment: a multi-word gloss reflects a transcribed word which cannot be translated with a single word in the metalanguage. (Note that the Leipzig Glossing Rules provide ways to reserve the hyphen for morpheme boundary by offering different typographical symbols for multi-word glosses; see footnote 16.) The glosses do not reflect any sub-word level of analysis because there is, as far as can be seen from the transcription, no access to that level.

When a segmented word receives a number of glosses that matches the segments, the situation described in d i) and illustrated in (3), the glosses are generally ordered according to the transcribed language. Both the corresponding number of segments and the order of the glosses suggest that this glossing scheme conveys a sub-word level analysis. Note however that situations are found where the segments in transcription and glossing do not correspond, even though they are of a matched number. This is the case, for example, with the following Kachari examples (Grierson 1903: 20): *hũ-ā-khũi* 'did-not-give', *sefai-ā-khũi* 'did-not-break'. The similar verb forms make it clear that the first segment is the verb root, and yet the glosses both have the verb root as the last element.

When a segmented word receives more glosses than it has transcribed segments, this signals that the glosses are made up of both single-word and multi-word glosses.²³

In this case, again, we find relatively consistent ordering within the glosses: the overall order of the glosses is that of the transcribed language, but the multi-word gloss's internal organization reflects the metalanguage word order. (This can be seen in (4), where bracketing the multi-word glosses leads to [of-him]²⁴-[for-the-sake]). There are, however, words that do not obey these principles: The following examples are from the same page of the Telugu parable (Grierson 1906: 592): *sammatintsa-lēdu* 'did-not-agree'; *mīra-lēdu* 'did-not-transgress'. One does not need knowledge of

²³ The number of words within the Parable specimen that fit this type vary considerably from language to language, for reasons that probably have to do with the morphology of the languages. The number of words with mismatched segmented transcription-gloss combinations are as follows: Kashmiri: 46 examples; Telugu: 29 examples; Kachari: 2 examples; Lepcha: 1 example; Singpho: 7 examples..

²⁴ The grammar sketch confirms that the first multi-word gloss, 'of-him', indeed corresponds to a genitive form of the animate unseen demonstrative (Grierson 1919: 280).

Telugu to see that the common element, *lēdu*, is not the verb root (according to the sketch grammar (ibid, 589), it is the negative auxiliary) and that the order of segments and glosses, while matched in number, do not correspond in these two words. These configurations of segment number and order are summarized in Figure 7.

	single-word gloss	multi-word gloss	multiple single word glosses	single-word gloss(es) +multi- word gloss(es)	multiple multi-word glosses
unsegmented transcription	1:1 correspondence	multi-word gloss internally follows word order of metalanguage	not found		
segmented transcription	rare (happens with languages segmented according to syllabic writing system or other (unidentified) phenomenon)		glosses follow word order of transcribed language	multi-word gloss internally follows word order of metalanguage; overall order of glosses follows word order of transcribed language	

Figure 7. Correspondences between segments in transcription and glosses

While mismatches in numbers of segments across the transcription and glossing tiers and different orders for glosses may make LSI glosses appear to be marked by inconsistency, they in fact constitute a cohesive system with underlying principles. It is interesting, insofar as they make up an innovative glossing system, that these principles are not made explicit in the introduction to the LSI, nor are they explained in any of the volumes. Perhaps this can be explained by virtue of their seeming quite natural: sub-word level structure is shown when possible, and when not, the focus is on semantics, in the word order of the metalanguage. There is only a conflict in a subset of glosses, in which case, again, these two principles are adopted to the extent possible, even in combination.

5. *Glossing practices in other materials*

Section 4.7 established a set of features found across the specimens examined in the LSI, and we now examine text specimens²⁵ from the same linguistic area produced in the period before and after the LSI, in order to the place of the LSI within this landscape.

5.1 *Glossing of South Asian languages before the LSI*

²⁵I use textual material, as opposed to examples within a grammatical description, because this limits one of the variables (see Introduction).

This section looks at two sources of descriptive material preceding the LSI: description from the 1850's by B. H. Hodgson, and a volume of translations of the Parable of the Prodigal Son into various languages of South Asia published in 1897.

5.1.1 *B. H. Hodgson*

Hodgson is acknowledged to be the first Western linguist to work on the description of Himalayan languages,²⁶ and produced the first grammars of Kiranti languages (van Driem 2004).

The grammatical descriptions Hodgson produced (Hodgson 1857a, on Vayu, Hodgson 1857b; Hodgson 1858, on Bahing) do not contain what I would consider glossed materials: forms that need to be made explicit, such as complex verb forms, are provided with a translation, but not glossed—there is only ever one English-language line accompanying materials. This is possibly a result of the fact that bulk of the descriptions are focused on the description of verbal morphology, but also applies to the text specimens that accompany the descriptions: they contain an unsegmented transcribed²⁷ text (see Fig. 8), with no glosses, and a free-standing translation.

A SPECIMEN OF THE VA'YU LANGUAGE.

Ang ming Páchya nom. Ang thoko Váyú nomi (or Gó Váyú gnom) Khásakhata Háyu itkem. Ungki dávo be Váyú ischikem. Go jekta dumsungmi. Hátha bong dumsungmi ghá má sengmi. Lé got kulup chhuyung* wanikhen. Dhankuta mu khakehling puehhum chupvikhata póguha háta vik páchikokni. Ang kó má nom. Ang távo Gajraj Thápa nung nomi. Gonha kóphe nakphe inang munang wathi yengkum. Wathim nárung gonha blektum. Wathim chho le pókum. Honko á thum rámi. Captánha thúm hánung hónpingkum. Ang dávo lit'nung blining chólo chupsit khen inhe gó gonha mutpingkum. Dávo chinggnak chamechem. Gon

Figure 8. Vayu specimen extract (Hodgson 1857: 483).

There is very little use of segmentation in the transcriptions, apart from a division (somewhat arbitrary, as it is not consistently applied across paradigms) into words using spaces.²⁸ Hodgson's annotation practices within examples and paradigms reveal a keen sense of the morphosyntax of these languages, as seen in the fine-grained free translations that accompany the material (for example, using archaic pronouns to specify number and role for 2nd person arguments—you vs *ye*, *thou* vs *thee*—and inclusive vs exclusive for non-singular 1st persons (Lahaussois forthcoming)), but he does not transfer his analytical knowledge with the help of any kind of glossing.

²⁶ Although one bibliography of Hodgson's life (Hunter 1896) and one collective work (Waterhouse 2004) describing facets of his polymathy exist, there is no historiography of his significant contributions to linguistic description. This is the topic of a PhD dissertation currently underway at Université de Paris by Jean-Baptiste Lamontre.

²⁷ Note that 'transcribed' here means a transcription from an oral source, as the language was unwritten.

²⁸ Hodgson's paradigms for Bahing do, in the initial pages on verb conjugation, contain hyphens for segmentation, a practice he abandons after a few pages. See Lahaussois (2020a) for more details on the verb paradigms in Hodgson.

The practices we find in Hodgson follow those seen in other materials on the languages of this area in the mid-19th century: The Tibetan grammar written by Csoma de Körös (1834) takes a similar approach in not providing sub-word-level glossing, either in grammatical examples or in text specimens, but makes use of translation (sometimes labeled 'literal'), in a text block below the source text, to annotate the materials. The annotation in this case is limited to translation, as the materials are only presented in Tibetan script.

5.1.2 Specimen translations

One rich source of pre-LSI annotated text specimens is an 1897 volume entitled *Specimen Translations in Various Indian Languages* (henceforth *Specimens*).²⁹ While the volume is edited by Grierson, his role was mostly that of a compiler of materials pulled from various sources, which are identified at the top of each language version of the Parable along with the date of collection. In some cases, the text specimen is collected by the same person across both the *Specimens* volume and the LSI, as is mentioned above (in Section 4.5) with respect to Singpho; in other cases, the versions appear to be identical, apart from some adjustments to the transliteration, even if they are not identified as such (because a different collector is named, or no name is provided). The existence of this collection makes it possible to trace the effect of Grierson's curating and editing on the LSI text specimens by comparing them with how the same texts are glossed in the *Specimens* volume.³⁰

What a close study of the *Specimens* volume reveals is a great variety of annotation and glossing techniques for South Asian languages in the period prior to the LSI. Looking at the same five languages as sampled in Section 4, we find a range of features among their glossing styles.

Most of the language specimens are presented with interlinear glossing (as is the case for the majority of language specimens in the volume). An example of the presentation is seen in the Kachari materials in Figure 9 (materials dated 1896). Note however that, the Kachari specimen is only glossed over its first four lines, although segmentation is used throughout the transcribed text.

Sá-se	mànsòi-há	fisá-zlá	sá-noi	dang-man.	Bí-ni	fisá-zlá
One	man-of	child-male	two	(there) were.	Him-of	child-male
mudoi-á	bifá-niáu	khithá-nai-sé,	He	áfá, nang-ni zi(a)	basthu-ni(a)	
little(one)-the	father-to	said,	O	father, thee-of what	property-of	
zi(a)	bhága(a)	áng-niáu	gaglai-a,	bi-kho	áng-na	hà;
what	portion	me-to	falls,	that	me-to	give;
gágai-ni	basthu	-kho(a)	bi	-sar	-na	rán-ná-noi
himself-of	property	(objective case)	them	(plural suffix)-to	having	divided (it)
gaglai-ni	gásenā	thákhá(a)	futhám-ná-noi			
gaján	tháu-niáu	tháng-ná-noi	hámá	hábá	máu-ná-noi	gásenā
khàm-naise.	Bi-há	gásenā	thákhá(a)	zāpnai-áuhá	bi	tháu-niáu
						ángkhá(a)
						gadāt zā-

²⁹ The *Specimens* volume can be found at Archive.org: <https://archive.org/details/dli.csl.8052/page/n111/mode/2up>

³⁰ The *Specimens* volume contains 65 different language versions of the Parable, offering plenty of material for comparison.

Figure 9. Sample of Kachari Parable in *Specimens* (53): the interlinear word-for-word glossing stops after four lines.³¹

Exceptions to the interlinear glossing scheme are the Lepcha materials and the Singpho materials (Fig. 11), each for different reasons. The Lepcha text (Figure 10; materials dated to 1896) is transcribed and segmented, but unglossed. (The Singpho glossing scheme is number-based, as seen below.)

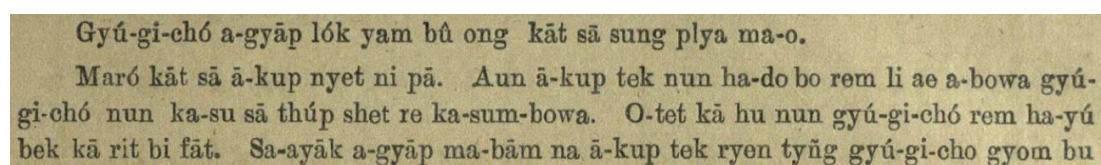


Figure 10. Sample of Lepcha Parable in *Specimens* (51).

It is unclear what the basis for segmentation is in the Lepcha in Figure 10. It resembles a syllable-level segmentation, except that a few words include unsegmented disyllables: *maró* (line 2) and *bowa* (line 2), probably invalidating the syllable hypothesis.

A variety of glossing types are found across the specimens, both number- and word-type glosses, and both word-for-word and literal translation. Singpho (Figure 11; materials dated 1876) uses a numbering glossing scheme.

SINGPHO.					ENGLISH.				
Singpho	aimá'ná	keshá	n'khong		A	man	had	two	sons.
2	1	5	4		1	2	3	4	5
ngádai.	Keshá	kaohí	gawá	fé ngá dai	(The)	younger	son	said to	(his) father,
8	7	6	10	9 8		6	7	8	9 10
nyê ná man	(1) gadê	má	ngá dī	mang jôú.	Give (me)	whatever	share	(of	property)
(my)	13		12	11	11	12	13		
Deng gáwá	gám dī	jô dai.			(is mine).	Then	(his)	father	dividing
4	15	16	17			14	15	16	
Gawá rai	gam	jôdí(2)	n'thomí	kashá,	gave (to him).				
18(a)	20	19	18	22	17				
kachí síbā	yá n'kringdí.	Nám	chang	goi	After	(his)	father	had	divided (his)
21	24 25	23	30	29 28	18		18(a)	19	
dám(3)	sá hádai.				property	the	younger	son	not
27	26				20	21	22	23	
					many	days	(at home)	went	to
					24	25	26	27	28
					a	far	country.		
					29	30			

Figure 11. Sample of Singpho Parable in *Specimens* (61): a numbering scheme is used to align Singpho and English words.

In the fragment of Singpho text in Figure 11, the transcribed material and the English translation are presented in side-by-side columns, with numbers used to

³¹ The volume has no page numbers; different versions of the Parable are given a number, which is what I list here.

connect the words across the two versions. Sometimes a number gloss covers more than one word in the Singpho (e.g. gloss number 12 which covers several Singpho words) or (more rarely) English (e.g. gloss number 19) text. In a sense, this is similar to the grouping of glosses that is achieved through hyphens in other texts. One interesting point to note is that the order of the numbers is that of the English text: the words of the English text are numbered sequentially, and not those of the Singpho.

Telugu (Figure 12; materials dated 1889) and Kachari use literal translation as a means of glossing.

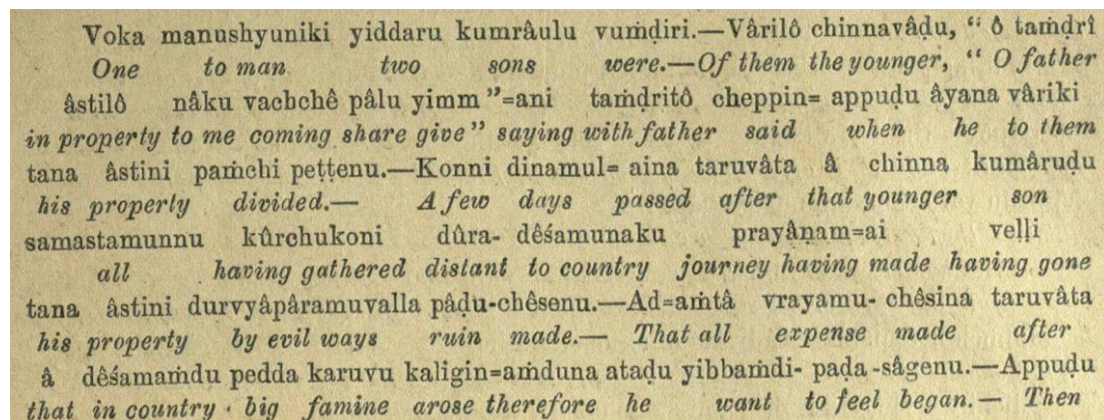


Figure 12. Sample of the Telugu Parable in *Specimens* (30): the glossing is accomplished through a literal translation.

For Telugu, the fact that the glossing line is occupied by a literal translation is all the more obvious when contrasted with the equivalent glossing in the LSI. Compare the first two lines of Telugu (Figure 12) with the LSI version (Figure 3):

"One to man two sons were.—Of them the younger, "O father in property to me coming share give" saying with father said when he to them..." (*Specimens* 1897)

"A man-to two sons were. Them-among the-younger 'O father property-in me-to coming share give'-so the father-to said-when he them-to..." (LSI)

The LSI retains characteristics that bring it close to a literal translation (punctuation, capitalization, the vocative particle), but also shows a shift towards representing the morphosyntactic workings of the language, with the case markers being ordered after the nouns they are associated with.

Another interesting feature of the Telugu text in Figure 12 is that some words are segmented with '=' signs (in addition to hyphens). These can be seen in line 2, line 3, line 4 and line 6 of Figure 12, and are also present in the rest of the text, where they appear to demarcate clausal morphology: quotatives, causal clauses, temporal clauses. It is interesting to note that these marks are only found in the *Specimens* texts of languages classified as Dravidian family suggesting that this is a mark used areally in text transliteration, although these marks are not found in the transcriptions in local scripts.

Word-for-word glosses consistent with those found in the LSI can be seen in the Kashmiri specimen (material dated 1882). The main difference is that in the latter, the order of material within glosses is English word order (and not that of the transcribed language, as in the LSI). This can be seen by comparing the same material as it appears in *Specimens* and in the LSI, especially in longer words involving multiple glosses. Compare:

tami-disakis-basawanis-nís 'near-to-of-that-country-an-inhabitant' (*Specimens*, 4a)

and

tami-dīshekis-basawanis-nishe 'to-of-that-country-an-inhabitant-near' (LSI);

or also

soran-handⁱ khōrakh-mahā-hyamav-sūty 'of-the-pigs with-the-(for)-food-bean-husks' (*Specimens*, 4a)

and

sōran-handi-khorakh-mahā-hemau-sūty 'by-the-swine's-food-bean-pods-with' (LSI).

In both of these complex words, the *Specimens* version orders the glosses according to the best approximation of English order, whereas the *LSI* version orders them according to the sub-word elements making up the transcribed word (see Section 4.6 for the principles underlying ordering of segments in the LSI).

In sum, the *Specimens* volume contains material, from the last quarter of the 19th century, which shows evidence of a wide range of glossing types and styles. Comparison of text specimens in the same languages across the *Specimens* volume and the LSI highlights how homogeneous glossing practices in the LSI are, with respect to what preceded.

5.2 Glossing of South Asian languages after the LSI

Some short descriptions of the languages Sangpang, Kulung and Thulung (all from Nepal) were produced in the 1930's by Wolfenden (1933; 1934; 1935). These descriptions begin with an annotated version of the Parable of Prodigal Son,³² before presenting sketch grammars. The article on Thulung, for example, provides information on the language, in the form of sections on 'Affinities'³³ (with other related languages), phonology ("Vowels, consonants, and consonant clusters"), 'Prefixes' (person indexes), 'Substantives' (including case marking), 'Adjectives', 'Numerals', 'Pronouns', 'Verbs', 'The verb from the conjugation', 'The verb in the story'. Throughout the grammatical sketch there is no glossing and examples are annotated with free translations: the material is rendered into fluent English, rather than provided with analytical morph-level glosses. This is perhaps surprising, considering Wolfenden's interest in morphology (he is the author of *Outlines of Tibeto-Burman Linguistic Morphology*, Wolfenden 1929), but the material is in fact

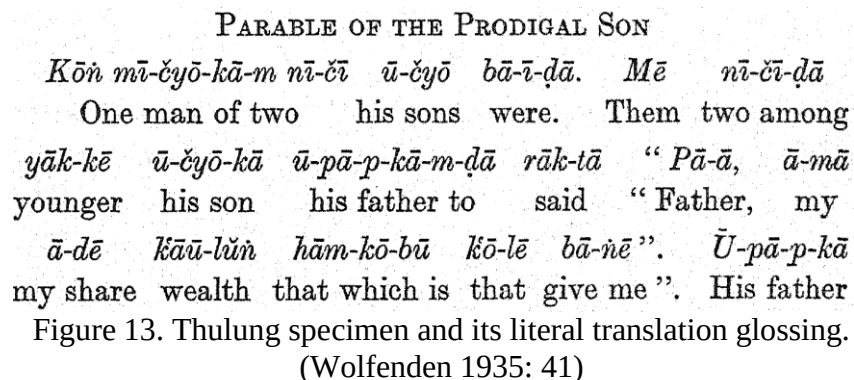
³² It is made clear that the specimens were collected by the author, thus constituting new data.

³³ All capitalized words are the actual titles for sections of the sketch.

presented in segmented form such that the value of the morphemes is transparent, despite the lack of glosses.³⁴

What are perhaps the more surprising elements of this short description are the use of segmentation and the type of glossing that accompanies the text specimen (in other words, different practices are used for examples and for the text specimen). As far as the segmentation is concerned, it appears to be based on the syllables of the language.³⁵ This decision on the part of Wolfenden is an interesting one: the language in question is certainly largely monosyllabic from a morphological point of view,³⁶ but the reduction of all words into segments made up of syllables suggests that Wolfenden is basing his analysis (as segmentation is clearly an element of an analytical process, as seen throughout this contribution) on his knowledge of related languages, and notably Tibetan. This is the case, for example, of the word 'man' *mī-cyō* rendered as two segments, on the basis not of speakers' sense of the word being made up of two morphemes (synchronically this word is felt to be a single morpheme by speakers), but on the basis of a reconstructed Tibeto-Burman form *mī*. Thus, rather than monosyllabic segments, we might say that Wolfenden's glossing treats as segments the morphemes that are useful for reconstruction across the family.

The second interesting point is to consider the glossing accompanying the text specimen: rather than provide glosses for each of the identified segments within the transcriptions—a challenging exercise, if we are right to suppose that the segments are based on etymological criteria rather than (synchronic) morphemic ones—Wolfenden provides glosses of the 'literal translation' type, similar to what was seen for Telugu and Kachari specimens in the 1897 *Specimens* volume (see Section 5.1.2), but distinct from what was seen in the LSI. This is illustrated in Figure 13.



The glossing practices found in Wolfenden's short grammatical sketches thus show an interesting juxtaposition of practices from the period preceding the LSI (the literal-type translation in the place of glosses, with resulting loss of segmentation in the

³⁴ An example of this is: "The genitive suffixes *-kā-m* as in *kōn mī-cyō-kā-m* 'of one man' (ibid, 640), where even though there is no gloss corresponding to each of the segments, they are analyzable from context and from the small phrases presented.

³⁵ The exceptions are words borrowed from Nepali, which are all identified and rendered, in footnotes, in their devanagari original.

³⁶ This is a known characteristic of Kiranti languages (of which Thulung is one example), particularly for verb roots. (Lahaussois 2020b: 48; Michailovsky 2017: 658)

glosses in order to map out correspondences between the transcription and glossing) as well as an LSI-like focus, at least as far as the transcription is concerned, with the sub-word level of the transcription. In the case of the latter, the segments go even below the morphemic level back to the cognate morphemes in related languages. An example of this is the word *pā-p* 'father', which is synchronically a single morpheme *pap*,³⁷ even though it is segmented by Wolfenden in order to resemble cognates in other Tibeto-Burman languages.

In this sense, the segments of the transcriptions here do not relate so much to the morphology and morphemes of the language in question, but rather to the morphemes of the proto-language. They thus retain the sub-word level analysis found in Grierson's work, but take it a step further in making the segments relevant to reconstruction.

Looking further afield, the practices seen in the LSI text specimens and in materials that follow are considerably different from the type of glossing found in the descriptions that make up the *Handbook of American Indian languages* (Boas 1911).³⁸ Like the LSI, Boas's *Handbook* is a survey work presenting grammatical sketches and text specimens from languages of a given geographical area and of multiple language families, in this case indigenous languages of North America. Furthermore, it was published within the same time frame as the LSI, appearing in 1911, at roughly the mid-point of the LSI project (which, recall, ran from 1894 and 1928). If we consider, as we have above, the presentation of a text specimen from one of the descriptions, in this case the Algonquian language Fox, we find that it is annotated as shown in Figure 14.

Inip¹³ It is said	acawaiy^{e4} long ago	negutenw¹⁵ it was once	ä'pepög¹⁶ when it was winter	ä·A·skime'pug¹⁷ when first it had snowed	
ä·A·skänwig¹⁸ while the first snow was on	neswi⁹ three	neniwag¹⁰ men	äcicāwāt^{c11} they went to hunt for game	māmai^{a.12} early	kegiceyāp^{a.13} in the morning.
Apatä'kig¹⁴ On the hillside	ä'pe'kwisasaga'k¹⁵ where it was thick with growth	ma'kwan¹⁶ bear	ä'pīci'kawānit^{c1.17} he went in making a trail.		

Figure 14. Text specimen in the language Fox (Jones & Michelson 1911: 868).

The glosses are based on the word unit of the transcribed language, which, as it is polysynthetic, sometimes functions as an entire clause. As such, the transcribed words of Fox are given glosses that behave as literal translations, conveying the content of the word or clause rather than its component parts. Consistent with this, the transcription features no segmentation, though the transcribed words are all accompanied by footnotes, where morphological explanations are provided.³⁹ The specimen is followed by a free translation presented as a single block.

³⁷ I base this statement on personal experience of more than 20 years of fieldwork on the Thulung language; see, for a summary of the scope of this work, see Lahaussais (2020b). (Note that *ä* in Wolfenden corresponds to IPA [a].)

³⁸ It is only after this contribution was accepted that I became aware of Solleveld's (2019) insightful article offering a comparative look at both Grierson's LSI and Boas's *Handbook*.

³⁹ The footnote for the fourth word in Fig. 14, for example, reads as follows:

This brief look at an Algonquian language reminds us that one of the significant characteristics of the LSI is that it deals with languages of the Indosphere, which share a number of linguistic characteristics, making it possible to devise glossing practices which can apply to the majority of them. Grierson's glossing practices may not have looked so consistent if forced to meet the challenge of also accommodating a language such as Fox.

6. Conclusion

In this article, I have presented the system that Grierson used to gloss text specimens in the LSI. This system is characterized by an interlinear presentation, with segmentation in both the transcription and glossing, using word-for-word glosses accessing the sub-word level of the material (See Section 4.6.) Though glosses in the LSI materials can superficially appear inconsistent in their treatment of the number and order of segments across transcribed and glossed material, when examined in detail, it becomes clear that LSI glossing practices constitute a cohesive system, the principles of which are laid out in Figure 7.

When contrasted with the work of scholars on South Asian languages in the period before the LSI (Hodgson in the 1850's and a compilation of texts published in 1897) and after the LSI (Wolfenden in the 1930's), we see that the LSI represents a turning point, in terms of consistency of rules and of analysis at the sub-word level. This suggests an evolution which paves the way for morphemic glossing, the basis for glossing as practiced today within the subfields of linguistic description and typology, and suggests a deepening analysis of the morphology of the languages, going hand-in-hand with the emergence of morphology as a field (François 2013). It is possible that the enhanced awareness of the morphological properties of the languages revealed by the glossing is a by-product of the survey process, through the generation of large collections of linguistic data from one language area.

It is my hope that this article constitutes a piece of a larger puzzle which is the historiography of linguistic glossing through time and space, and that a sufficient number of similar pieces will eventually give us a better understanding of the evolution of this wide-spread practice.

REFERENCES

- Alberizzi, Valerio. 2014. An introduction to kunten glossed texts and their study in Japan. *Les dossiers de HEL: [supplément électronique à la revue Histoire Epistémologie Langage]* 7. 1–9.
- Bailey, T. Grahame. 1921a. Review of: Linguistic Survey of India: The Dardic or Pisachā Languages. By Sir George Grierson, KCIE., D.Litt, Ph.D. 14x10 1/4, xi + 567 pp. Calcutta: Government Press, 1919. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 467–471.
- Bailey, T. Grahame. 1921b. Review of: Linguistic Survey of India: Sindhi and Lahnda. By Sir George Grierson, KCIE, D.Litt, Ph.D. 14x10 1/4, viii + 584 pp. Calcutta:

"ä'pepō'gⁱ third person singular inanimate intransitive verb of the indefinite conjunctive mode (ä- temporal augment; pep- initial stem used to express notions of WINTER, COLD, SNOW [§ 16]; -gi suffix with a locative sense [§ 42])."

- Government Press, 1919. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 471–475.
- Barbe, Katharina. 1996. The Dichotomy Free and Literal Translation. *Meta* 41(3). 328–337.
- Boas, Franz (ed.). 1911. *Handbook of American Indian Languages* (Smithsonian Institution Bureau of American Ethnology Bulletin 40). Vol. 1. Washington: Government Printing Office.
- Borin, Lars, Shafqat Mumtaz Virk & Anju Saxena. 2018. Language Technology for Digital Linguistics: Turning the Linguistic Survey of India into a Rich Source of Linguistic Information. In A Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing. CICLing 2017* (Lecture Notes in Computer Science 10761), 550–563. Cham: Springer.
- Bradley, David, Randy LaPolla, Boyd Michailovsky & Graham Thurgood (eds.). 2003. *Language variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff* (Pacific Linguistics 555). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Campbell, G. 1874. *Specimens of Languages of India including the Aboriginal Tribes of Bengal, the Central Provinces, and the Eastern Frontier*. Calcutta: Bengal Secretariat Press.
- Chelliah, Shobhanna & Willem de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.
- Cinato, Franck. 2015. *Priscien glosé : l'Ars grammatica de Priscien vue à travers les gloses carolingiennes* (Studia Artistarum 41). Turnhout (Belgium): Brepols.
- Csoma de Körös, Alexander. 1834. *A grammar of the Tibetan language*. Calcutta: Baptist Mission Press.
- Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF - Language Typology and Universals* 60(2). 95–99.
- Driem, George L. van. 2004. Hodgson's Tibeto-Burman and Tibeto-Burman today. In David Waterhouse (ed.), *The Origins of Himalayan Studies: Brian Houghton Hodgson in Nepal and Darjeeling 1820-1858*, 227–248. London and New York: RoutledgeCurzon.
- Finck, Franz Nikolaus. 1909. *Die Haupttypen des Sprachbaus*. Leipzig: B. G. Teubner.
- François, Alexandre. 2019. A proposal for conversational questionnaires. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological Tools for Linguistic Description and Typology* (Language Documentation & Conservation Special Publications 16), 155–196. Honolulu: University of Hawai'i Press.
- François, Jacques. 2013. Deux pionniers de la formalisation en morphologie linguistique au XIXe siècle: August Schleicher et Hugo Schuchardt. *Histoire Epistémologie Langage* 35(1). 111–142.
- Grierson, G.A. 1903. *Linguistic Survey of India Vol. 3.2 Tibeto-Burman Family, Specimens of the Bodo, Naga, and Kachin Groups*. Vol. 3.2. Calcutta: Office of the Superintendent of Government Printing, India.
- Grierson, G.A. 1906. *Linguistic Survey of India Vol. 4 Munda and Dravidian Languages*. Vol. 4. Calcutta: Office of the Superintendent of Government Printing, India.
- Grierson, G.A. 1909. *Linguistic Survey of India Vol. 3.1 Tibeto-Burman Family, General Introduction, Specimens of the Tibetan Dialects, the Himalayan Dialects, and the North Assam Group*. Vol. 3.1. Calcutta: Superintendent of Government Printing, India.
- Grierson, G.A. 1919. *Linguistic Survey of India Vol. 8.2 Specimens of the Dardic or Pisacha Languages (including Kashmiri)*. Vol. 8.2. Calcutta: Superintendent Government Printing, India.
- Grierson, G.A. 1927. *Linguistic Survey of India Vol. 1.1 Introductory*. Vol. 1.1. Calcutta: Government of India, Central Publication Branch.
- Hamans, Camiel. 2017. The Return of the Prodigal Son. *Scripta Neophilologica Posnaniensia* 17. 103–116.

- Hodgson, Brian H. 1857a. Comparative vocabulary of the languages of the broken tribes of Nepal [Vocabulary and grammar of the Vayu tribe]. *Journal of the Asiatic Society of Bengal* 26. 317-522 [372-485].
- Hodgson, Brian H. 1857b. Comparative vocabulary of the languages of the broken tribes of Nepal [Vocabulary of the Bahing tribe]. *Journal of the Asiatic Society of Bengal* 26. 317-522 [486-522].
- Hodgson, Brian H. 1858. Comparative vocabulary of the languages of the broken tribes of Nepal (continued from vol 26) [Grammar of the Bahing tribe]. *Journal of the Asiatic Society of Bengal* 27. 393-456 [393-442].
- Humboldt, Wilhelm von. 1968. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Bonn: Dümmler.
<https://archive.org/details/berdieverschied00humbgoog/mode/2up>.
- Hunter, William Wilson. 1896. *The Life of Brian Houghton Hodgson, British resident at the court of Nepal, member of the Institute of France; fellow of the Royal society; a vice-president of the Royal Asiatic society, etc.* London: John Murray.
<https://archive.org/details/lifeofbrianhough00hunt/page/n8>.
- Jones, William & Truman Michelson. 1911. Algonquian (Fox). In Franz Boas (ed.), *Handbook of American Indian Languages* (Smithsonian Institution Bureau of American Ethnology Bulletin 40), vol. 1, 735-874. Washington: Government Printing Office.
- Kosukegawa, Teiji. 2014. Explaining what kundoku is in the premodern Sinosphere. *Les dossiers de HEL: [supplément électronique à la revue Histoire Epistémologie Langage]* 7. 1-20.
- Koul, Omar N. & Kashi Wali. 2006. *Modern Kashmiri Grammar*. Springfield: Dunwoody Press.
- Lahaussais, Aimée. 2016. The translation tier in Interlinear Glossed Text: Changing practices in the description of endangered languages. In Patricia M. Phillips-Batoma & Florence Xiangyun Zhang (eds.), *Translation as Innovation: Bridging the Sciences and the Humanities*, 261-278. Victoria, TX: Dalkey Archive Press.
- Lahaussais, Aimée. 2020a. The shapes of verbal paradigms in Kiranti languages. *Faits de Langues* 50(2). 71-93.
- Lahaussais, Aimée. 2020b. *Descriptive and methodological issues in Kiranti grammar(s)*. Paris, France: Université de Paris Habilitation à diriger des recherches.
- Lahaussais, Aimée. forthcoming. Glossing practices in 1850-1911 descriptions of languages with complex verbal morphology. In Franck Cinato, Aimée Lahaussais & John Whitman (eds.), *Glossing Practice: Comparative Perspectives*. Lexington.
- Lehmann, Christian. 1982. Directions for interlinear morphemic translations. *Folia Linguistica* 16. 199-224.
- Lehmann, Christian. 2004. Interlinear morphemic glossing. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*. (Handbücher Der Sprach- Und Kommunikationswissenschaft 17/2), 1834-1857. Berlin & New York: W. de Gruyter.
- Majeed, Javed. 2019a. *Colonialism and Knowledge in Grierson's Linguistic Survey of India*. London and New York: Routledge.
- Majeed, Javed. 2019b. *Nation and Region in Grierson's Linguistic Survey of India*. London and New York: Routledge.
- Majeed, Javed. 2019c. The politics of Grierson's Linguistic Survey of India. In. Edinburgh: University of Edinburgh.
- Masica, Colin. 1991. *The Indo-Aryan Languages* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Matisoff, James. 1990. On megalocomparison. *Language* 66(1). 106-120.
- Michailovsky, Boyd. 2017. Kiranti languages. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan Languages* (Routledge Language Family Series 3), 646-679. 2nd edn. Oxford & New York: Routledge.
- Newmark, Peter. 1988. *A textbook of translation*. Hertfordshire: Prentice Hall International.

- Ohannessian, Sirarpi & Gilbert Ansre. 1975. Some reflections on the use of sociolinguistic surveys. In Sirarpi Ohannessian, Charles A. Ferguson & Edgar C. Polomé (eds.), *Language Surveys in Developing Nations: Papers and Reports on Sociolinguistic Surveys*, 51–69. Arlington: Center for Applied Linguistics.
- Solleveld, Floris. 2019. Language, People, and Maps: The Ethnolinguistics of George Grierson and Franz Boas. *History of Humanities* 4(2). 461–471.
- Teeuwen, Mariken & Irene van Renswoude (eds.). 2017. *The Annotated Book in the Early Middle Ages: Practices of Reading and Writing* (Utrecht Studies in Medieval Literacy 38). Turnhout (Belgium): Brepols.
- Unsigned. 1928. Sir George Grierson and the “Linguistic Survey of India.” *Journal of the Royal Asiatic Society of Great Britain and Ireland* 3. 711–718.
- Waterhouse, David (ed.). 2004. *The Origins of Himalayan Studies: Brian Houghton Hodgson in Nepal and Darjeeling 1820-1858*. London and New York: RoutledgeCurzon.
- Whitman, John. 2011. The ubiquity of the gloss. *Scripta* 3. 95–121.
- Whitman, John. 2014. Présentation du projet Reading Classical texts in the Vernacular. *Les dossiers de HEL : [supplément électronique à la revue Histoire Epistémologie Langage]* 7. 1–21.
- Whitman, John, Miyoung Oh, Jinho Park, Valerio Alberizzi, Masayuki Tsukimoto, Teiji Kosukegawa & Tomokazu Takada. 2010. Toward an International Vocabulary for Research on Vernacular Readings of Chinese Texts. *Scripta* 2. 61–83.
- Wieland, Gernot. 1983. *The Latin Glosses on Arator and Prudentius in Cambridge University Library MS Gg. 5.35* (Studies and Texts 61). Toronto: Pontifical Institute of Medieval Studies.
- Wolfenden, Stuart. 1929. *Outlines of Tibeto-Burman Linguistic Morphology, with special reference to the Prefixes, Infixes and Suffixes of Classical Tibetan and the Languages of the Kachin, Bodo, Naga, Kuki-Chin and Burma Groups*. London: Royal Asiatic Society.
- Wolfenden, Stuart. 1933. A specimen of the Sāngpāng Dialect. *Acta Orientalia* 12. 71–79.
- Wolfenden, Stuart. 1934. A Specimen of the Kūlung Dialect. *Acta Orientalia* 13. 35–43.
- Wolfenden, Stuart. 1935. A Specimen of the Thūlung Dialect. *Journal of the Royal Asiatic Society* 67(4). 629–653.

SUMMARY

In this article, I explore glossing practices in the period surrounding the publication of the *Linguistic Survey of India* (henceforth LSI), the large-scale survey of languages spoken on the Indian subcontinent at the turn of the 20th century, under the stewardship of George Abraham Grierson.

After a brief discussion of the reasons that the LSI constitutes a useful corpus for studying glossing practices, I provide a detailed examination of the glossing practices used in the text specimens which accompany language descriptions in the LSI. I then contrast these practices with glossing in materials produced both prior to and subsequent to the LSI, in order to place the glossing practices established by Grierson within a historical context, thereby contributing a description of one step in the history of glossing of descriptive linguistic materials.

RÉSUMÉ

Cet article propose d'étudier le glosage de données textuelles tel que pratiqué dans le *Linguistic Study of India* (ci-après LSI), le grand projet de recensement et de description des langues du sous-continent indien mené au début du 20ème siècle par

George Abraham Grierson. Après une exposition des raisons pour lesquelles le LSI constitue un important corpus pour l'étude du glosage à cette période, je détaille les pratiques de glosage qui caractérisent les spécimens textuels qui y accompagnent les descriptions linguistiques. Dans un deuxième temps, ces pratiques sont contrastées avec celles des périodes précédant et suivant la publication du LSI, afin de les placer dans un cadre historique. L'objectif de l'article est de contribuer à l'historiographie du glosage interlinéaire la description détaillée de pratiques correspondant à une phase de cette histoire.

ZUSAMMENFASSUNG

Author's address:

Aimée Lahaussais
Histoire des théories linguistiques (UMR 7597), C.N.R.S. / Université de Paris
Case 7034, 5 rue Thomas Mann
FR-75013 PARIS
F r a n c e
e-mail: aimee.lahaussais@cnrs.fr