



The Translation Tier in Interlinear Glossed Text: Changing Practices in the Description of Endangered Languages

Aimée Lahaussais

► To cite this version:

Aimée Lahaussais. The Translation Tier in Interlinear Glossed Text: Changing Practices in the Description of Endangered Languages. Patricia Phillips-Batoma; Florence Zhang. Translation as Innovation: Bridging the Sciences and the Humanities, Dalkey Archive Press, pp.261-278, 2016, 9781564784094. hal-01895439

HAL Id: hal-01895439

<https://hal.science/hal-01895439>

Submitted on 15 Oct 2018

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.

à paraître dans Phillips-Batoma, Patricia, and Florence Zhang, eds. 2016. *Translation as Innovation: Bridging the Sciences and the Humanities*. Normal, IL: Dalkey Archive Press, pp. 261-278.

The Translation Tier in Interlinear Glossed Text: Changing Practices in the Description of Endangered Languages

Aimée Lahaussais

Abstract

As a field linguist working on several exclusively oral endangered languages, I am constantly faced with the challenges of putting my data into written form. There are models for data annotation, the most common being to render it as Interlinear Glossed Text, a system generally involving three tiers: a transcription tier, a glossing tier, and a translation tier. There are standards associated with the first two of these tiers (International Phonetic Alphabet, Leipzig Glossing Rules), but practical recommendations are scarce for the translation tier. While manuals and general practice suggest that this tier provide a "free translation", recently published data in oral archives suggests that translations are very frequently literal. In this paper I examine translations found in an oral archive. Using examples from the archive, I propose a typology of some of the features of these literal translations, as well as a hypothesis which would explain why translation tiers are becoming more literal, specifically in the case of digital archives.

Résumé

En tant que linguiste de terrain travaillant sur des langues en danger exclusivement orales, je suis régulièrement confrontée aux difficultés de mise par écrit de mes données. Il existe des modèles d'annotation, le système le plus commun étant de générer du Interlinear Glossed Text qui comprend trois niveaux : transcription, glose, et traduction. Des normes sont associées aux deux premiers niveaux, notamment l'alphabet phonétique international et les Leipzig Glossing Rules, mais il n'existe aucune recommandation pratique en ce qui concerne le niveau traduction. Les manuels de terrain mettent en avant, quand ils traitent du sujet, le concept de 'traduction libre' mais les traductions trouvées dans des archives orales numériques suggèrent plutôt que les traductions sont souvent littérales. Dans cet article j'examine des traductions associées à des données de terrain recueillies dans une archive orale, et je propose une typologie des traits non-littéraux dans ces traductions, ainsi qu'une hypothèse touchant aux raisons pour lesquelles ces traductions sont de moins en moins libres.

Key words: language documentation, endangered languages, interlinearization, oral archives ;

documentation linguistique, langues en danger, interlinéarisation, archives orales

Introduction

In documenting endangered languages in the course of fieldwork, the standard practice for presenting the data in writing is called interlinearization. The EMELD (Electronic Metastructure for Endangered Languages Data) website provides the following description of Interlinear Glossed Text (henceforth IGT): "The model typically used in IGT incorporates three tiers: a transcribed (either phonetically or orthographically) text, a gloss and a free translation." An example of IGT, to which I have added tier labels for clarity, is presented below:

(1) ¹	k ^h uly p ^h eri k ^h rem-s-ɖa ʔe	[transcription tier]
	earth again cover-mm-3sg.pst hs	[glossing tier]
	'The Earth covered (itself) up again.'	[translation tier]

As the above example illustrates, the transcription tier involves the rendering of oral source material into a written form. Depending on the linguistic tradition and country in which the documentation is taking place, this transcription tier can be rendered in any one among a number of different systems, the most common being phonetic (using the international phonetic alphabet), phonological (also based on the international phonetic alphabet but taking into account the phonological rules of the language), or orthographic, if the language has either its own written tradition or a tradition of using the writing system for the dominant language of the region.

The glossing tier involves the identification of each of the elements found in the transcription tier, typically on a morpheme by morpheme basis. Lexical items are translated and grammatical morphemes are signalled by abbreviations. The glossing tier, like the transcription tier, is governed by well-known standards, making it possible for readers encountering the data to interpret the content of the glossing tier.² A system which has become very widely used is the Leipzig Glossing Rules, a set of recommendations on how to align and present the tier and a suggested list of abbreviations for very common grammatical concepts. The adoption of such conventions for the glossing tier makes the data easier to compare cross-linguistically, as data from different languages is encoded in a similar way.

Thus both the transcription and glossing tiers are governed by standards, making the annotation of these tiers fairly straightforward (but only, of course, once an analysis of the language has led to an understanding of how to transcribe and gloss it). This is not, however, the case when it comes to the translation tier. It is generally acknowledged that the translation accompanying a glossed example should be a "free translation," which, according to Bow et al. (2003: 2), means a translation "in which more emphasis is given to the overall meaning of the

text than to the exact wording". Yet a cursory look at language description materials suggests that in many cases, the translation tier is not free but instead reflects the structure of the source language.

Note for example the translation tier in example (2), taken from Shibatani's (1990:85) description of Ainu:

(2)	Kamuy kat casi casi-upsor a-i-o-resu
	god build castle castle-inside PASS-1sg/O-APPL-raise
	'The god-built mountain castle, inside the mountain castle, I was raised.'

Bow et al. (2003:5) note, referring to this segment of text: "The free translation doesn't always read fluently, which may be an attempt to reproduce the phrase structure of the source text in the English translation." This situation is of course not limited to the sample presented above, and one frequently encounters examples of translation tiers which do not read as free but instead seem to be written in such a way as to convey grammatical information about the source language, particularly in textual³ material presented in isolation. The reason for this is that in text collections (appendices to grammars, volumes devoted to texts, oral archives), interlinearized material is disconnected from grammatical explanations that help the reader make sense of them. Mosel (2006:47), in discussing the format of grammars in an article on grammaticography, points out how collections of texts in grammars are often marginalized by their physical situation: "Corresponding to the back matter of dictionaries, the text collection does not fit into the organization of the content of the main part; it is an appendix which helps the reader understand how the language is used in various contexts." As a result of the commonly adopted structure for descriptive grammars, which relegates texts to an appendix, the glossing and translation tiers of these texts take on the role of explicating the material, not only semantically but also morphosyntactically.

This observation has led me to formulate the following hypothesis about the translation tier: the presentation of linguistic material on endangered languages in the form of interlinear glossed text in collections of texts (removed from grammatical description) makes it necessary for linguists to use the translation tier to explicate morphosyntactic features of the language. The result is a tendency towards a literal⁴ translation in the translation tier.

In this paper, I explore the validity of the proposed hypothesis by looking at some of the ways the translation tier is used in IGT. I first consider what explicit instructions are provided for field linguists, in the absence of recommendations and standards such as those found for the transcription and glossing tiers, by looking at field manuals. I next consider the uses which are made of the translation tier, by both the resource producer and resource user, to examine how these uses influence the choices made in generating a translation tier. I then proceed with an investigation of a digital oral archive for endangered language materials, the Pangloss Collection. I look at the range and types of texts found in the archive, and set up, for those which are in three-tiered interlinear form, a typology of issues found in the translation tier making them examples of literal rather than free translation.

The translation tier in field manuals

Mosel (2006: 50) points out that "[i]n spite of its long tradition and pervasive application, interlinear glossing has neither received attention in (meta-) grammaticography beyond Lehmann's (1982) article and the guidelines by editors, nor is it usually discussed in text books on grammatical analysis. The only exception I am aware of is Haspelmath (2002)."

It is indeed true that the practice of interlinearization is rarely discussed in books or manuals on morphological analysis, field methods or typology, all fields for which it is highly relevant to have an understanding of how to present language data. The following discussion is intended to show that the advice presently found in the literature is very limited.

Haspelmath (2002: 4, footnote 2) briefly lays out the basic principles of the translation tier ("idiomatic translation") and the glossing tier ("literal translation"): "For each example sentence from an unfamiliar language, not only an idiomatic translation is provided, but also a literal ('morpheme-by-morpheme') translation." An appendix to Haspelmath's chapter on Basic Concepts (2002: 34) contains a description of some of the conventions usually applied to the glossing tier (one-to-one correspondence between elements in transcription and glossing tiers, use of abbreviations for grammatical categories, use of hyphens and periods, alignment), but does not explicitly state what is to be done with the translation tier.

Payne's popular field methods manual, *Describing Morphosyntax* (1997), makes absolutely no mention of what kind of translation to associate with IGT. An abundance of examples drawn from a variety of grammars provide (passive) pointers for readers, but the book does not address translation in descriptive situations. The omission is understandable considering the goal of the book is to present a typology of morphosyntactic features found in the world's languages, and propose a system which can be used to describe them. This is regrettable, however, because a field manual is a perfect venue for introducing students to techniques related to the physical presentation of data they aim to collect.

In *Studying and Describing Unwritten Languages* (1992), (a translation of the 1976 second French edition) Bouquiaux and Thomas address the production of IGT. The advice is quite practical, providing examples of how to physically align transcription and translation on the page (63), and emphasizing the importance of getting glossing and translation done while the speaker is still available (58). Nonetheless, the term "free translation" is used with no clear explanation of what this entails

Nida (1964: 159) is one of the rare linguists involved in endangered language description to clearly lay out the translation options, distinguishing between formal and dynamic equivalence, and connecting the former to glossing:

Formal equivalence focuses attention on the message itself, in both form and content. [...] The type of translation which most completely typifies this structural equivalence might be called a 'gloss translation', in which the translator attempts to reproduce as literally and meaningfully as possible the form and content of the original. [...] A gloss translation of this type is designed to allow the reader to identify himself as fully as possible with a person in the source-language context. [...] In contrast, a translation which attempts to produce a dynamic rather than a formal equivalence is based upon "the principle of equivalent effect" (Rieu and Phillips, 1954). In such a translation, there is less emphasis on matching the receptor-language message with the source-language message, and more emphasis on the dynamic relationship [...], that the relationship between receptor and message should be substantially the same as that which existed between the original receptors and the message. A translation of dynamic equivalence aims at complete naturalness of expression, and tries to relate the receptor to modes of behavior relevant within the context of the receptor's own culture.

But these types of translation equivalence represent the extremes, and Nida (1964: 160) points out that "[b]etween the two poles of translating (i.e. between strict formal equivalence and complete dynamic equivalence) there are a number of intervening grades, representing various acceptable standards of literary translating. During the past fifty years, however, there has been a marked shift of emphasis from the formal to the dynamic dimension." While the shift from formal to dynamic may indeed be what is found in literary translation, this does not appear to hold true for translation for the purposes of linguistic research.

Use of the translation tier in practice

The two main categories of linguists interacting with interlinearized materials are, unsurprisingly, resource users and resource creators, and in each case, the relationship with the translation tier is somewhat different. The resource user can be somewhat difficult to identify, particularly in the context of linguistic documentation projects, which aim to record a language as fully as possible with the idea that materials be used by specialists from fields such as anthropology, ethnobotany, religious studies. But in the context of linguistic research, we will consider that the most likely user will be a typologist or an areal specialist, in both cases researchers who probably do not know the language in question. If we consider the following example from that perspective, it gives us a sense of the use of the translation tier in IGT.

(3)	Limbu ⁵ (Nepal): "On the Road"
	pəhile uhile inga cu:kt-aŋ-elle akkhen bahrə tehrə bərsə
	first before 1SG be.small-1SG.SO.PA-SUB how.many 12 13 year
	yor-aŋ khəmbhelle kə a-pa sintola bəgan lept-ε
	suffice.S2-1SG.SO.PA then TOP 1SG-father tangerine garden contract-PA
	<i>Before when I was little — I was about 12 or 13 — my father took a contract on some</i>
	<i>orange groves</i>

It is quite clear that a resource user with no knowledge of either Limbu or the contact language Nepali will not be able to make much of the example if only the transcription and glossing tiers are provided. In this context, the translation tier serves three important purposes: 1) It conveys the narrative content of the linguistic material in question; 2) It serves as a key to how the glossing tier works 3) It can help the resource user identify whether a certain construction is present. This last reason is of particular significance to linguists using the data for comparative purposes, as they will search the material for the presence of particular constructions, and this is precisely where the nature of the translation becomes relevant: in order for the constructions in a

linguistic resource to be identifiable, the translation tier must reflect the characteristic features of the source language and therefore be literal rather than free.

Example (4) below from my own work on Khaling (Nepal)⁶ will illustrate this point further. Generating the top two tiers of the IGT is, while challenging in its own way, not particularly problematic in any theoretical sense, once an analysis of the language has been carried out: the standards associated with these tiers make creating them a fairly objective matter. Generating the translation tier for the material is considerably more difficult, because of the multiple possibilities. In my work, I have four different translation tiers associated with the sentence below, each exemplifying different features of the language and produced according to what I needed to highlight for a given use of the data. The parentheses after each translation make explicit the reasons behind the choice made in the wording and word order of the sentence.

(4)	sô:-ʔε	mat-tε-na	kumîn-ʔε	mat-tε-na	ʔip-dək-tε-m
	hunger-INSTR have.to-PST-after thirst-INSTR have.to-PST sleep-AUX-PST-NMLZ				
	a) 'He was hungry and thirsty and had fallen asleep.' ("free") b) 'After suffering from hunger and thirst, he fell asleep.' (highlighting instrumental case with 'hunger' and 'thirst') c) 'Obliged by hunger, obliged by thirst, he fell asleep.' (highlighting use of verb 'have to' in this construction) d) 'The fact is, he fell asleep from hunger and thirst.' (highlighting nominalization of entire sentence)				

Each of these translations was generated with a different purpose in mind: the most neutral translation tier is that in a), which I have labelled "free". Its sole purpose is to translate the narrative content into something close to natural English. This can be of use in comparing various forms of the same narrative, and identifying whether certain passages are found in all versions of a story. It can also be useful in order to identify the presence of certain lexical items, and to see how they are used in a natural context. It is less useful, linguistically speaking, than the three other translations I have used in various contexts for the same sentence, and which have specific

goals, because they attempt to highlight unusual features of the language, all of which are found in the same sentence. Thus depending on the use I might make of a piece of IGT (whether, for example, the sentence is used to exemplify sentential nominalization, as in d), or the instrumental, as in b)), I have altered the translation to better suit my purposes. Within the context of papers or chapters describing specific aspects of a language system, the decision as to what kind of translation to use in the translation tier is not a particularly difficult issue to resolve, but within the context of text collections divorced from grammatical explanations, the choice of how best to use the translation tier—and what should be highlighted for any sentence-- becomes considerably more difficult.

Investigation of the translation tier in a digital archive of endangered languages

In order to test my hypothesis with respect to why translation tiers in the descriptions of endangered languages tend to be literal, I have examined materials found in the Pangloss Collection. The following is the description of the collection found on the Web site for this archive⁷: "The Pangloss Collection provides free access to documents of connected, spontaneous speech, mostly in 'rare' or endangered languages, recorded in their cultural context and transcribed in consultation with native speakers." The materials in the collection are surprisingly varied, not only in terms of geographical coverage, but also in terms of time span and interlinearization practices. The earliest texts date from the 1960's, and new texts are being added regularly. Over this time span, practices in interlinearization appear to have changed, and these changes affect the use that is made of the translation tier. The IGT is either made up of a word for word translation in the glossing tier, essentially made up of lexical items, and accompanied by a free translation; or, in IGT prepared by younger researchers, the glossing tier uses technical morphosyntactic glosses and is accompanied by a more literal translation. Even in situations where the translation in the second type looks fairly free, there are traces of literalness in the form

of the use of parentheses to add grammatical material, or alternative translations proposed. My explanation for this trend is that technical morphosyntactic glosses, which are found in more recent texts and presumably represent efforts to standardize abbreviations and glossing practices, render the glossing tier less accessible to the resource user and therefore tend to be accompanied by a more explicative rendering of the translation tier.

•Typology of features of literal translations in the Pangloss Collection

Within the Pangloss collection's IGT texts with more literal translation tiers, I have identified a typology of features which result in the translations being classified as literal. These are exemplified in turn.

○Word order issues

In sentences whose translations have word order issues, the translation tier reflects the word order of the source language. This is a very common issue found in the translation tier.

(5)	Langi8 (Tanzania): “Story”				
	ma	dʒira	ŋɡɔ	kara	katʃiɦɪ kwɪrɪrɛ
	ma				
	then	dp10-dem.d	np10-clothes	dp12-dem.d	np12-bird vp17-acc-darken-perf then
	kakamɔwɛkɪra		dʒira	ŋɡɔ	
	vp12-narr-obj1-close-appl-appli-sfx	dp10-dem.d	np10-clothes		
	‘Then those clothes, the little bird landed and covered her with them.’				

(6)	Thulung9 (Nepal): Kakcilip				
	hum-saɖ-dytsi-m		pətsʰi	ritsi	lat-tsi-m
	pətsʰi	murtɕip			
	spread-ASP-3DU/3SG.PST.NOM	after	sour.fruit	see-3DU/3SG.PST-NOM	after 3DU
	pʰeri	ri-tɕip	dzyrpa-ku	mina-ka	swaɬ-ka

	gərdak ^h eri muɾtsip p ^h eri dʒ ^h uk-
	again sibling-DU bitter-GEN [filler word]-INSTR taste-INSTR doing 3DU again land- tsɪ ʔe 3DU.PST HS
	‘After she spread out the fruit, and after they saw it, the two, with the bitter taste, they landed again.’

While a free translation would have rearranged the word order to reflect that of the target language, in (5) and (6) above, the translation reflects the order of segments found in the source language.

○Tense concordance issues

Some texts have translations which show inconsistency in the tenses used within a same sentence, reflecting the fact that the use of tenses is different in the source language; the result of conveying this tense divergence in the translation tier is a literal translation.

(7)	Araki ¹⁰ (Vanuatu): “The rat, the hawk and the octopus”
	Hadiv mo dogo mo de mo m̄arahu
	rat 3:R feel 3:R say 3:R fear
	mo de "Ale! Codo o sna, codo o sna o vici !"
	3:R say alright okay 2S:I come okay 2S:I come 2S:I embark
	Mo lig-i-a mo vari-a mo v̄a
	3:R embark-TR-3S 3:R take-3S 3:R go
	‘The rat felt seized with fear and exclaimed "Okay! It's alright, come in, you may come aboard" takes them on board, carries them along...’

(8)	Nepali, Accham dialect ¹¹ (Nepal): “Tiger and Jackal” ¹²
	bago rə syalo satə bhəyapəchi nə əbə ... syal— bagəi khana— syaləkənə bagəi khanəinə rəechə. mama rəechə, syaləko. syal.
	‘When they were together... the tiger doesn’t eat the jackal, he’s his maternal uncle — the jackal’s — the jackal.’

(9)	Na-našu ¹³ (Italy): « Le voyage à Palata »
	Alor Tonin sa vračaša na mala
	Thus Tonin.NOM.SG.M REFL return.IPFV-IPRF-3SG ART.INDF- ACC.SG.N little-ACC.SG.N
	naza, vidim ka biša Profesor.
	back see-PRS-1SG COMP be-IPRF-SG professor-NOM.SG.M
	‘So Tonino drove back a bit, I see (=saw) that the professor was there.’

While we expect a free translation to smoothe out issues related to tense concordance, the translations in (7), (8) and (9) above reflect the actual tenses used, and give us a better sense of the oral nature of the data.

○Morphosyntactic issues

In texts identified as having morphosyntactic issues, there are a number of translation choices that do not read as free, and as a result these reveal the structure of the source language.

(10)	Romani ¹⁴ (Greece): “The louse and the Rom”
	koleski kor tʃinen kor kirisni lafen
	that.one.DAT neck cut.3PL neck his leave.3PL
	‘They cut this one's neck, that one's they leave.’

In the above example, it seems that the choice of translation into English serves to highlight the expression of possession in Romani, something that cannot be rendered with a natural English translation.

(11)	Araki (Vanuatu): “The rat, the hawk and the octopus”
	Hede na pa p̄isu cudug-i-a aka-ca!
	today 1S:I SEQ prick pierce-TR-3S canoe-1IN:P
	‘I am going to prick and make a hole in your boat’

The Araki example above highlights, through an unusual translation, the fact that the verb in the sentence is a serial verb construction, something which would not come through if a more natural English translation were proposed.

(12)	Nelemwa15 (New Caledonia): “Kaavo Dela”
	hla ma aaxiik kaari-n xe yaara-n i Hiixe,
	3PL COMIT un(e) soeur cadette-POSS.3SG THEM nom-POSS.3SG REL Hiixe
	na aaxiik khia-hli xe bwaaxolat,
	CONTR un(e) soeur ainée-POSS.3DU THEM infirme
	khabwe kia kua-n bu u fagaut hada xa shit.
	c'est-à-dire il n'y a pas jambe-POSS.3SG EXPL RESTR corps seul et aussi bras
	‘Elles et sa soeur cadette du nom de Hiixe mais leur soeur ainée est infirme, elle n'a pas de jambes mais seulement un corps et un bras.’

In the Nelemwa sentence, it appears that the source sentence is able to express existential predication not requiring copula, something which cannot be rendered naturally in French.

- Alternative explanations for literal translations in corpus

I believe that the choices made in rendering the source language sentences into the target languages in all the cases above were conscious, and that the intent of the linguist was to use the translation tier to highlight typologically significant features of the language. But we must also raise the possibility that the literalness of the translation tiers may have been unintentional. One fact about descriptive linguistics is that linguists carrying out language documentation projects are inherently involved in translation, but without necessarily having received training in translation. It is thus possible, although I shall discuss why I do not believe this to be the case, that the overly literal translations found in some descriptive work are not intentional.

There are however clues in the translation tier as to the intentionality of translation choices. The use of parentheses to add material to the translation suggests that the linguist is aware that a completely literal translation is not always possible to convey all the data contained in the target language. The addition of linguistic material with parentheses (as in 13) and brackets (14), or alternative translations (15) are proof that the linguist in question is perfectly aware of the fact that a literal translation alone will not convey all the information needed by the resource user, and yet is still not willing to provide a free translation in its stead.

(13)	Langi (Tanzania): "The Hyena and the lion"
	εε va-dʒokolo va-a-ni ni-mo-sim-ir-ε
	Yes np2-grand.child dp2-det-poss1sg pv1sg-obj1-tell-appl-sfx
	‘Hey, grandchildren, let me tell you (a story).’

(14)	Hayu16 (Nepal): "The merchant"
	ã:ki-thik-mo na siŋtoŋ-ha ithara beŋ-ta noktshuŋ po-yi
	1PL.IN-same-GEN INTENS man-ERG this.much wide-PASS.PART ear do-ACT.PART
	‘[Go where there are] men like us but with ears this big – ‘

(15)	Na17 (China): "Sister: the sister's wedding"
	tʂʰuʈniʈ gyʈ tsuʈ myʈ
	like.this happen °rep °affirm
	‘c’est comme ça (que ça s’est passé), à ce qu’on raconte: / on dit que c’est ainsi’

The translation tier in (15) goes further than (13) and (14), actually providing both a literal and a free translation, suggesting a clear awareness of the need to both explain and translate the source structure.

Conclusion

An examination of the translation tier of IGT from the Pangloss Collection showed evidence of what I consider to be intentionally literal translations in the translation tier. This translation choice serves to highlight typologically unusual or interesting features of the language in question, by having the structure of the language reflected in the wording of the translation tier, even if the translation as a result deviates from what is considered “free.” My argument is that the context of an oral archive, where language materials are presented independently of a reference grammar, where explanations for characteristic features might be found, increases the need for a mechanism to signal such linguistic elements. Such signaling is achieved using a style of translation which is literal as opposed to free, even though the manuals which address translation all seem to suggest that a free translation is required. I also suggest that recent trends in morphosyntactic glossing, while rendering data more informative in terms of linguistic analysis, in fact make it difficult to scan the material in search of specific constructions, because there are inevitably additional, non-standard abbreviations that must be used and because of the difficulty in accessing glosses for a language one is not familiar with. In such a context, a literal translation in the translation tier facilitates the reading of the glossing tier and the identification of material of interest.

One question of interest is whether it might be useful to set up standards or recommendations regarding the translation tier of IGT. These recommendations might quite simply involve an additional tier of material, maintaining the transcription tier, a glossing tier involving all morphemes and grammatical information, and then two translation tiers, one where the words are rendered literally (to enable users to scan for particular constructions) and another involving a free translation, which can be used to look for narrative and semantic content. Recent

work developing a system of Advanced Glossing (Drude and Lieb 2000, Drude 2002) points to the fact that some researchers are aware of the need for additional linguistic information in interlinearized texts, but this additional encoding comes at the cost of time and efficiency.

References

- Adamou, Evangelia. Romani corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11
- Bouquiaux, Luc, and Jacqueline Thomas. *Studying and Describing Unwritten Languages*. Dallas: SIL International, 1992.
- Bow, C., B. Hughes, and S. Bird. 'Towards a general model of interlinear text'. In *Proceedings of EMELD Workshop*:1-47, 2003. <http://emeld.org/workshop/2003/bowbadenbird-paper.pdf>
- Breu, Walter. Na-našu corpus, Programme franco-allemand Euroslav 2010 : ANR(09-FASHS-025) - DFG(BR 1228/4-1), <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11
- Bril, Isabelle. Nelemwa corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11
- Dunham, Margaret. Langi corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11
- Drude, Sebastian. 'Advanced glossing--A language documentation format and its implementation with shoebox'. In *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics (Las Palmas 26-27 May 2002)*, Peter Austin, Helen Dry and Peter Wittenburg (eds.), 2002.
- Drude, Sebastian, and Hans-Heinrich Lieb. 'Advanced Glossing: A Language Documentation Format.' Working paper, 2000.
- EMELD website, <http://emeld.org/school/classroom/text/levels.html>
- François, Alexandre. Araki corpus, CNRS/Fédération de recherche 2559: Typologie et universaux linguistiques, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11
- Haspelmath, Martin. *Understanding Morphology*. London: Arnold, 2002.
- Leipzig Glossing rules, http://www.eva.mpg.de/lingua/tools-at-lingboard/glossing_rules.php
- Lahaussais, Aimée. Thulung Rai corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11

Lecomte-Tilouine, Marie; Michailovsky, Boyd. Nepali corpus, Agence National de la Recherche ANR-06-CORP-030-01, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11

Mazaudon, Martine; Michailovsky, Boyd. Hayu corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11

Michaud, Alexis. Na corpus, <http://lacito.vjf.cnrs.fr/pangloss/>. Accessed 2014-07-11

Mosel, Ulrike 2006. 'Grammaticography', In Ameka, Felix, Dench, Alan, and Evans, Nicholas (eds.) *Catching Language: The Standing Challenge of Grammar Writing*, pp. 41-68.

Nida, Eugene. *Towards a science of translating*. Leiden: Brill, 1964.

Pangloss Collection, <http://lacito.vjf.cnrs.fr/archivage/index.htm>

Payne, Thomas. *Describing morphosyntax*. Cambridge: Cambridge University Press, 1997.

Shibatani, Masayoshi. *The languages of Japan*. Cambridge: Cambridge University Press, 1990.

Notes

¹ This example is from a story called Miyapma in Thulung (Nepal), http://lacito.vjf.cnrs.fr/archivage/languages/Thulung_Rai_en.htm

² As will be shown in a later section, it can in fact be quite challenging to make sense of the glossing tier if one is not familiar with the language. Proposals for a more informative interlinearization system exist, such as Advanced Glossing (see Drude 2002, Drude and Lieb 2000.)

³ In using the term "text" I adopt the same definition as proposed in Payne (1997: 366): "any sample of language that accomplishes a non-hypothetical communicative task" and contrasting with elicitation data, which is hypothetical and non-spontaneous. For oral languages with no written tradition, spontaneous narrative is considered "text".

⁴ There is no judgment value whatsoever associated with the term 'literal translation': it simply refers to a translation which is more closely aligned with the source language rather than the usual practices of the target language.

⁵ http://lacito.vjf.cnrs.fr/archivage/languages/Limbu_en.htm

⁶ Work on Khaling is being carried out in collaboration with Guillaume Jacques via the ANR-funded HimalCo project.

⁷ http://lacito.vjf.cnrs.fr/archivage/index_en.htm

⁸ http://lacito.vjf.cnrs.fr/pangloss/languages/Langi_en.htm

⁹ http://lacito.vjf.cnrs.fr/pangloss/languages/Thulung_Rai_en.htm

¹⁰ http://lacito.vjf.cnrs.fr/pangloss/languages/Araki_en.htm

¹¹ http://lacito.vjf.cnrs.fr/pangloss/tools/list_rsc.php?lg=Nepali&aff=nepali

¹² Note that this resource is not glossed in the archive.

¹³ http://lacito.vjf.cnrs.fr/pangloss/tools/list_rsc_lieu_en.php?lg=Na-na%C5%A1u&aff=Na-na%C5%A1u&lieu=Montemitro

¹⁴ http://lacito.vjf.cnrs.fr/pangloss/languages/Romani_en.htm

¹⁵ http://lacito.vjf.cnrs.fr/pangloss/tools/list_rsc.php?lg=Nelemwa&aff=n%C3%AAI%C3%AAmwa

¹⁶ http://lacito.vjf.cnrs.fr/archivage/tools/list_rsc_en.php?lg=Hayu&aff=Hayu

¹⁷ http://lacito.vjf.cnrs.fr/pangloss/tools/list_rsc_en.php?lg=Na&aff=Na