



HAL
open science

Descriptive and methodological issues in Kiranti grammar(s)

Aimée Lahaussais

► **To cite this version:**

Aimée Lahaussais. Descriptive and methodological issues in Kiranti grammar(s). Humanities and Social Sciences. Université de Paris / Université Paris Diderot (Paris 7), 2020. tel-03030562

HAL Id: tel-03030562

<https://shs.hal.science/tel-03030562>

Submitted on 30 Nov 2020

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.

Université de Paris

Ecole Doctorale "Sciences du langage", ED 622

MÉMOIRE DÉPOSÉ EN VUE DE L'HABILITATION À DIRIGER DES RECHERCHES

**DESCRIPTIVE AND METHODOLOGICAL ISSUES
IN KIRANTI GRAMMAR(S)**

AIMÉE LAHAUSSOIS

JURY

M. Olivier Bonami, PU, Université de Paris (garant)

Mme. Hilary Chappell, DE, EHESS (examineur)

M. George van Driem, Professor, Université de Berne (rapporteur)

M. Enrique Palancar, DR, CNRS-Sedyl (examineur)

M. Nick Riemer, Professor, University of Sydney (examineur)

Mme. Françoise Rose, DR, CNRS-DDL (présidente)

Mme. Margaret Thomas, Professor, Boston College (rapporteur)

20 novembre 2020

TABLE OF CONTENTS

DOCUMENT DE SYNTHÈSE	4
Acknowledgments	5
Abbreviations	6
Lists of figures and tables	7
1 INTRODUCTION	9
1.1 Personal context	9
1.2 Organization of this document	11
1.3 A note on the Thulung corpus	13
2 DESCRIPTION OF THE THULUNG VERB COMPLEX	15
2.1 Introduction to Kiranti and Thulung	15
2.1.1 Classification	15
2.1.2 Grammar sketch of Thulung	19
2.1.2.1 Phonology	20
2.1.2.1.1 Phoneme inventory	20
2.1.2.1.2 Vowel length	21
2.1.2.1.3 Syllable structure	21
2.1.2.2 Morphosyntax	21
2.1.2.2.1 Personal and possessive pronouns	21
2.1.2.2.2 Number	23
2.1.2.2.3 Demonstratives and deixis	24
2.1.2.2.4 Numerals	24
2.1.2.2.5 Adjectives and adverbs	25
2.1.2.2.6 Case	26
2.1.2.2.7 Nominalization and relativization	31
2.1.2.2.8 Discourse markers	33
2.2 The Thulung verbal template	39
2.2.1 Prefixal slot	42
2.2.1.1 Negation	42
2.2.1.2 Ideophones	43
2.2.2 Root	48
2.2.2.1 Verb classes and stem alternation patterns	48
2.2.2.2 Contrasting paradigms within verb classes	57
2.2.2.3 The influence of postfinals -t and -s	60

2.2.3 Suffixal slots	62
2.2.3.1 Suffixal slot 1: partial person indexes	62
2.2.3.2 Suffixal slot 2: derivational markers	63
2.2.3.2.1 Multiple exponence	66
2.2.3.2.2 Allomorphy of derivational formatives	68
2.2.3.2.3 Associated Motion	72
2.2.3.2.4 Valence changes	79
2.2.3.2.5 Aktionsart and aspect	82
2.2.3.3 Suffixal slot 3: indexation markers	91
2.2.3.3.1 Tense marking	94
2.2.3.3.2 Cell syncretism	96
2.2.3.3.3 Allomorphy	96
2.2.3.3.4 Inflectional gaps	101
2.2.3.3.5 Evidence of referential hierarchies within the paradigm?	102
2.2.3.4 Suffixal slot 4: mood markers	108
2.2.3.4.1 Optative mood marker	109
2.2.3.4.2 Irrealis marker	113
3 EPISTEMOLOGICAL AND HISTORICAL ISSUES SURROUNDING DESCRIPTIVE LINGUISTICS	120
3.1 Grammaticographical models	120
3.1.1 Missing categories: Interjections and ideophones	121
3.1.2 Verbal paradigms	125
3.1.3 Organizational schemes in descriptive grammars	134
3.2 Annotation practices	142
3.2.1 Generalities regarding the study of annotation	142
3.2.2 Annotation studies: methodology and materials	144
3.3 Methodological tools	157
3.3.1 Corpora: Kiranti comparable corpus	157
3.3.2 Questionnaires	168
3.3.3 Khaling verb dictionary	171
4 FUTURE DIRECTIONS	175
Bibliography	178

DOCUMENT DE SYNTHÈSE

Acknowledgments

I have benefitted immensely from the support and inspiration provided by all my HTL colleagues (including the much-missed Valérie Raby), and have been particularly moved by their open-mindedness in not only accepting but even encouraging the field linguist in their midst. Special mentions go to Bernard Colombat, who has taught me so much, through our joint project on the history of the parts of speech and beyond, and who shines for his combination of humility and tremendous erudition; and to Jacqueline Léon, who has been a wonderful mentor, patiently explaining and illustrating how historians approach linguistics, in a way which has opened new worlds to me.

I am also very grateful for the awesome linguists who inhabit the concentric rings around HTL: at Paris Diderot, Olivier Bonami, Bernard Fradin and Marie-Claude Paris have been the source of intellectual stimulation, entertainment and friendship; at Paris-area CNRS labs specialized in field linguistics, Katharina Haude, Guillaume Jacques, Enrique Palancar, Stéphane Robert, Yvonne Treis (among many others) have gotten me so excited about linguistics that I can barely believe it is my day-job, in addition to providing many laughs and good times; further afield, Lyon (and affiliated) colleagues Françoise Rose, Alice Vittrant and Marine Vuillermet are reminders that not all the great people and projects are in Paris. And that is without getting into the joys of interacting and collaborating with internationally-based colleagues (though I cannot not mention Barb Kelly, in Melbourne, who truly blurs the line between friend and colleague).

Every time I vaguely consider applying for a job elsewhere, it strikes me that a professional situation just doesn't get better than what I already have... For that, I thank everyone, mentioned or not, who makes linguistics such a very good place to spend one's life.

Abbreviations

>	shows relationship between A (on left) and P (on right) in transitive scenario	IMP	imperative
ABL	ablative	IMPERS.NMLZ	impersonal nominalizer
ACT.PTCP	active participle	INF	infinitive
ADV	adverbial	INSTR	instrumental
AMB	ambulative	INT	intensifier
ANT.CVB	anterior converb	INV	inverse
ANTIC	anticipatory Aktionsart	IRR	irrealis
APPL	applicative	IRREV	irreversible Aktionsart
AT	attention particle	LIM	liminal Aktionsart
AUX	auxiliary	LOC	locative
CAUS	causative	LOW.LOC	low-locative
CL	classifier	N	noun
COM	comitative	NEG	negative
COND	conditional	NEG.DEON	negative deontic
CONJ	conjunction	NMLZ	nominalizer
CONTR	contrastive focus	NOM	nominative
DAT	dative	NPST	non-past
DEON	deontic	OPT	optative
DIST.DEM	demonstrative (distal)	PLU	plural
DU	dual	POSS	possessive
DUR	durative	PROX.DEM	proximal demonstrative
EMPH	emphasis particle	PST	past
EP.AUTH	epistemic authority particle	PURP	purposive
ERG	ergative	QUOT	quotative
FILLER	filler material (eg. hesitation)	RED	reduplication
FOC	focus	REFL	reflexive
GEN	genitive	SEQ	temporal sequencer
HAB	habitual	SG	singular
HIGH.LOC	high locative	SIM.CVB	simultaneous converb
HON	honorific	STAT	stative
HS	hearsay	TEMP	temporal
IDEO	ideophone	TOP	topic
		VOC	vocative

Lists of figures and tables

Figure 1. Van driem's Fallen Leaves model	16
Figure 2. The Kiranti language area	18
Figure 3. The verbal template of Thulung	41
Figure 4. Khaling ideophones	46
Figure 5. Parallel development of the optative and the impersonal negative deontic necessity markers	112
Figure 6. Verbal template (partially completed) for Khaling	114
Figure 7. Toba 1984's chapter on Khaling interjections	123
Figure 8. Map of Kiranti area and languages	124
Figure 9. Table of contents of toba's <i>The Khaling verb</i>	126
Figure 10. Personal endings in Khaling in Toba 1973	127
Figure 11. Hodgson's presentation of the Vayu transitive verb 'to kill'	128
Figure 12. Schematic representation of scenarios in Hodgson's initial paradigms	129
Figure 13. Partial paradigm for indicative present/future 'eat'	130
Figure 14. Schematic representation of scenarios in paradigm in fig. 13.	131
Figure 15. Paradigms with numbers instead of glosses	132
Figure 16. Transitive paradigm from Allen 1975	132
Figure 17. Paradigm showing indexes for transitive verbs	133
Figure 18. Metadata database for grammars and authors in corpus	135
Figure 19. Screenshot of part of the full tables of contents for the grammars in corpus	136
Figure 20. Partial view of the macro-level database	136
Figure 21. Notice in JASB	144
Figure 22. Archaic pronouns of english and their modern equivalents	149
Figure 23. Transcription of text specimen in Vayu	150
Figure 24. Annotation in Wolfenden 1935 showing segmentation	151
Figure 25. Analytical grid for tracking evolution of annotation practices	154
Figure 26. Completed grid contrasting 1897 <i>Specimens</i> volume and same languages in <i>LSI</i>	155
Figure 27. Table of similarities	160
Figure 28. Partial view of corpus selection interface	161
Figure 29. Integral text view for 3 versions of the ur-mother story	162
Figure 30. Screenshot of similarity view for event 5 of the Ur-mother story	163
Figure 31. Similarity view based on concordance, highlighting the search term in the glosses	166
Figure 32. Lexical entries for the verbs 'sleep' and 'fall asleep'	172
Figure 33. Sample entry from Khaling verb dictionary	173
Figure 34. 1sg>2 section of the devanagari-transcribed paradigm	173
Figure 35. List-style paradigm	174
Table 1. Initial consonants in Thulung	20
Table 2. Vowels in Thulung	20

Table 3. Personal pronouns of Thulung (post-2000)	22
Table 4. Possessive pronouns (prefixal and free)	22
Table 5. Nominalized deictic adverbs in Thulung	24
Table 6. Case markers in Thulung	26
Table 7. Relativization strategies for the role of head nouns	33
Table 8. The verb classes of thulung, with root final and alternating stem types	50
Table 9. Paradigm for a verb with no stem alternation	50
Table 10. V-class verbs <i>bomu</i> 'do' and <i>bumu</i> 'be'	51
Table 11. Stem distribution patterns for transitive (left) and intransitive (right) k-class verbs	52
Table 12. Stem distribution patterns for transitive (left) and intransitive (right) p-class verbs	52
Table 13. Stem distribution patterns for transitive (left) and intransitive (right) t-class verbs	53
Table 14. Stem distribution patterns for transitive (left) and intransitive (right) n-class verbs	54
Table 15. Stem distribution patterns for transitive (left) and intransitive (right) s-class verbs	54
Table 16. Paradigm for <i>banmu~ba:mu</i> , with variants	56
Table 17. Transitive verb <i>tse:mu</i> , 'pick through'	58
Table 18. Transitive verb <i>sjanmu</i> , 'clean house'	58
Table 19. Paradigm for <i>lɔ:mu</i> , take away	59
Table 20. Allomorphs, functions and source verbs for derivational formatives	69
Table 21. Allomorphs of the durative marker	69
Table 22. Associated motion markers in Khaling	76
Table 23. Thulung motion verbs	78
Table 24. Thulung Aktionsart markers	84
Table 25. Khaling Aktionsart markers	88
Table 26. Intransitive paradigm	92
Table 27. Transitive paradigm: non-past indexes (all suffixal)	92
Table 28. Transitive paradigm: past indexes	94
Table 29. Breakdown of past indexes	95
Table 30. Allomorphs of 1sg>3 according to verb class	98
Table 31. Allomorphy for 1pi and 3sg non-past indexes according to verb class	98
Table 32. Phonologically-conditioned allomorphs	100
Table 33. Features of markers within non-past indexes	104
Table 34. Analysis of the order of exponents in non-past transitive indexes	105
Table 35. Local and mixed indexation (singular, non-past) in Thulung	107
Table 36. Features (in the order in which they appear) in the local indexes	107
Table 37. Contrastive paradigms	113
Table 38. Non-past and past affirmative and negative verb forms in Khaling	115
Table 39. Koyi affirmative and negative non-past and past forms for verb <i>mumu</i> 'do'	116

1 INTRODUCTION

1.1 Personal context

I will begin this thesis with a little personal context in order to better situate my work for readers.

I began work on languages of Eastern Nepal in 1999. A Fulbright scholarship enabled me to spend ten months in Nepal, during which time I collected materials for my dissertation. The resulting document, *Aspects of the grammar of Thulung Rai*, was submitted to the University of California, Berkeley in 2002, under the supervision of Jim Matisoff. It focused on different grammatical aspects of the Thulung language, some of them in the context of contact with Nepali and others from the point of view of language change over the previous 30 years, thanks to work by a British anthropologist, Nick Allen (1975), on the Thulung culture and language.

A postdoctoral contract made it possible for me to spend a year at LACITO in 2002-2003, working at the French epicenter for research on languages of Nepal, thanks to the presence of Martine Mazaudon and Boyd Michailovsky. I spent the next year with an author's contract at UNESCO, attempting to assemble a handbook for linguistic auto-description (Lahaussais 2006): a guide that would enable non-linguist speakers of endangered languages to undertake basic description of their language in the absence of linguists. The handbook was never widely disseminated, but helped hone an interest--which continues to this day--in linguistic field manuals and questionnaires, seen for example Lahaussais (2019a) and Lahaussais & Vuillermet (2019).

A 2004 Hans Rausing Endangered Language Project from SOAS (an "Independent Postdoctoral Fellowship") allowed me to start working a new Kiranti language, Koyi. The civil war was at its peak in Nepal at the time (even the Peace Corps workers were sent home), and the period was not at all conducive to real fieldwork. I managed to connect with the Koyi community in Kathmandu, and worked for three (non-continuous) months with a very knowledgeable young man (who would sometimes fall asleep during sessions on verb paradigms--attesting both to his state of physical exhaustion from Army training sessions and to the methodological perils of eliciting verbal morphology!). My knowledge of Thulung helped me work on Koyi relatively efficiently, thanks to the typological similarities and

cognates of the two languages, and within a few months I managed to build an annotated corpus of two hours worth of audio recording, available in the PanGloss Collection.¹

By the end of 2004, it was too difficult to carry out descriptive work in Nepal, and I accepted a research engineer position to serve as head of linguistic coordination at the directorates of international relations and communications at CNRS headquarters. My role was to create a management system for the translation needs of both services, and to set up a network of translators, CNRS-specific terminology, and tools to ensure that translations were consistent.

This position, although removed from actual research, held several advantages: it allowed me to discover the academic field of translation studies, corpus linguistics and tools for the automatization of translation. Through connections with researchers at conferences on translation studies, I became a member of CRATIL (Center for Applied Research on Translation, Interpretation and Language), a research group of ISIT (Institut supérieur d'interprétation et traduction, at the time; now Institut de Management et de Communication Interculturels). The exposure to computer-assisted translation (CAT) software and to the concept of alignment of translation memories was ultimately the inspiration behind the project setting up a parallel/comparable corpus of Himalayan mythological narratives (the Kiranti Comparable Corpus), described in 3.3.1. This is reassuring in that it suggests that any experience, even when more distantly related to linguistics, can still contribute to forward motion in a field closer to one's main preoccupations.

I was hired in 2010 by the Histoire des théories linguistiques (HTL) research group to co-direct a long-standing project that had lost some momentum: a historical dictionary of linguistic terminology. Along with Bernard Colombat and 20-some colleagues, the project was eventually redefined: instead of attempting to provide historical sketches of *all* linguistic terminology, we would focus on the history of the parts of the speech from the Greco-Latin tradition, and trace the evolution of their terminology through modern times. The collaborative enterprise, with participants contributing to the chapters for each part of speech based on their expertise in a given historical tradition, eventually yielded a 600-page manuscript, published by Peeters in the Orbis/Supplementa collection (Colombat & Lahaussais 2019a).

I eventually became involved in a number of other projects at HTL (a program on translation, directed by Emilie Aussant; Jacqueline Léon's program on the linguistic annotation of American languages) and beyond (the ANR Corpus project "HimalCo", as well as

¹ https://www.vjf.cnrs.fr/lacito/pangloss/corpus/list_rsc.php?lg=Koyi%20Rai

TULQuest, a program on linguistics questionnaires that I ran for the Federation of Typology and Linguistic Universals), marking a transition back into linguistics research, including fieldwork.

The details of said research makes up the rest of the present document, but it seems significant that it is through involvement in projects further from my initial linguistic preoccupations -- I had never even had a class in the history of linguistics, as the course was discarded from the Berkeley PhD syllabus the year I started there -- that I discovered how a historical and epistemological outlook can significantly enhance a descriptive linguistics approach. It is also very fortunate that HTL is physically located within the Paris Diderot (now Université de Paris) linguistics department, a source of very stimulating interactions with colleagues from outside my immediate research group. It also provides an opportunity to teach: after teaching a 16-hour field methods course as part of a larger one-off MA course on linguistic methodology in 2014 at Diderot, I took on a MA typology course from 2016 until now, which I team-taught with Olivier Bonami in 2016-2018, and which I now teach with Otto Zwartjes, as of 2019. I have also been given a methodology course in the new MA syllabus: a half-semester course on descriptive grammar writing. The contact with MA students has resulted in my supervising a few students for their first- or second-year MA theses (Amalia Hidayat on the *basa walikan* 'inverted language' in Javanese slang; Alvaro Diaz on ergativity in Basque; Mina Yi on grammaticographical models in Mongolian grammars) and others for MA internships.

1.2 Organization of this document

The themes I take up in this document are all centered around the two main areas that make up my research activity: on the one hand, fieldwork on Kiranti languages (primarily Thulung, secondarily Khaling, and lastly Koyi), with contributions to description and typology; on the other, research of a more historical and epistemological nature, centered around the topics of grammaticography, data annotation, and the methodological tools that shape the descriptive enterprise.

Taking as a starting point the undeniable centrality of the verb to Kiranti syntax, with many utterances made up exclusively of a single finite verb,² the goal of this thesis is to bring together work I have contributed to the description of the verb complex in Kiranti languages,

² For a study on the low referential density (in other words, the low prevalence of overt nouns and pronouns) of the Kiranti language Belhare, see Bickel (2003a).

mostly through research on Thulung, on the one hand, and research I have carried out over the last decade into the history and epistemology of language description, especially as relates to verbal morphology, on the other. The document is thus organized into two main chapters: Chapter 2, "Description of the Thulung verb complex", begins with a sketch of Thulung grammar, in order to linguistically and typologically situate the main content that follows. The main section of the chapter is structured around the verbal template of Thulung. For Kiranti languages with fixed root and affix ordering (exceptions being Chintang (Bickel et al. 2007) and Bantawa (Doornenbal 2009)), the verb is often presented in the form of template, with a slot for the root and a number of prefixal and suffixal slots, showing the order of the morphological material that contributes to the assembly of the verb complex.

Because the purpose of the template model is to show the linear ordering of different affixes with respect to the root and each other, it concerns morphological material only. I will nonetheless use the material associated with each slot as an entry point to the discussion of a variety of phenomena from the same functional domain, even in cases where these are based on constructions rather than specific formatives. I will also use the template slots of Thulung to bring in data, when relevant, from other Kiranti languages. As an example, suffixal slot 2 is devoted to derivational material, which in Thulung includes an associated motion marker; this slot will provide me with the opportunity to describe the larger context of Associated Motion in Thulung as well in Khaling, another Kiranti language I have worked on, and to bring in additional data from other Kiranti languages, when relevant. Similarly, suffixal slot 4, which hosts an optative mood marker, will be the opportunity for an excursus into the apprehensional domain, because the Thulung avertive construction, which makes up a large part of that domain, is built from the optative mood marker.

Chapter 3 will then present my contributions to epistemological and historical issues surrounding descriptive linguistics. Topics include grammaticographical models used to produce contemporary grammatical descriptions and what they reveal about influences on descriptive linguistics; annotation practices (such as glossing) for descriptive linguistic data, and changes said practices have undergone over the past 150 years; methodological tools that shape language description, such as aligned corpora, linguistic questionnaires, and electronic dictionaries. The questions explored in this chapter are ones that I have come to in the last decade, through exchanges and collaborations with colleagues at the HTL research group. Even though these are relatively recent interests of mine, they have greatly influenced how I carry out descriptive work on Kiranti languages and have proved very stimulating areas of exploration for a linguist with no previous experience in the history of linguistics.

I shall conclude, in chapter 4, by briefly presenting plans for future research.

1.3 A note on the Thulung corpus

Before launching into chapter 2, which is based on first-hand Thulung data collected in Nepal from 1999-2018, I feel it is important to provide a brief discussion of the narrative corpus of Thulung that I have relied on in assembling this document and many of the studies it mentions.

My Thulung corpus consists of approximately 9 hours of annotated data, comprising narratives, conversations, procedural texts, chants. This data was originally annotated in Toolbox, and a large part of it is archived in the PanGloss collection,³ where most of the material is sound-synchronized. In preparation for writing this document, however, I decided to convert my entire corpus into ELAN.⁴ This involved starting anew from the original sound files, segmenting them anew, completely retranscribing all the data, and in most cases retranslating it as well.

As time-consuming as the process was, it was an absolutely crucial exercise in approaching the language data anew, as someone who now understands it considerably better than in my initial forays. I have also developed a much better sense of Thulung prosody and grammar, and the re-annotation process was an opportunity to capture this personal development. I had also been dissatisfied with my use of Toolbox, and wanted to be able to have easy access to the sound recordings associated with the annotations, as I reanalyzed material in light of attempting to write about the language. Only some of this new corpus is interlinearized (thanks to CorpA's interlinearization module for ELAN),⁵ but my original interlinear glosses are still accessible through Toolbox files, and at this stage, I feel that my understanding of Thulung is such that I do not need interlinear glosses to make use of the data. Additionally, I have come to realize that the glosses and translations that accompany examples are often reworked for the purposes a specific use of the material (see section 3.2.2 for more discussion). What is essential to me is the ability to easily access the sound file -- which is after

³ https://lacito.vjf.cnrs.fr/pangloss/corpus/list_rsc_en.php?lg=Thulung%20Rai&name=Thulung%20Rai

⁴ This was in part out of a sense of frustration that the Toolbox-annotated corpus I was using for searches (PanGloss cannot currently be used for searches) and concordances did not include sound files, something which raises the issue of what indeed we are working with when data from oral languages is 'reduced' to a somewhat decontextualized transcription (often made shortly after encountering the language for the first time).

⁵ See http://lacan.vjf.cnrs.fr/res_ELAN-CorpA.php

all the only part of the data which has any claim to objectivity -- when using the data for the purposes of linguistic analysis.

Another advantage of ELAN is the possibility of adding annotation tiers, and as a result my corpus is now tagged for lexemes/formatives/constructions of interest to me. These include deixis of an elevation-encoding nature; unusual inflection patterns (specifically differences across the two dialects of Deusa and Mukli); derivational formatives; interjections (which are notoriously difficult to retrieve through concordances, as they tend to be somewhat idiosyncratic in their production); ideophones (which, despite easily identifiable patterns, cannot be searched for on the basis of their phonological form); reported speech.

2 DESCRIPTION OF THE THULUNG VERB COMPLEX

This chapter aims to provide as complete as possible a description of the verb complex of Thulung, based on work I have carried out on Thulung over the past twenty years, either been published or presented orally, or representing a new analysis or topic for me. In order to structure the discussion of the material, I use the concept of the verbal template, which in Thulung is made up of one prefixal slot, a root slot, and a number of suffixal slots. The material that occurs in each of these ordered slots is described in turn.

First, however, I present in 2.1 a grammatical sketch of Thulung, in order to give some linguistic context to the material that makes up and accompanies the verb complex. This sketch is based on my own descriptive work, and refers to publications on various topics of relevance. I then describe the verbal template and its content in section 2.2.

2.1 Introduction to Kiranti and Thulung

This section will provide a brief presentation of Kiranti languages and their philological context, before presenting a short grammatical sketch of Thulung.

2.1.1 Classification

The Kiranti languages are part of the Sino-Tibetan family. Various classification schemes have been proposed for the family, among which Matisoff (2003), Thurgood (2003), and Bradley (1997) for the Tibeto-Burman branch. Recently, the application of computational techniques has resulted in different proposals, such as Sagart et al. (2019). One significant argument affecting the proposals is whether the person agreement morphology found in Kiranti, Gyalrongic, West Himalayan, Kham and Nungish is considered to be a feature of the proto-language, later lost in the isolating languages of the family (a position argued for by a number of scholars (DeLancey 1989; van Driem 1993a; DeLancey 2010)), or an innovation limited to the branches where it is currently found (LaPolla 1992a; Matisoff 1993; Nishi 1995; LaPolla 2003).

Another recent debate in the field concerns the idea that the Sino-Tibetan family should in fact be rechristened Trans-Himalayan. According to van Driem (2018), the labels for both Sino-Tibetan and Tibeto-Burman are problematic for phylogenetic and historical reasons. The name Tibeto-Burman has been used in different senses by different scholars: to

refer to the entire family (eg. Hodgson, see van Driem (2014: 12)), and to refer to the major branch remaining after the branching off of the Sinitic languages (eg. Matisoff (2003)). As for the label Sino-Tibetan, from a phylogenetic point of view, it suggests a tree model, with two major branches which are the Sinitic languages on the one hand, and the Tibeto-Burman languages on the other. Van Driem has suggested that, given our current understanding of the relationship between the languages of the family, with notorious difficulties in top-level groupings, a better model for the family is the Fallen Leaves model (van Driem 2014): the model is "bottom up, i.e. starting with the tangible leaves that have fallen from the tips of the branches, and then moving upward to gain an understanding of the nodes in the tree" (ibid, 22), and represented visually as a 'leaf' for each subgroup, organized in roughly geographic terms (as shown in Figure 1, from ibid: 19), with no claims as to the phylogenetic connections between the leaves.



Figure 1. Van Driem's Fallen Leaves model

While recognizing an underlying tree-type relationship between subgroups, the model makes no pronouncements about higher order grouping. From a historiographical point of view, Sino-Tibetan is problematic in that it first surfaced as a terminological stand-in for "Indo-Chinese", a family which included all languages of Asia and Oceania. Van Driem has proposed the term Trans-Himalayan, considered to be "purely geographical in inspiration for

this language family straddles the highest land barrier on the planet and is disseminated both north and south of the Himalayas" (2018: 118).⁶

A number of classification schemes exist for the Kiranti languages as well, the main ones being described and compared in Gerber and Grollman (2018). The term "Kiranti" is generally used by linguists to encompass the Rai languages and Limbu (Michailovsky 2017: 646), whereas anthropologists tend to focus on the Rai as a distinct ethnic subgroup (Schlemmer 2019). The linguistic groupings are largely based on geographical distribution, backed by sound correspondences and lexical isoglosses. Van Driem (2001: 615)'s scheme is presented here:

Limbu

Eastern Kiranti

Greater Yakkha: Yakkha, Chiling, Athpare

Upper Arun: Lohorung, Yamphu, Mewahang

Central Kiranti

Khambu: Kulung, Nachiring, Sampang, Sam

Southern: Chamling, Puma, Bantawa, Chintang, Dungmali

Western Kiranti

Midwestern: Thulung

Chaurasiya: Ombule, Jero

Upper Dudhkosi: Dumi, Khaling, Kohi

Northeastern: Bahing, Sunwar, Hayu

Note that various scholars have nonetheless questioned whether these languages constitute a valid phylogenetic group (eg. (Ebert 2003: 516; Michailovsky 2010)). The issue of whether or not these languages form a valid linguistic grouping does not strike me as particularly important to the work presented here, as the focus is not on diachrony, even though the Thulung verb complex is compared and contrasted in several places with Khaling and Koyi. On the other hand, the Kiranti Comparable Corpus discussed below only has validity if there is such a thing as a Kiranti group, although, as will be seen, one of the goals of the project was

⁶ One problem with rechristening is that raises issues of loyalty to mentors, familiarity, and habit; this possibly explains LaPolla's vehement rejection of "Trans-Himalayan" (eg. (LaPolla 2016)), which he presumably sees as disloyal to Matisoff's classification scheme (Matisoff 2003).

precisely the assemblage of a corpus allowing lexical and morphosyntactic comparison with an aim to getting a better sense of the relatedness of the languages on the basis of shared myths.

The languages are geographically distributed as shown in the map in Figure 2.

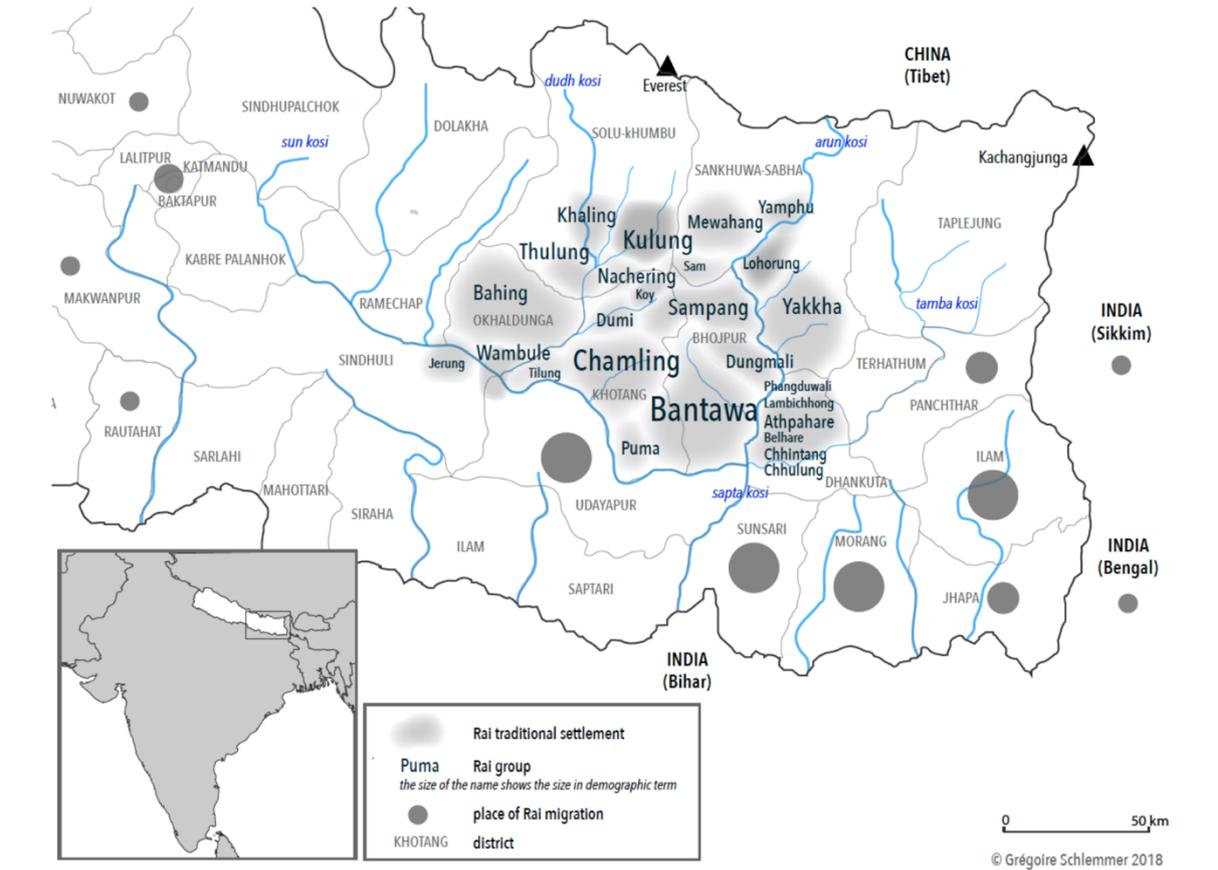


Figure 2. The Kiranti language area (map from (Schlemmer 2019))

The first documented writings in English on Kiranti languages are by Hodgson, with descriptions of the grammars of Bahing (Hodgson 1858) and Vayu [Hayu] (Hodgson 1857a), as well as comparative word lists from a number of other Kiranti languages, listed as Chamling, Rungchhenbung (Bontawa), Chhintang (Bontawa), Nacchereng, Waling, Yakha, Chourasya, Kulung, Thulung, Bahing, Lohorong, Lambichhong (Waling), Balali, Sangpang, Dumi, Khaling, Dunmali in Hodgson's spelling and nomenclature (Hodgson 1857b: 318). There followed a period during which contributions to Kiranti language research were in the form of additional comparative word and phrase lists (Hunter 1868; G. Campbell 1874), of which the former reorganized material collected by Hodgson.

The next influx of new material was with the *Linguistic Survey of India*, volume 3 of which dealt with the Tibeto-Burman languages (Grierson 1909). It is made up of individual languages sketches which were assembled, as far as can be discerned, on the basis of a word and phrase list (which was that in Campbell (1874), collected anew) and text specimens, among which the Parable of the Prodigal Son featured prominently. The Kiranti languages in the *Linguistic Study of India* were classified as 'complex pronominalized languages' based on their verbal morphology: Limbu, Yakha, Khambu, Bahing, Minor Khambu dialects (Balali, Sangpang, Lambichhong, Waling, Chhingtang, Runchhenbung, Dungmali, Rodong, Nacchereng, Kulung, Thulung, Chourasya, Khaling, Dumi), Rai, Vayu.

In the 1930's, Wolfenden produced a few grammar sketches, accompanied by (and apparently assembled on the basis of data from) annotated texts: Thulung (Wolfenden 1935), Kulung (Wolfenden 1934) and Sampang (Wolfenden 1933). Work on Kiranti languages was pursued in the 1960's and 1970's with material on Limbu (Sprigg), Khaling (Toba 1973; Toba 1984) and Thulung (Allen 1975), before a period of great effervescence which yielded a number of detailed descriptions of Limbu (van Driem 1987), Hayu (Michailovsky 1988), Dumi (van Driem 1993b), Belhare (Bickel 1996), Athpare, (Ebert 1997a), Camling (Ebert 1997b). This list only mentions book-sized treatments, but work also appeared in the form of articles and other publications (notably work on Chintang and Puma by Bickel and his team), and work comparing Kiranti languages (Ebert 1994).

This wave of linguistic interest in the 1980's and 1990's generated a number of published grammars (originally written as dissertations) by students of van Driem, Ebert, and Bickel, leading to a much more complete picture of Kiranti languages, such as Yamphu (Rutgers 1998), Wambule (Opgenort 2004), Jero (Opgenort 2005), Kulung (Tolsma 2006), Sunwar (Borchers 2008), Bantawa (Doornenbal 2009), Yakkha (Schackow 2015). A useful overview of the linguistic features of the Kiranti languages can be found in Michailovsky (2017).

2.1.2 Grammar sketch of Thulung

The following sketch of Thulung is intended to provide a general understanding of the typological features of the language, in order to place the description of the verb complex in perspective. More details can be found in the general presentation of Thulung published as Lahaussais (2004).

2.1.2.1 Phonology

2.1.2.1.1 Phoneme inventory

Initial consonants found in native Thulung words are the following.

p	t		ts	k	[ʔ] ⁷
ph ⁸	th		tsh	kh	
b	d	dʰ	dz	g	
bh	dh		dzh	gh	
m	n			ŋ	
	s				h
	l	r			
w			j		

Table 1. Initial consonants in Thulung

Final consonants are more limited (because, in part, of devoicing), and are the following: p t k m n ŋ l r s.

Thulung vowels are the following¹⁰.

i	ɯ	u
e	ɵ	o
	ʌ	
	a	

Table 2. Vowels in Thulung

⁷ The glottal stop is very marginal in Thulung: it is found with the hearsay particle ʔe, with the color term 'yellow' ʌmʔʌm (in which transcription it essentially serves to mark a syllable boundary between m and ʌ); and with a number of interjections. Some Kiranti scholars transcribe all vowel-initial words as ʔV (this system was adopted for Khaling (Jacques 2015; Jacques et al. 2012); others do not but claim that the glottal initial is present (Bickel et al. 2007); Michailovsky concludes that "it is not clear to [him] whether such an element is a phonological segment." (2017: 649)

⁸ I adopt the keyboard-friendly convention of writing aspiration not with a superscript (as in t^h); similarly, I write affricates with two glyphs--ts, dz-- instead of as tʃ, dʒ.

⁹ This is the only commonly-used retroflex in Thulung. A single supposedly native lexeme is found to begin with [ɬ], the verb *ɬhwa:mu*, 'hit with something'. Loan words have initials gh, ʈ, ɬh, dɬ.

¹⁰ Note that I have changed some of these transcriptions since earlier work: ɯ for y, ɵ for u, ʌ for ɔ.

Note that the glide+vowel combination *ja* is sometimes pronounced ε , but I do not consider it (in the current state of the language) to be a distinct phoneme. Similarly, the combination of *w* + *a* sometimes shows a parallel development: While there is synchronically a difference between *wa* and λ (seen in the minimal pair *thwa:mu*, 'squeeze' vs *th λ :mu*, 'hide'), there are also instances where λ is a variant of *wa*, as in: *nwa:-le* ~ *n λ :-le* (2-CL); *rwakta* ~ *r λ akta* (say-3SG.PST).

2.1.2.1.2 Vowel length

I reported in my dissertation that I did not find any evidence of vowel length, contrary to what had been found in previous work (Wolfenden 1935; Allen 1975). I now feel that such a claim was wrong, possibly a result of what appears to be the loss of vowel length contrasts among younger some speakers. Even among older speakers, vowel length contrasts in some nouns are not particularly consistent, with a certain amount of variation found, even for the same speaker. The one area where vowel length can be said with certainty to be present is the compensatory vowel lengthening in transitive verbs with stems alternating between CVk- and CV:-. Intransitive verbs are less consistent than transitive (some of them appear to alternate between CVk- and CV-, rather than the expected CV:-). It is interesting to note that the community, in developing a writing system based on devanagari, has apparently had some debate about whether or not Thulung has vowel length: the conclusion, for now, is that they are not sure, and will not attempt to transcribe it (Chandrakala Thulung, pc, May 2018). This confirms the idea that a certain amount of variation is found within the community.

2.1.2.1.3 Syllable structure

As is characteristic of Kiranti languages, the syllable structure of Thulung can be given as $(C_i)V(C_f)$, where C_i is an initial consonant (but can also be a combination of initial consonant and a liquid or glide), *V* is a vowel, and C_f is a final consonant. When occurring word-initially, vowel-initial syllables have a glottal onset, which I do not transcribe.¹¹

2.1.2.2 Morphosyntax

2.1.2.2.1 Personal and possessive pronouns

Personal pronouns exhibit person, number (singular, dual, plural) and clusivity contrasts.

¹¹ I do, however, transcribe the hearsay marker as ʔe .

Additionally, there is currently a formality contrast in 2nd and 3rd person pronouns: the earlier system had no formal pronouns (based on data from Allen (1975)), but a two-way contrast was created, most likely because of contact with Nepali (which has various levels of formality in 2nd and 3rd person pronouns). The former 2nd and 3rd plural pronouns were imported into the singular to serve as the formal pronoun forms; new plural forms were created, through the suffixation of the nominal pluralizer.

	SG	DU	PL
1	go	excl. gutsuku	excl. guku
		incl. gutsi	incl. gui
2	gana	gatsi	gani
	gani (formal)		ganimim (formal)
3	gu	gutsi	gumi
	gumi (formal)		gumimim (formal)

Table 3. Personal pronouns of Thulung (post-2000)

The creation of formal pronouns and resulting creation of new plural pronouns has had consequences in the case-marking system, as will be seen below. A more complete discussion can also be found in Lahaussais (2003a).

There are two series of possessive pronouns, one prefixed (the top line in each cell), the other free (the lower line in each cell). They are used interchangeably, except that the free possessive pronouns can be used independently in the absence of the noun they modify.

	SG	DU	PL
1		<i>excl.</i>	<i>incl.</i>
	a-	atsi-	itsi-
	ama	atsima	itsima
2	i-	itsi-	ini-
	ima	itsima	inima
3	u-	utsi-	uni-
	uma	utsima	unima

Table 4. Possessive pronouns (prefixal and free)

Note that the formality contrast found in the personal pronouns (Table 3) is not maintained in the possessive pronouns. This is exemplified in (1), in which the possessor marked by 2SG.POSS *i-* and the P of the 1SG>2SG.HON are the same argument, indexed differently in terms of formality.

- (1) mima, go i-ser sen-sa-ni
 Grandmother, 1SG 2SG.POSS-1ouse kill-APPL-1SG>2SG.HON
 'Grandmother, I will kill your lice for you.'

2.1.2.2.2 Number

Nouns can be optionally marked for number (dual, *-tsip*, or plural, *-mim*; singular is unmarked), with humans of high status being likely to be marked, humans of low status and other animates less likely, and inanimates unlikely. This optionality of number marking extends to verb indexation, as can be seen in (3), where the plural-marked P (*a-tsho:ra-mim*) is indexed on the verb as a singular.

- (2) nwa:-le tsettse-tsip ba-ttsi ?e
 two-CL child-DU be-3DU.PST HS
 'There were two children.'

- (3) a-tsho:ra-mim hala-ŋa khet-to
 1SG.POSS-child-PLU up-INT bring-3SG>3SG.PST
 'I brought my children up above.'

Example (4) shows the optionality of marking on inanimates, as 'fish' is not plural-marked.

- (4) Δ sinḁa tsai su:-le khatsha-ra ŋo bu
 here CONTR 3-CL pot-LOC fish be.3SG
 'Here, in three pots, there are fish.' [description of image in which the three pots are overflowing with fish]

One particularity of languages of the region is that plural-marking can code a collective made up of the referent and other similar entities, as in 'cakes and such' in (5):

- (5) pe-pa-ʈhok kek-mim ba-ira
 eat-ACT.PTCP-NMLZ cake-PLU be-3SG.PST
 'There was food and cakes and the like.'

2.1.2.2.3 Demonstratives and deixis

Thulung has proximal (*o/oram*) and distal (*me/meram*) demonstratives, with the distal being much more frequent than the proximal, and sometimes substituted for the 3SG pronoun.

The language has a series of elevational suffixes (with a four-way contrast: high (*-la*), low (*-ju~-jɥ*), same level (*-nu~-no*), and unspecified (*-ra*)) which can be combined (often with some assimilation) with the deictic bases *o* and *me* in order to generate elevation-coding adverbs. These adverbs can then be nominalized, as illustrated in Table 5, to yield 'the place/the one characterized by the corresponding adverb', as in (6).

elevation	unspecified	high	low	same-level
proximal	o-ra-m	ha-la-m	hu-ju-m	hu-nu-m/hu-no-m
distal	me-ra-m	me-la-m	me-ju-m	me-nu-m/me-no-m

Table 5. Nominalized deictic adverbs in Thulung

- (6) **meno-m** meno-laŋka rʌksi khen-mi-m
over.there-NMLZ over.there-ABL raksi bring.up-3PL>3SG-NMLZ
 dʌm
 become.3SG
 'That place over there, they must bring up raksi from over there.'

2.1.2.2.4 Numerals

Apart from numerals for 'one' (*ko:*), 'two' (*nwa:*) and 'three' (*su:*), Thulung speakers for the most part use Nepali numerals. Native numerals can be still be elicited from some speakers, and are occasionally found in texts. The Thulung data can be found in my contribution to the Denumerals database¹². Numerals are used with classifiers, of which a single native one, *-le*, is in common usage, in addition to two classifiers borrowed from Nepali, the generic classifier *-waŋa*, and the classifier for humans, *-dzana*. The distribution between the two generic classifiers

¹² The database, <http://drehu.linguist.univ-paris-diderot.fr/denumerals/>, is coordinated by Bernard Fradin.

(native and borrowed, respectively) is that *-le* is used with Thulung numerals (in other words, with 'one', 'two' and 'three') and Nepali *-waṭa* is used with borrowed numerals. An additional classifier has been collected through elicitation (*-phe* for round-shaped objects, such as coins, bananas, bread), and Allen (1975: 113) mentions that Hodgson reported Thulung classifiers for round objects (*-bop*), elongated object (*-səl*) and flat objects (*-phe*).¹³ Example (7) illustrates the three common numerals, combined with the generic classifier *-le*.

- (7) dilimduŋ-kam **su:-le** tsettso-mim ba-mri ʔe
 Dilimdung-GEN **three-CL** child-PLU be-3PL.PST HS
nwa:-le dzaumo **ko:-le** tarapa
two-CL daughter **one-CL** son
 'Dilimdung had three children: two daughters and one son'

2.1.2.2.5 Adjectives and adverbs

Most adjectives are deverbal, based on productive derivation from Thulung verbs with the active participial marker *-pa*. Exceptionally (as with *dzəpa*, 'good') no source verb can be found for the derivation. There is additionally a very small class of lexical adjectives which is not derived in this way, such as *dokpu*, 'big' and *jakke*, 'small'.

Another means of creating adjectives is to derive them from the language's very limited set of adverbs (*dala*, 'fast', *wakha*, 'slow', *malo*, 'recently', *ɲado*, 'early') through nominalization, as in (8).

- (8) **ɲado-m** dewa si:-ra-m-ka delpa-ka koŋmi dewa
early-NMLZ priest die-3SG.PST-NMLZ-ERG villager-ERG other priest
 tuŋ-miri
 elect-3PL>3SG.PST
 'Because the former priest died, the villagers elected another.'

Additionally, there is a relatively large class of ideophones: they are mostly characterized by reduplication, and when nominalized, serve as attributes for nouns. They are

¹³ Note the semantic shift from the form noted by Allen and the equivalent I collected.

discussed in the section below on ideophones (2.2.1.2). Color terms are a subclass of this category.¹⁴

2.1.2.2.6 Case

The following cases are found in Thulung. They mark nouns, pronouns, and nominalized elements (which function nominally). Some are also used as subordinators with nominalized clauses: the ergative/instrumental is used to subordinate a causal clause¹⁵ and the locative is used to mark a purpose clause¹⁶ (although in the case of the latter, I re-gloss it PURP); subordination with case-markers is found throughout the Kiranti group (see eg. Ebert (1994: 135)).

Nominative	-∅ ¹⁷
ergative	-ka
dative	-lai
instrumental	-ka
genitive	-ku, -kam
comitative	-num~-nuŋ
locative	-ra~-ɕa
high-level locative	-la
low-locative	-ju~-ju
same.level-locative	-nu~-no
ablative	-laŋka
mediative	-ram~-ramma

Table 6. Case markers in Thulung

¹⁴ The color terms are the following: *lalam*, red, *gigim* green, *kekem* black, *ʔoʔom* yellow *bubum*, white, *nunum* blue.

¹⁵ go ba:so dzam pe-uto-m-ka homlo nɛpa bu-ŋu, 1SG leftover rice eat-1SG>3SG.PST-NMLZ-ERG now ill be-1SG, 'Because I ate old rice, I am now sick.'

¹⁶ ghrenem thep-ɕa laS-ta-m bu, nettle pick-PURP go-3SG.PST-NMLZ be, 'He has gone to pick nettle.'

¹⁷ In this section, the examples of different types of basic verb frames are glossed with -∅ for the unmarked nominative; in the rest of this document (as in the data I have archived in the Pangloss collection), I do not gloss the nominative case.

Core argument marking makes use of case markers for nominative, ergative and dative.¹⁸

Alignment patterns are described in the following sections.

Intransitive scenarios

S is nominative-marked, verb indexes S.

(9) go- \emptyset $\lambda\eta$ - η u
1SG- \emptyset sleep-1SG
'I sleep'

(10) khlea- \emptyset λ m
dog- \emptyset sleep.3SG
'The dog sleeps'

Experiencer construction:

S is dative-marked, verb indexes impersonal 3SG.

(11) go-lai kwa:ra λ :-ra
1SG-DAT thirst feel-3SG.PST
'I am thirsty '

In some cases, like (12), the construction is more complex (involving possession by the experiencer in addition to the dative-marking), while still conforming to the basic pattern:

(12) u-pe-pe dwak-ta me seruli-lai
3SG.POSS-eat-RED like-3SG.PST DIST.DEM jackal-DAT
'The jackal wanted to eat it.' (Jackal and Crow)

¹⁸ More neutral case labels might be 'absolute', 'ergative' and 'antiergative', based on Creissels (2009), which would get around the problem of using the label 'nominative' or 'dative' for a P (see (14) and (16) respectively).

with *animacy*.¹⁹ Additionally, we shall see below (2.2.3.3.5, 'Evidence of referential hierarchies within the paradigm?') that indexation markers also show signs of tapping into the same referential hierarchy.

Monotransitive scenarios

There are four alignment patterns found with monotransitive scenarios, based on the nature of the A and P:

-A: nominative-marked (ie. unmarked) for 1st person and 2nd person singular and dual; ergative-marked for 2nd plural, 3rd person and other NP's;

-P: nominative-marked (ie. unmarked) for inanimates; dative-marked for high-status humans. Non-human animates and low-status animates are optionally marked.²⁰

Argument indexation on the verb marks both A and P.

A- \emptyset P- \emptyset V-A>P:

(14) go- \emptyset khlea- \emptyset jal-u
 1SG- \emptyset dog- \emptyset strike-1SG>3SG
 'I strike the dog.'

A-ka P- \emptyset V-A>P:

(15) gu-ka khlea- \emptyset jal-u
 3SG-ERG dog- \emptyset strike-3SG>3SG
 'He strikes the dog.'

A- \emptyset P-lai V-A>P:

(16) go- \emptyset me mʉtsʉ-lai jal-u
 1SG- \emptyset DIST.DEM man-DAT strike-1SG>3SG
 'I strike that man.'

¹⁹ For an introduction to the history of research into referential hierarchies (and their decomposition into sub-hierarchies based on features, such as person and animacy), see Haude and Witzlack-Marakevich (2016).

²⁰ Inanimates are rarely marked, and when so, it is usually for emphasis.

A-ka P-lai V-A>P:

- (17) gu-ka go-lai jal-ŋi
 3SG-ERG 1SG-DAT strike-3SG>1SG
 'He strikes me.'

Ditransitive scenarios

In ditransitive scenarios, there are three arguments: an A, and a G (the most goal-like argument) and a T (the most theme-like argument) (the labels G and T follow Bickel (2011)). The A can be differentially marked, as seen above, taking nominative or ergative case marking. The G is dative marked, and the T is unmarked.

Argument indexation on the verb depends on whether the verb is of a secundative (A>G argument indexation) or an indirective (A>T indexation) type, the two types being a lexical property of verbs.

Secundative verbs

A-∅ T-∅ G-lai V-A>G:

- (18) go-∅ gana-lai kuṭuma-∅ gwa:-ni
 1SG-∅ 2SG-DAT puppy-∅ give-1SG>2SG
 'I give you a puppy.'

Even when the T is human, it does not take dative-marking:

- (19) go a:ma dzaumotse hanu-m mʉtsu-lai
 1SG 1SG.POSS daughter other-NMLZ person-DAT
 gwak-pu
 give-1SG>3SG
 'I give my daughter to the other person.'

A-ka T-∅ G-lai V-A>G:

- (20) gu-ka u:ma tsettse-mim-lai tsʌkleŋ-∅ gwak-tumi

3SG-ERG 3SG.POSS child-PLU-DAT chocolate- \emptyset give-3SG>3PL.PST
 'She gave chocolate to her children.'

Indirective verbs

A- \emptyset T- \emptyset G-lai V-A>T:

(21) go- \emptyset gana-lai bwa- \emptyset thur-pu
 1SG- \emptyset 2SG-DAT pig- \emptyset send-1SG>3SG
 'I send you a pig.'

A-ka T- \emptyset G-lai V-A>T:

(22) pʌsʌle-ka go-lai sor- \emptyset bhreʈ-dʰu
 store.keeper-ERG 1SG-DAT rice - \emptyset sell-3SG>3SG.PST
 'The storekeeper sold me rice.'

There are additional argument frames that are found in the language (with interesting cases of case-marking when a verb's valency is reduced or increased), and a useful model to follow for their identification is Schikowski et al (2015). The examples given above are nonetheless thought to cover the main types of core argument marking encountered.

2.1.2.2.7 Nominalization and relativization

Like many other Kiranti languages, Thulung has a nominalizing formative, *-m* (and allomorph *-mim*), which is multifunctional: it reflects what has been labeled the Standard Sino-Tibetan Nominalization (SSTN) pattern (Bickel 1999), whereby the same formative can be used:

a) to nominalize clauses (for complementation, as in (23) and (24)) and entire sentences ((25) and (26)):

(23) bhansa go khok-to-**m** rwak-ta
 food 1SG cook-1SG>3SG.PST-NMLZ say-3SG.PST
 'She said "I cooked the food."'

(24) dhwa:-lanʒka ko:-le seruli-ka tsai ga:pu-ka u-si-dʒa ɲo

ground-ABL 1-CL jackal-ERG CONTR crow-ERG 3SG.POSS-mouth-LOC fish
 per-thʌʈ-dʊ-m lwas-tʊ
 pinch-DUR-3SG>3SG.PST-NMLZ see-3SG>3SG.PST

'From the ground a jackal saw that the crow was pinching a fish in his mouth.'

(25) homu-ku bwi theptso dʊs-ta-m ʔe
 bear-GEN head flat become-3SG.PST-NMLZ HS

'It is said that the bear's head became flat.'

(26) on-laŋka gu-ka po-mu-lai khu-irʊ-m
 over.there-ABL 3SG-ERG eat-INF-DAT steal-3SG>3SG.PST-NMLZ
 hola
 maybe

'He probably stole it from there to eat.'

b) for attribution (as in (27)):

(27) he:ka hala u-si-dʒa-m ɲo kwa:-jʊ al-mu
 how up 3SG.POSS-mouth-LOC-NMLZ fish mud-LOW.LOC fall.down-INF

'How to have the fish fall down to the ground from his mouth.'

c) for relativization (as in (28)):

(28) go khok-to-m dzam brʌpa ba-iɖa
 1SG cook-1SG>3SG.PST-NMLZ rice delicious be-3SG.PST

'The food I cooked was delicious.'

As can be seen in (28), the relative clause is externally headed, and prenominal. Other examples of the use of *-m* can be found in Lahaussais (2003b).

As far as the relativization function is concerned, the nominalizer that participates in the SSTN pattern (*-m*, with allomorph *-mim*) is limited in terms of its ability to relativize on all roles: the A is blocked from relativization with *-m* (something which is not uncommon for ergative languages, see Creissels (2006: 215)), in which case an alternative relativization strategy, using an active participial marker, is available. Table 7 illustrates which formatives

are available for relativization on different roles (*-pa* is the active participial marker; *-ma* is the passive participial marker): while *-m* shows a pattern consistent with ergative behavior ($A \neq S = P$), the participial makers follow an accusative pattern ($S = A \neq P$). This is discussed and exemplified further in Lahaussais (2017a).

	S	A	P	G/T	Obl	Gen	Obj Comp
<i>-pa</i>	✓	✓	x	x	mostly yes (depends on type)	✓ (with restrictions)	x
<i>-ma</i>	x	x	✓	✓	mostly yes (depends on type)	x	x
<i>-m/-mim</i>	✓	x	✓	✓	✓	✓	x

Table 7. Relativization strategies for the role of head nouns

2.1.2.2.8 Discourse markers

Thulung has discourse markers operating at a number of different levels: word-, phrase- and sentence.

At the word-level is the intensive marker *-ŋa*, which is most often used with adverbs to intensify them, but can also be found with other word classes. The marker functions in very much the same way as Nepali *-nai*, and is illustrated below.

- (29) [hikai-ŋa bu-mi dzøpa-ŋa bu-mi
 okay-INT be-3PL good-INT be-3PL

'They are really okay, they are really good'. [in context of asking about someone's health]

(30) nep-ra roget-tsi-lo mankantshi-num pap tsai ηado-ηa
house-LOC arrive-3DU.PST-TEMP stepmother-COM father CONTR first-INT
rok-tsi
come-3DU.PST

'When they arrived at the house, their stepmother and father had already come first.'

At the phrase-level, Thulung has topic marker *ne*; focus marker *re*, used to introduce new information; and contrastive focus marker *tsai*. The markers are found after the nouns that they mark. The first two are shown in contrast in the same sentences in (31) and (32).

(31) go dzΛηgΛli bu-ηu go **ne** goϥhalare
1SG forest.man be-1SG 1SG **TOP** herder **FOC**
'I am a forest-man, I am a herder.'

(32) gatsi **ne** bju-ka **re** salpo-tsi
2DU **TOP** vulture-ERG **FOC** devour-3SG>3DU
'A vulture will devour you.'

Note in (32) that the focus marker follows a case-marked noun, something which is also found with topic and contrastive focus markers.

Sentence (33) shows the contrastive focus marker: the sentence was collected while using a picture book,²¹ with *Λsinda* referring to the picture in question, and the contrastive focus marking that picture as distinct from others.

(33) Λsinda **tsai** su:-le khatsha-ra ηo bu retsΛ
here **CONTR**3-CL pot-LOC fish be.3SG it.seems
'Here, in three pots, there are fish.'

Occasionally a core case-marker will be dropped and a discourse marker found instead. In (34) below, the noun *ga:pu* is the A of the verb *lwastu*, from which it is separated by a (bracketed for clarity) complement clause, and does not take the expected ergative marking.

²¹ Jackal and Crow: <http://tulquest.huma-num.fr/en/node/161>

Note however that the complement clause also includes an A, in this case ergative-marked; it is possible that the omission of the ergative marking on the A of the matrix clause, and the use of a contrastive focus marker, is a disambiguation strategy, rather than a sign of the optionality of core argument marking.

- (34) mekotima ga:pu tsai [me seruli-ka u-ŋo
 then crow CONTRDIST.DEM jackal-ERG 3SG.POSS-fish
 pʉ:-saʈ-ɖu-m] lwas-tʉ ma u-ŋali
 eat-APPL-3SG>3SG.PST-NMLZ see-3SG>3SG.PST CONJ 3SG.POSS-face
 ɖʌs-thʌ-rʉ ma dʌksa-ku grem-ɖʌ:la
 drop-DUR-3SG>3SG.PST CONJ tree-GEN branch-above
 gʌn-ta-si
 sit-STAT-REFL.3SG
 'Then the crow realized that jackal ate his fish, and kept his face lowered and sat up on
 the tree branch.'

Thulung also has discourse markers with scope at the sentence level: epistemic authority marker *sə*, and hearsay marker *ʔe*. Additionally, at the sentence-level, we find loan particles *retsʌ*, *ni*, *tʌ*, which will not be discussed here.

The epistemic authority marker *sə* is very often used with directives²² to strengthen their illocutionary force, and in some contexts can take on an apprehensive reading, as in (36) and (37).

- (35) ku phir-a sə
 water fetch-2SG>3SG.IMP EP.AUTH
 'Go get water!'

- (36) khlea-ka khre-na sə
 dog-ERG bite-3SG>2SG EP.AUTH
 'The dog will bite you!' (so be careful)

²² Significantly, it cannot be used with directives addressed to an elder (in which case the speaker's epistemic authority presumably cannot be expressed).

- (37) ma:ma sə ta:-na sə
 careful EP.AUTH fall-2SG EP.AUTH
 'Careful! You will fall!'

The apprehensive reading can even be interpreted as a prohibitive:²³

- (38) Λm-na sə
 sleep-2SG EP.AUTH
 Don't sleep! [context: so you don't miss an important event]

The utterance in (38) is considered a more natural way of expressing prohibition than the corresponding negative imperative (*me-Λms-a sə*, NEG-sleep-2SG.IMP EP.AUTH, 'Don't sleep!'), even though the latter is also possible. This suggests that the apprehensive reading of the epistemic authority marker is extended to the expression of prohibition, even though it involves no negation marker. This ties in with the grammaticalization pathway described in Pakendorf & Schalley (2007) from epistemic marking to prohibitive marking. Interestingly, of the few other languages for which this pathway was described in the article in question, two others are spoken in Nepal (Belhare and Chepang). More details can be found in Lahaussais (In press). [See also section 2.2.3.4 which discusses the avertive construction, also found in the apprehensive domain.]

Thulung, like other Kiranti languages, has a rather impoverished evidential system when one compares it with other languages in the area (Tibetan and its dialects are very rich in evidential markers, see eg. LaPolla (2003: 35)). The only native marker that serves an evidential function is the hearsay marker *ʔe*.

The sequence in (39) illustrates how in a narrative, it is often every utterance that ends with the hearsay marker, the marker serving almost to punctuate the narrative. The translation accompanying (39) does not indicate the hearsay, which is understood from the fact that the sentences are produced within the context of recounting a narrative.

²³ The definition of prohibitive in Van Olmen (2010: 471) of "a construction of grammar which typically presents a state of affairs as undesirable and calls on the hearer not to actualize it" fits this situation.

- (39) a) hala-laŋka u-mam dzes-ta ʔe
 above-ABL 3SG.POSS-mother call-3SG.PST HS
- b) lu ge-tsi ge-tsi ge-tsi ʔe
 hey come.up-2DU come.up-2DU come.up-2DU HS
 bu:-ru me tukisa:le-ŋa
 do-3SG>3SG.PST DIST.DEM spool.thread-INT
 tsar-saŋ-dʒa ʔe
 hrow-APPL-3SG>3SG.PST HS
- c) me tukisa:le dzheŋ-to dzheŋ-to dzheŋ-to
 DIST.DEM spool.thread hold-SIM.CVB hold-SIM.CVB hold-SIM.CVB
 laŋk-tsi ʔe
 go-3DU HS

'From above, their mother called out. (a) 'Hey, come up, come up, come up', she said and she threw them that spool of thread. (b) They went, holding, holding that spool of thread. (c)'

One issue of interest is the relationship between the hearsay marker *ʔe* and the expression of reported speech. A number of discussions about quotative verbs mention that one of the functions can be the expression of hearsay evidentiality (Matic & Pakendorf 2013; Chappell 2008). There is no plausible diachronic connection in Thulung between *ʔe* and the main quotative verb, which is *rwa:mu*, used to embed reported speech.

Similarly to Chantyal (Noonan 2006), Thulung uses a direct speech to embed quoted material, as in (40).

- (40) go mi-bi-ŋu rwak-ta
 1SG NEG-come-1SG say-3SG.PST
 'He said he wouldn't come.' [lit. He said "I won't come."]

In some cases, a participial form of the quotative verb can be used to embed the direct speech (including thought),²⁴ as in (41).

²⁴ This rejoins Chappell (2002: 348; 2008)'s proposal for possible grammaticalizations from *verba dicendi*.

- (41) gu-ka ne set-to **rwak-pa** mim-d̥a-m
 3SG-ERG TOP kill-1SG>3SG.PST **say-ACT.PTCP** think-3SG>3SG.PST-NMLZ
 ba-ira
 be-3SG.PST
 'She had thought she killed him.' [lit. She had thought, "I killed him"]

In other cases, and this is probably the relevant path of development for the complementizer in (41), *rwakpa* (and the entire quoted phrase) is a relative clause with head noun *lwa*, 'story'. In (42), this construction is shown, with the quoted material marked with *ʔe* (the use of the hearsay marker is not obligatory in this construction, but found regularly):

- (42) ʌno meno ʌ-mri ʔe **rwak-pa** lwa be-m-thal-miri
 here there go-3PL.PST HS **say-ACT.PTCP** story do-3PL-HAB-3PL>3SG.PST
 'They did [=told] the said story that they went here and there.'

The grammatical sketch presented above is by design brief, in order to give readers a sense of the morphosyntax relating essentially to the nominal domain. For additional information, see Allen (1975), Lahaussais (2002; 2004) as well as the more specialized articles mentioned above.

2.2 The Thulung verbal template

The verb complex of Thulung is described here by making use of the Thulung verbal template, shown in Figure 3 below. This is a convenient structuring device, which allows me to discuss, in addition to the grammatical material that fills various slots of the template, functional domains related to the affixal material. This artifice (of sorts) will thus allow me to cover the following topics, which have been the subject of previous work (in the form of publications or oral presentations; the latter are listed only when no relevant publications exist):

- the expression of negation (Lahaussais 2013, conference presentation)
- ideophones (Lahaussais 2017b, for Khaling), (Lahaussais & Treis 2019, conference presentation)
- the inflectional classes of Thulung (Lahaussais 2011; Lahaussais submitted)
- the multiple exponence exhibited by some derived verbs (Lahaussais submitted), (Lahaussais 2019e, conference presentation)
- polypersonal indexation (Lahaussais 2011; Lahaussais 2020a)
- the expression of aspect and Aktionsart (Lahaussais 2002; Lahaussais submitted)
- Associated Motion (Jacques, Lahaussais & Zhang In press), (Lahaussais 2018b; Jacques & Lahaussais 2018, conference presentations)
- reflexive constructions (Lahaussais 2016a; Lahaussais Submitted), (Jacques, Lahaussais & Rai 2016, for Khaling)
- valence changes, such as applicatives and causatives (Lahaussais submitted)
- optative mood and the expression of apprehension (Lahaussais In press)
- irrealis mood, and counterfactual constructions (Lahaussais 2013, conference presentation)

I additionally discuss topics which represent new areas of exploration, such as the presence, among Thulung inflectional classes, of traces of older derivational suffixes, and possible evidence of an alignment hierarchy within the indexes of Thulung. Both are treated in subsections of 2.2.2.

The notion of template has been frequently used for the description of the verbal morphology of languages in which verbal affixes occur in a fixed order, such as Athapaskan languages (Kari 1989). For a typology of the use of this conceptual tool, as well as for additional terms used to refer to it, see Good (2011; 2016).

The definition provided in Good (2016: 7) for a template is "an analytical device used to characterize the linear realization of a linguistic constituent whose linear stipulations are

unexpected from the point of view of a given linguist's approach to linguistic analysis." The mention of unexpectedness in this definition reasserts how the use of the template as an analytical tool is particularly relevant to Kiranti languages: the description of verb forms in terms of ordered slots reveals the phenomenon of multiple exponence, whereby arguments are indexed more than once in a single verb form, with these additional indexes occurring in an unexpected linear order.²⁵ This is discussed in 2.2.3.2.1 below.

In the description of Kiranti verbal morphology, many authors have used templates in order to describe the order in which various prefixes and suffixes attach to the verb root. Among these are van Driem (1987; 1993b; 1990), for Limbu, Dumi and proto-Kiranti, respectively; Bickel (1996: 52), for Belhare; Rutgers (1998: 110), for Yamphu; Jacques et al. (2012), for Khaling; Schackow (2015: 207), for Yakkha. In the case of languages with free prefix ordering, the template only features verbal suffixes; see Doornenbal (2009: 166), for Bantawa²⁶, Bickel (2007: 49 ff), for Chintang.

As far as Thulung is concerned, prefixes and suffixes that attach to the verb root occur in a fixed order, which is represented in Figure 3 below. The presentation in Figure 3 takes a few short-cuts, with respect to the number of slots presented: as an example, the material within suffixal slot 3 is made of up indexes specifying the person/number/clusivity/role of a single argument (for intransitives) or two arguments (for transitives), yet I have adopted the convention of conceiving of transitive indexes as global *scenarios* expressing the features of the two arguments, with the formula A>P.²⁷ Other linguists working on Kiranti languages often break down these indexes into the elements that can be found within (see 2.2.3.3 for details), resulting in templates with a great many indexation slots, as there are multiple suffixes (and sometimes prefixes, depending on the language) for the indexation of transitive arguments, whereas I posit a single one (in addition to suffixal slot where the formatives involved in multiple exponence are found).²⁸

²⁵ For a rather different, but very interesting, case of unexpectedness in the ordering of affixal material, see Jacques (2013) on Japhug categories grammaticalizing into affixes unexpected on account of basic word order.

²⁶ Note that the slot analysis presented p. 166 is only for suffixes, as there is free ordering for verb prefixes (ibid, 167 ff).

²⁷ As past indexes are most often portmanteaux, I gloss them A>P.PST; non-past indexes do not have any explicit glossing of their tense, as it is unmarked.

²⁸ Kari (1989) breaks down the template proposed for the Athapaskan verb complex -- which is considerably more complex than the Kiranti, with 253 prefixes and 16 positions posited by Young and Morgan (1987) for Navajo

prefixal slot	root	suffixal slot 1: agreement suffixes	suffixal slot 2: derivational suffixes ²⁹			suffixal slot 3: agreement suffixes; non-finite suffixes	suffixal slot 4: mood markers
negative (2.2.1.1)	stem alternations (2.2.2)	partial indexation markers (2.2.3.2.1)	Associated Motion (2.2.3.2.3)	valence changes (2.2.3.2.4)	Aktionsart (2.2.3.2.5)	indexation markers (2.2.3.3)	Optative (2.2.3.4.1)
							Irrealis (2.2.3.4.2)
ideophones of type A (2.2.1.2)						Infinitive	
						Anterior converb	
						Simultaneous converb	

Figure 3. The verbal template of Thulung

(material discussed in chapter 2 is in bold, along with relevant sections)

(Kari 1989: 426) -- into *zones*, such as the pronominal zone (ibid, 438). The 'suffixal slots' in Figure 2 thus probably correspond more closely to Kari's zones, as they are areas of the template where functional domains are expressed. Kari also lays out a proposal for methodology in ensuring terminological and methodological consistency in order to ensure that the position class analyses can be use for linguistic comparison; these practices could presumably be successfully applied to Kiranti language descriptions (in line with what (van Driem 1990) has done for a few languages).

²⁹ The order of the derivational suffixes for this suffixal slot is as presented: Associated Motion suffixes occur before valence changing suffixes which, in turn, occur before Aktionsart suffixes. This can be seen in examples (61) and (62).

There is additionally a prohibitive construction using a non-imperative verb form and epistemic authority marker *sə* which expresses negative polarity even though it involves no negative marker (see the section on discourse markers in 1.3.2.2 above, and specifically discussion of (38)).

2.2.1.2 Ideophones

The reason ideophones are placed in the same prefixal slot as the negative prefix is that negated ideophonic forms are not found in Thulung. This features concords with what has been claimed in the typological literature (such as Klamer (2001: 168)). Note that the ideophones which are in the prefixal slot are a single type (type A) of Thulung ideophones; other ideophone types, described below, are not considered to have a slot in the verbal template.

In Lahaussais (2017b), I have described the three types of ideophones I identified in Khaling: a monosyllabic type, often characterized by falling tone, always found preverbally and contributing semantics of suddenness and forcefulness to the predicate, labeled type A; a trisyllabic type, labeled type B, often with a $C_1VC_2VC_2V$ pattern and frequently conveying sound and motion; and a third type exhibiting reduplication, labeled type C, more lexicalized than the other two types and used with an adverbial/adjectival function. Together, these three types cover the entire spectrum of sensory perceptions found in Dingemanse (2012)'s implicational hierarchy (reproduced in Figure 4 below, along with the ranges of the three types of Khaling ideophones).

Through new, unpublished material on Thulung as well, it is apparent that ideophones are a large word class. This may be surprising considering that they have frequently been omitted from grammatical descriptions of Kiranti languages (the exceptions are Bantawa, Chintang and Yakkha (Doornenbal 2009; Rai & Winter 1997; Rai et al. 2005; Schackow 2015)). While the distribution in descriptions makes it look like ideophones are only present in the Central and Eastern Kiranti languages, it appears that their absence elsewhere is a grammaticographical issue rather than a linguistic one. This is discussed at greater length in 3.1.1 below.

Thulung ideophones, like Khaling ideophones, can be divided into three types according to the following phonotactic, morphological and semantic characteristics:

Type A:

CV, CVŋ, CVp, CVr, CVCV

Associated with a single verb: set collocations of IDEO + V

Cannot occur with negated verbs

Meaning always +/- 'suddenly', 'forcefully'

No derivation possible

No obvious lexical sources

- (47) **hu-bak-ta** tsʌŋ-lan̄ka las-ta **hu-bak-ta** ma
 IDEO-rise-3SG.PST back-ABL go-3SG.PST IDEO-rise-3SG.PST CONJ
 me thʌ:sa-ku thʌŋki **dzhoŋ-kos-tu** ma
 DIST.DEM pine-GEN resin IDEO-pick.up-3SG>3SG.PST CONJ
thoŋ-kot-d̄u ʔe u-bui-d̄ʌ:la
 IDEO-pour-3SG>3SG.PST HS 3SG.POSS-head-above

'He **suddenly** got up, he went from her backside and got up **suddenly**, and picked up the pine resin **suddenly**, and poured it **suddenly** on the top of her head.'

Type B

CV₁-CV₁; CV₁CV₁- CV₁CV₁

Involves total reduplication, of either a single syllable or two syllables

Can take derivational suffix *-ja* (>ADV) or *-m* (>NMLZ)

Have specific meanings

(Color terms are of this type)

- (48) **khurukhuru** mi-dzəpa lapdi mi-lak-tsi dzəpa lapdi
continuously NEG-good road NEG-go-2DU good road
 lak-tsi **khrekhre-ja** be-pa lapdi mi-lak-tsi **plʌplʌ-ja**
 go-2DU **bumpy-ADV** do-ACT.PTCP road NEG-go-2DU smooth-ADV
 be-pa lapdi lak-tsi
 do-ACT.PTCP road go-2DU

'Do not go **continuously** on the bad road, go on the good road, do not go on the **bumpy** road, go on the **smooth** road.'

- (49) **tsiri-maksi** khreŋ-d̄u-lo ne mertsip-ka
biting.by.small.insect-ADV bite-3SG>3SG.PST-SEQ TOP 3DU-ERG
 me se:r tsəm-tsi

DIST.DEM louse crush-3DU>3SG.PST

'When [**the louse**] **bit**, they crushed that louse.'

Type C

C₁V₁C₂V₁C₂V₁(C₂V₁)....

Involves partial reduplication (and can be lengthened for dramatic effect)

Can take derivational suffix *-maks* (>ADV)

Have specific meanings

Ideophones of type C are often very close to Nepali ideophones of the same shape; see Dahal (1999) for some examples.³⁰

- (50) u-sem ne **tolololole** thʉs-tʉ
 3SG.POSS-hair TOP **so.long** pull-3SG>3SG.PST
 'He pulled his hair **so, so long**.'

The three ideophones types found in Thulung and their semantics, morphology, phonotactics are very close those of neighboring Khaling, with the significant difference that many fewer ideophones have been found in Thulung than in Khaling. Using the Kiranti Comparable Corpus, made up of narrative data from Thulung, Khaling and Koyi (described in 3.3.1) to study ideophones reveals that the Khaling texts in the corpus have many more ideophones than the Thulung (about 2:1, consistently across different speakers; interestingly, the Koyi data reveals no ideophones at all) when considering utterances in both languages referring to the same events in a shared myth. Individual differences across speakers are of course to be expected with a word class such as ideophones, claimed to be very sensitive to register (Haiman 2018: 131; Kilian-Hatz 2001: 156) and to inter-speaker variation (Bernard Wälchli, LingTyp discussion list, March 22nd, 2019: "Ideophones are maybe the most sociolinguistic part of speech there is"), and the conditions of data collection may well have influenced their use across the two languages. The differences in numbers of ideophones across

³⁰ The larger question of the effect of language contact on ideophones is something I plan to explore through a LABEX EFL project on ideophones and interjections, led with Yvonne Treis, and is particularly interesting in the context of Kiranti languages.

languages nonetheless appears significant, and correlates with the general proportion of ideophones in my glossaries for both languages. (It is also interesting that no ideophones at all were found in the Koyi part of the comparable corpus.) In Khaling, I spent considerable time doing elicitation work on ideophones -- something which was very enjoyable to the consultants I was working with -- and perhaps this somehow triggered their increased use in narratives, with the community realizing that these "funny sounds" were a legitimate part of the grammar of their language. As I have not yet carried out elicitation on ideophones in Thulung, I cannot say whether a focus on this word class will make them more salient and increase their numbers in spontaneous narratives.

Dingemanse's (2012: 663) implicational hierarchy allows us to map the range of sensory domains covered by the three types of ideophones in Khaling (Lahaussois 2017b: 20): Type A is mostly focused on movement, but also covers sound (insofar as the two are very often inseparable, with certain ideophones sound-symbolically capturing the sound which is affiliated with a type of movement; see for example Hinton et al. (1994: 9)); type B covers both sound and movement, and additionally extends to other domains; type C covers movement, visual patterns, other sensory perceptions, inner feelings and cognitive states. The three types of Khaling are shown mapped onto sensory domains in Figure 4.

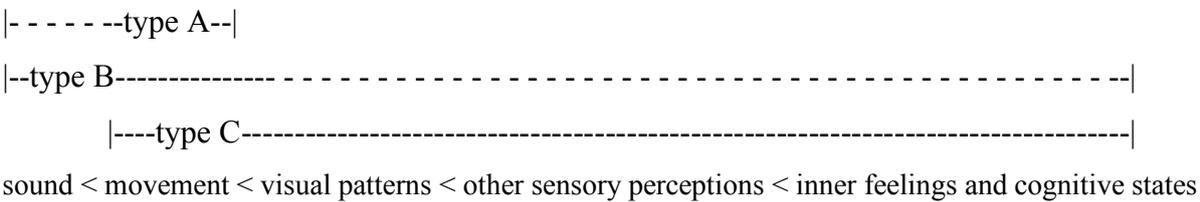


Figure 4. Khaling ideophones with respect to
Dingemanse 2012's implicational hierarchy

While further research is needed in order to confirm similar patterns with Thulung, and specifically whether all domains, in addition to sound and movement, are covered by ideophones of type C, the situation in Khaling suggests the extent to which the implicational hierarchy is a powerful methodological and analytical tool.

As to whether to consider that Type A ideophones in Thulung are prefixal, they have characteristics that are quite different from the two other types of ideophones in the language: Nothing can be inserted between the ideophone and the verb, and ideophones cannot take derivational morphology like other ideophones types do: type B can be used as the bare ideophone or undergo derivation with *-ja* or nominalizer *-m*; type C can be used as the bare

ideophone or be augmented by *-maksī*. Both *-maksī* and *-ja* are adverbializers, even though the bare noun and the derived forms have the same properties and distributions. As for nominalization, this is necessary before an attributive use with type B ideophones, and is thus a source of derived adjectives.

Another particularity of Type A ideophones which distinguishes them from other types is their fixed semantic contribution, which applies to the entire class, conveying a quality of suddenness and/or forcefulness to the action. No lexical sources for Thulung Type A ideophones have been found (although this is not true of Khaling; see Lahaussais (2017b)). This is in opposition to Type B and C ideophones, which contribute distinct semantics (and are thus given lexical glosses).

This semantic particularity can be combined with the fact that, as far as my corpus allows me to determine, the ideophone-verb pairs are fixed collocations, with a single ideophone used with a single verb. My considerably larger ideophone corpus in Khaling revealed that while this was a strong tendency, there were sometimes exceptions: labile verbs and verbs related through productive derivation generally share ideophones (such as *phlotso leine/lênne* 'emerge suddenly/take out suddenly'; *pak thoômne/thoôm sine* 'beat forcefully/bump into something forcefully'); homophonous verbs share ideophones in about 50% of cases (such as *wî: sânnē* 'cover quickly/pull grain off stalks suddenly'; *piti khlênne* 'burn suddenly/misstep suddenly'); occasionally verbs with similar or connected semantics combine with the same ideophones (*soro hoômne* 'slurp quickly'; *soro phâmne* 'drink up quickly'). It remains to be seen what the distribution of Type A ideophones is in Thulung with respect to forming consistent collocations with single verbs.

In (51), I provide a final example, which includes two ideophones of Type A:

- (51) u-khel phuttai **tsha**-ghrok-tu ni tʌ ʌni
 3SG.POSS-leg suddenly **IDEO**-throw-3SG>3SG.PST AT EMPH³¹ and
 pakha-nu **lep**-bok-tu
 outside-LEV.LOC **IDEO**-throw.down-3SG>3SG.PST
 'He threw his leg suddenly, he threw it down suddenly outside.' [after being bitten by a fish when dangling his leg in the river]

³¹ The glosses for these two particles, both from Nepali, come from Yliniemi (2016) and Genetti (1994), respectively, where they stand for attention and emphasis.

Their prosody strongly suggests that these ideophones are part of the same word as the verb they combine with.

Because the Thulung data presented above represents an analysis based on corpus materials, but has involved no targeted work (such as elicitation), this is a topic I am very eager to work on during future field trips. I believe the study of ideophones in Thulung would additionally very much benefit from a better understanding of the ideophonic systems present in Nepali, especially insofar as the Type C ideophones described above are very similar to what is found in Nepali (sometimes even using the same syllables for the same meaning). There as, as far as I am aware, no studies of how language contact affects ideophones, and the case of Thulung and Nepali would be an excellent place to look into the phenomenon.

2.2.2 Root

The material presented in section 2.2.2 concerns verb roots from the point of view of two features: the stem alternations found in Thulung and the distribution patterns of these stems for different verb classes, and traces in Thulung of no-longer productive post-final suffixes which have a strong presence in other languages. Section 2.2.2.1 relies on data presented most explicitly in Lahaussais (2011), while sections 2.2.2.2 and 2.2.2.3 present new analyses of data.

2.2.2.1 Verb classes and stem alternation patterns

As is typical of Kiranti languages, underived Thulung verbs³² have a monosyllabic verb root of the form (C_i)V(C_f). The final consonant (C_f) of the root is the basis for sorting verbs into verb classes, and determines the stems for any given verb. In spite of the Kiranti-wide template³³ for the verb root of simplex verbs, in other words those which are not augmented by derivational formatives (discussed in 2.2.3.2), it is worth questioning how one assigns material to the stem and to the inflectional suffixes. This question is a central one in inflectional morphology, and discussed in a great many publications (Bonami 2012; Baerman & Corbett 2012; Spencer 2012; Corbett 2015). The rule of hand applied here is to take maximally recurring inflectional material

³² This is not the case for borrowed verbs, which tend to have a bisyllabic root, or for lexicalized verbs incorporating no-longer productive derivational material.

³³ Note that this abbreviated template only applies to the verb root and not to the verb complex. There is great diversity in the templates for verb complexes in different languages, complicating reconstruction across the group.

-- which is relatively stable (although see 2.2.3.3 for discussion of allomorphy in indexes) -- and to claim that this material makes up the indexes (which code person, number, clusivity, role, and tense). All remaining material is assigned to the verb stem.

The main Thulung verb classes, as described in Lahaussais (2011), are listed in Table 8 below, along with their stem alternations, with the root initial and vowel abstracted as 'CV'.

verb class	root final ³⁴ (C _f)	stems	number (in dictionary)
p-class	p	CVp, CVm	51
t-class	t	CV ^t ³⁵ , CVn, CV	99
k-class	k	CVk, CV: ³⁶	94
m-class	m	CVm	34
n-class	n	CVn	8
ŋ-class	ŋ	CVŋ, CV:	14
l-class	l	CVl	25
r-class	r	CVr	31
s-class ³⁷	s	CVs, CV:	31

³⁴ Because of stem alternation with some classes, root final C_f is not (necessarily) found in all stems. The IPI (A/S) form always uses the stem with the root final, and can reliably be used for verb class identification. The C_f, once determined, always makes it possible to predict the stems for the verb--except for the patterns found with the sub-classes discussed in 2.2.2.2.

³⁵ Allomorphs for this stem are CVt, CVd and CVr.

³⁶ The long vowel is a result of compensatory lengthening due to the loss of the root final, one of the morphophonological processes giving rise to stem alternations proposed by Michailovsky (1975: 187). The V is shortened when it is a closed syllable, through re-syllabification with the initial of the suffix. This applies to all verb classes with a stem with a long vowel due to compensatory lengthening.

³⁷ Positing a verb class with -s as the C_f is not a neutral decision, as there is good evidence that the -s in verbs of this class comes not from a true final consonant but rather from what was originally a post-final -s with a reflexivizing/middle/detransitivizing function. For example, Michailovsky (1975: 187), writing about Bahing, claims that "no roots are identified with final /s/; all cases of /s/ preceding a verbal suffix are interpreted as the /s/ augment of the reflexive stem rather than as the inherent final of the root". The very high number of intransitive verbs in the s-class suggests that this is true, but the organization reflected in Table 8 is a means of accounting *synchronically* for the stem patterns that are found in Thulung.

Table 8. The verb classes of Thulung, with root final and alternating stem types

The paradigmatic distribution of stems is fairly straightforward, grouping into the following behaviors: verb classes with no stem alternation; and verb classes with stem alternations. Additionally, transitive and intransitive verbs of the same class do not have the same stem distributions, although similarities in the distribution patterns are visible. As a result, each verb class has an intransitive and transitive stem distribution associated with it.

Verb classes with a single stem throughout the paradigm, for both intransitive and transitive, are *m*-, *n*-, *l*- and *r*-classes. This is exemplified with the *l*-class transitive verb *jalmu*, 'strike', in Table 9 below.³⁸ Note that in the paradigms in this section, all paradigms are presented with a default 3SG object. This is to avoid the very large number of cells (11x11) that are necessary to illustrate full paradigms of all person/number scenarios; Section 2.2.3.3 shows full paradigms.

	NPST	PST
1SG>3SG	jal-u	jal-to
1DI>3SG	jal-tsi	gwak-tsi
1PI>3SG	jal-i	jal-li
1DE>3SG	jal-tsuku	jal-tsoko
1PE>3SG	jal-ku	jal-toko
2SG>3SG	jal-na	jal-na
2DU>3SG	jal-tsi	jal-tsi
2PL>3SG	jal-ni	jal-ni
3SG>3SG	jal- u	jal-lu
3DU>3SG	jal-tsi	jal-tsi
3PL>3SG	jal-mi	jal-miri

Table 9. Paradigm for a verb with no stem alternation

³⁸ For all verb classes, the citation form uses the stem which is found in the non-past 1DI cell (intransitive or transitive).

Verb classes which have a vowel-final root generally do not have stem alternation. There are only three exceptions to this: *bomu*, 'do', *pomu*, 'eat', and *bumu*, 'be'. Table 10 below illustrates the stem alternations for *bomu*³⁹ and *bumu*. Note that as the former is transitive and the latter intransitive, the indexes are not the same, nor are the stem distributions across the paradigms.

	NPST	PST		NPST	PST
1SG>3SG	be-u	be-uto	1SG	bu-ŋu	ba-ŋro
1DI>3SG	bo-tsi	be-ttsi	1DI	bu-tsi	ba-ttsi
1PI>3SG	bi-i	bi-iri	1PI	bu-i	ba-iri
1DE>3SG	bo-tsuku	be-ttsoko	1DE	bu-tsuku	ba-ttsoko
1PE>3SG	bo-ku	be-ttoko	1PE	bu-ku	ba-ttoko
2SG>3SG	bo-na	be-nna	2SG	bu-na	ba-nna
2DU>3SG	bo-tsi	be-ttsi	2DU	bu-tsi	ba-ttsi
2PL>3SG	bo-ni	be-nni	2PL	bu-ni	ba-nni
3SG>3SG	bu-u	bu-uru	3SG	bu	ba-ira
3DU>3SG	bo-tsi	be-ttsi	3DU	bu-tsi	ba-ttsi
3PL>3SG	bo-mi	be-mri	3PL	bu-mi	ba-mri

Table 10. V-class verbs *bomu* 'do' and *bumu* 'be'

Verbs with a C_f of *k* follow the patterns in Table 11 for transitive *gwa:mu* 'give' (left-hand side) and intransitive *bimu* 'come' (right-hand side).

	NPST	PST		NPST	PST
1SG>3SG	gwak-pu	gwak-to	1SG	bi-ŋu	bi-ŋro
1DI>3SG	gwa:-tsi	gwak-tsi	1DI	bi-tsi	bik-tsi
1PI>3SG	gwak-i	gwak-ti	1PI	bik-i	bik-ti
1DE>3SG	gwa:-tsuku	gwak-tsoko	1DE	bi-tsuku	bik-tsoko
1PE>3SG	gwa:-ku	gwak-toko	1PE	bi-ku	bik-toko
2SG>3SG	gwa:-na	gwak-na	2SG	bi-na	bik-na
2DU>3SG	gwa:-tsi	gwak-tsi	2DU	bi-tsi	bik-tsi

³⁹ The verb *pomu*, 'eat' has stems which pattern as those of *bomu*; the only difference is the initial consonant.

2PL>3SG	gwa:-ni	gwak-ni	2PL	bi-ni	bik-ni
3SG>3SG	gwak-ʉ	gwak-tʉ	3SG	bi	bik-ta
3DU>3SG	gwa:-tsi	gwak-tsi	3DU	bi-tsi	bik-tsi
3PL>3SG	gwa:-mi	gwak-miri/ gwa-mri ⁴⁰	3PL	bi-mi	bi-mri

Table 11. Stem distribution patterns for transitive (left) and intransitive (right) k-class verbs

Verbs with a C_f of *p* follow the patterns in Table 12 for transitive *remmu* 'see' (left-hand side) and intransitive *khrammu* cry (right-hand side).

	NPST	PST		NPST	PST
1SG>3SG	rep-u	rep-to	1SG	khram-ŋu	khrap-ŋoro
1DI>3SG	rem-tsi	rep-tsi	1DI	khram-tsi	khrap-tsi
1PI>3SG	rep-ri	rep-ri	1PI	khrap-ri	khrap-ri
1DE>3SG	rem-tsuku	rep-tsoko	1DE	khram-tsuku	khrap-tsoko
1PE>3SG	rem-ku	rep-toko	1PE	khram-ku	khrap-toko
2SG>3SG	rem-na	rep-na	2SG	khram-na	khrap-na
2DU>3SG	rem-tsi	rep-tsi	2DU	khram-tsi	khrap-tsi
2PL>3SG	rem-ni	rep-ni	2PL	khram-ni	khrap-ni
3SG>3SG	rep-rʉ	rep-rʉ	3SG	khram	khrap-ta
3DU>3SG	rem-tsi	rep-tsi	3DU	khram-tsi	khrap-tsi
3PL>3SG	rem-mi	rep-miri	3PL	khram-mi	khrap-miri

Table 12. Stem distribution patterns for transitive (left) and intransitive (right) p-class verbs

⁴⁰ Note that this cell features two possible stems: gwak- and gwa:-. There is competition in the 3PL>3SG cell between a form with the expected stem--the one including the root C_r , which occurs throughout the past tense in verb classes with similar distribution patterns (see Tables 12 and 13)--and a stem which is phonologically reduced to avoid the -km- sequence (this is a matter of phonotactics). Verbal forms with a -km- always appear to have a phonotactically simpler variant as an option: another example is *rwakma~rwa:ma* (say-PST.PTCP). Note however that the same 'simplification' is not found with nouns: the pluralizer -mim combines with -k final nouns, without any evidence of reduced forms.

The *gwa* stem (where we might expect *gwa:*) is due to resyllabification with the initial m in the suffix, which reduces vowel length: divided according to syllables, the form is *gwam-ri*.

Verbs with a C_f of *t* follow the patterns in Table 13 for transitive *senmu* 'kill' (left-hand side) and intransitive *hunmu* 'fly' (right-hand side).

	NPST	PST		NPST	PST
1SG>3SG	seṭ-pu	seṭ-to	1SG	hun-ŋu	huṭ-ŋoro
1DI>3SG	sen-tsi	seṭ-tsi	1DI	hun-tsi	huṭ-tsi
1PI>3SG	seḍ-i	seṭ-dḷi	1PI	huḍ-i	huṭ-dḷi
1DE>3SG	sen-tsuku	seṭ-tsoko	1DE	hun-tsuku	huṭ-tsoko
1PE>3SG	sen-ku	seṭ-toko	1PE	hun-ku	huṭ-toko
2SG>3SG	sen-na	seṭ-na	2SG	hun-na	huṭ-na
2DU>3SG	sen-tsi	seṭ-tsi	2DU	hun-tsi	huṭ-tsi
2PL>3SG	sen-ni	seṭ-ni	2PL	hun-ni	huṭ-ni
3SG>3SG	seḍ-ḁ	seṭ-dḁ	3SG	hun	huṭ-dḁ
3DU>3SG	sen-tsi	seṭ-tsi	3DU	hun-tsi	huṭ-tsi
3PL>3SG	sen-mi	seṭ-miri	3PL	hun-mi	huṭ-miri

Table 13. Stem distribution patterns for transitive (left) and intransitive (right) *t*-class verbs

Tables 11-13 pattern very similarly: the stem with the root final is found throughout the past forms (with a few exceptions for *k*-class verbs) and in the transitive non-past 1SG, 1PI and 3SG forms. With intransitives, the 'past' stem is only found with non-past 1PI forms.

As with verbs of the *p*-, *t*- and *k*-classes, verbs of the *s*- and *ŋ*-stem classes share a stem distribution pattern. This is seen in the similar pattern in the greyed cells in Tables 14 and 15 below, showing that most of the paradigms use a single stem, apart from 1PI and 3SG (but not 3SG.NPST for intransitives), which have a different stem.

Verbs with a C_f of *ŋ* follow the patterns in Table 14 for transitive *ḍo:mu*, 'pound rice' (left-hand side) and intransitive *swa:mu* 'flee' (right-hand side).

	NPST	PST		NPST	PST
1SG>3SG	ḍo:-u	ḍo:-uto	1SG	swa-ŋu	swa-ŋro
1DI>3SG	ḍo:-tsi	ḍo-ttsi	1DI	swa:-tsi	swa-ttsi
1PI>3SG	ḍoŋ-i	ḍoŋ-ri	1PI	swaŋ-i	swaŋ-dḷi
1DE>3SG	ḍo:-tsuku	ḍo-ttsoko	1DE	swa:-tsuku	swa-ttsoko

1PE>3SG	ɖo:-ku	ɖo-ttoko	1PE	swa:-ku	swa-ttoko
2SG>3SG	ɖo:-na	ɖo-nna	2SG	swa:-na	swa-nna
2DU>3SG	ɖo:-tsi	ɖo-ttsi	2DU	swa:-tsi	swa-ttsi
2PL>3SG	ɖo:-ni	ɖo-nni	2PL	swa:-ni	swa-nni
3SG>3SG	ɖoŋ-ɸ	ɖoŋ-rɸ	3SG	swa:	swaŋ-ɖa
3DU>3SG	ɖo:-tsi	ɖo-ttsi	3DU	swa:-tsi	swa-ttsi
3PL>3SG	ɖo:-mi	ɖo-mri	3PL	swa:-mi	swa-mri

Table 14. Stem distribution patterns for transitive (left) and intransitive (right) ŋ-class verbs

Verbs with a C_r of *s* follow the patterns in Table 15 for transitive *lwa:mu*, 'see' (left-hand side) and intransitive *semu* 'fart' (right-hand side).

	NPST	PST		NPST	PST
1SG>3SG	la-u	la-uto	1SG	se-ŋu	se-ŋro
1DI>3SG	lwa:-tsi	lwa-ttsi	1DI	se-tsi	se-ttsi
1PI>3SG	lwas-i	lwas-ti	1PI	ses-i	ses-ti
1DE>3SG	lwa:-tsuku	lwa-ttsoko	1DE	se-tsuku	se-ttsoko
1PE>3SG	lwa:-ku	lwa-ttoko	1PE	se-ku	se-ttoko
2SG>3SG	lwa:-na	lwa-nna	2SG	se-na	se-nna
2DU>3SG	lwa:-tsi	lwa-ttsi	2DU	se-tsi	se-ttsi
2PL>3SG	lwa:-ni	lwa-nni	2PL	se-ni	se-nni
3SG>3SG	lwas-ɸ	lwas-tɸ	3SG	se	ses-ta
3DU>3SG	lwa:-tsi	lwa-ttsi	3DU	se-tsi	se-ttsi
3PL>3SG	lwa:-mi	lwa-mri	3PL	se-mi	se-mri

Table 15. Stem distribution patterns for transitive (left) and intransitive (right) *s*-class verbs

It is noteworthy that the existence of a *s*-class of verbs is a matter of debate in various Kiranti languages (see, eg., Jacques (2017)). The *s*-class is different from other verb classes, in that some verbs clearly have the *-s* in addition to another C_r (see, for example, in a form like *laks-i*, 'go-1DI'). This suggests strongly that the origin for verbs with stems patterning like Table 15 is that they have what was originally a post-final suffix *-s*, and that this has resulted in the distribution of stems we currently find. Among the factors pointing in this direction is the fact that *s*-class verbs are more typically intransitive than transitive (see

Lahaussais (2016a)), suggesting a derivational origin with a detransitivizing function. Though the *-s postfinal is sometimes associated with causativization (this is the case for Limbu; see Michailovsky (1985), van Driem (1987)), the reflex is associated with reflexive derivation (or detransitivization) in some Kiranti languages and beyond (Benedict 1972: 97–98; van Driem 1988). It is not however clear what mechanism could result in the s-class's CVs stem only surfacing in 1PI and 3SG (non-past and past) cells. While the historical development of the current arrangement of verb inflections is relevant and interesting, it is also important to look at the data through a synchronic lens that allows us insight into how speakers classify verbs and inflect them. In such a perspective, Thulung is considered to have an s-class of verbs in the same way as it does other verb classes.

In attempts to describe the stem distribution patterns presented in Tables 9-15, I have used sometimes awkward formulations, highlighting a basic distinction between non-past and past stems, and then adding the proviso that scenarios with 1SG, 1PI, 3SG S/A arguments (the exact details depending on the verb class and transitivity) also used the past stem (Lahaussais 2011; Lahaussais 2004). This results from part of the pattern being motivated (by tense), and the other unmotivated (the few non-past forms using the same stem as the past forms).

An interesting way of expressing the data is to refer to the notion of morpheme, a morphological pattern which unites a set of paradigm cells yet which is not motivated (synchronously) by any phonological or morphosyntactic features. Hecce (2020) is a recent typological overview presenting the history of the concept, which first appears in Aronoff (1994). The idea of morpheme might better help us capture the patterns connected with the verb classes of Thulung, and the interconnections between them.⁴¹ It is clear that there is something different about the stems used for 1SG, 1PI and 3SG forms (for which non-past and past often match), when compared to other person/number combinations, but describing this in terms other than lists of person/number combinations is difficult.⁴²

The verb classes and stem distribution patterns described above apply to most Thulung verbs, but as my corpus has increased in size, the data has revealed itself to be somewhat

⁴¹ Hecce presents (my) data from Thulung (*ibid*, 266-267) in terms of morphemes, albeit limited to two verb classes.

⁴² In keeping with the tradition, which comes from Maiden (2005), of giving morphemes labels based on the letter shapes they map out, Bernard Fradin has suggested (somewhat in jest) the label "broken comb morpheme" for the pattern found with verb classes -p, -t, -k: while the comb base is stable (the past forms), some of the teeth (the non-past forms) are missing.

messier (a phenomenon that I have not yet described in publications). There are two ways in which patterns differ from what has been described above:

a) Some verbs alternate between stems associated with two (or more) different verb classes, often varying freely, with their behavior suggesting they are simultaneously felt to belong to several verb classes.

One example is the verb *banmu~ba:mu* 'wear ornaments', illustrated in Table 16, for which forms have variants reflecting the stems of t-class (non-past 1PI, 3SG and all past forms), V-class (all non-past forms except for 1PI and 3SG) and n-class verbs (all left-most forms in both non-past and past).

	NPST	PST
1SG>3SG	ba:pu~banpu	baṭto~banto
1PI>3SG	ba:ḍi~bandi	baṭḍi~bandi
1PE>3SG	ba:ku~banku	baṭtoko~bantoko
2SG>3SG	ba:na~banna	baṭna~banna
2DU>3SG	ba:tsi~bantsi	baṭtsi~bantsi
2PL>3SG	ba:ni~banni	baṭni~banni
3SG>3SG	ba:ḍu~bandu	baṭḍu~bandu
3PL>3SG	ba:mi~banmi	baṭmiri~banmiri

Table 16. Paradigm for *banmu~ba:mu*, with variants

The situation described in a) will not be discussed further herein, because the patterns are as yet unclear to me and appear to be very much a matter of inter-speaker variation (but not always, as I also encounter cases where a same speaker has freely varying stems for what appears to be the same verb). One possible explanation for the verb class 'instability' is the natural variability found in large paradigms, suggestive of the artificiality of collection methods (which, short of a corpus of a size very rare in endangered languages documented by a single person or small team, are inevitable, as far as I can tell, in order to get complete paradigms); another hypothesis is that there are two phonologically very close verbs, whose semantics cannot be distinguished by speakers, and that speakers are alternating between the two. There are examples of this in the verb lexicon: Thulung has verbs *ja:mu* (C_f k) and *jalmu*, both meaning 'hit' with no identifiable differences in semantics, or *hormu* and *holmu*, both 'open'. The verbal paradigms are distinct, but one can imagine how if the C_f's were such

that part of the paradigm used similar stems (eg C_f -n and -t both use a CVn stem), confusion could arise. These hypotheses are captured by Thornton (2019a; 2019b)'s notion of overabundance: defined as "the situation in which two (or more) inflectional forms are available to realize the same cell in an inflectional paradigm" (Thornton 2019a: 223), there are a number of ways in which this situation can manifest. "Certain cell mates can be used interchangeably in the same context, even by the same speaker, with no difference in style and meaning; other cell mates are subject to various kinds of conditions, so that different forms are used in different styles, registers, social or geographic dialects, or in different semantic, syntactic or pragmatic contexts" (Thornton 2019b). Given the small size of the Thulung corpus, understanding whether verb forms share a cell with equal frequency or whether their use is subject to conditions (be they grammatical or socio-linguistic, as I have hypothesized) is difficult to determine.

b) Some verbs exhibit different stem distribution patterns from those previously associated with verb classes; these new patterns are only found with very small numbers of verbs (per C_f), and were presumably missed because of their infrequency. The patterns associated with these verbs are discussed in the following section.

2.2.2.2 Contrasting paradigms within verb classes

The new patterns that have been identified for stem distributions concern only three C_f 's, and result in subdivisions within t-, η- and k-verb classes, discussed in that order in what follows.

As for **t-class verbs**, the verb *tse:mu* 'pick through' patterns like *senmu* for past forms (see Table 13), using stem CVt throughout, as well as for non-past 1PI and 3SG. Other non-past forms, however, do not pattern as expected: they use the stem CV:- instead of the CVn expected of a t-class verb; note that this is also the case of the citation form of the verb, which uses the CV: stem. The pattern is reminiscent of a combination of V-stem and t-stem class verbs. This is illustrated in Table 17⁴³.

	NPST	PST
1SG>3SG	tse:-pu	tseɬ-to
1PI>3SG	tser-i	tseɬ-qi

⁴³ Table 17 is an abbreviated paradigm, leaving out dual forms for the 1st and 3rd persons, as these pattern predictably (ie. like the 2DU forms).

1PE>3SG	tse:-ku	tseʔ-toko
2SG>3SG	tse:-na	tseʔ-na
2DU>3SG	tse:-tsi	tseʔ-tsi
2PL>3SG	tse:-ni	tseʔ-ni
3SG>3SG	tser-ʉ	tseʔ-dʉ
3PL>3SG	tse:-mi	tseʔ-miri

Table 17. Transitive verb *tse:mu*, 'pick through'

One diachronic explanation, which Michailovsky proposes to explain some of the patterns of stem distribution for Dumii (Michailovsky 2012), is that the stems in Table 17 reflect a dental post-final suffix, a reflex of a no-longer productive causativizing (but labeled 'directive' by Michailovsky, following Wolfenden (1929)) derivational affix (Michailovsky 1975; Michailovsky 1985; Michailovsky 2012). The existence of two different types of verbs, on the one hand with a C_f of -t (in other words, CVt, like *senmu*) and on the other hand with a CV root and postfinal suffix t (in other words, CV-t), could very well explain the different stem distribution patterns found among verbs which have, among their stems, a stem CVt. This explanation is of course not of any relevance to speakers, but is nonetheless interesting from a comparative point of view.

Similarly, with **ŋ-class verbs**, we find two sub-patterns in stem distributions involving a stem CVŋ: on the one hand, we find the stem distribution pattern described in Table 14; on the other hand, we find a distinct pattern, also involving a single stem CVŋ, shown in Table 18.

	NPST	PST
1SG>3SG	sjaŋ-pu	sjaŋ-to
1PI>3SG	sjaŋ-i	sjaŋ-dʒi
1PE>3SG	sjaŋ-ku	sjaŋ-toko
2SG>3SG	sjaŋ-na	sjaŋ-na
2DU>3SG	sjaŋ-tsi	sjaŋ-tsi
2PL>3SG	sjaŋ-ni	sjaŋ-ni
3SG>3SG	sjaŋ-ʉ	sjaŋ-dʉ
3PL>3SG	sjaŋ-mi	sjaŋ-miri

Table 18. Transitive verb *sjaŋmu*, 'clean house'

In the case of the paradigm in Table 18, we have a CV η verb that does not show stem alternation, using the CV η stem throughout its paradigm. Only one other verb patterns in this way, so this is a very small inflectional class, but it nonetheless needs to be accounted for in the description of the verbal system.

The **k-class** also presents two sub-patterns: we find the standard distribution of stems described in Table 11, with alternation between stems of the form CVk and CV:. Table 19 below illustrates yet another distribution pattern for the same two stems:

	NPST	PST
1SG>3SG	l Δ k-pu	l Δ k-to
1PI>3SG	l Δ :	l Δ :-ri
1PE>3SG	l Δ :-ku	l Δ k-toko
2SG>3SG	l Δ :-na	l Δ k-na
2DU>3SG	l Δ :-tsi	l Δ k-tsi
2PL>3SG	l Δ :-ni	l Δ k-ni
3SG>3SG	l Δ :	l Δ :-r Δ
3PL>3SG	l Δ :-mi	l Δ -mri

Table 19. Paradigm for *l Δ :mu*, take away

As with other verb classes, the 1PI and 3SG stand out, but in this case, they do so by having the weaker stem (the one that has lost the final k and undergone compensatory lengthening). Even more unusual is the fact that the non-past 1PI and 3SG forms have no indexes, and are thus both bare stems (and the unexpected stem, at that, given the scenario at hand). As with the sub-type of η -class verb described immediately above, this sub-type of k-class verb only concerns two verbs (the other is *t Δ :mu*, 'know, understand a concept').

To sum up the divergences found within these three verb classes, we find the following subdivisions of three verb classes:

t-class: verbs patterning like *senmu* (table 13); verbs patterning like *tse:mu* (table 17)

η -class: verbs patterning like *qo:mu* (table 14); verbs patterning like *sjanmu* (table 18)

k-class: verbs patterning like *gwa:mu* (table 11); verbs patterning like *l Δ :mu* (table 19)

Considering that the vast number of verbs in the class pattern like the first in each set in the list, this is considered the primary subclass and is not sub-specified. The marginality of the second

subclass associated with each C_f , in terms of exemplars of verbs within that subclass, mean that it can be assigned a special label with a prime sign for identification (t'-class; η '-class; k'-class).

2.2.2.3 The influence of postfinals -t and -s

For a number of Kiranti languages, a significant development in the analysis of verb stem alternation has been work by Boyd Michailovsky (Michailovsky 1975; Michailovsky 1985; Michailovsky 2012): his "general theory of the structure of the Kiranti verb" (2012: 85) posits the basic verb structure CVC_f (the C_f of the root can be p, t, k, m, n, η , r, l and in some languages, s), as seen in 2.2.2.1 above, to which are optionally suffixed post-finals -t or -s. While scholars have long been aware of the importance of two no-longer productive causative suffixes *-t and *-s in shaping pairs of verbs⁴⁴ (Wolfenden 1929; Benedict 1972: 97 ff; Sprigg 1984; van Driem 1988), Michailovsky's work showed that the post-final suffixes can account for verb stem distributions in some Kiranti languages, making it possible to explain the existence of different stem distribution patterns between verbs that ostensibly have the same C_f .⁴⁵

In Khaling, for example, verbs of root types CVp, CVpt, CVt, CVtt, CVk, CVkt, CVm, CVmt, CVn, CVnt, CV η , CV η t, CVr, CVrt, CVl, CVlt have been posited (Jacques et al 2012), reflecting the occurrence of postfinal -t with all C_f 's. These verbs all have distinct stem paradigms, even though the difference between paradigms for transitive CVC and CVCt verbs is neutralized across large swaths of the paradigm, with the reflex of postfinal -t only surfacing in a few cells (Jacques 2015: 80). The postfinal is nonetheless an essential element to acknowledge in accounting for the various paradigms that occur for the verbs in the language, with many intransitive and applicative/causative pairs which differ by virtue of the reflex of postfinal -t ((Jacques 2015; Jacques 2017)).

Working with Thulung, one is struck by the fact that organization into verb classes does not reflect the post-final -t in any obvious way (the exception to this possibly being the t'-class discussed above, as well as, although it is not clear what the mechanism would be, η '- and k'-classes). Cognate intransitive/transitive pairs do exist in Thulung, but in many cases, the verb

⁴⁴ These pairs manifest as either transitive/causative~applicative or intransitive/transitive sets, the second verb of each set reflecting the addition of the postfinal; in Michailovsky (2010), the suffixes are relabeled "applicative -t and causative -s".

⁴⁵ This is similar to the situation described in 2.2.2.2, where the t'-class of Thulung is posited as being CV-t, thus explaining the different stem paradigms from CVt verbs; the situation is however different in that the two verb classes do *not* share a C_f .

roots are labile, able to take intransitive or transitive indexes without changes to the root. An example of this is *da:mu*, 'flood' (vi or vt); if associated with intransitive indexes, the interpretation is as an intransitive verb; if with transitive indexes, then there is a P (expressed or understood). Other types of intransitive/transitive pairings are found in Thulung are, in order of decreasing frequency:

a) a transitive base verb and a reflexivized (either productively with *-sit*, or at an earlier stage, in which case the verb has become an s-verb class) equivalent. An example of the former is *blanmu*, 'dry' (vt) and *blansimu*, 'dry oneself' (vi); of the latter, *jenmu*, 'call' (vt) and *je:mu* (an s-class verb), 'make a sound (of animals)' (vi);

b) verb pairs reflecting not the addition of a postfinal, but of an *s- causative *prefix*, as found in most Sino-Tibetan languages and surfacing as voicing alternations in the initials of the verb roots (LaPolla 2003: 33; Opgenort 2005; Michailovsky 2017: 658). An example is the pair *bimu*, 'break' (vi); *pimu*, 'break' (vt). More of these pairs are listed in Lahaussais (2002: 133) and Allen (1975: 43–35);

c) intransitive-transitive verb pairs which appear to reflect the addition of postfinal *-t*. They are identified as such because one of the verbs in the pair has at least one stem in *-t*, whereas the other does not. This difference in stems suggests that the *-t* is not the C_f but rather a postfinal *-t*. Examples of such pairs are: *bo:mu* ([bok]), 'open, blossom' (vi) vs *bonmu* ([bot]) 'blossom' (vt indexes but vi semantics and case-marking); *ga_nmu* ([g_nan]), 'sit' (vi) vs *ga_nmu* ([g_nat]), 'set' (vt).

Even though Thulung has few verbs of the type in c), there are other potential clues to the earlier influence of the postfinal, in the form of variation in some indexation markers: Verbs of the *p*-, *m*- and *n*- classes take different indexes for non-past 1PI and 3SG S/A arguments than other verb classes, namely *-qi* (vs *-i* elsewhere) and *-q_n* (vs *-_n* elsewhere) respectively (see Table 32 below). We thus have *rep-qi* 'see-1PI>3SG.NPST' but *gwak-i* 'give-1PI>3SG.NPST'.

In order to describe this in a synchronically viable way, I present *-qi* and *-q_n* (1PI>3SG.NPST and 3SG>3SG.NPST respectively) as allomorphs to *-i* and *-_n*. It seems likely that a better way to account for these indexes is to posit that the initials in fact are reflexes of the postfinal *-t*. In such a scenario, we would have had a stage of the language in which CVp-t, CVm-t and CVn-t verbs had a stem reflecting the presence of the postfinal *-t* in non-past 1PI and 3SG forms. The *-t* would then have been reanalyzed at some point as being part of the indexes for those arguments, rather than as a postfinal augment on the verb stem.

One issue with this hypothesis is that not only transitive but also intransitive verbs in the *p*-, *m*- and *n*-classes use these indexes, even though intransitivity is not compatible with a

causative postfinal *-t*. The explanation is perhaps that once the *-t* was reanalyzed as part of the index, it spread, through analogy, to all verbs with a C_f of *-p*, *-m* or *-n*, regardless of transitivity. (Note that we find something similar with the s-stem class, the origins of which are derivational, based on a reflexive voice marker, and which now includes transitive verbs as well).

In this section describing the roots of Thulung verbs, with a focus on verb classes and associated stem alternations (2.2.2.1), additional subclasses (2.2.2.2) and traces of historical transitivizing postfinal suffixes (2.2.2.3), only material making up the first of these sections has been the subject of publications. This is in part because as my corpus evolves, new verb classes (or rather, subclasses, as I hope to have made clear in 2.2.2.2) emerge from the data (even though they are very marginal). As far as the novelty, in my work, of data relating to 2.2.2.3 is concerned, this is because although the diachrony of Kiranti verbal morphology is a very rich topic, which generates a great number of publications, reconstructions, and debates, as far as the relationship to proto-Tibeto-Burman/Trans-Himalayan is concerned, it is not one on which I feel I have had much to contribute. This is partly personal -- I find it more relevant to my analysis of Thulung to focus on synchronic phenomena⁴⁶ -- and partly because Thulung appears to me to be quite different from other Kiranti languages in how it manifests traces of diachronically relevant material: as 2.2.2.3 hopefully shows, while neighboring languages, such as Khaling and Koyi, have clear, synchronically tangible traces of older derivational material, this is not the case for Thulung, where the influences are much more difficult to discern. This is in and of itself a very interesting result of the consideration of material at hand, and one that can be useful from a comparative point of view.

2.2.3 Suffixal slots

2.2.3.1 Suffixal slot 1: partial person indexes

Suffixal slot 1 is where we find phonologically reduced indexes which occur with some types of complex predicates. As these phonologically reduced indexes inevitably co-occur with full indexes (slot 3, see below), they are instances of multiple exponence. They are not very common, as suffixal slot 1 is only filled in the case of compound verbs, and only with a subset

⁴⁶ Another reason is that, having worked as a research assistant to Jim Matisoff on his STEDT (Sino-Tibetan Etymological Dictionary and Thesaurus) project, I have seen first-hand how much knowledge is necessary to propose reconstructions.

of compound verbs at that, in circumstances dictated by the morphophonology of the verbs in question.

The material that can occur in this slot, as well as the relevant conditions for its appearance, is intimately tied with slot 2,⁴⁷ and is therefore discussed in the next section.

2.2.3.2 Suffixal slot 2: derivational markers

Suffixal slot 2 is where we find markers of Associated Motion, valence change, and Aktionsart. As these markers can co-occur, in a fixed order, they are assigned linearly organized sub-slots.

In a recent article (Lahaussais, submitted), I describe what I refer to as Predicate derivations in Thulung. These are forms with derivational material marking Associated Motion, valence changes, or Aktionsart, and they form a subset of complex predicates, which have been treated at length for a number of South Asian languages (Hook 1974; Butt 1995; Slade 2016; Bickel 1996; Jacques 2018). Before describing Predicate derivations in more detail, I will discuss complex predicates in Thulung. The reason for this detour is that there is, in South Asian languages, a continuum going from complex predicates to derived verbs. Some familiarity with other constructions along the continuum helps reveal how the diachronic development of predicate derivations--our main topic of interest--has contributed to their special morphological characteristics.

Two examples illustrate complex predicates in Thulung: the first involves a noun + verb combination. In (51), neither the nominal element *na:le* nor the verbal element *qanmu* has an independent synchronic lexical meaning, and they only ever occur as a complex predicate with the meaning 'taunt'. Yet, as can be seen in the example below, the negation marker occurs between the nominal and verbal elements. In some other cases, both elements have independent lexical meanings, but lexicalize into a complex predicate when taken together; in yet other cases, only one of the elements has an independent lexical meaning.

(51)	bo:bop-ka	wo	thama	ne	memsaka	na:le
	owl-ERG	also	later	TOP	like.that	N.taunt
	me- dat -nji				rwak-ta	
	NEG- v.taunt -2SG>1SG				say-3SG.PST	

⁴⁷ As a reminder, the relevant part of the template for sections 2.2.3.1 and 2.2.3.2 concerns the slots for: root-[partial person index]-derivational formative-indexation marker

'The owl also later said "Do not taunt me like that".'

Example (52) illustrates a complex predicate involving two verbs:

- (52) meʔema:la go a-phjaphjaksa-ka **phjak-seʔ-pu**
 otherwise 1SG 1SG.POSS-bow.arrow-INSTR snap.in.two-kill-1SG>3SG
 [Mini told his mother: Be angry with him, tell him to hold still] 'Otherwise, I will
 snap and kill him with my bow and arrow.'

In (52), both verbs also occur as independent lexical items, and the interpretation of the compound verb is the sum of the two verbs that make it up, in essence functioning as a serial verb construction. The difference from a serial verb construction is that there is a single indexation marker on the compound verb: arguments are indexed only once, suffixally, and not on both verbs independently.

Verbs in parataxis, without any conjoining material, also exist, as in (53), but this situation is rare, and the prosody marks these as distinct clauses.

- (53) me hunuk-the gai-kam gonu huŋ-ɖa ʌms-ta
 DIST.DEM there-near cow-GEN inside enter-3SG.PST sleep-3SG.PST
 'He went in with the cows and slept.'

It is more common, in cases of parataxis, to have to fully finite verbs conjoined with *ma*, as in (54), a sentence taken from the same story as (52), and occurring 11 sentences [1min18] later. This suggests a certain amount of morphological flexibility in assembling verbs, as both (52) and (54) make use of the same lexical verbs, but combine them differently:

- (54) ko:lem gupsu-ka u-mam-kam samrip **phjas-tu**
 one.day tiger-ERG 3SG.POSS-mother-GEN throat **snap.in.two-3SG>3SG.PST**
 ma **seʔ-ɖu** ʔe
 CONJ **kill-3SG>3SG.PST** HS
 'One day, the tiger snapped his mother's throat and killed her.'

Returning to the question of predicate derivations, an example is shown in (55) below: the root *tsar* is followed by a formative glossed 'ANTIC' (for 'anticipatory', as it conveys the

tradition in South Asian languages, seems to tap into, as these formatives in many cases have motion verbs as their sources and express directionality).

I will therefore discuss relevant aspects of the morphology of these derived verbs in Thulung, touching upon their semantics and placing them within the context of similar constructions in South Asian languages. I will conclude this section by describing the specific functional domains (Associated Motion, valence changes, Aktionsart) in turn, and bringing in data from Khaling as well.

2.2.3.2.1 *Multiple exponence*

One of the main characteristics of predicate derivations is that in some instances they exhibit multiple exponence, with suffixal slot 1 filled with additional, albeit partial, indexation markers.

Only vowel-final stems exhibit multiple exponence, and the phenomenon is thus limited to k-, η-, s- and V-class verbs (in other words, those for which at least one of the stems is vowel-final). Some examples of this multiple exponence are given below:

k-class verb thΛ:mu |thΛk|, ‘hide’

- (56) sʌ: thΛ-**n-si-na**
 who hide-2SG-REFL-2SG
 ‘Who are you, who are hiding?’

η-class verb dzo:mu |dzoŋ|, ‘plough’

- (57) wostsø ri:pap-ka dzo-**m-thal-miri**
 boy man-ERG plough-3PL>3SG.PST-HAB-3PL>3SG.PST
 ‘The boy and man ploughed regularly.’

V-class verb bomu |be|, ‘do’

- (58) sʌ:-ka be-**η-sa-ηri**
 who-ERG do-3SG>1SG-APPL-3SG>1SG.PST
 ‘Who did this for me?’

The additional exponent--occurring between the verb root and the derivational formative--is made up of a single phoneme, identical, in examples (56)-(58), to the initial of the full indexes

(the second of the bolded formatives in the examples⁴⁸). This has resulted in its being called a 'copy' in some Kiranti descriptions of this phenomenon (eg. (Ebert 1997a: 71)). For Thulung, however, while the additional indexation is always a single phoneme, it is not always a copy of the initial: for V-class verbs, some forms have a different formative in suffixal slot 1.

- (59) ʔã ko:-din-ka be-**u**-thal-**to**⁴⁹
 yes one-day-TEMP do-**1SG>3SG.PST**-HAB-**1SG>3SG.PST**
 'Yes, I usually did [that much weaving] in one day.'

The partial indexation marker is based on the initial phoneme of the appropriate allomorph of the formative based on the verb stem or derivational suffix. The formula can be described as the following:

- (60) Verb stem - initial of appropriate allomorph of the indexation (with respect to the verb) -
 DERIV.SUFF - appropriate allomorph of the index (with respect to final of derivational
 suffix)

As for how (60) explains the form of the verb in (59), the (underived) verb *bomu* 'do' takes the 1SG>3SG.PST *-uto*, resulting in a form *be-uto*. According to (60), we keep only the initial of the index, and this yields the form *be-u-*. To this we add the derivational suffix *-thal* (for habitual), yielding *be-u-thal-*. The appropriate 1SG>3SG.PST index that needs to be added to *be-u-thal-* is that for an l-class verb (as *-thal* is l-final), in other words *-to* (see Table 9).

When assembled, this gives us the form *be-u-thal-to*, as in (59).

This form can be contrasted with the equivalent non-past verb: the 1SG>3SG form of *bomu* is *be-u*; when habitual *-thal* is added to this, we get the form *be-u-thal-*. The 1SG>3SG index for l-class verb (again, the derivational suffixes behave like verbs in their selection of allomorphs for the indexes) is *-u*, and the whole form is thus *be-u-thal-u*.

What is interesting about this pair is that, depending on the inflectional class of the verb and the final of the derivational suffix, the multiple exponence can manifest as the same

⁴⁸ The verbs in (56)-(58) would be *tha:-na*, *dzo:-mri*, and *be-ŋri*, respectively, if underived.

⁴⁹ This verb form would be *be-uto*, if underived.

allomorph (or part thereof) as the full index, thus sometimes getting the label "copy", or as (part of) a different allomorph, looking nothing like the full index.

In sum, while the partial indexation material in suffixal slot 1 is only ever a single phoneme long, it is always the initial of the appropriate allomorph of the full indexation that would occur if the verb were an underived verb rather than a predicate derivation.

2.2.3.2.2 Allomorphy of derivational formatives

Another particularity of the derivational formatives is that they display allomorphy: of the nine listed in Table 20, the three formatives ending in *-l* are invariant (matching the invariance of l-class verbs, which have a single CV1 stem throughout their paradigm). The remaining six formatives have allomorphs ending in *-t*, *-n*, *-k* and *V*, resembling the stems of both k-class and t-class verbs combined.

derivational suffix (+ allomorphs)	label/gloss	function(s)	lexical source
-bal	ambulative/AMB	associated motion/orientation; spatially distributed action	<i>balmu</i> , wind around (vt)
-bet, -ben, -be, -bek	causative/CAUS	valence-increase, by one agentive argument	<i>benmu</i> , transitive auxiliary
-sit, -sin, -si, -sik	reflexive/REFL	valence-decrease	(no identifiable source)
-sat, -san, -sa, -sak	applicative/APPL	valence-increase, by one patientive argument	(no identifiable source)
-th Δ t, -th Δ n, -th Δ , -th Δ k	durative/DUR	expresses fact that action lasts in time	(no identifiable source)
-so Δ , -son, -so, -sok	irreversible/IRREV	focuses on irreversability of an action, often accompanied by the disappearance of the patient argument	(no identifiable source)
-let, -len, -le, -lek	liminal/LIM	focuses on start or finish of action	<i>lenmu</i> , intransitive auxiliary

-dzəl	anticipatory/ANTIC	highlights outcome of action, often in anticipation of another and through placing the patient argument somewhere	<i>dzəlmu</i> , set aside
-thal	habitual/HAB	expresses regularity of occurrence of action	(no identifiable source)

Table 20. Allomorphs, functions and source verbs for derivational formatives

The paradigm for durative *thət-* is shown in Table 21. The formative *thət-* mostly patterns like a t-class verb, with occasional forms that resemble those of a k-class verb. These patterns are not phonologically conditioned.

	NPST	PST
1SG>3SG	gwak- thək -pu	gwak- thək -to
1DE>3SG	gwa:- thən -tsuku	gwak- thək -tsoko
1PE>3SG	gwa:- thən -ku	gwak- thək -toko
1PI>3SG	gwak- thək -i	gwak- thət -d̥i
2SG>3SG	gwa-n-thΛ-na	gwak- thət -na
2DU>3SG	gwa:- thən -tsi	gwak- thət -tsi
2PL>3SG	gwa-n-thΛ-ni	gwak- thət -ni
3SG>3SG	gwak- thək -u	gwak- thət -d̥u
3DU>3SG	gwa:- thən -tsi	gwak- thət -tsi
3PL>3SG	gwa-m- thΛ -mi	gwa-m- thΛ -mri

Table 21. Allomorphs of the durative marker

The fact that the pattern maps quite well onto the distribution pattern of verb stems (albeit those of two different verb classes combined) suggests that these formatives behave like elements that were originally verbal; that the forms look like stems of two different verb classes suggests a breakdown of an original verbal stem distribution into something with a less predictable pattern.

An additional element to consider with respect to the nature of the derivational formatives is that in some cases (four of the nine formatives) there are associated lexical source verbs, in the form of two auxiliaries and two full verbs.

Taken together, these three factors--the multiple exponence using 'stem-appropriate' allomorphs, the formative allomorphy patterning like verb classes (albeit muddled), the occasional lexical source verbs--suggest strongly that the diachrony of these constructions is development from a serial-verb like construction similar to what is seen in (54). The multiple exponence is a trace of the indexation originally associated with each of the verbs, but reduced as they evolved into compounds, until the second verb eventually became derivational, and the first verb's indexation was reduced to a single trace phoneme.

Within the context of South Asian languages and research on complex predication and specifically on compound verbs, one might wonder to what extent it is possible to consider these constructions comparatively. There are very large differences even within Kiranti regarding the finiteness of the initial verbal element (mostly finite, but not in some Western languages and Hayu (Michailovsky 2017); transitivity matching (derivational formatives generally match the transitivity of the verb, but this is not the case in Thulung); grammaticalization pathways followed by verbs; the degree of grammaticalization into derivational morphology, among others.

In an article comparing compound verbs ('CVs') in Nepali and Hindi, Slade (2013) points out some major differences between the two languages, notably in terms of the features of a) interruption; b) recursion; c) construction-specific restrictions; d) dominance and transitivity issues (Slade 2013: 552–557). While a Hindi compound verb can be interrupted by words such as the emphasis marker *to*, Nepali CVs cannot; this difference is reflected in the graphic form of CVs in the two languages: Nepali CVs are written as a single word, while Hindi CVs are written as separate words. Regarding the feature of recursion, Hindi allows only a single light verb per main verb in the CV, while Nepali allows up to two (Slade (2013: 553, ex. 42) gives the example *saphā gar-i-di-i-hāl-in*, clean do-ABS-give-ABS-put-PAST.3SG.FEM). With respect to the third feature, construction-specific restrictions, Slade points out that Nepali CVs can occur in converbial form (*kha-i-sak-era*, eat-ABS-finish-CONV), something which does not happen in Hindi, where converbs are only found with simplex verbs. As for dominance, in Nepali CVs, "it is always the main verb which determines the transitivity of the entire verbal compound" (2013: 555), whereas in Hindi CVs, the transitivity of the light verb is generally what determines the argument structure of the clause.

When one examines these four features in Thulung, one finds that they align nicely with the behavior of Nepali CVs: Thulung predicate derivations can never be interrupted by additional material. Regarding recursion, the formatives found in slot 2 can co-occur, as can be

seen in (61)-(62) below (no examples of all three sub-slots being filled simultaneously were found in the corpus, although there is no reason for them not to).⁵⁰

- (61) beula beuli dzul-toko mem-ŋa be-ttoko memsa-ŋa
 groom bride put-1PE>3S.PST like.that-INT do-1PE>3SG.PST like.that-INT
 tsam-**bal-sit**-toko
 play-**AMB-REFL**-1PE.PST
 'We set up the [mock] bride and groom, we did that, we played around like that.'

- (62) mela burkhum-go-la thak-**dzol-thal**-lu ʔe
 up.there cave-inside-HIGH.LOC hide-**ANTIC-HAB**-3SG.PST HS
 'He was always hiding it inside a cave up there.'

Like in Nepali (and unlike Hindi), Thulung predicate derivations are found frequently with converbial forms, as exemplified in (63):

- (63) sokmu-ra khir-sik-to-ŋa ler-**u**-mim hopmam
 forest-LOC turn-REFL-SIM.CVB-EMPH AUX-3SG>3SG-NMLZ like.this
 lwa:-si
 see-REFL.3SG
 'We see that he is turning around in the forest like this.'

Converbial forms with predicate derivations are however limited to derivation with the ambulative marker *-bal* and with valence changing markers, and are never (in my corpus) found with aspectual derivational formatives.

Compound verbs also pattern like Nepali with respect to the last feature, namely dominance: it is always the main verb that determines the transitivity of the predicate. There are, however, differences between Thulung and Khaling insofar as the transitivity of the derivational formative is concerned: while Khaling (and other Kiranti languages) select a derivational suffix that matches the main verb in transitivity (see (64), where the completive

⁵⁰ With respect to the status of slot 1 in cases with multiple derivational suffixes, I have no corpus examples which illustrate whether the double exponence (slot 1 and slot 3) can in fact be increased further, with an additional exponent for each derivational suffix. This will require elicitation work during a future field trip.

alternates between an 'intransitive' form *-khΛ* and a 'transitive' form *-khΛt*, Thulung does not, instead using the same set of allomorphs for both transitive and intransitive main verbs.

- | | | |
|------|----------------------------------------------------------|------------------------------------------------------------------|
| (64) | dzhΛi-khΛ-ŋΛ
wet(vi)-GO-1SG
'I am completely wet.' | dzhΛt-khΛt-u
wet(vt)-GO-1SG>3SG
'I got it completely wet.' |
|------|----------------------------------------------------------|------------------------------------------------------------------|

The predicate derivations of Nepali thus behave very much like those of Thulung and other Kiranti languages, with respect to the four features proposed by (Slade 2013). This suggests that language contact has very likely have shaped the predicate derivations of Nepali and Kiranti along areal rather than genetic lines.

2.2.3.2.3 *Associated Motion*

I shall base the discussion of Associated Motion (henceforth 'AM') in this section on contrasts between Thulung and Khaling. The systems in the two languages are discussed in Jacques et al. (In press), which compares data from Kiranti and Gyalrongic languages, as well as from a few additional Trans-Himalayan languages. AM is defined here, per Guillaume (2016: 13), in the following way:

'An AM marker is a grammatical morpheme that is associated with the verb and that has among its possible functions the coding of translational motion.'

A Khaling example is given in (65), in which the action of the main verb (here represented by stem *ŋô* 'cry') is accompanied by a translational motion, which in (65) is a round-about motion.

- | | |
|------|----------------------------------------------------------------------------------------------|
| (65) | ʔi-ŋô-n-le-ni
2/INV-cry-2PL:NPST-go.around.doing-2PL:NPST
'You.PL went around crying.' |
|------|----------------------------------------------------------------------------------------------|

An important difference between the Thulung and Khaling is that they have a different number of AM formatives: Khaling has seven, many of which are quite transparently grammaticalized from motion verbs, whereas Thulung has a single marker, grammaticalized from the transitive verb 'wind around' and primarily encoding an ambulative motion, as well as different functions which are extensions from this primary meaning.

AM has generally been considered from the point of view of three parameters, namely deixis, temporal relation (which can be prior, concurrent and subsequent) and argument of motion (Guillaume 2016; Wilkins 1991; Koch 1984), both of which need to be expanded to account for Kiranti data, but it also seems that additional parameters, namely those of vertical dimension and degree of event integration, are useful for discussion of AM in Kiranti and Trans-Himalayan, (Jacques, Lahaussais & Zhang In press).

The single AM marker in Thulung expresses ambulative (the term is borrowed from Ebert (1997a: 50)) motion as its core function, but has additional functions which I will describe below. (These functions were presented orally at several conferences; (Lahaussais 2018b; Jacques & Lahaussais 2018)). The associated motion function of *-bal* is illustrated in (66), describing a round-about motion accompanying the main action.

- (66) me pel sam-**bal**-lu ?e
 DIST.DEM shadow catch-AMB-3SG>3SG.PST HS
 'He goes around catching that shadow.'

In terms of the main AM parameters, this function of the ambulative marker has (in at least one interpretation) a concurrent temporal relation between the main verb ('catch') and the motion; the argument of motion is the A; and the deixis, which in most descriptions of AM around the world is either trans- or cislocative, in this case can be said to be circumlocative, in that the motion radiates around a deictic center.

The other functions found for the ambulative marker are extensions from this core scenario; they are discussed below in an order that represents a gradual bleaching of features with respect to the associated motion ambulative in (66).

The function illustrated in (67) is the expression of *spatially distributed action*, where the associated motion is not a round-about motion, but rather the performance of the main action in different places.

- (67) oram-ka bante meno si-**bal**-lu retsΛ
 this-ERG where there teach-AMB-3SG>3SG.PST it.seems
 'This one has taught around everywhere.'

With respect to core properties, the temporal relation has changed: it is no longer concurrent, but rather punctuated, with the motion interspersed with bouts of the main action. Other

properties, namely deixis and argument of motion, remain the same. This is the main function described for the equivalent of the ambulative⁵¹ in Belhare (Bickel 1996: 164), where it apparently has both AM (68) and non-AM (69) uses:

(68) rot-d-e i-baŋ-ŋa chap-**kon**-u
 road-LOC one-HUM-OBL write-SDT-3U
 ‘Somebody is walking around taking notes on the road.’

(69) thali khore wat chi-**gon**
 plate cup clean clean-SDT
 ‘He is cleaning plates and cups.’

The function illustrated in (70) below represents a further extension, with the expression of the action no longer *spatially* but *orientationally* distributed, in other words with no associated motion. Note that the data relevant to analyzing AM systems is quite often obscured in descriptions because of the use of terminology encompassing both motion and orientation, such as Wolfenden's 'directives', taken up by DeLancey (1985) in an influential paper.⁵²

(70) memsaka etha utha rep-**bal-lu** mɯ:tsu wo
 like.that this.way that.way look-AMB-3SG>3SG.PST man also
 me-lwas-u-ba
 NEG-see-3SG>3SG-IRR
 ‘He looked here and there and saw no one (not even a human).’

The function illustrated in (71) abstracts further away from AM, no longer encoding motion or orientation, and instead expressing *recurring action* (what remains of spatially distributed action which has no spatial or orientational component): the action is repeated over time, but not necessarily at regular intervals, and not necessarily in different places:

⁵¹ The formative is glossed SDT by Bickel, for 'spatially-distributed temporary'.

⁵² Consider, for instance, the difference between upward AM 'look and come up' and upward orientation 'look upwards', often treated as the same thing in descriptions.

- (71) go me-llai mamtha-m-bili ne khau
 1SG DIST.DEM-DAT last.year-NMLZ-time TOP money
 gwak-**bal**-to atha ne gwa:-mu d̥it-to
 give-AMB-1SG>3SG.PST now TOP give-INF stop-1SG>3SG.PST
 'I gave him (every now and then) a little money last year, and now I stopped giving.'

A final function is the expression of *characteristic action*, shown in (72), signaling a defining characteristic of a particular entity, with both spatial (and orientational) and temporal properties having been bleached.

- (72) noksu prok-**bal**-pa phal-**bal**-pa
 monkey jump-AMB-ACT.PTCP destroy-AMB-ACT.PTCP
 mini sewari bha-pa ma khole rep-**bal**-pa...
 human headdress wear-ACT.PTCP CONJ all look-AMB-ACT.PTCP
 'The monkey, who was jumping around and destroyed things, the human, who was wearing a headdress and looking around at everything...'

The ambulative marker in Khaling, *-le*, illustrated in (65), is not cognate with Thulung *-bal* and has no obvious lexical source. While I do not have Khaling examples illustrating all the same functions as those described for the Thulung ambulative, I do have instances of its being used as a proper associated motion marker, as in (73), which can be interpreted as either concurrent or punctuated motion with the main action, and of the final function described above, namely expressing characteristic action, as in (74).

- (73) woŋâmʔu-kɛm mu-gə khɛbi mɛ jo tshoəm-le
 other 3POSS-work NEG-be where NEG even dance-AMB
 'She has no other work so she dances around everywhere.'

- (74) mʌ-sêŋpɛ sendi kɛ-n-le-nɛ mu-nɛ
 NEG-clean nail bite-INF-AMB-INF NEG-be.good
 'It is not acceptable to go around biting unclean nails.' (describing a habit)

We can assume that the grammaticalization pathway to get from (73) to (74) is similar to that described for Thulung, and yet the input for this pathway in both languages is presumably different ('wind around' (trans.), in Thulung, and no known source verb for Khaling). While many Kiranti languages have what can be described as an ambulative marker, they mostly seem to have different source verbs.

Khaling, with its seven AM markers, has six additional markers compared to Thulung, as shows in Table 22 below (reproduced from Jacques, Lahaussais & Zhang (In press)).

Table 2: The Khaling AM system

Source	AM marker	Temp.	Deixis	Vertical	Argument
?	-le(t)- 'go around doing X'	C	CIRCUM	∅	S/A
k ^h ot 'go'	-k ^h o(t)- 'do X and go'	S	TRANS	∅	S/A(+P)
?	-pe(t)- 'go and do X'	P	TRANS	∅	S/A
fiŋ 'come'	-fiŋ(t)- 'do X and come'	S	CIS	∅	S/A(+P)
k ^h oŋ 'come up'	-k ^h oŋ- 'do X and come up'	S	CIS	up	S/A
pi 'come (same level)'	-pi(t)- 'do X and come (same level)'	S	CIS	same level	S/A(+P)
ten 'fall'	-te(ŋt)- 'do X and come/bring down'	S/P	CIS	down	S/A(+P)

Table 22. Associated Motion markers in Khaling

The parameters used to compare AM markers are listed in the various columns of the table: temporal relation, deixis, verticality, and argument. In terms of deixis, apart from the ambulative with circumlocative deixis, Khaling can express both trans- and cislocative deixis; in the case of cislocative AM, note that there is a verticality feature, with a 3-way distinction: upwards motion, downwards motion, level motion (the language also features a cislocative marker which is neutral with respect to verticality, |fiŋ|).

The argument of motion, which in languages of South America, can be S, A or P,⁵³ in Khaling involves the S or the A. One characteristic of Khaling, which we have not found described elsewhere in Kiranti or beyond, is that the P is often also an argument of motion in transitive scenarios, in addition to the A. This is because the AM marker can involve (even though this is not necessarily the case) the displacement of the P, as a result of transitivity

⁵³ For an example of a P as the moving argument, see the following example, taken from Guillaume (2016: 113), where it is numbered (22a) (the original is from Guillaume 2008:234, A69a):

Peadya ekwita=tuke=∅ ba-**dadi**-wa...

one person=3SG=1SG see-GO(O)-PFV

'I saw a man going away from me.'

matching, where a transitive main verb will trigger the use of a transitive version of the AM marker: in (75), the AM maker can be interpreted (despite its gloss) as 'do and **bring** down'.

- (75) tukkâm tsulo-tu khir
 up.above-NMLZ stove-HIGH.LOC porridge
 khip-tɛn-tɔ
 cook-DO&COME.DOWN-1SG>3SG.PST
 'I cooked porridge on the stove up there (on the roof) and brought it down.'

Whether the AM marker involves the motion of the P as well as the A appears to be context-dependent. Example (76), in the context in which it was uttered, was claimed by our consultant to involve only motion of the A:

- (76) ?in khɔs-ther-e ?uŋʌ ?ʌ-dʌɾʌm
 2SG go-CONT-2SG.IMP 1SG.ERG 1SG.POSS-friend
 hũŋ-khond-u⁵⁴
 wait-DO&COME.UP-1SG>3
 'You keep going, I will wait for my friends and come up then.'

While Khaling has a contrastive temporal relation with translocative AM markers, with *khot* expressing subsequent motion and *pɛt* expressing prior motion, Thulung, which does not have any AM marker beyond the ambulative (whose action is simultaneous with the associated motion, in other words expressing concurrent motion), must use other constructions in order to temporally link motion verbs with other actions: these are constructions involving converbial motion verbs to express simultaneous ((77), using a simultaneous converb) and anterior ((78), using an anterior converb) motions. Alternatively, a construction with two finite verbs linked by a conjunction can be used, with the order of actions determined by the position of the motion verb relative to the other action; this can be used for prior motion (79), and is the only strategy for the expression of subsequent motion (80).

- (77) me sɔpdi-ra lɔk-to lɔk-to ku ɖar-ru

⁵⁴ The verb 'wait' is transitive, and khond is the transitive-matched version of DO&COME.UP, but in this context it nonetheless means 'wait for them and come up' rather than 'wait for them and bring them up'.

DIST.DEM forest-LOC go-SIM.CVB go-SIM.CVB water meet-3SG>3SG.PST
 'Going and going through the forest, he found water.'

(78) hano hano babante lək-saka ro-mi ʔe
 hereabouts hereabouts everywhere go-ANT.CVB come-3PL HS
 'Having gone to many different places, they come.'

(79) dui tin mλina dəs-ta λsinɔa ro-ŋro ma ba-ŋro
 two three month be.3SG.PST here come-1SG.PST CONJ stay-1SG.PST
 'It's been two or three months since I came here and stayed.'

(80) me səkhli bu:-ru ma dər kur-ru
 that hunting do-3SG>3SG.PST CONJ deer carry-3SG>3SG.PST
 ma bik-ta
 CONJ come-3SG.PST
 'He hunted and carried the deer and came.'

While Khaling AM markers show a with a three-way verticality distinction in cislocative verbs (upward, level, downward, as opposed to unspecified verticality), the single AM marker in Thulung does not express verticality. Verticality is nonetheless present in other domains, with the same 3-way contrast which can be expressed in motion verbs, case-markers (as seen in the section on Case markers above), and adverbs (which can in turn be nominalized and used adjectivally, as seen in Table 5):

translocative	cislocative
lλ:mu, go	romu, come (unspecified elev.)
	genmu, come up
	bimu, come across
	jomu, come down
lλ:mu, take away	renmu, bring here (unspecified elev.)
	khenmu, bring up
	phinmu, bring across
	sənmu, bring down

Table 23. Thulung motion verbs

2.2.3.2.4 Valence changes

Valence changes through suffixation of the relevant formative are an extremely productive process, and include causativization, applicativization and reflexivization.

2.2.3.2.4.1 Causativization

Causative verbs can be derived by suffixation of *-bet* (and allomorphs) from both intransitive (81) and transitive (82) input scenarios, adding an agent-like argument to both. The new argument is marked as any A in Thulung: it takes ergative marking if 2PL, 3rd person or other nominals, and is unmarked elsewhere. The original S from the underived intransitive scenario takes primary object marking if human (or salient animate protagonist), and is unmarked elsewhere (eg. the *ku* immediately preceding the causativized verb in (81)). The A from the corresponding underived scenario takes primary object marking *-lai*, if present in the derived scenario (*u-bɔini* in (82)).

- (81) gani bi-m basta ku-ku mam phiṭ-ni
2PL come-INF DEON water-GEN mother bring-2PL>3SG
aki-del-ra ku pun-**bet**-ni
1PI.POSS-village-LOC water spring.out-CAUS-2PL>3SG
'You must come, bring the sources of the water, and make the water spring forth in our village.'

- (82) thama u-bɔini-lai sɔbem tsa-**ben**-mu-lai
later 3SG.POSS-sister-DAT bread cook-CAUS-INF-PURP
mu hon-**ber**-u
fire light-CAUS-3SG>3SG.PST
'Later she made his sister light the fire to cook the bread.'

2.2.3.2.4.2 Applicativization

Applicative derivations add an additional object argument to a predicate. The extra argument can be encoded with the primary object marker *-lai* (83); via a possessive construction, as in (84); or it can be marked solely through indexation on the verb (as in (85)). The additional argument is indexed in the P slot of the inflectional marking.

(83) tsettse-ka saŋ u:ma **mam-lai** khʌl-saʈ-dɛ
 child-ERG wood 3SG.POSS **mother-DAT** fetch-APPL-3SG>3SG.PST
 'The child fetched wood for his mother.'

(84) lu mima: go i-se:r se-n-sa-ni
 hey grandmother.VOC 1SG **2SG.POSS-lice** kill-1SG>2PL-**APPL-1SG>2PL**
 hʌi rwak-ta retsa ʔe
 EMPH say-3SG.PST it.seems HS
 'Then he said "Hey, grandmother, I will kill your lice for you".'

(85) hopmam mʉtsʉ-num hawo bia
 like.this man-COM why marriage
 be-n-saʈ-ni-m
 do-2PL>3SG-**APPL-2PL>3SG.PST-NMLZ**
 'Why did you have her marry a man like this?'

2.2.3.2.4.3 Reflexivization (and related functions)

The reflexive voice marker *-siʈ* (and allomorphs *-si*, *-sin*, *-sik*) is used to express co-reference between two arguments of a derived verb. This is illustrated in (86), which stands in contrast to (87), which contains the underived source verb:

(86) memma thʌ:-**si**-m thʌ:-**si**-m-ŋa sintha koŋŋa
 then hide-**REFL-NMLZ** hide-**REFL-NMLZ-INT** night only
 je
 come.out.3SG
 'Then he hides and hides and only comes out at night.' [referring to behavior characteristic of the owl]

(87) tshimori wo thʌk-a me-gonu-ŋa
 broom also hide-2SG>3SG.IMP DIST.DEM-inside-INT
 'Hide the broom in there as well.'

It is apt that neither (86) or (87) features an overtly expressed subject NP, as this represents a very frequent situation in Thulung (and other Kiranti languages). The absence of overt NP

arguments is probably relevant to the fact that the primary strategy for the expression of coreference is a voice marker rather than a reflexive nominal. Thulung does sometimes make use of a possessively-marked emphatic nominal to remove ambiguity about the co-reference, and in certain cases, it even appears to be an essential part of the construction. This is the case, for example, with examples (88) and (89): without the emphatic nominal (in bold), these are said to be difficult to parse.

(88) meram mʊtsʊ **u-twap-ŋa** sen-s-ta
 DIST.DEM man **3SG.POSS-self-INT** kill-REFL-3SG.PST
 'The man killed himself.' (elicited)

(89) khlea **u-twap-ŋa** khren-si
 dog **3SG.POSS-self-INT** bite-REFL.3SG
 'The dog bites itself.' (elicited)

In addition to the core function of marking reflexivity, *-siŋ* is also used to mark reciprocals, antipassives, anticausatives, autobenefactives (and is glossed as REFL throughout, as this is the core function).

Reciprocal function:

(90) məmim twap-twap bal-**si**-mi
 3PL self-RED wind.around-**REFL**-3PL
 'They are intertwined/they are tangled together.'

(91) mər-tsip mamtha phwa-**siŋ**-tsi
 that-DU last.year separate-**REFL**-3DU.PST
 'They separated last year.'

Antipassive:

(92) gumi bira-nuŋ/-ram⁵⁵ ghram-**si**-mi
 3PL leech-COM/-ABL be.disgusted-**REFL**-3PL

⁵⁵ Both of these oblique case-markers can be used for the demoted object.

'They are disgusted by leeches.'

- (93) gu mandir-ra khir-si
3SG temple-LOC circumambulate-REFL.3SG
'He circled around at the temple.'

Anticausative:

- (94) dɔksa tsar-siŋ-ɖa
tree fall-REFL-3SG.PST
'The tree fell.'

- (95) basta go on-ŋoro-lo taŋ-ro ma
yesterday 1SG run-1SG.PST-TEMP fall-1SG.PST CONJ
a-khel nər-siŋ-ɖa
1SG.POSS-leg twist-REFL-3SG.PST
'When I was running yesterday, I fell and my leg got twisted.'

These same functions are also marked with the reflexive marker in Khaling, in addition to an auto-benefactive use and an impersonal use. For additional examples and discussion, see Lahaussais (2016a; Submitted).

2.2.3.2.5 *Aktionsart and aspect*

The markers in Kiranti that confer Aktionsart are distinctly different from language to language in their semantics, even when they are grammaticalized from cognate verbs⁵⁶. Ebert (1994: 61) provides a table in which she provides comparative data on what she labels "second verbs" in Limbu, Athpare, Bantawa, Camling and Thulung, most of which are used for the expression of Aktionsart, contrasting their meanings as full verbs and as "V2". Among the verbs in this list, Ebert discusses the grammaticalization of the following: *go; send; come; put; (make) fall; reach; come out; finish; die, kill; lie; stay, keep; eat; want*. These are associated with the

⁵⁶ The cognacy of the verbs involved in grammaticalizing into aspectual (and related) formatives spreads across the South Asian linguistic area, often involving verbs such as "GIVE, and often GO, TAKE, THROW/PUT, RISE, FALL and COME" [caps in original, as often the case when writing 'V2s'] (Slade 2016: 565).

expression of telicity, inchoative, completive, durative, progressive, continuative, habitual (none of which labels are usually accompanied by an explanation). This makes it difficult, on the basis of labels, explanations and examples provided, to reasonably contrast grammaticalization paths across different, even closely related, languages. Further complicating analysis and comparison is the fact that in some languages, the formatives are polysemous, such as Khaling *-khot* (and allomorphs) which can be used to indicate associated motion (subsequent translocative motion) or telicity. This polysemy is illustrated in (96):

- (96) t̪ŋ-kʰʌ-tɛ
 drink-DO&GO/AWAY-3SG.PST
 ‘He drank it and left’ or ‘He drank it all up.’

Thulung Aktionsart markers are grouped into two types: those that reduce the telicity of verbs, and those which add telicity to verbs. Of the markers in Table 24 below (which is a subset of the material in Table 20), the first two are in the first category, while the last three are in the second category. One important thing to note is that while the labels and groupings make it seem like the semantic contribution of these markers is clear, this is not necessarily the case: in some cases, speakers claim that there is absolutely no difference between a marked and unmarked verb, but that adding the marker makes the sentence more natural. The inherently complicated semantics of these markers is apparent across descriptions of Kiranti languages, both in the variety of labels of the markers themselves, but also of the category of derivational marker. Van Driem's label "aspectivizer" (van Driem 1987; van Driem 1993b) is a fairly successful attempt at pointing out the fuzzy nature of what the markers contribute to a given predicate.

-thʌt̪, -thʌn, -thʌ, -thʌk	durative/DUR	expresses fact that an action lasts in time
-thal	habitual/HAB	expresses regularity of occurrence of action
-soʈ, -son, -so, -sok	irreversible/IRREV	focuses on irreversability of an action, often accompanied by the disappearance of the patient argument
-leʈ, -len, -le, -lek	liminal/LIM	focuses on start or finish of action

-dzəl	anticipatory/ANTIC	highlights outcome of action, often in anticipation of another and through placing the patient argument somewhere
-------	--------------------	-------------------------------------------------------------------------------------------------------------------

Table 24. Thulung Aktionsart markers

2.2.3.2.5.1 *Markers which downshift the telicity of the verb*

The durative and habitual markers are illustrated below. They are characterized, as a group, by defocusing the inherent telicity of the verbs they mark. In the case of both markers, even though one possible translation for marked predicates (and sometimes the best one) is with an English perfect, the semantics are in fact quite different; furthermore, there is a separate construction for expressing perfect aspect in Thulung.

Durative marker *-thət* is used to express that the state resulting from the action continues after the action is completed. This is seen in examples (97)-(100):

- (97) dhwa:-lanka ko:-le seruli-ka tsai ga:pu-ka u-si-dɕa
ground-ABL one-CL jackal-ERG CONTR crow-ERG 3SG.POSS-mouth-LOC
ŋo per-**thət**-ɕu-m lwas-tu
fish pinch-DUR-3SG>3SG.PST-NMLZ see-3SG>3SG>PST
'From the ground a jackal saw the crow pinched the fish in his mouth (and kept it pinched).'

- (98) me gu-ka seɕ-ɕu-m der tsai
DIST.DEM 3SG-ERG kill-3SG>3SG.PST-NMLZ deer CONTR
tebal-dʌ:la tsar-**thar**-u
table-above drop-DUR-3SG>3SG
'He sets ('drops and leaves dropped') the killed deer on the table.'

- (99) bhwa:-mi bhwa:-mi bhwa:-mi mepətshi bha:ra-ra
boil-3PL>3SG boil-3PL>3SG boil-3PL>3SG afterwards pot-LOC
gan-**th**-mi ta:to səl-mu menu
set-DUR-3PL>3SG hot wash-INF NEG.DEON
'They boil, boil, boil it and then set it (and leave it set) in a pot; it must not be washed hot.'

- (100) u-ŋali dʌs-**th**Λ-rɐ
 3SG.POSS-face drop-**DUR**-3SG>3SG.PST
 'He dropped his face (and kept it down).'

Habitual marker *-thal* expresses the regularity of occurrence of an action. This is illustrated in (101)-(103).

- (101) bhansa bΛ-n-**th**al-na
 meal make-2SG>3SG.PST-**HAB**-2SG>3SG.PST
 'You have been making the meals.'

- (102) inima-ra hamham pudza bo-m-**th**al-mi nep-ra
 2PL.POSS-LOC which prayersdo-3PL>3PL-**HAB**-3PL>3PL house-LOC
 'What pujas do they do at your place, in the house?'

- (103) dwa: be-ttoko me-ŋŋa phurka bΛne-ttoko
 loom do-1PE>3SG.PST DIST.DEM-INT decorative.flower make-1PE>3SG.PST
 ma ka:m be-ttoko me:sem-ka wostsø ripap-ka
 CONJ work do-1PE>3SG.PST woman-ERG man male.relative-ERG
 dzo-m-**th**al-miri
 plow-3PL-**HAB**-3PL>3SG.PST
 'We wove and we made decorative flowers and we women worked, and the boys and men ploughed.'

Example (103) is interesting in that it sets up a contrast between women's work and men's work, but only the latter is marked with *-thal*; this is perhaps because the women's work is a list of different activities, whereas the men's is a single activity which occurs repetitively.

2.2.3.2.5.2 Markers which add telicity to the verb

The three other markers in Table 24 which contribute Aktionsart to verbs add a telic dimension, focusing on different aspects of the action's completion. Both *-soŋ* and *-dzəl* focus on the resulting effect on the object of the verb, while *-leŋ* focuses on one of the boundaries of the action, usually the end, but occasionally the beginning. The expression of the effect on the

say-3PL>3SG CONJ that.much alcohol-COM DIST.DEM

thɯr-mi

send-3PL>3SG

'The bride's side sends off what they call Runche Phila (pig's thigh), they send it off with alcohol.'

Examples (107)-(109) illustrate the effect of adding *-dzəl* to the verb, conveying that the object is set out in anticipation of a further action.

(107) metthamma udi:kam homlo-ŋa kho-m-**dzəl**-mu ba:si
 then next.day today-INT cook-INF-ANTIC-INF DEON
 'It needs to be completely cooked today (for a set-up tomorrow).'

(108) oram sen-**dzəl**-tsi la:-tsi
 PROX.DEM kill-ANTIC-1DI>3SG go-1DI
 'We'll kill him off (and leave his body) and go.'

(109) ʌni memlo pheri mə bha:le-ka ko:-le u-mer
 and then again DIST.DEM cock-ERG one-CL 3SG.POSS-tail
 dʌs-**dzəl**-lu ʔe mesinɖa
 drop-ANTIC-3SG>3SG.PST HS then
 'And then the cock dropped (and left) one of his tail [feathers].'

Examples (110)-(112) illustrate the effect of adding *-leɿ* to the verb: it usually adds a focus to the end-point of an action, although in some cases, the focus is on the starting point. This is the case with *dum-le-mu*, (drink-LIM-INF) which can be either 'be about to drink' or 'finish drinking'. (Note that even though the focus on the starting point comes up in elicitation, I have not found it in corpus examples.) It commonly combines with verbs like 'sleep' (see (112)) and 'die' to express a change of state, and especially with past forms, there is a focus on the immutability of the action, bringing it close to a perfect interpretation.

(110) lalam sisi hoŋŋa dum-**lek**-pa bʌnɯ:-rɯ ʔe
 red blood like become-LIM-ACT.PTCP make-3SG>3SG.PST HS
 'He made (the water) completely red like blood.'

(111) ʔǎ pom-**le**-mu ba:si waŋ-mim-ka pe-si menɯ
 yes eat-**LIM**-INF DEON other-PLU-ERG eat-NMLZ NEG.DEON
 gu-ka-ŋa pɯ:
 3SG-ERG-INT eat.3SG>3SG
 'Yes, he must finish eating everything, other people cannot eat [the special offerings],
 only he eats.'

(112) uni-khlea wo lamtsoko-ra ʌm-**le**
 3PL.POSS-dog also door-LOC sleep-**LIM**.3SG
 'Their dog is asleep/has fallen asleep at the door.'

Khaling has a similar distribution of Aktionsart formatives, again largely divisible into those focusing and those reducing the telicity of the verb.

+ telic	- telic
-tə (P remains)	-dze (continuous)
-sə (P gone)	-ther (habitual)
-tʌ (APPL -sʌ + P remains -tə)	-de (continuous, with expectation of subsequent action; state resulting from anterior action)
-khʌ (AM motion/telic)	

Table 25. Khaling Aktionsart markers

Of interest among these, insofar as they differ from what is found in Thulung, are *-tʌ* and *-khʌ*.

The former combines the functions of two markers, namely applicative *-sʌ* and the telic marker involving the leaving behind of the P: the result is the expression of a telicized action, involving depositing a P somewhere and having been performed on behalf of someone else. This is seen in (113).

(113) ʔuŋʌ ʔi-sʌikʌl khoŋn-tʌ-nɛ
 1SG.ERG 2SG.POSS-bicycle bring-LEAVE+APPL-1SG>2SG

'I will bring and leave the bicycle there for you (before you arrive) [note that the object indexed on the verb is 2sg, namely the recipient, as is usual with applicatives]

This can be contrasted with the non-applicativized version of the same derived verb:

- (114) ʔuŋʌ ʔi-sʌikʌl khoɔt-tʌ-ŋʌ
 1SG.ERG 2SG.POSS-bicycle bring-LEAVE-1SG>3SG
 'I will bring your bicycle (and leave it there).'

Another interesting situation in Khaling is *-kha* [khot], which is polysemous, used to express either associated motion function (as in (115)) or telic Aktionsart (as in (116)). Speakers say that context is relevant in determining which of the two interpretations is the dominant one.

- (115) ʔuŋʌ dzhola plep-khʌ-tʌ.
 1SG.ERG bag hold.under.arm-DO&GO-1SG>3SG.PST
 'I held the bag under my arm and left.'

- (116) dzhūŋ-ʔɛ ʔʌ-khʌr tʌm-khʌ-tɛ.
 cold-INSTR 1SG.POSS-hand be.cold-AWAY-3SG.PST
 'My hands became numb from the cold.'

The meaning of the sentence in (116) is the same as if it contained an underived verb, but *-kha* adds a focus on the thoroughness of the resulting state.

2.2.3.2.5.3 *The expression of aspect*

Progressive and perfect aspect can be expressed in Thulung, and this is done periphrastically.

Progressive aspect is expressed in two ways, both involving converb forms of the main verb:

-with the anterior converb *-sa*, usually followed by an intensifier, and a finite form of the verb *bumu*, 'be'. This is seen in (117);

- (117) ham be-sa-ŋa bu-ni athambili
 what do-ANT.CVB-INT be-2PL nowadays
 'What are you doing these days?'

-with the simultaneous converb *-to* and a finite form of the auxiliary *lenmu*, as in (118).

- (118) liser ʌbʌ siʔ-to ler-ʉ
 millet now bear.fruit-SIM.CVB AUX-3SG>3SG
 'The millet is now bearing fruit.'

Perfect aspect is expressed through a nominalized past form of the main verb, followed by a copula, which can be in the non-past, as in (119), or past, as in (120).

- (119) hunu gre:nem-ra ʌs-ta-m bu
 over.there nettle-LOC go-3SG.PST-NMLZ be.3SG
 'He has gone for nettles over there.'

- (120) po-ku dʒi kur-rʉ-m ba-iɖa ʔe
 hen-GEN egg carry-3SG>3SG.PST-NMLZ be-3SG.PST HS
 'He had carried eggs.'

To close this section on the discussion of suffixal slot 2, I offer some thoughts on how the Kiranti material presented above is relevant to a wider linguistic context. As far as Associated Motion is concerned, this is currently a topic of great interest to typologists and Kiranti languages have a considerable amount to contribute to the definition of the domain: the traditional definitional criteria require some extensions in order to properly describe Kiranti data. Deixis, generally described in terms of trans- vs. cis-locative, should also include circumlocative deixis, because the ambulative marker (which involves translational motion but with round-about directionality) is omnipresent in this language area; the argument of motion needs to be extended from S, A or P to the possibility of two arguments undergoing the motion simultaneously (as Kiranti languages have transitive equivalents to the motion verbs used to express Associated Motion). And another significant contribution is the importance of distinguishing between orientation and Associated Motion: both can be expressed through the same formative, yet the semantics are very different (as orientation involves no translational motion) and linguists must be aware of the difference when producing descriptions of these markers.

As far as Aktionsart markers are concerned, these remain very challenging to describe across Kiranti languages, in part because the markers are polysemous in some languages (for example expressing Associated Motion or Aktionsart), and in part because they are

considered by many speakers to be optional, from a semantic point of view, and to be essentially stylistic devices. There is also the fact that the interpretations are linked with tense, and that the majority of examples in grammars (and indeed corpora, which tend to be made up of stories) are in the past tense, leading to only a partial view of the function of the markers. Furthermore, language contact with Nepali, as in the case of ideophones, is of great interest in studying the material in slot 2, as Predicate derivation-like forms are very common in Nepali (and throughout South Asia), yet do not have exactly the same functions as in Kiranti, and it seems likely that part of the challenge of describing these forms in Thulung and other Kiranti languages is due to the influence of Nepali (on both speakers and fieldworkers).

2.2.3.3 Suffixal slot 3: indexation markers

This section discusses the indexes used for person/number/clusivity contrasts on Thulung verbs. In addition to presenting the paradigms of indexes, it also presents some of the characteristics of indexes in Thulung: tense marking, syncretism, inflectional gaps, allomorphy, evidence for underlying person hierarchy, each of which is discussed in turn. This data has most extensively been presented in Lahaussais (2011), but is also treated in Lahaussais (2004). Relevant discussion of the presentation of transitive paradigms can be found in Lahaussais (2020a).

Thulung has two full sets of indexes, based on the transitivity of the verb: intransitive indexes, and transitive indexes. As discussed above, these indexes sometimes occur with the same, labile root. Table 26 presents both non-past (NPST) and past (PST) intransitive indexes, arranged according to the person and number of the S.

	NPST	PST
1SG	-ŋu	-ŋoro/-ŋro
1DI	-tsi	-tsi/-ttsi
1DE	-tsuku	-tsoko/-ttsoko
1PI	-i [-d̪i]	-d̪i/-iri
1PE	-ku	-toko /-ttoko
2SG	-na	-na /-nna
2DU	-tsi	-tsi /-ttsi
2PL	-ni	-ni /-nni
3SG		-d̪a /-ira

3DU	-tsi	-tsi /-ttsi
3PL	-mi	-miri/-mri

Table 26. Intransitive paradigm:

Non-past (middle column) and past (right column) indexes

Tables 27 and 28 present transitive indexes for non-past forms and past forms respectively. The A argument is in the vertical column, and the P argument is in the horizontal row; individual cells correspond to the various possible A+P scenarios.⁵⁹

A /P→ ↓	1SG	1DI	1DE	1PI	1PE	2SG	2DU	2PL	3SG	3DU	3PL
1SG						-ni	-nitsi	-nini	-u [- pu]	-utsi [- putsi]	-umi [- pumi]
1DI									-tsi		-tsimi
1DE						-tsuku					-tsukumi
1PI									-i [- ɖi]	-itsi [- ɖitsi]	-imi [- ɖimi]
1PE						-ku				-kutsi	-kumi
2SG	-ŋi		-tsiki		-ki				-na	-natsi	-nami
2DU	-ŋitsi		-tsiki		-kitsi				-tsi		-tsimi
2PL	-ŋini		-tsikini		-ki				-ni	-nitsi	-nimi
3SG	-ŋi	-sa	-tsiki	-sa	-ki	-na	-tsi	-ni	-u [- ɖu]	-utsi [- ɖutsi]	-umi [- ɖumi]
3DU	-ŋitsi	-satsi	-tsiki	-satsi	-kitsi	-natsi	-tsi	-nitsi	-tsi		-tsimi
3PL	-ŋimi	-sami	-tsikimi	-sami	-kimi	-nami	-tsini	-nimi	-mi	-mitsi	-mi

⁵⁹ Discussion of the evolution of the presentation of these paradigms from 1850 to the present can be found in the section on Grammaticographical models in Section 3.1.2 of this document.

Table 27. Transitive paradigm: Non-past indexes⁶⁰ (all suffixal)

	1SG	1DI	1DE	1PI	1PE	2SG	2DU	2PL	3SG	3DU	3PL
1SG						-ni / -nni	-nitsi / - nnits i	-nini / - nnin i	-to / -uto	-totsi / - utotsi	-tomi / - utomi
1DI									-tsi / -ttsi		-tsimi / - ttsimi
1DE						-tsoko / -ttsoko				- tsokomi / - ttsokom i	
1PI									-d̥i/ id̥i	-d̥itsi / - id̥itsi	-d̥imi / - id̥imi
1PE						-toko / -ttoko			-tokotsi / -ttokotsi	-toko / - ttoko	
2SG	-ŋiri / -ŋri		-tsiki / -ttsiki		-tiki / -ttiki				-na / -nna	-natsi / - nnatsi	-nami / - nnami
2DU	- ŋirits i / - ŋritsi		-tsiki / -ttsiki		-tikitsi / - ttikitsi				-tsi / -ttsi		-tsimi / - ttsimi

2PL	- ɲirini /- ɲrini		- tsikini /- ttsikin i		-tikini / - ttikini				-ni / -nni	-nitsi / - nnitsi	-ni / - nni
3SG	-ɲiri /-ɲri	-saɖɖa	-tsiki / -ttsiki	- saɖɖa	-tiki / -ttiki	-na / -nna	-tsi / -ttsi	-ni / -nni	-ɖa /- uɖa	-ɖatsi /- uɖatsi	-ɖami / -uɖami
3DU	- ɲirits i / - ɲritsi	- saɖɖats i	-tsiki / -ttsiki	- saɖɖa tsi	-tikitsi / - ttikitsi	- natsi /- nna si	-tsi / -ttsi	- nitsi /- nnits i	-tsi / -ttsi		-tsimi / - ttsimi
3PL	- ɲirim i / - ɲrimi	- saɖɖa mi	- tsikim i / - ttsiki mi	- saɖɖa mi	-tikimi / - ttikimi	- nami /- nna mi	- tsimi /- ttsim i	- nitsi (mi) /- nnits i(mi)	-miri /-mri	-miritsi /- mritsi	-miri /- mri

Table 28. Transitive paradigm: Past indexes

2.2.3.3.1 Tense marking

Tense is expressed through an unmarked non-past, and a marked past which is fused into portmanteaux in the past indexes. The unmarked nature of the non-past can be gleaned from the intransitive 3SG.NPST form of verbs, which is a bare stem. As this suggests that both 3SG and non-past are unmarked, the intransitive 3SG.PST marker provides clues as to the form of the past marker at an earlier stage (when it presumably had its own slot in the template, as a distinct formative), namely *-ɖa*. Based on Michailovsky's (1994) reconstruction, Thulung *ɖ* develops from Proto-Kiranti **t*; on the basis of a past marker **-t*, it is possible to analyze the portmanteaux indexes in the past paradigm and to assign values to various phonemes. Such values are listed in Table 29.

	NPST	PST	*PST	*PST breakdown
1SG>3SG	-u/-pu	-uto	-u-t-o	-u =1SG (agent), -t =PST, -o = possibly, additional 1SG marking
1SG	-ŋu	-ŋro	-ŋ-t-o	-ŋ(u) =1SG (subject), -t = PST, -o = possibly, additional 1SG marking
1DI>3SG, 1DI	-tsi	-ttsi	-t-tsi	-t = PST, -tsi = DU
1DE>3SG, 1DE	-tsuku	-ttsoko	-t-tso-ko	-t = PST, -tso = DU, -ko = exclusive
1PI>3SG, 1PI	-i/-dji	-dji	-t-i	-t = PST, -i = 1PI
1PE>3SG, 1PE	-ku	-ttoko	-t-to-ko	-t = PST, -ko = exclusive, -to = possibly, additional past morpheme
2SG>3SG, 2SG	-na	-nna	-t-na	-t = PST, -na = 2SG
2DU>3SG, 2DU	-tsi	-ttsi	-t-tsi	-t = PST, -tsi = DU
2PL>3SG, 2PL	-ni	-nni	-t-ni	-t = PST, -ni =2PL
3SG>3SG	-y/-dy	-dy	-t-y	-t =PST, -y = 3SG (agent)
3DU>3SG, 3DU	-tsi	-ttsi	-t-tsi	-t = PST, -tsi = DU
3PL>3SG, 3PL	-mi	-mri	-m-t-i	-m + -i = 3PL, -t = PST

Table 29. Breakdown of past indexes (from (Lahaussois 2011: 202))⁶¹

What is unusual about the breakdown of indexes in Table 29 is that the position of the past marker is not always consistent with respect to markers for arguments' person/number/clusivity, sometimes occurring before them, sometimes between. Because this type of analysis is presumed not to be synchronically available to speakers (as it is based on an abstract form of the past marker that speakers are unaware of), I believe the best solution for glossing indexes (both past and non-past) is to treat them as units expressive of the various possible scenarios that trigger their use: they index person, number, clusivity for either an S, or a combination of an A and a P, as well as tense. As such, I adopt the convention of glossing them as either S(.NPST) or S.PST, or A>P(.NPST) or A>P.PST (S, A and P are replaced in actual glosses with the person, number and clusivity of the relevant arguments; I do not use .NPST in glosses, except when necessary for clarification), and I choose not to break them down into small units. This can be equated with treating indexes as 'co-argument sensitive' (Witzlack-Marakevich et al. 2016) to the extent that every index is associated with a scenario.

⁶¹ Table 29 reflects a different transcription convention than the one I currently use: what is transcribed y corresponds to #.

More discussion of this will follow, with respect to whether person hierarchies can be identified as operating within the indexes.

2.2.3.3.2 Cell syncretism

Some cells exhibit syncretism, with the same index found in multiple cells. This is notably the case for the 1DI, 2DU and 3DU forms (both past and non-past) in Table 26, which can be analyzed as indexing only (dual) number and not the person of the single argument. In the transitive paradigms (Tables 27 and 28), syncretic forms are sometimes in contiguous cells, in which case the cells are extended over several A+P scenarios. The syncretism is interesting because it highlights the fact that in certain scenarios a feature is not marked redundantly: this is the case when both arguments are dual, yet the index only has a single dual feature (this is the case, for example, of 3DU>3SG.PST and 3DU>3DU.PST, which are syncretic). The glosses assigned to the scenarios nonetheless specify the person/number/clusivity of the A and P in question, even though this is technically redundant; this decision was taken in order to maintain the same glossing conventions for all polypersonal indexes. The fact that arguments are often not overtly expressed as case-marked NP's means that the only data available on person/number/clusivity/role of the arguments is on the verb (and the context for the utterance); as a result, it is useful that the glosses for indexes encode the information.⁶²

2.2.3.3.3 Allomorphy

The indexes in Tables 26, 27, 28 exhibit a considerable amount of allomorphy. This allomorphy can be broken down into three types, in all cases conditioned by the tense of the form in question. They are summarized here, and then described and exemplified in greater detail.

⁶² One additional reason to fully express person/number/role in the glosses for the indexes is that simplifying the glosses to reflect syncretic forms (and the neutralization of some features) might not always be 'simpler': imagine an utterance made up of the verb *jal-tsi*; this would usually be glossed *hit-3DU>3SG.PST* or *hit-3DU>3DU.PST* (or even *hit-3DU>3SG.NPST* or *3DU>3DU.NPST*, depending on the case). Taking into account the neutralization of dual marking when occurring on both arguments, one might be tempted to reduce the gloss to *hit-3DU>3* (reducing it to also take into account the fact that in this precise case NPST and PST indexes are syncretic); what then to do if the actual context called for a 3SG A and a 3DU P? Would the solution be to gloss that scenario *hit-3>3DU*? Or perhaps *hit-DU*, in order to fully take into account only the feature actually marked. I believe that the gloss should make explicit the scenario in question, because applying the idea of glossing only present features is only possible for a limited subset of indexes (and many of the features are synchronically inaccessible), and would result in considerable inconsistency.

-only affecting non-past forms: some verb classes have variant forms of specific indexes. While this is synchronically unconditioned, there are diachronic processes at work here, with reanalysis of a stem-index boundary and analogy; (these are indicated with '[]' in Tables 26 and 27)

-only for past forms: the allomorph of the index is selected based on the stem-final, with one set of indexes for consonant-final stems, and another set for vowel-final stems; (these are indicated with / in Tables 26 and 28)

-as a subset of the previous type, there is also phonologically-conditioned allomorphy, only for past forms, with the initial consonant of indexes which are -d-initial assimilating to the verb stem final; (these are not indicated in Tables 26-28, as the tables only shows indexes and not stem+index combinations)

Concerning the first type of allomorphy, limited to non-past forms, and indicated in Tables 26 and 27 with square brackets, it is determined by verb class (but is not a matter of phonological conditioning). This allomorphy only affects non-past indexes for specific scenarios, with distinct patterns for 1SG A indexes, on the one hand, and for the 1PI and 3SG S/A indexes on the other. Even though the allomorphy is different (in conditions and realization), in both cases it can be linked to diachronic processes.

The 1SG A indexes use the variant in *-pu* (and *-putsi* (1SG>3DU) and *-pumi* (1SG>3PL)) with verbs of the t-, k-, m-, n- classes; the *-u* variant is found with all other verb classes, as is presented in Table 30. The analysis proposed for these variant forms in (Lahaussois 2011) is that the form in *-p* is a result of reanalysis at an earlier stage of the language of the boundary between stem and index on the basis of p-class verbs, with a CVp-u form being reanalyzed as CV-pu, and this new 1SG A index spreading to other verb classes (but crucially, not all).

	t-class verb	p-class verb
	k-class verb	ŋ-class verb
	m-class verb	r-class verb
	n-class verb	l-class verb
		s-class verb
		V-class verb

1SG>3.NPST ⁶³	-pu	-u
--------------------------	-----	----

Table 30. Allomorphs of 1SG>3 according to verb class

These data can be exemplified with the following forms from each set of verbs:

k-class verb *bo:mu*, 'push down', has the form *bok-pu* for a 1SG>3SG scenario;

l-class verb *jalmu*, 'strike', has the form *jal-u* for a 1sg>3SG scenario.

The variation in the 1PI and 3SG S/A indexes is an alternation between vowel- and *ɔ̄*-initial forms. The verbs taking *ɔ̄*-initial forms in the non-past for 1PI and 3SG S/A are *p*-, *m*- and *n*-classes, with other classes using vowel-initial forms.

	p-class verb m-class verb n-class verb	t-class verb k-class verb ŋ-class verb r-class verb l-class verb s-class verb V-class verb
1PI.NPST	-ɔ̄i	-i
1PI>3 ⁶⁴ .NPST		
3SG>3.NPST	-ɔ̄u	-u

Table 31. Allomorphy for 1PI and 3SG non-past indexes according to verb class

These data can be exemplified through the following verb forms.

p-class verb *remmu*, 'see,' has the form *rep-ɔ̄i* for a 1PI>3SG scenario;

k-class verb *ki:mu*, 'tie', has the form *kik-i* for a 1PI>3SG scenario.

One possible explanation for the variant forms with *ɔ̄*-initial is that they are connected to postfinal -t (discussed in 2.2.2.2 above). The presence of the postfinal is not transparent in Thulung currently, but it is likely that it was at an earlier stage of the language (similar to

⁶³ Here, as in Table 31, >3 is a cover for 3SG, 3DU and 3PL; Number marking must be added to the form in the table for the dual and plural P indexes.

⁶⁴ See previous footnote.

Khaling, where verbs can be divided into CVC and CVC-t, with paradigms largely similar apart from a few distinct forms (for contrastive paradigms, see Jacques (2015 table 3)).

Thulung may then have neutralized the difference between CVC and CVC-t verb types, but in collapsing them, maintained indexes consistent with a CVC root for some verbs, and with a CVCT root for other verbs. In other words, for a hypothetical verb form *CVp-t-i, with a 1PI>3SG.NPST index, the post-suffix -t could have been reanalyzed as part of the index, as opposed to an element related to the verb stem. In a way similar to the situation with the reanalysis of CVp-u into CV-pu, this reanalysis was--for some unknown reason--only associated with certain verb stem classes.

The second type of allomorphy for indexes involves past forms only, and is phonological insofar as index selection is based on the phonological environment of the preceding stem. These allomorphs are indicated in Tables 26 and 28 with a '/' separating the two possible forms: allomorphs listed before the '/' occur after stems which are consonant-final; allomorphs listed after the '/' occur after stems which are vowel-final.

The various paradigms illustrating stem distribution for the different verb classes reveal that there are in fact few vowel-final stems in Thulung past forms (remember that this allomorphy only concerns past stems): the only verb classes that are involved are the η-, s- and V-classes. V-class verbs have exclusively vowel-final stems; η- and s-stem classes have vowel-final stems in all but 1PI and 3SG S/A forms in the non-past. Because of the number of cells with this type of stem-final conditioned allomorphy, I only illustrate the phenomenon with a few examples:

V-class *bhremu*, 'buy', has the form *bhre-uto* for a 1SG>3SG.PST scenario;

η-class *qu:mu*, 'drink', has the form *qu-uto* for a 1SG>3SG.PST scenario;

s-class *phromu*, 'untie', has the form *phro-uto* for a 1SG>3SG.PST scenario.

These forms stand in contrast, for example, to an m-class verb *brammu*, 'scratch', which has the form *bram-to* for a 1SG>3SG.PST scenario.

As an intransitive example, V-class verb *lu:mu*, 'feel', has the form *lu-ηro* for a 1SG.PST scenario. In contrast, the m-class verb *ammu*, 'sleep', has the form *am-ηoro* for the 1SG.PST scenario.

The third type of allomorphy in the indexes is a subset of the previous type: it concerns past indexes which exhibit assimilation of their initials based on the preceding stem. These allomorphs are not indicated in Tables 26-28 as they concern stem + index combinations, whereas the Tables only show indexes without their stems. The only indexes concerned by this allomorphy are those which are -q-initial. The allomorphy is exhibited in

Table 32 below, showing the four scenarios which are concerned and the allomorphs found with the different verb classes. The form in -*ɖ* is the one listed in the paradigms of indexes for the simple reason that it is the one that does not show assimilation and cannot be explained through phonological conditioning.

	t-class verb m-class verb n-class verb ŋ-class verb	k-class verb; s-class verb; p-class verb	r-class verb	l-class verb
1PI.PST, 1PI>3.PST ⁶⁵	- <i>ɖ</i> i~ri	-ti	-ri	-li
3SG.PST	- <i>ɖ</i> a~-ra	-ta	-ra	-la
3SG>3.PST	- <i>ɖ</i> u~-ru	-tu	-ru	-lu

Table 32. Phonologically-conditioned allomorphs for
1PI and 3SG S/A indexes with different verb classes

The phonological changes in question are the following:

ɖ > t/_s, k, p (voicing assimilation);

ɖ > r/_r (assimilation);

ɖ > l/_l (assimilation), with conditioning based on the nature of the final consonant of the verb stem, as the indexes assimilate to the stem final.

Thus the 3SG.PST index *-ɖa* (which alternates freely with *-ra*), seen in *swaŋ-ɖa*, disappear-3SG.PST, has allomorphs *-ta*, *-ra*, *-la* conditioned by the final consonant of the verb stem, as seen in:

thok-ta, ripen-3SG.PST;

bhur-ra, be.angry-3SG.PST;

dzal-la, be.warm-3SG.PST.

The same applies to indexes for 1PI.PST, 1PI>3.PST, 3SG.PST and 3SG>3.PST, the allomorphs being illustrated in Table 31, according to verb class.

⁶⁵ Even though I do not use 1PI>3 as a gloss for person indexes, this formulation covers the indexes for all scenarios with 1PI A and 3 P. The indexes for 3DU and 3PL P have number markers as well, namely *-tsi* and *-mi* respectively. The same applies to forms listed in the table as 3SG>3. Note also that the allomorphy described is true of forms which are non-past as well as past.

Other indexes in Tables 26-28 which are not -q-initial are not affected by this particular allomorphy.

2.2.3.3.4 Inflectional gaps

The transitive paradigms in Tables 27 and 28 contain empty cells, which are found with scenarios of exact coreference of arguments (1>1 and 2>2), as well as with partial coreference⁶⁶ (2<>INCL). These inflectional gaps can be traced to two different sources.

Exact coreference of arguments is expressed through reflexive voice marking, marked with *-sit* (see 2.2.3.2.4.3). This affects scenarios of 1>1 and 2>2. As reflexivization also brings about the use of intransitive indexes, these forms are not part of the transitive paradigm, resulting in empty cells for said scenarios.

The cells corresponding to 2<>INCL scenarios (in other words all scenarios simultaneously involving 2nd person and inclusive arguments) involve partial co-reference (as the 2nd person is included in inclusive referents.) Partial coreference cannot be encoded in Thulung using reflexivization, and other strategies are employed to express these scenarios, despite the inflectional gaps:

a) changing the scenario to exact coreference and using a reflexive construction:

- (121) gutsi dzu:-sit-tsi-lo ko:-le-ŋa je hum-sin-tsi
 1DI be.cold-REFL-1DI-TEMP one-CL-INT cloth wrap-REFL-1DI
 'When we are cold, we wrap ourselves in the same cloth.'
 Intended: You covered us (incl) in the same blanket.

- (122) gutsi tshatta-ka rim-sin-tsi
 1DI umbrella-INSTR cover-REFL-1DI
 'We cover ourselves with the umbrella.'
 Intended: You covered us (incl) both with the umbrella.

b) changing the reference to 2SG>1SG:

⁶⁶ The terms 'exact' and 'partial' coreference are used in Haspelmath (2019).

(123) gana go-lai wakha lamdi-beṭ-ŋi
 2SG 1SG-DAT slowly walk-CAUS-2SG>1SG.PST
 'You made me walk slowly.'
 Intended: You slowed us (incl) down.

c) using a semantically related intransitive verb (instead of a transitive one):

(124) gana me-dzopa gari thən-na ma:la gui si-i
 2SG NEG-good car drive-2SG>2SG.PST COND 1PI die-1PI
 'If you drive a car so badly, we will die.'
 Intended: You will kill us all. (scenario: driver driving dangerously).

2.2.3.3.5 Evidence of referential hierarchies within the paradigm?

As mentioned above, I have chosen to use global glosses for indexes, expressing the person/number/clusivity/role of combinations of arguments, in order to facilitate the interpretation of the arguments involved in a given scenario. This involves analyzing as a single formative material that can in fact in some cases be broken down into distinct elements.

As an example, take the form *jal-nimi*, which can be glossed, based on the cells where it occurs in Table 27, as 'strike-2PL>3PL' or as 'strike-3PL>2PL'. The same form could also be segmented further and glossed as *jal-ni-mi*, 'strike-2PL-3PL'. The values of *-ni* and *-mi* can be inferred from non-past intransitive forms. This analytical approach, rather than the global, scenario-based one I use, is found in many Kiranti language descriptions: it is typical of those found in descriptions by van Driem and his students; see for example (Opgenort 2004: 309 ff) for a break-down of Wambule indexes (then reconstructed into proto-morphemes). The approach is made complicated for Thulung by the fact that most past indexes are portmanteaux (see Table 28).⁶⁷ Another issue is that we do not know to what degree speakers conceive of these indexes as global versus compositional. In some descriptions of Kiranti languages, the analysis of indexes descends to level that seems unlikely to be synchronically accessible to speakers.⁶⁸

⁶⁷ In other words, I feel it is important to have a consistent glossing system for all indexes, regardless of whether they are synchronically analyzable or not.

⁶⁸ The partial template in Figure 5 above (from (Jacques et al. 2012: 1101)) is presented below in its complete form. Note how applying a compositional analysis of this sorts means that, while some formatives are easily identifiable (presumably also by speakers), others are considerably less so: how convincing is the gloss in the

Nonetheless, non-past indexes *can* for the most part be broken down into the component formatives and associated with glossable features, and one advantage of this type of analysis is that it reveals information about an underlying referential hierarchy. Table 33 lists all of the formatives that I feel can be identified within the global indexes, divided according to role. The annotation system adopted in Table 33 is the following:

'#>#' indicates special forms for an A>P scenario

'#>' indicates a form for an A argument of person #

'>#' indicates a form for a P argument of person #

(When no indication of role is given, the marker in question is unspecified for role and can be used for either. Also, note that some markers in Table 33 are polysemous: this is the case with *-ni* which marks both 1>2 and 2PL.)

	feature(s)	form
special forms (A>P)	1>2	-ni
	1>3	-u
forms indicating A only	3SG >	-u
forms indicating P only	>1	-ŋi
	>INCL	-sa
	>1PE	-ki
forms unspecified for role	1PI	-i
	1PE	-ku
	2SG	-na
	2PL	-ni

slot "person (I)" for -n- (2PL, 1>2, INF), or that in the "reflexive" slot for -n (3PL.PST): these formatives are probably only identifiable by speakers when they occur in combination with other elements, raising the question of how much sense an analytical breakdown makes. The answer is of course dependent on the language, and on the type of analysis one chooses.

prefix	stem	person (I)	auxiliary	person (II)	TAM	person (III)	reflexive ⁵	number
-1	0	+1	+2	+3	+4	+5	+6	+7
?i- (2 or INVERSE)	Σ	-ŋ- (1SG)	-de(t-)	-ŋʌ/-ʌ (1SG.S/O)	-t- past	-e (2/3)	-si	-su (2/3DU)
mu- (NEG)		-i- (DU)	-t ^h er(d-)	(1SG.S/O)	-w- irrealis	-ŋʌ/-ʌ (1SG.S/O)	-su (1SG.PST)	-nu (2/3PL)
ma- (2.NEG)		-k- (1PL)	-k ^h ʌ(t)- etc	-u (1SG>3)		-u (1SG>3)	-n (3PL.PST)	-ni (2PL NPST, 1>2.PST)
		-n- (2PL, 1>2, INF)		-i (1DU)		-i (1DI)		
				-k (1PL.PST)		-u (1DE)		
				-ne (1>2, INF)		-(i)ki (1PI)		
						-(ʌ)kʌ (1PE)		

	DU	-tsi (sometimes with vowel harmony ⁶⁹)
	PL	-mi

Table 33. Features of markers within non-past indexes

The data in Table 33 can be used to reconsider the entire paradigm of non-past transitive indexes, and to examine the order in which these component formatives are ordered within the indexes. When the formatives are converted into glosses indicating the features they code, the order in which the components appear reveals interesting information about the organization of the paradigm and the underlying hierarchy. Table 34 is an attempt to represent the resulting "paradigm": glosses (from Table 33) are used for the component formatives that make up the full indexes, in the order in which they occur (separated by + signs); they are color-coded with respect to whether they are associated with a feature of the A (in red), the P (in blue), both roles (in black). (The few markers which have the same feature in both arguments are in a purple font).

⁶⁹ This is the case when in some past indexes, *-tsi* becomes *-tso*.

	1SG	1DI	1DE	1PI	1PE	2SG	2DU	2PL	3SG	3DU	3PL
1SG						1>2	1>2 +DU	1>2+2 PL	1>3	1>3 +DU	1>3+PL
1DI									DU		DU+PL
1DE						DU+1PE					DU+1P E+PL
1PI									1PI	1PI+ DU	1PI+PL
1PE						1PE				1PE +DU	1PE+PL
2SG	>1		DU+>1PE		>1PE				2SG	2SG +DU	2SG+PL
2DU	>1+ DU		DU+>1PE		>1PE +DU				DU		DU+PL
2PL	>1+ 2PL		DU+>1PE +2PL		>1PE				2PL	2PL +DU	2PL+PL
3SG	>1	>INCL	DU+>1PE	>INCL	>1PE	2SG	DU	2PL	3	3+D U	3+PL
3DU	>1+ DU	>INCL+ DU	DU+>1PE	>INCL +DU	>1PE +DU	2SG +DU	DU	2PL+ DU	DU	DU	DU+PL
3PL	>1+ PL	>INCL+ PL	DU+>1PE +PL	>INCL +PL	>1PE +PL	2SG +PL	DU+ 2PL	2PL+ PL	PL	PL+ DU	PL

Table 34. Analysis of the order of exponents in non-past transitive indexes

What emerges from this table is a picture of the order of the formatives associated with different arguments:

- in the green parts of the paradigm, material associated with the P argument (blue glosses) occurs first, either before material associated with the A argument (red glosses), or alone;

- in the orange parts of the paradigm, material associated with the A argument (red glosses) occurs first, either before material associated with the P argument (blue glosses), or alone.⁷⁰

Noteworthy is the fact that features associated with 3rd person P's are limited to number indexes (for dual and plural; the singular is not marked, and there is therefore no feature associated with 3SG P within the indexes).

The green and orange parts of the paradigm in Table 34 perfectly match the sections of paradigms traditionally labeled inverse and direct in Kiranti (Michailovsky 1988: 82; Ebert 1994: 26 ff; Jacques et al. 2012), even though Thulung has no overt direction marking (unlike, say, Khaling, which has an inverse marker which is used on any form where the A is lower-ranked on a 1>2>3 hierarchy than the P). One might expect that a language with no direction marking is not sensitive to a referential hierarchy where indexes are concerned, and that indexes for transitive scenarios potentially encode the formatives for each argument in an A>P order. Instead, the indexes of Thulung show that this is not the case: indexes are made up of formatives which are always ordered with the higher-ranked person first, showing that Thulung's indexes follow a hierarchy. The hierarchy is the same as was posited for differential case marking in (13), and can be summed up as 1>2>3⁷¹. Even when these formatives only code number, they are organized according to the hierarchy: the index for 1DI>3PL is, according to Table 34, made up of formatives DU+PL, with the formative associated with the 1st person occurring *before* the one associated with the 3rd person; similarly, the index for 3PL>2PL is made up of formatives 2PL+PL, with the patient formative occurring *ahead* of the agent formative because it is higher on the hierarchy.

Among the ample recent work on referential hierarchies (Cristofaro & Zúñiga 2018; Rose 2018; Konnerth & DeLancey 2019), DeLancey (2018a) is of particular relevance to the situation in Thulung. Having extensively studied Kiranti indexical data relevant to 'local' and 'mixed' domains (those parts of the paradigm involving speech act participants only, for the former, and speech act participant + other, for the latter as per Zúñiga (2006)), DeLancey has

⁷⁰ Also colored orange are cells where material associated with A>P (black glosses) either occurs alone or occurs before material associated with the P argument (blue glosses).

⁷¹ Given that Thulung has the same indexes for 3>3 scenarios, no matter the nature of the 3rd persons, the language is not sensitive to any kind of proximate/obviative distinction (unlike Gyalrongic languages; (Jacques & Antonov 2014)).

suggested the importance of sociopragmatic factors -- namely avoiding reference to the addressee, which "is the realm of bickering and badinage, of promises, threats, request and cajolery, negotiations, offers and rejections" (DeLancey 2018a: 370) -- in determining which features are marked in different cells of the paradigm.⁷²

Considering only singular (and non-past) scenarios, the data for Thulung can be boiled down to Table 35, for local and mixed domains.

	1	2	3
1		-ni	-u
2	-ŋi		-na
3	-ŋi	-na	

Table 35. Local and mixed indexation (singular, non-past) in Thulung

This approach has allowed DeLancey to uncover a number of trends in the marking of these two domains across Tibeto-Burman languages with polypersonal indexation, and Thulung data follows a certain number of them: it has a special index for 1>2 scenarios (probably diachronically connected with the 2PL index; see DeLancey (2018b: 361)); the same marker is used for both 2>1 and 3>1 scenarios (DeLancey 2018a: 366), essentially functioning as an inverse index (as what is marked is thus a 1st person P).

Another question raised by DeLancey (2018a: 368) is whether there is any basis for claiming that 1st person or 2nd person is higher ranked than the other across these languages. As far as Thulund data is concerned, and as suggested above, non-singular local forms are illuminating with respect to this question⁷³:

	1SG	2SG	2DU	2PL
1SG		1>2	1>2+DU	1>2+2PL
2SG	>1			
2DU	>1+DU			
2PL	>1+2PL			

⁷² DeLancey only considers singular person markers in his simplified paradigms, in order to abstract away number, which can make the data more opaque for purposes of comparison.

⁷³ Number seems like a crucial category to take into account in this data, despite being ignored in some analyses of Kiranti languages (for entirely justified reasons); see Witzlack-Marakevich et al. (2016: 547).

Table 36. Features (in the order in which they appear) in the local indexes

While the portmanteau indexes involving 1SG and 2SG scenarios make it difficult to determine the hierarchy between them, the fact that the number-marking formative related to the 2nd person always follows the portmanteau indexes, regardless of the role of the 2nd person, suggests that we do have some insight into ranking between the two persons, and that 1st person is ranked higher than 2nd person.

Note that while I am suggesting that evidence for the hierarchy in (13) can be found when decomposing the transitive indexes (which I conceive of, as mentioned above, as global markers for scenarios involving two arguments), and that many interesting insights can be gained with this approach, they are essentially relevant only from a diachronic perspective.

Describing the indexation markers of slot 3 raises a number of questions, one of the larger ones being the relationship between diachronic and synchronic approaches. My (synchronically-oriented) decision to describe the formatives as global scenarios indicating, for transitive indexes, two arguments (in terms of their person, number and clusivity), rather than breaking them down into constituent parts is not a common approach for Kiranti languages. This seems, however, partly dictated by the more fusional nature of these markers, compared to those of other languages in the group, for example those which distribute indexal material among prefixes and suffixes. The decision has a number of consequences in terms of glossing the indexes: it results in distinct glosses even in cases of syncretism (for example -miri is glossed as either 3PL>3SG.PST or 3PL>3PL.PST, depending on the context, even though the information distinguishing the two cannot be recovered from the formatives themselves), in order to maintain consistency across all glosses for indexes. It also means that discussion such as that in section 2.2.3.3.5, on evidence for a referential hierarchy *within* indexes (in the absence of an inverse marker that would make point to a hierarchy explicitly), means abandoning the glosses used elsewhere and discussing formatives in terms of their component parts (even though this does not entirely make sense; why not, in that case, consider those component parts to be the relevant formatives?). Nonetheless, it seems to me that even though it is relevant for comparative purposes to resort to a diachronic approach, a synchronic approach makes sense for indexation material such as is found in Thulung, for which the breakdown of formatives is fairly speculative.

2.2.3.4 Suffixal slot 4: mood markers

Thulung has two grammatical mood markers, for optative and irrealis moods. In addition to these, Thulung has an imperative mood, which does not have any grammatical marker but can be identified through a unique 2SG S/A index -a, often listed in paradigms of indexes.⁷⁴ Imperative forms with 2DU and 2PL S/A arguments take the standard non-past indexes.

2.2.3.4.1 Optative mood marker

The optative mood is marked with *-nu*, which is affixed to the finite form of the verb. The optative mood is frequently found with incantations and prayers, expressing a desired outcome. Material presented in this section is discussed in (Lahaussais In press), covering not only the optative marker but the avertive construction which makes use of it.

Examples are provided in (126) and (127), with the optative-marked verb bolded.

- (126) go oram-nuŋ tseŋ-si-ŋu he: d̄əm-ŋu-lo
 1SG PROX.DEM-COM hang-REFL-1SG how become-1SG -TEMP
d̄əm-ŋu-nu rwak-ta ʔe
become-1SG-OPT say-3SG.PST HS

'And he said "I will hang myself onto this, and however I become, **may I become** like that".' [context: man trying to escape from cave by grabbing onto monkey tail and accepting whatever may happen as a result.]

- (127) bu:dzi-ŋa **me-d̄əm-nu** ottha tsiri-ŋa bo:-tsi
 shy-INT **NEG-become.3SG -OPT** here tsiri.ritual-INT do-1DU
 gana banta ba-nna taŋma du: ba-nna
 2SG where be-2SG.PST wild.banana grove be-2SG.PST

'**May it not be shy**, we are doing the 'tsiri' ritual here, where did you stay, did you stay in the wild banana grove?' [ritual chant in which one calls to the wild ginger to reveal itself]

⁷⁴ For the 1-class verb *jalmu*, 'strike', the 2SG>3SG (negative) imperative form is *me-jal-a* (NEG-strike-2SG>3SG.IMP), in contrast with the indicative non-past form *me-jal-na* (NEG-strike-2SG>3SG.NPST).

It is possible to posit a development pathway for the optative marker, from a lexical verb *n̄mtu*, 'be auspicious, be good'. This involves a change whereby two independent clauses, [gu dalaŋa sʌ:] and [n̄], originally in a paratactic relationship, are reinterpreted as making up a single clause, with the original second verb being reinterpreted as an expression of optative mood.

(128) gu dala-ŋa sʌ: n̄
 3SG fast-INT heal.3SG **be.good.3SG**
 'He will heal fast, that is good.' [fieldnotes]

(129) gu dala-ŋa sʌ:-n̄
 3SG fast-INT heal.3SG-OPT
 'May he heal fast.'

One of the functions of the optative mood is in the first clause of purposive and avertive constructions (in the clause expressing desired or undesired outcome), illustrated in (130) and (131) respectively. (More details of the grammaticalization pathway and on avertive constructions can be found in a submitted paper, which is included in the accompanying volume of publications; (Lahaussais In press)).

(130) ma:ke dzəmka si:-ru-n̄
 grain carefully mature-3SG>3SG.PST-OPT
 rwaksaka go biruwa dzəmka khor-pu
 QUOT 1SG small.weeds carefully hoe-1SG>3SG
 'So that the grain matures nicely, I hoe the small weeds with care.' [fieldnotes]

(131) tsett̄s̄o-mim-ka pi:masi me-bi-mi-n̄ rwaksaka
 child-PLU-ERG fruit NEG-beg-3PL>3SG-OPT QUOT
 thʌk-t̄
 hide-3SG>3SG.PST
 'So that the children wouldn't beg for the fruit, he hid it.' [elicited]

While Thulung has no dedicated morphology for the marking of the apprehension domain, the avertive clause is one means of doing so. The construction is made up of two clauses: a negated

optative clause, subordinated by *rwaksaka* (glossed "QUOT"), and a main clause, which expresses the action which is taken to avoid the apprehension-causing situation described by the first, subordinated clause.⁷⁵

Note that the use of *rwaksaka*, the anterior converbal form of 'say', as a subordinator hints strongly at the development of the construction. The intermediary stage, where the converb was interpreted literally, is shown in the bracketed translation in (132); it has grammaticalized into a subordinator essential to the construction:

- (132) *dzu-kka me-si:-nɯ rwak-saka*
 cold-INSTR NEG-die.3SG-OPT say-ANT.CVB
je phɰ-bet-to
 clothes dress-CAUS-1SG>3SG.PST
 [Having said 'May he not die of cold', I dressed him in clothes]
 'So he wouldn't die of cold, I dressed him.' [elicitation]

The source verb for the optative mood marker, *nɯmu*, 'be auspicious', is also involved in a construction for the expression of impersonal negative deontic necessity, illustrated in (133). The construction is impersonal and is used to express general interdictions against certain behaviors (generally those going against cultural norms).

- (133) *hapa kɰrmar be-si menu*
 lots hard.work do-IMPERS.NMLZ NEG.DEON
 'One must not do lots of hard work [while pregnant].'

The construction is made up of a verb nominalized with *-si* (glossed 'IMPERS.NMLZ')⁷⁶ or, less frequently, of an infinitive form of a verb (also considered to be a subtype of nominalization; see Lahaussais (2003b)). The nominalized verb form is followed by the particle *menu* (~ *mɯnɯ*). Presumably an original construction made up of a nominalized verb

⁷⁵ The use of a quotative as the linker between clauses in an avertive construction is found cross-linguistically; see Spronck & Nikitina (2019: 140, 151).

⁷⁶ This impersonal nominalizer is possibly related to the reflexive marker *-si*, used, among other things, to mark impersonal forms of verbs (Lahaussais 2016a)).

(INF or *-si*) and of the negated finite form of the verb 'be auspicious' was reinterpreted as 'One shouldn't V', as suggested in (134).

- (134) nominalized verb (INF or *-si*) + NEG-*nu*
 'V-ing is not auspicious'
 >
 'One shouldn't V'

Figure 5 illustrates the parallel developments involving the verb *numu*.

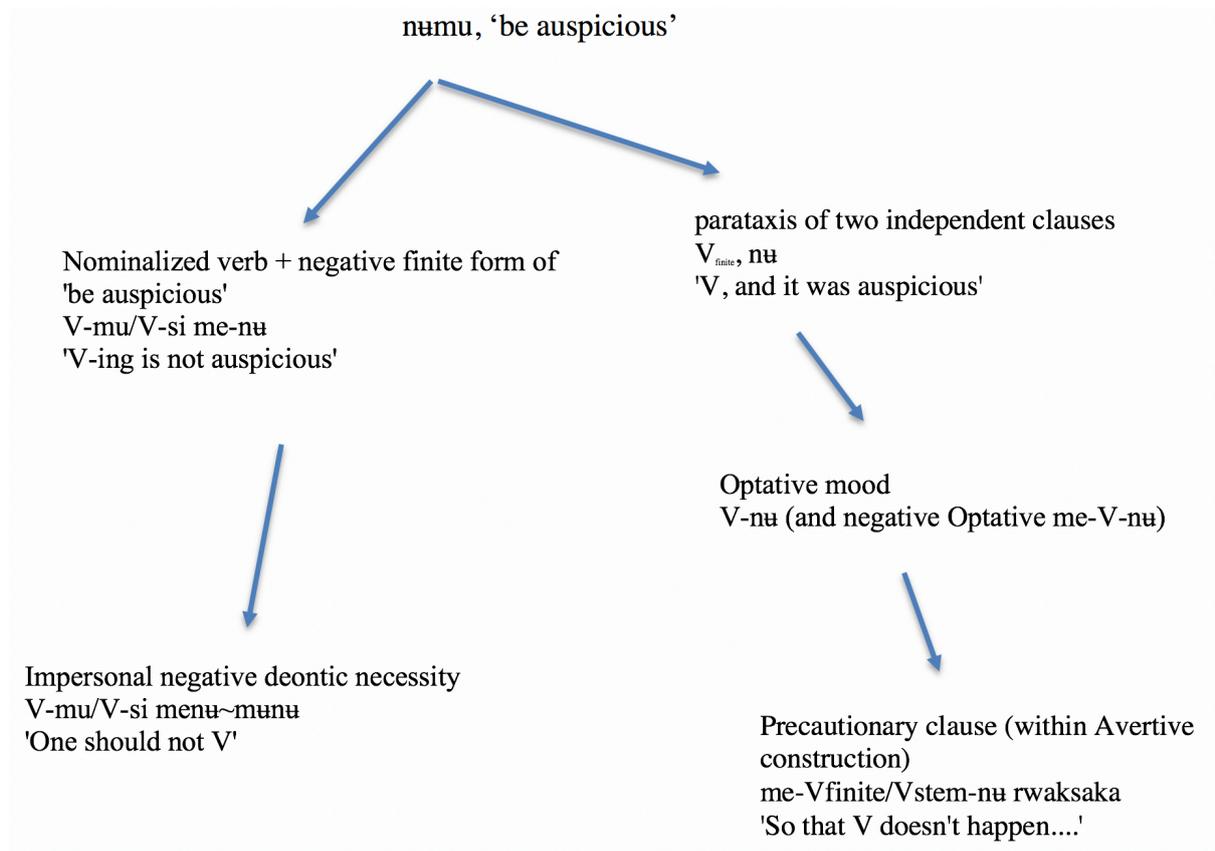


Figure 5. Parallel development of the optative and the impersonal negative deontic necessity markers from the same verb *numu*, 'be auspicious'

While Thulung, like most Asian languages, does not have much any morphology exclusively used for the expression of apprehension and fear, other morphological markers, such as the optative mood marker and the epistemic authority marker (presented in the section on Discourse markers in 2.1.2.2), have been coopted for such purposes.

2.2.3.4.2 Irrealis marker

The irrealis mood marker is found with negative past forms of verbs as well as with counterfactual conditionals. Material in this section has been presented orally (Lahaussois 2013) but has not been published.

Table 37 shows negated forms for the verb *bomu* 'do', contrasting affirmative non-past and past forms with negative non-past and past forms. (As I have done elsewhere, I use a default 3SG object, in order to reduce the number of cells).

>3SG	NPST	PST	NEG.NPST	NEG.PST
1sg	be-u	be-uto	mi-be-u	mi-be-u-wa
1di	bo-tsi	be-ttsi	mi-bo-tsi	mi-bo-tsi-ja
1de	bo-tsuku	be-ttsoko	mi-bo-tsuku	mi-bo-tsuku-wa
1pi	bi-i	bi-idi	mi-bi-i	mi-bi-i-ja
1pe	bo-ku	be-ttoko	mi-bo-ku	mi-bo-ku-wa
2sg	bo-na	be-nna	mi-bo-na	mi-bo-na-wa
2du	bo-tsi	be-ttsi	mi-bo-tsi	mi-bo-tsi-ja
2pl	bo-ni	be-nni	mi-bo-ni	mi-bo-ni-ja
3sg	b u -u	b u -u q	mi-b u -u	mi-b u -u-ja
3du	bo-tsi	be-ttsi	mi-bo-tsi	mi-bo-tsi-ja
3pl	bo-mi	bem-dji	mi-bo-mi	mi-bo-mi-ja

Table 37. Contrastive paradigms for affirmative and negative forms of the verb *bomu*, 'do'

The negative forms are characterized by the negative prefix *mi-*, while the negative past forms are additionally marked with the irrealis marker *-wa* (allomorph *-ja* after *i*). Note that the negated past forms are made from the non-past forms, rather than the past forms: both stem selection and indexes are those found with non-past verb forms.

The irrealis marker is also found in counterfactual conditional clauses. Real conditionals have non-past verb forms in both clauses, as seen in (135). This can be contrasted with counterfactual conditionals in (136) and (137), where irrealis marking is found in both clauses.

(135) go mukli **mi-bi-ŋu** ma:la ama mam-ka dukha
 1SG mukli NEG-come-1SGCOND 1SG.POSS mother-ERG hardship

bo-mi

do-3PL>3SG

'If I do not come to Mukli, my mother will suffer.'

(136) go mukli **mi-bi-ŋ-wa-m** ma:la ama mam-ka
 1SG mukli NEG-come-1SG-IRR-NMLZ COND 1SG.POSS mother-ERG

dukha **bo-mi-ja**

hardship do-3PL>3SG-IRR

'If I had not gone to Mukli, my mother would have suffered.'

(137) me nem **di-sət-pu-wa** ma:la dzhjal-laŋka batse-ŋa
 DIST.DEM day leave-IRREV-1SG-IRR COND window-ABL survive-INT

mi-dupsa-wa

NEG-become.3SG-IRR

'If I had left her that day, she would not have survived (the fall) from the window.'

In Khaling, the irrealis marker does not occupy the same slot in the verbal template as in Thulung: it in the Tense-Aspect-Mood slot (see template in Figure 6; for a more complete template, see Jacques et al. (2012: 1101)), occupying the same position as the past marker.

prefix	stem	person	auxiliary	person	TAM	person	reflexive	number
?i- : 2/inverse	Σ				-t- : past			
mu- ~ mΛ- : negative					-w- : irrealis			

Figure 6. Verbal template (partially completed) for Khaling

The contrastive verb paradigms in Table 38 for Khaling affirmative and negative non-past and past forms of the verb *mune* 'do' show that the differences between a past affirmative (3rd column) and negative past verb form: the latter has the negative prefix *mu-* ~ *mΛ-*, and shows the substitution of irrealis marker *-w* for past marker *-t*.

>3SG NPST PST NEG.NPST NEG.PST

1SG	mΛ-ŋΛ	mu-ŋ-t-Λ	mu-mΛ-ŋΛ	mu-mu-ŋ-w-Λ
1DI	mʌ-ji	mʌ-i-t-i	mu-mʌ-ji	mu-mʌ-i-w-i
1DE	mʌ-ju	mʌ-i-t-u	mu-mʌ-ju	mu-mʌ-i-w-u
1PI	mʌ-ki	mʌ-k-t-iki	mu-mʌ-ki	mu-mʌ-k-w-iki
1PE	mʌ-kΛ	mʌ-k-t-ΛkΛ	mu-mʌ-kΛ	mu-mʌ-k-w-ΛkΛ
2SG	i-mε	i-mʌ-tε	mΛ-mε	mΛ-mʌ-w-ε
2DU	i-mʌ-ji	i-mʌ-iti	mΛ-mʌ-ji	mΛ-mʌ-i-w-i
2PL	i-mʌ-ni	i-mo-t-nu	mΛ-mo-t-nu	mΛ-mo-n-w-ε
3SG	mε	mʌ-t-ε	mu-mε	mu-mʌ-w-ε
3DU	mε-su	mʌ-s-su	mu-mε-su	mu-mʌ-w-εsu
3PL	mε-nu	mʌ-t-nu	mu-mε-nu	mu-mʌ-w-εnu

Table 38. non-past and past affirmative and negative verb forms in Khaling
(verb *mʌne* 'do', with 3SG default object)

Similarly to Thulung, the Khaling irrealis marker is found in counterfactual conditional constructions. The examples below contrast a real conditional clause (138) with counterfactual conditionals (139, 140). As in Thulung, the irrealis marker is found on both the protasis (marked with conditional marker *-kʰo*) and the apodosis.

- (138) ʔine jû: ʔi-dzε-kʰo ʔuŋΛ kʰib-u
 2SG.ERG meal INV/2-eat-COND 1SG.ERG cook-1SG
 'If you eat a meal, I will cook it.'

- (139) ʔin mu-hə:-wε-kʰo ʔûŋ jo mu-hə-ŋΛ-wΛ
 2SG NEG-come-IRR-COND 1SG also NEG-come-1SG-IRR
 'If you had not come, I would not have come either.'

- (140) ʔuŋΛ jΛŋΛja:s jâ: mu-ŋ-wΛ-kʰo tʰεbelo
 1SG.ERG peach plant do-1SG-IRR-COND this.time
 sî:t-wε
 bear.fruit-IRR
 'If I had planted peaches, they would have borne fruit by now.'

considerably more interactive (whereas the bulk of my current corpus is stories told by a single speaker).

The work presented in 2.2.1.2 on Thulung ideophones is based on what I able to extract from my corpus, and will need to be supplemented by elicitation in order to develop a more nuanced analysis of ideophonic material in Thulung and how it compares to Khaling.

Another interesting topic to pursue is that of dialectal differences in verbal indexes. I have some evidence of different indexes for certain person/number scenarios in my data, and would very much like to explore the variation more systematically. This is particularly interesting with predicate derivations, because the multiple exponence can take on unusual forms. See, for example, the contrast between the verbs in (144) and (145) below: they are combinations of the same verb root and derivational formative, with the same subject, and yet they look quite different:

Mukli

- (144) ko:dep **ba-m-thal-miri** ?e
 together stay-3PL-HAB-3PL.PST HS
 'They stayed together.'

Deusa

- (145) mekotima mi:jakma mam mini-nuŋ irma nwa:-le
 then Miyakma mother Mini-COM Irma two-CL
 tse-tsip-nuŋ **ba-m-thal-li** ?e
 child-DU-COM stay-3PL-HAB-3PL.PST HS
 'Then Miyakma Mother and her children Mini and Irma stayed together.'

In (145), the 3PL.PST index is *-li* (an allomorph of *-ri*), a variant form of *-mri*. Yet the additional exponent following the root *ba* is *-m*, making it look like the derivational formative *-thal* is an infix within the standard 3PL.PST index.

Because I only have a few stories in my corpus which are told by Deusa speakers, I have no sense of the relative frequency of *-ri* vs *-mri* as a 3PL.PST marker. Additionally, as *-ri* is the 1PI.PST index in Mukli, I am very curious about whether there have been changes in the Deusa 1PI.PST marker, in order to maintain two distinct indexes for the different scenarios. This is but one example of the interest of collecting data (both narrative and elicited) on other dialects of Thulung.

Finally, some of the material within slot 3 (indexation markers) has not been discussed here: converbs and infinitive markers, the latter being particularly interesting because they sometimes include indexing of the P.

3 EPISTEMOLOGICAL AND HISTORICAL ISSUES SURROUNDING DESCRIPTIVE LINGUISTICS

This chapter of the present thesis is of a very different nature from Chapter 2. Whereas the former presented material towards a description of the verbal morphology of Thulung, with some data from other Kiranti languages, Chapter 3 is dedicated to the use I have made of methodological tools. The three sections within this chapter concern grammaticographical models (3.1), by which I mean the models underlying the organization of descriptive grammars; annotation practices (3.2), and specifically what can be said about the history of the standardization of the different tiers that make up what is now known as interlinear morphemic glossing; methodological tools (3.3), in which I discuss work I have carried out on the alignment of narrative corpora, on linguistic questionnaires, and my contribution to speaking dictionaries. While it may seem that these topics are somewhat distant from the Thulung verb complex, all of the sections touch upon at least some aspects of research into the verbal morphology of Kiranti languages.

3.1 Grammaticographical models

Work that I have engaged in under the label 'grammaticography' has very strong ties with the preoccupations of the HTL research group, and specifically with two projects: one led by Emilie Aussant and entitled "Grammaires étendues", and another which I co-led with Bernard Colombat, called "Dictionnaire historique de la terminologie linguistique" (the result of which has been published as *Histoire des parties des discours*, (Colombat & Lahaussais 2019a)). The basis of the *Grammaires étendues* project is to pursue research around a phenomenon discussed by Sylvain Auroux (Auroux 1992: 19; Auroux 1994: 82), namely the adaptation of a grammar-writing model, originally developed for a specific language, to other languages. A classic example of this is the Latin grammar model adapted by missionaries to accommodate the description of very different languages (Zwartjes 2011; Zwartjes 2012). One outcome of the *Grammaires étendues* project is a special issue of *Faits de langues* exploring verbal morphology in the context of extended grammars, namely what evidence one finds in descriptions of verbal morphology of the influence of grammatical traditions developed for other languages (Aussant & Lahaussais 2020).

In what follows, I shall focus on three topics that relate to grammaticographical models for language description, namely the question of how certain categories may be omitted or excessively highlighted on the basis of a chosen (or unconscious) grammaticographical model (3.1.1); the presentation of polypersonal verbal paradigms, and how their evolution can be traced over a historical corpus (3.1.2); the organizational schemes found in contemporary descriptive grammars (3.1.3).

3.1.1 Missing categories: Interjections and ideophones

The grammaticographical role of interjections was brought to my attention through work on the main chapter I contributed (Lahaussais 2019b) to *Histoire des parties du discours*, a collaborative volume retracing the history of the traditional parts of speech from the Greek grammatical tradition until now. Interjections, since their 'invention' to fill a gap left by the absence of articles in Latin, have, despite their marginality (they are often defined in negative terms: phonologically aberrant, not participating in morphology or syntax), been significant in a number of debates about language, such as the origin of language (was the first part of speech interjections?), universality of language (are interjections understood cross-linguistically?), and others.

This interest in interjections from a historiographical point of view led to me to explore how they are treated in contemporary grammars, by focusing on a corpus of various English-medium grammars of Nepal of the last 30 years (Lahaussais 2016b). To my surprise, interjections received almost no mention across the corpus, the exception being grammars written by linguists with missionary affiliations: interjections have their own chapter in Khaling (Toba 1984) and in a section of the chapter on 'Minor word classes' in Kham (Watters 2009), and are mentioned in the text of the grammar, but not given chapter or section billing, in Bantawa (Doornenbal 2009). Others made no mention of interjections in the text of the grammars, although some authors included interjections in their glossary or list of abbreviations (this was the case for Camling (Ebert 1997b), Thangmi (Turin 2012), Wambule (Oppenort 2004), Yamphu (Rutgers 1998), Dumi (van Driem 1993b), Limbu (van Driem 1987)). Three grammars in the corpus used the term 'interjection' nowhere (Dolakha Newar (Genetti 2007), Sunwar (Borchers 2008), Jero (Oppenort 2005).)

In the study, I discuss a number of possible explanations for the lack of interjections in the descriptions. One suggestion is that as interjections do not generally participate in morphosyntactic processes (they cannot, for example, usually undergo derivation), they are

reasonably left out from descriptive grammars whose central focus tends to be morphosyntax.⁷⁸ Other possibilities revolve around the correlation between a linguist with missionary goals and the presence of interjections: one hypothesis is to see this correlation as the legacy of the parts-of-speech grammaticographical model, applied to a variety of languages by missionaries throughout the Renaissance and modern period (Hovdhaugen 1996; Zwartjes 2011; Zwartjes 2012). Linguists applying the parts of speech model often encountered languages which did not appear to have a given category⁷⁹, in which case one option was to interpret the available language data according to the available model: an example of this is the classification of what are now recognized as ideophones into a category of interjections (Dingemans 2011; Lahaussais 2019b). In the case of interjections in descriptions of languages of Nepal by missionaries, it is possible that the chosen organizational scheme for a grammar resulted in the prominence of interjections where they may not have otherwise been brought to the fore with a different another grammaticographical model. A possible example of this is Toba's Khaling grammar, which is included in the corpus. The grammar (which is very short, at 40-some pages) is organized according to a parts-of-speech scheme, with the following chapters: 0. Introduction; 1. Phonology; 2. Word order; 3. Nouns; 4. Pronouns; 5. Numerals; 6. Verbs; 7. Adjectives; 8. Adverbs; 9. Conjunctions; 10. Particles; 11. Interjections.

The chapter on interjections is very limited: see Figure 7 below for the entirety of its contents. The type of information it contains could just as easily have been put into the glossary that accompanies the grammar, an option apparently chosen by a good number of linguists (listed above: these are the grammars that have interjections in their glossaries but do not have any discussion of them in the text of the grammars)⁸⁰.

⁷⁸ For an unusual example of a grammar which does not focus on morphosyntax but still, by my count, qualifies, see Michaud (2017).

⁷⁹ Note that Noonan (2007: 116) reports considerable diversity of opinion in his survey of typologists and their assessment on whether the absence of a category should be explicitly highlighted in a grammar. He reports that Bickel, Comrie, Croft, and Haspelmath concur that it is good practice in an index to note “in boldface type the main entry [or entries] for a given feature, and noting the absence of a given feature directly in the index with the dash. So, for example, one could indicate the absence of tone as follows: tone: —.” For another example of this, see grammars written using the grammaticographical template in (Comrie & Smith 1977).

⁸⁰ Note that chapters 8, 9 and 10 also consist almost exclusively of lists of lexemes from the relevant category.

11. INTERJECTIONS

minaa 'well now' introducing an utterance or a new topic of importance.

mäki, maakaa, mana 'right', 'I see' response of the listener, not necessarily agreement.

äyi, äye 'is that so!' surprise over something seen or heard.

aakaave 'it hurts!'

ehoy 'hey!' calling some one far away.

hou 'hey!' calling some one, always used together with the name of the person called. The name precedes the interjection.

aa 'yes'.

mo 'no'

aalaasö 'thank you'

nüwo, ne 'please take this!' (giving a gift, offering food to a guest).

do 'let's go!' always followed by a finite verb form.

Figure 7. Toba 1984's chapter on Khaling interjections

A very different hypothesis to consider with respect to the presence of the treatment (or the listing) of interjections in grammars by missionaries is the fact that missionaries spend extended amounts of time in the field, often living with communities for years on end. This linguistic situation familiarizes them with interjections which tend to arise in more familiar registers of language (especially compared with the registers one finds in the production of narrative and elicitation sessions). This is not to say that secular linguists are not also exposed to a variety of registers over years of fieldwork, but one could speculate that a missionary linguist--often with family in tow--will have a uniquely immersive (and often monolingual, see Thomas (2019)) experience.⁸¹

Another study (Lahaussois 2017b) considered another marginal part of speech, namely ideophones, in Khaling. While the study was very much descriptive, it was striking to find, in looking into descriptions of ideophones in neighboring languages, that very little has been written on this topic. Consider the map in Figure 8 below, which indicates with black dots languages for which I found descriptions of ideophones: a typologist carrying out a large-scale study of ideophones in languages of the world might well come across this distribution and assume that ideophones are found only in Central and Eastern Kiranti languages.

⁸¹ For a popularized description of a missionary field experience, see Everett (2009).

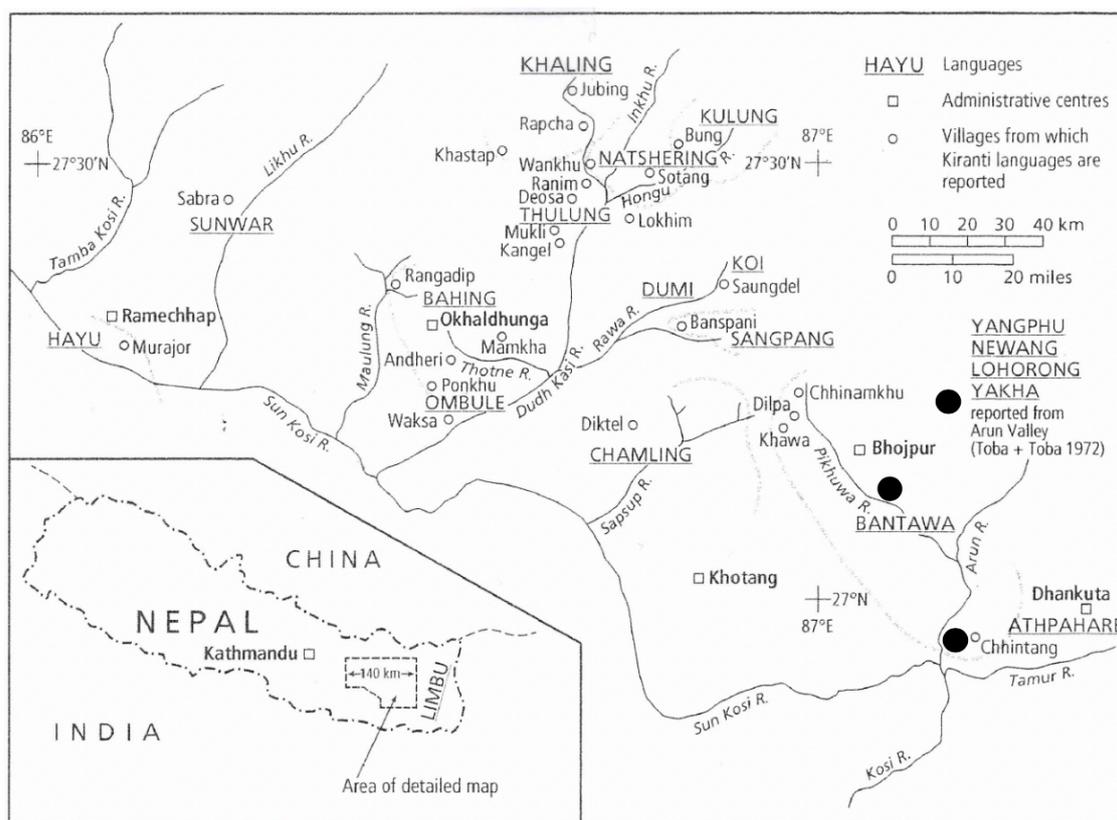


Figure 8. Map of Kiranti area and languages (based on (Michailovsky 1975)):

black dots for languages with ideophones described

It seems likely, though, that the issue is one of grammaticography rather than of linguistics: ideophones are definitively found in at least Khaling and Thulung, as shown above, and presumably in other Kiranti languages as well, especially when one considers their presence in Nepali, the contact language throughout the area (Dahal 1999). In looking at the languages with descriptions of ideophones, what emerges is the connections between the descriptions: Rai & Winter (1997) describe ideophones in Bantawa and their article is, as far as I can tell, the first description of the word class in languages of Nepal. Doornenbal's (2009) Bantawa grammar also features a description of ideophones, a topic he clearly encountered in looking at previously work on Bantawa.⁸² A later article on ideophones in Chintang (Rai et al. 2005) features the same first author as the first study on Bantawa;⁸³ the second author is Bickel, who was the dissertation supervisor for Schackow, author of the very thorough grammar of Yakkha (Schackow 2015). While I suspect that a scholar like Schackow would

⁸² The Rai & Winter (1997) article is cited several times by Doornenbal in his grammar.

⁸³ This is similar to my carrying out work on ideophones in Khaling and then 'discovering' them in Thulung-- although to be fair, I had mentioned them briefly in my dissertation, albeit with different terminology.

have mentioned ideophones regardless of her dissertation author's previous work on the topic, there is nonetheless a 'genealogical' connection between the descriptions for Chintang and Yakkha.

With no similar precedents in the description of Western Kiranti languages, it is perhaps not surprising that existing grammars should not mention ideophones. This is possibly compounded by an organizational scheme used by van Driem for Limbu and Dumi (1987; 1993b) and his students, which leaves no obvious section in which to present ideophones. These questions are discussed in Kelly & Lahaussais (in press) and in section 3.1.3 below.

Another far from insignificant issue is the impact of Dingemanse's work on ideophones: the implicational hierarchy he has proposed (Dingemanse 2012) has resulted in an effervescence of work on ideophones⁸⁴ and questions as to the universality of the word class generate ample discussion on lists such as LingTyp.⁸⁵ It is possible that, as a result, new descriptions of Kiranti languages, should there be any, will include ideophones (and perhaps also interjections) and correct the current skewed geographical distribution.

3.1.2 Verbal paradigms

Another topic related to grammaticography that is of interest to me is the evolution in the presentation of verbal paradigms from early descriptions of Kiranti languages until the present. All the languages of the group have polypersonal indexation, with two arguments marked on the verb; this cannot be represented easily with the types of verbal conjugations familiar to speakers of European languages (which index a single argument). My goal in the study was to identify the stages of development of the types of paradigms currently used to present Kiranti verb stems and indexes. Two articles look into this topic: Lahaussais (2018a) examines different approaches used to represent the verbal morphology of Khaling; Lahaussais (2020a) looks at a wider range of languages, beginning with the verbal paradigms found in Hodgson's 1857-1858 descriptions, and extending the study to the present day.

⁸⁴ This is presumably in part due to a similar mechanism as the great number of descriptions of relative clause strategies that followed the publication of Keenan & Comrie (1977) 's article on noun phrase accessibility in relative clauses: an implicational hierarchy provides descriptive linguists with a tool to test their languages against, something that functions somewhat like a checklist.

⁸⁵ This is the very active mailing list for the Association for Linguistic Typology.

The first mention of the Khaling language in a Western publication appears to be in Hodgson's "Comparative vocabulary of the languages of the broken tribes of Nepal" (1857b: 350 ff). The next, again in the form of one of the languages a comparative wordlist, is in (Hunter 1868), based on data taken from Hodgson's work. The first description of the language is in volume 3.1 of the *Linguistic Study of India* (Grierson 1909: 370–372), with a very brief outline of 4 word classes in Khaling, namely nouns, numerals, pronouns, and verbs. While very little information is provided in this one-and-a-half page description, some examples are provided, including verb forms and discussion that suggest that the author is aware of the indexation of two arguments on transitive verbs (eg. "The suffix ngā is added in order to denote an object of the first person singular in *bi-ngā-ye*, give me", p. 371)

The first description dedicated to Khaling verbal morphology (Toba 1973) is a 14-page long article with the following layout:

- I. Suffixes
 - A Finite suffixes
 - B Infinitival suffixes
- II. Enclitics
- III. Compound stems
- IV. Other auxiliary constructions
- V. Verb stem alternations
 - A Formation of the infinitive stem
 - B Formation of the finite stem
 - C Formation of the finite forms
- VI. Personal endings

Figure 9. Table of contents of Toba's *The Khaling Verb*

The sections which present paradigms are sections V and VI on verb stem alternations and person indexes. Section VI presenting personal endings illustrates them with two tables: The table reproduced in Figure 10 below, and a second table which gives 1st person forms for sample verbs (showing the stems used for 1st person forms). Figure 10 shows (non-past) suffixes for 'agentive' and 'nonagentive' forms for 1st, 2nd and 3rd persons, with number and clusivity distinctions.

		Agentive	Nonagentive	
<u>1st person</u>				
singular		-u	-ngaa	
dual	inclusive	-si	-yi	
	exclusive	-su	-yu	
plural	inclusive	-ki	-ki	
	exclusive	-kaa	-kaa	
<u>2nd person</u>		(interrogative)	(imperative)	
singular		-ü	-ɟ	-eye
dual		-si	-yi	-siye
plural		-ni	-ni	-nuye
<u>3rd person</u>				
singular		-ü	-ɟ	
dual		-su	-su	
plural		-nu	-nu	

Figure 10. Personal endings in Khaling in Toba 1973

There are a number of puzzling features of this presentation of person indexes:

a) The only mention of object marking on the verb is a brief sentence (“Agreement between person and number for subject as well as for object is indicated everywhere except third person singular”, 1973:3), which occurs at the very beginning of the text, and is not referred to again.

b) Second persons and inverse forms in Khaling make use of a prefix (-?i) of which there is no mention in Toba's article; this makes sense in that what is presented are the "personal endings", but suffixes alone do not make it possible to index all scenarios.

c) The terms "Agentive" and "Nonagentive" are not explained, and have two possible interpretations here. The first is that they refer to the indexes for transitive and intransitive forms, respectively.⁸⁶ The other possibility is that they refer to the indexes for A and for S/P respectively, in which case the presentation may be interpreted as follows: transitive forms require taking the relevant A suffix from the "Agentive" column, and then the relevant P suffix from the "Nonagentive" column. Because of the configuration of the Khaling paradigm, whereby S/P indexes are often the same, this works for a number of forms (including the most common inverse forms, namely with 3SG as A), but not for all.

⁸⁶ A later publication by S. Toba actually does provide an explanation of the two terms: "The majority of Khaling verbs are agentive (transitive). Nonagentive verbs (intransitive) include verbs of motion, body functions..." (Toba 1984: 16).

One might wonder what is gained from analyzing a 14-page article about verbal morphology which the author herself has said is not complete (pc, 2005), and this is precisely where grammaticography comes in: I. Toba, along with her husband S. Toba, spent many years living among the Khaling, and was no doubt a speaker of the language, presumably able to conjugate verbs. The question is then to what extent her presentation--which suffers from the omission of a number of important features of the language (clear acknowledgement of object marking and of the prefixes for 2nd person and inverse forms) -- is the result of a European bias in terms of her presentation of conjugation charts: European languages have suffixal person indexation, notably marking a single argument, which is how Toba chose to present the data for Khaling. In a sense, the idea of setting up a paradigm as a matrix, with cells indicating each of the possible argument scenarios, is not one that is intuitive,⁸⁷ and reverting to a familiar presentation pattern is not surprising (even if fails to account for all the dimensions of the data).

A second article, Lahaussais (2020a), explores the evolution of shapes of paradigms used to present Kiranti verbal data from the 1850's until now. The data from the 1850's is from Hodgson (Hodgson 1857b; Hodgson 1858). In the first description within this article (which spans volumes 26 and 27 of the *Journal of the Asiatic Society of Bengal*), Hodgson initially lays out his verbal paradigms of Vayu as in Figure 11 below.

<i>Singular.</i>	<i>Indicative present.</i>	
	<i>Dual.</i>	<i>Plural.</i>
1. Sinmi.*	+ { Sischhokmi, excl. Sischhikmi, incl.	+ { Sitkokmi, excl. Sitkem, incl.
	<i>Dual and plural of agent.</i>	
1. Sinchhem.	I kill them two.	
1. Sinnem.	I kill them all.	
2. +Sitmi.	+Sischhikmi.	+Sitmem.
	<i>Dual and plural of object.</i>	
2. Sischhikmi.	Thou killest them two.	
2. Sitmem.	Thou killest them all.	
3. +Sitmi.†	+Sischhikmi.	+Sitmem.
	<i>Dual and plural of object.</i>	
3. +Sischhikmi.	He kills them two.	
3. +Sitmem.	He kills them all.	

Figure 11. Hodgson's presentation of the Vayu transitive verb 'to kill' (1857a: 448–449)

⁸⁷ I test this each year on students at Paris Diderot when presenting Thulung verbal morphology by asking them how to represent the data; they do not spontaneously come up with the matrix solution which is now the standard presentation format for both Kiranti and Algonquian languages.

If one were to represent this data schematically, it would correspond to the following organization:

1SG>3SG	1DE>3SG	1PE>3SG
	1DI>3SG	1PI>3SG
1SG>3DU		
1SG>3PL		
2SG>3SG	2DU>3SG	2PL>3SG
2SG>3DU		
2SG>3PL		
3SG>3SG	3SG>3SG	3PL>3SG
3SG>3DU		
3SG>3PL		

Figure 12. Schematic representation of scenarios in Hodgson's initial paradigms

One thing to note is that, unless otherwise specified, the forms are for a 3SG default object argument, and thus this presentation does not provide forms with two non-singular arguments, or object arguments other than 3rd person.⁸⁸

Hodgson changes course partway through the article, and his verbal paradigms become, in his words, "seemingly more accommodated to the genius of the language" (1857a: 470). The new presentation style is illustrated in Figure 13 below:

⁸⁸ Other scenarios are treated later in the presentation: 3DU/PL > 1SG/2SG/3SG are under "Passive voice, indicative present" ((1857a: 450–451), while forms involving 1st and 2nd persons are treated under "Special forms of action between the first 2 persons" (ibid, 452).

Indicative Mood.

Present and Future.

Singular.

I eat or will eat.	1. Jánchungmi, self, as agent or object.
	2. Jángmi, it, him.
	3. Játungmi, his or for him.
	4. Jángchhem, them two.
	5. Játungchhem, their two.
	6. Jángmem, them all.
	7. Játungmem, their all.
	8. Jónum, thee.
	9. Jánum, thine or for thee.
	10. Jónochhem, you two.
	11. Jánochhem, your two.
	12. Jónonem, you all.
	13. Jánonem, your all.

Dual.

We two eat or will eat.	1. { Jánachokmi, excl. } selves.
	{ Jánachbikmi, incl. }
	2. { Jáchhokmi, excl. } it.
	{ Jáchhikmi, incl. }
	3. Jáchhokmi-chhikmi, his, for him.
	4. Jáchhokmi-chhikmi, them two.
	5. Jáchhokmi-chhikmi, their two.
	6. Jáchhokmi-chhikmi, them all.
	7. Jáchhokmi-chhikmi, their all.
	8. Jómi, thee.
	9. Jáchhokmi, thine.
	10. Jáchhikmi, you two.
	11. Jáchhokmi, your two.
12. Jónem, you all.	
13. Jánem, your all.	

Plural.

We all eat or will eat.	1. { Jánchikokmi, excl. } selves.
	{ Jánchikem, incl. }
	2. { Jákokmi, excl. } it.
	{ Jákem, incl. }
	3. { Játikokmi, excl. } his, for him.
	{ Játikem, incl. }
	4. Jákokmi-kem, them two.
	5. Játí-kokmi-kem, their two.
	6. Ják-kokmi-kem, them all.
	7. Játí-kokmi-kem, their all.
	8. Jomi or Jókokmi, thee.
	9. Jákokmi, thine.
	10. Jáchhikmi, you two.
11. Jákokmi, your two.	
12. Jónem or Jókokmi, you all.	
13. Jánem or Jákokmi, your all.	

Figure 13. Partial paradigm for indicative present/future 'eat', Hodgson 1857: 472

Schematically, this corresponds to the following A and P scenarios, with all combinations of subject and objection listed in the paradigms:

subject:	object ⁸⁹ :
1SG	reflexive
1DI	3SG
1PI	3DU
2SG	3PL
2DU	2SG
2PL	2DU
3SG	2PL
3DU	1SG

⁸⁹ The glosses for some of the objects are genitive or dative forms; this suggests that Hodgson is possibly listing applicative verb forms in the paradigms.

Figure 14. Schematic representation of scenarios in paradigm in Fig. 13.

I have called this type of paradigm presentation the 'list format' (Lahaussais 2020a), and a similar presentation style is still found in the work of a number of Kiranti scholars, as a way to present a large amount of paradigmatic data in a limited amount of space (van Driem 1987 appendix ii; van Driem 1993b appendix ii; Doornenbal 2009: 397–407; Opgenort 2004 Appendix Four; Rutgers 1998: 605–610).

The type of presentation in Figure 13 incorporates inverse forms, which were previously treated in different sections of the paradigm by Hodgson (and labeled 'passive forms' and 'special forms'), and an important advantage of this type of paradigm is that all the information pertaining to the different scenarios can be captured.

More modifications to the presentation style for verb paradigms appear in Hodgson's (1858) article on Bahing: numbers are used to identify verb forms. Interestingly, they correspond not to the person of the subject but to the *number of the object* (1. for a singular P, 2. for a dual P, and 3. for a plural P), resulting in a grid-like presentation from which no forms are omitted, facilitating comparison between the forms, and removing the need for glosses (once the pattern is established for readers within a few paradigms).

<i>Present and Future Tenses.</i>		
<i>Singular of Agent.</i>	<i>Dual of Agent.</i>	<i>Plural of Agent.</i>
<i>First Person.</i>		
1. { <i>Já-gna</i> , I eat or will eat it.	1. { <i>Já-sa</i> , inclusive. <i>Ja-suku</i> , exclusive. We two eat it.	1. { <i>Já-ya</i> , inclusive. <i>Já-ka</i> , exclusive. We all eat it.
<i>Dual of Object.</i>		
2. { <i>Ja-gna-si</i> , I eat them two.	2. { <i>Ja-sa-si</i> , inclusive, <i>Ja-sukusi</i> , exclusive, We two eat them two.	2. { <i>Já-ya-si</i> , inclusive. <i>Já-ka-si</i> , exclusive. We all eat them two.
<i>Plural of Object.</i>		
3. { <i>Ja-gna-mi</i> , I eat them all.	3. { <i>Ja-sa-mi</i> , inclusive, <i>Ja-suku-mi</i> , excl. We two eat them all.	3. { <i>Ja-yami</i> , incl. <i>Ja-ka-mi</i> , excl. We all eat them all.*
<i>Second Person.</i>		
1. <i>Já-(y) f.</i> 2. <i>Já-(y)-i-si.</i> 3. <i>Já (y)-i-mi.</i>	1. <i>Já-si.</i> 2. <i>Já-si-si.</i> 3. <i>Já-si-mi.</i>	1. <i>Ja-ni.</i> 2. <i>Já-ni-si.</i> 3. <i>Já-ni-mi.</i>
<i>Third Person.</i>		
1. <i>Ja-wa.</i> 2. <i>Já-wa-si.</i> 3. <i>Já-wa-mi.</i>	1. <i>Já-se.</i> 2. <i>Já-se-si.</i> 3. <i>Já-se-mi.</i>	1. <i>Já-me.</i> 2. <i>Já-me-si.</i> 3. <i>Ja-me-mi.</i>

Figure 15. Paradigms with numbers instead of glosses (Hodgson 1858: 407–408)

The next Kiranti paradigms to appear, chronologically, are those in I. Toba's description of Khaling verbal morphology, described above. Following Toba, Allen's work on Thulung yielded a different type of paradigm presentation, something I label "matrix paradigms", following Pike (1962), and shown in Figure 16.

Pers. of Aff of Cs		1st person					2nd person			3rd person		
		sing	incl dual	incl plur	excl dual	excl plur	sing	dual	plur	sing	dual	plur
1st	sing						ni ni	nici nici	ni ni	u to	uci toci	umi tomi
	incl dual									ci ci	ci ci	ci ci
	incl plur									i li	i(ci) li(ci)	i(mi) li(mi)
	excl dual						naci naci	naci naci	nicimi nicimi	cuku coko	cuku coko	cuku coko
	excl plur						nami nami	nacimi nacimi	nimi nimi	ku toko	ku toko	ku toko
2nd	sing	ŋi ŋiri			ciki ciki	kimi tiki				na na	na(ci) na(ci)	na(mi) na(mi)
	dual	ŋici ŋirici			ciki ciki	kimi tiki				ci ci	ci ci	ci ci
	plur	ŋini ŋirini			ciki ciki	kimi tiki				ni ni	ni(ci) ni(ci)	ni(mi) ni(mi)
3rd	sing	ŋi ŋiri	saci saraci	sa sara	ciki ciki	kimi tiki	na na	naci naci	nimi nimi	iu liu	iu(ci) liu(ci)	iu(mi) liu(mi)
	dual	ŋici ŋirici	saci saraci	saci sara	ciki ciki	kimi tiki	naci naci	naci naci	nimi nimi	ci ci	ci ci	ci(mi) ci(mi)
	plur	ŋimi ŋirimi	sami sarami	sami sarami	ciki ciki	kimi tiki	nami nami	nacimi nacimi	nimi nimi	mi miri	mi(ci) miri(ci)	mi miri

Figure 16. Transitive paradigm from (Allen 1975)

The innovations found in this type of paradigm presentation are the arrangement of arguments (A in the vertical axis, P in the horizontal)⁹⁰ and the empty cells for 1>1, 2>2 forms (these are reflexive, and therefore use reflexive marker *-siŋ* and intransitive indexes) and 2<>INCL forms.

Van Driem (van Driem 1987; van Driem 1993b) appears to favor the list-style paradigm (the innovative presentation that Hodgson felt best reflected the language's 'genius' and seen in

⁹⁰ Allen mentioned (Allen, pc) that his inspiration was that this seemed to him to be "the most natural way of presenting the material". This appears to be universally intuitive as an arrangement (as well as analogical with the arrangement of intransitive paradigms, where the S is in the vertical axis), although one wonders whether a linguist native in an ergative language might arrange the axes differently.

Fig. 11), as do his students (Opgenort 2004; Opgenort 2005; Doornenbal 2009) when presenting the paradigms representative of different verb classes, showing verb stem alternations; person indexes are however introduced with a matrix paradigm, showing the indexes in each cell.

Michailovsky, in his grammar of Hayu (1988), opted for a matrix paradigm similar to Allen's in Figure 16, but with some subtle differences:

		(formes directes)							
		Objet :							
		2S	2D	2P	3S	3D	3P		
Légende :	Non-passé / applicatif * passé	Sujet : 1S		no Nno	notshe Nnotshe	none Nnone	η ~ N ~ sɔŋ / tɔŋ kɔŋ	η ~ N ~ sɔŋtshe / tɔŋtshe kɔŋtshe	η ~ N ~ sɔŋme / tɔŋme kɔŋme
		1DE		tshok tshoŋ					
		1DI		tshik tshiŋ					
		1PE		kok / tikok kikoŋ / tikoŋ					
		1PI		ke / tike kikoŋ / tikoŋ					
		1PI		ke / tike kikoŋ / tikoŋ					
		(formes inverses)			(formes directes)				
Objet :		Sujet :							
IS	1DE	1DI	1PE	1PI	1PI				
2S	ηo sɔŋ	tshok tshoŋ	tshik tshiŋ	kok (ki)kɔŋ	ke (ki)keŋ			θ ko / to	me kome / tome
2D	ηotshe sɔŋtshe							tshik tshe	
2P	ηone sɔŋne							ne Nne	
3S	ηo sɔŋ					θ N	tshik tshe	ne Nne	θ ko / to
3D	ηotshe sɔŋtshe								tshik kotshe / totshe
3P	ηome sɔŋme					me Nme			me kome / tome

Figure 17. Paradigm showing indexes for transitive verbs (Michailovsky 1988: 82)

- He explicitly labels the parts of the paradigm 'inverse'⁹¹ vs 'direct' forms;
- The matrix is designed in such a way as to not have empty cells: 1>1 and 2>2 slots are simply not present. This format is adopted by Opgenort (2005: 144–145) when laying out person indexes.

Another variant on the matrix paradigm is found in what we might call the "Zurich School": Ebert, in her grammar of Athpare, refers to van Driem's paradigm presentations, which she follows, but innovates by inserting intransitive affixes into the transitive paradigm in an additional column on the right-hand side (Ebert 1997a: 30–31). This has the advantage of laying out intransitive and transitive person indexes in close proximity, making it possible to compare them in ways which previous layouts do not allow. This is particularly significant

⁹¹ He defines inverse those forms where the action is directed towards the speaker (3 --> 1, 2 --> 1) or from outside towards the interlocutor (3 --> 2) (ibid, p. 83)

for languages for which the P argument indexes are close to the intransitive indexes. Bickel and Schackow (2015: 218) both adopt this paradigm style: a matrix presentation with both transitive and intransitive indexes.⁹²

The two broad types of paradigms -- matrix and list -- are used by different authors, or combined for different purposes, often with the matrix used to represent all person indexes for transitive verbs (sometimes combined with intransitive) and the list to show the different stems with different argument scenarios. A number of additional features of these paradigms are of great interest in studying the evolution of paradigm presentation styles for these languages: how explicit the authors are about the paradigm style and its interpretation; the labels given to the two arguments that can be indexed; and for older paradigm styles, the unconscious influence of how paradigms are traditionally presented in other languages known to the linguist.

This type of study would be interesting to replicate with other language groups with polypersonal indexation, to see what patterns emerge in the presentation of paradigms. Algonquian languages come to mind, and in fact are a source of very interesting material in that their descriptive tradition dates back to 1666 (Eliot's grammar of Massachusset).

3.1.3 Organizational schemes in descriptive grammars

One important and as yet little-studied question in descriptive linguistics and typology is how grammars are organized, and the effects of different organizational choices on the use that is made of said grammars. This question ties in in obvious ways with the previous section and the discussion of models underlying the presentation of data on languages which mark two arguments on transitive verbs. This is the topic addressed by a recent article (Kelly & Lahaussais in press): through an examination of a corpus of 18 grammars of Nepal of the last 30 years, we explored the types of structural schemes found in the tables of contents of the grammars, organizing them into a taxonomy (the first of its sort that we know of; although the question of how grammars are organized is a central interest to many historians of linguistics, this does not (yet) appear to be the case for typologists). We also attempted to cross structural information about the grammars with bibliographical and biographical data on the grammars and their authors, in order to retrace potential influences the grammarians may have been

⁹² Michailovsky (1988: 83–85) points out the transitive P/intransitive S index identity ("Nous remarquons que les suddixes transitifs "inverses" sont identiques aux suffixes intransitifs.[...] Le tableau 3-3 est présenté avec la personne et le nombre de l'objet sur l'axe vertical (à la différence du tableau 3-2 [Fig 15]) pour faire ressortir cette identité" but his 'tableau 3-3' somehow renders his intuition opaque.

subjected to, even grouping the authors into 'schools' of grammar writing, on the basis of terminological and structural choices.

The study is based on a series of interlinked databases: one contains biographical and bibliographical metadata concerning the authors and grammars of the corpus; another contains full tables of contents (for each numbered and/or labelled section of the grammar); yet another is a relational database, in which similar content in the grammars has been identified and tagged, resulting in an alignment of materials on the same topic across the grammars.

Screenshots from these different databases are given below. Figure 18 shows part of the database containing biographical and bibliographical metadata on the authors and grammars that make up the corpus. We considered that both categories of metadata could be relevant to how an author structured their grammar. Biographical metadata included the author's name, current and previous academic affiliations (including non-academic, when known, such as SIL), the institution where the Ph.D. was obtained and dissertation supervisor; bibliographical metadata included the language name, the date of publication, the publisher and collection, the number of pages of the grammar, and whether texts and/or glossaries were included (and if so, how many pages were taken up by one or the other). All of these factors appear to be potentially significant in terms of influences undergone and exerted which could affect how the grammars are structured: as many grammars were initially written as dissertations, the institution where the Ph.D. was obtained and the dissertation supervisor were likely influential factors;⁹³ as for bibliographical information, if the grammar was produced within a collection by a publisher, this could influence the organization of the grammar.

	Language	n	Last name	First name	Book title	date (of pub	current	past affil	past affiliatic	university (w	dissertation	publisher	collection	number of p	link to TOC	Texts: numb	Glossary: Y	
Gawne 2016	Yolmo	Gawne	Lauren		A sketch grammar	2016	LaTrobe	SOAS	Nanyang Technical U	University o Kelly, Barba	Asia-Pacific	A-PL 30	160 (171)			14	N	
Schackow 2016	Yakkha	Schackow	Diana		A grammar of Yak	2015				University o Bickel, Ball'	Language S	Studies in D527	603)			30	N	
Regmi 2012	Bhujel	Regmi	Dan Raj		A Grammar of Bh	2012	Tribhuvan University				Lincom	Languages	161 (206)			19	Y	
Turin 2012	Thangmi	Turin	Mark		A grammar of the	2012		Yale	Cambridge		Leiden Univ van Driem, (Brill	Languages	484 (vol 1)			270	Y	
Hari 2010	Yohimo	Hari	Anna Maria		Yohimo Grammar	2010	RIP	SIL			University of Edinburgh	SIL International/Central	127		N		Y (chapter)	
Doomenbal 2009	Bantawa	Doomenbal	Marius		A Grammar of Bar	2009					Leiden Univ Booij, GE; CLOT		338 (497)			29	Y	
King 2009	Dhimal	King	John		A grammar of Dhii	2009					Leiden University	Brill	Languages	288 (612)			181	Y
Borchers 2008	Sunwar	Borchers	Dörte		A Grammar of Sur	2008	University of Graz	University of Muenster			Leiden University	Brill	Languages	315				
Genetti 2007	Dolakha	Ne Genetti	Carol		A Grammar of Dol	2007					University o De Lancy, Mouton	Mouton Gra	515 (578)			18	Y	
Poudel 2006	Dhankute	Ti Poudel	Kedar Prass		Dhankute Tamang	2006						Lincom	Languages	178 (181)		N		
Hildebrandt 2004	Manange	Hildebrandt	Kristine		A grammar and gl	2004	University of Soutl Univer;	University of Leipzig			University o Genetti, Car	Pacific Linguistics	121 (192)			9	Y	
Kelly 2004	Sherpa	Kelly	Barbara		A grammar and gl	2004	University of Melbourne				University o Genetti, Car	Pacific Linguistics	75 (131)			7	Y	
Watters 2002	Kham	Watters	David		A grammar of Kha	2002	RIP		Tribhuvan University/SIL		University o De Lancy, Cambridge	Cambridge	1456 (467)			24	Y (chapter)	
Rutgers 1998	Yamphu	Rutgers	Roland		Yamphu: Gramma	1998					Leiden Univ van Driem, (Research S	Languages	318 (632)			196	Y	
Ebert 1997a	Camling	Ebert	Karen		Camling (Chamlin	1997a	RIP		University of Zurich		Kiel University	Lincom	Languages	68 (76)		N	N	
Ebert 1997b	Athpare	Ebert	Karen		A grammar of Ath	1997b	RIP		University of Zurich		Kiel University	Lincom	Studies in A	157 (283)			32	Y
van Driem 1993	Dumi	van Driem	George		A Grammar of Dui	1993	University of Bern	Leiden Universit			Leiden Univ Kortland, Fr	Mouton	Mouton Gra	275 (452)			52	Y
Michailovsky 1988	Hayu	Michailovski	Boyd		La langue hayu	1988	CNRS				University o Matisoff, Jiri	Editions du CNRS	204 (234)			22	N	
van Driem 1987	Limbu	van Driem	George		A Grammar of Lir	1987	University of Bern	Leiden Universit			Leiden Univ Kortland, Fr	Mouton	Mouton Gra	276 (565)			90	Y

Figure 18. Metadata database for grammars and authors in corpus

⁹³ Apart from the influence of a specific dissertation advisor, there is a larger pattern to be found in whether the university where the linguist trains takes a functionalist or formalist perspective, as these are associated with different curricula, methodology and acceptable dissertation topics; see Thomas (2020).

The macro-level database made it possible to determine a taxonomy of the grammars in the corpus. The major types we found were the following schemes:

- Part of speech prominence
- Nominal-verbal opposition
- Linguistic subfield organization
- Topics of areal prominence
- Grammatical category

The first two types, Part-of-speech prominence and Nominal-verbal opposition, are based on word classes and a grammarian's choice to either focus on a basic opposition between nominal and verbal elements, or to treat word classes at the top-level of the grammar, given them each a chapter. The third type, Linguistic subfield organization, refers to a structural scheme with chapters such as Phonology, Morphology, Syntax, Pragmatics, seen in a great many grammars. The other two schemes are sub-schemes: 'Topics of areal prominence' refers to a scheme that highlights topics which are significant from an areal point of view. An example is nominalization in Sino-Tibetan languages, a cover term for a number of functions (relativization, attribution, complementation) covered by the same morphology (Matisoff 1972; Bickel 1999; Genetti et al. 2008; Yap, Grunow-Hårsta & Wrona 2011). The inclusion of chapters covering these kinds of topics might be seen as consistent with an effort to describe the language "in its own terms"⁹⁴ (a notion that goes back to Boas), in addition to facilitating language comparison within a family or group. 'Grammatical category' is also a sub-scheme in that we have not found in the corpus any grammars exclusively organized around grammatical categories. Grammars of this type are those that highlight categories such as aspect, tense, case, and the like by giving them top-level chapter billing, where the topics might have been discussed within chapters on verbal and nominal morphology respectively.

⁹⁴ See for example this statement in Genetti's grammar of Dolakha Newar: "An advantage of the functionalist framework in the writing of descriptive grammars is that it is sufficiently flexible to **allow the language to be seen in its own terms**. There are no theoretical requirements that certain types of categories be present in every language--even lexical categories as common as noun and verb may not be found in all languages. One needs to look at the details of the language to see if the presence of such categories is justified." (Genetti 2007: 27); Note that Genetti has a chapter in her grammar called "Nominalization and related structures".

The taxonomy of structuring schemes (and sub-schemes) can, once established, be correlated with criteria found in the metadata database, most significantly with the date of publication of grammars and their size. The types that the grammars represent can also be explored with respect to biographical data, which leads to the suggestion of linguistic schools. One rather striking example is the nominal-verbal opposition grammars written by van Driem, and taken up by his students. They are not alone in doing this (Ebert also favors a nominal-verbal opposition in her grammars, but interestingly, she orders her discussion of verbs before that of nominals), but one of the consequences of the choice of structure is that there is no discussion of adverbs in van Driem grammars, and therefore there is no place where ideophones (often considered a subcategory of adverbs) could be discussed.

The taxonomy of grammatical organizational types can, and certainly will, be refined: the version described herein is based on the corpus of only 18 grammars written over a 30-year period and is therefore in no way representative. We nonetheless feel that establishing taxonomies of this type is an important tool for thinking about how grammars are written and how they can be used, something which has been surprisingly absent (as far as we can tell) from contemporary grammar writing and typology. One of the uses of the taxonomy, beyond smaller-scale influences (such as those by supervisors, by practices in a language area), is to recognize historical influences: the Greco-Latin traditional grammar scheme is easily identifiable as a parts-of-speech organization; the nominal-verbal scheme reflects the age-old prominence given to nouns and verbs in grammars, going as far back as Plato (Colombat & Lahaussais 2019b: 32). But what of the other 'types' we have identified? And is it possible to connect them to historical models? Is it possible, for instance, to identify a type associated with Paninian grammar? The Paninian tradition is said to be characterized by an emphasis on derivational morphology, with specific terminology for derivational procedures (see for instance (Lambert-Brétière in press), but how can this influence be identified in subsequent grammars? One scheme of particular interest is what we have called "linguistic subfield organization". While this appeared to be the dominant scheme in the 1980's and 1990's, it now seems that a "grammatical categories" scheme is more common. How can we account for this apparent change? Is this the influence of the development of typology over the last 20 years, fed by better descriptions and methods? Is this the result of a change in the organization of the curricula of linguistics departments, which students then replicate when they are assembling grammatical descriptions? And beyond the organization of curricula, what is the deeper origin of the linguistic subfield model?

Many schemes that descriptive linguists assume are contemporary in fact have considerably deeper roots: the "so-called Boasian trilogy" (see for instance Evans & Dench (2006: 10)) of grammar, glossary and texts in fact dates back considerably further than Boas: Smith-Stark (2009) discusses the 'trilogía catequística' that is made up of grammatical descriptions ('artes'), dictionaries ('vocabularios') and text collections ('doctrinas') -- even though the latter were translation-based renderings of Catholic texts rather than autochthonous narratives -- in 16th century language descriptions from New Spain. (Smith-Stark himself points out the parallel with the Boasian trilogy (2009: 471)). Note that the pieces of the trilogy were produced at different times and by different authors, although cross-references make clear that they were materials that built on each other, and it is therefore possible that the doctrinal texts served as the corpus for the grammar. From the second half of the 17th century, we find all three parts of the trilogy published together in one volume (Otto Zwartjes, pc. May 2019).

Similarly, the linguistic subfield scheme has roots which are deeper than we might suspect: The organization of grammars around chapters on phonology and syntax goes back to early stages of the Western tradition.⁹⁵ The treatment of morphology, while present in the very first Western grammars, was often within chapters on individual parts of speech, describing their 'accidents' (Lallot 2019; Colombat 2019a; Colombat 2019b). The term 'morphology' as applied to linguistics (as the German "Morphologie") seems to date back to Schleicher and to a work published in (1859), *Zur Morphologie der Sprache* (François 2013). In his voluminous *Compendium* (1861), there are only two large sections: *Phonologie* and *Morphologie*.

Yet despite the existence of earlier practices, it seems that the decision by contemporary linguists to structure grammars based on subfields of linguistics most likely results not from historical precursors but from the core courses in linguistics departments. Thus Nida's (1949) textbook on morphology was probably a significant, if indirect, influence on this 'subfield'-type of grammaticographical scheme.

One less developed part of the overarching database is the relational database: the intent is to align all sections of the grammar dealing with similar content, in order to visualize easily where different linguists have decided to discuss certain topics. This is analogous with a parallel corpus, or rather a comparable corpus, as the materials are not identical, of the tables of contents of the grammars in the corpus.

⁹⁵ Discussion of syntax appears to trace back to Apollonius Dyscolus in the 2nd century AD.

One of the major issues in comparing grammars is the question of dealing with different terminological choices. Aligning the tables of contents of the grammars allows us insight into the choices made by different linguists. In order to develop a tool for automatic alignment (or rather partially automatized) of the tables of contents requires a comparative terminological database. While we have not yet built this terminological database, an initial alignment of the grammars on the basis of a few categories (parts of speech and grammatical categories) yields some interesting results which can be correlated with the structural schemes.

The project is only beginning, but it seems to us that this approach, of considering the structure of descriptive grammars, assigning them to types within a taxonomy (which will certainly need to be refined), and correlating aspects of their structures with other metadata, and eventually aligning their content so it can be visualized easily (in other words, a search for a particular term, say 'participle', would reveal where the notion was discussed within all the grammars, even if using a different term) may well shed light on the more unconscious aspects of grammar writing. It is additionally an interesting pedagogical tool to have students code data for this type of study, and attempt to align content: it brings to the fore the actual shape that grammars take, and the variation in terminological use. It also helps students develop an awareness of the subjectivity of the content and structure of grammars, and the challenges in using them for typological studies--precisely because a choice of a particular structure may make it difficult to include material about a topic.⁹⁶

One other related study I have carried out is looking at field manuals and what methods they promote for structuring grammars. Lahaussais (2019a) considers a number of field guides to determine what use they make of the parts of speech as a tool for investigation or method for organizing the resulting analysis. The various field guides under study in fact group into two (sometimes overlapping) types: manuals aimed at data collection (Samarin 1967; Bouquiaux & Thomas 1976a; Bouquiaux & Thomas 1976b; Bouquiaux & Thomas

⁹⁶ One second-year master's student I supervised wrote a thesis ('Une étude de grammaticographie— Influences exercées par le métier des auteurs et le temps dans les grammaires mongoles', submitted by Mina Yi in 2019) on the different models found in grammars of Mongolian from 1845 to now. She was particularly interested in how grammarians would organize a discussion of evidentiality in grammars written before the invention of the category. Even though nothing earth-shattering emerged from her study, it became clear that having students carefully read and compare grammars in terms of their structures (comparing the actual textual content only for specific topics) can be a very useful way to have them learn about the history of categorization and the development of terminology.

1976c; Vaux & Cooper 2003; Bower 2008; Crowley 2007), and manuals focused on analysis and grammar-writing (Payne 1997; Chelliah & de Reuse 2011; Aikhenvald 2015). A conclusion from this study is that while data collection manuals do not focus on the parts of speech as a heuristic tool, they do feature more prominently in analysis-oriented guides. There is of course much to study in considering the influence of field manuals: Samarin (1967)'s *Field Linguistics* appears to be the first guide of this type, followed a decade later by Bouquiaux & Thomas (1976a)'s *Enquête et description des langues à tradition orale*, but despite this slow start, publications in the field picked up greatly in the 1990's and show no signs of slowing (the most recent as of this writing appears to be Meakins, Green & Turpin (2018)). How do these manuals shape the collection of data by students, and more significant to the topic at hand, how do they (perhaps subliminally) influence the organization of the grammars written? The conclusion I reached was that the traditional parts of speech do not feature prominently in field manuals, suggesting that one can conceivably successfully carry out a descriptive project without making use of them. On the other hand, the notion of parts of speech (or of word classes) is very much present in grammar-writing manuals, as a metalinguistic tool for organizing one's analysis of collected data.

3.2 Annotation practices

3.2.1 Generalities regarding the study of annotation

"Glossing", in the context of language description, is a polysemous term, used at once to refer to the overall process of generating multi-leveled annotation of linguistic data as well as to the production of word-for-word or morpheme-for-morpheme translations accompanying transcriptions. I use the term in both senses herein: alone, or in the collocations with 'practices', it refers to the former, namely the entire process of annotating a text so it can be read by a speaker of another language (this is the sense found in much research into manuscript glossing, such as Whitman (2011) and Cinato (2017); this holistic perspective makes it possible to establish connections to glossing practices in different time periods and linguistic traditions, and to attempt to determine what are common elements, of either a continuous or discontinuous nature. I also use 'glossing', in collocations with 'tier', 'interlinear' or 'morphemic' for the more technical sense, in which it refers to providing morphemic translations for segmented elements within the transcription tier (this is the sense used by descriptive linguists, and discussed by Lehmann (1982; 2004) from a historical as well as programmatic perspective.)

For linguists collecting data on languages, creating a glossed text represents a significant analytical step: it presupposes that the linguist has a unified vision of the morphology of the language. As these views are at least in part shaped by external factors, studying glossed texts can potentially provide insight into influences such as prevailing theories of morphology at the time.

Even without going so far as claiming to be able to detect theoretical leanings based on glossing choices, seemingly simple questions about the evolution of physical aspects of glosses remains a challenge. Some open questions are the following:

- When is the first instance of a vertically-aligned three-leveled system of glossing, with a transcription tier, a glossing tier, and a translation tier?
- When does segmentation, marked with dedicated segmentation signs, appear?
- How does the development of grammatical (vs lexical) glosses tie in with the emergence of morphemic glossing?

Part of the difficulty in finding answers to these kinds of questions stems from issues which obscure the trends that might otherwise emerge. A few significant issues are detailed below:

a) It is unclear to what extent changes found in glossing are linear: in many cases, no clear dates can be identified because changes occur at a micro-level, and are not picked up by the larger

community. These micro-levels can be individual authors, or traditions that become established within linguistic areas at a given point in time. There is always a period of ebb and flow before a new practice is embraced by the community at large, and this makes it difficult to pinpoint periods of transition, or individual texts or authors who are responsible for a change, with any precision.

b) The audience for the glossed material affects the output. Lehmann sees the choice of lexical as opposed to grammatical glosses as determined by intended readership: for example, he suggests that Finck (1909)'s use of glosses in order to render Turkish data accessible to his 'non-specialist readership' 'falsifies the working of the language by attributing lexical meanings to its grammatical morphemes' (Lehmann 2004: 1836). But it seems that it may not be a matter of choice in all cases: grammatical glosses have not always been part of linguists' toolkit--or rather, while they have existed at different times and places through history, and in many different forms (such as medieval Japanese morphosyntactic glosses (Whitman 2011; Whitman & Cinato 2014), their appearance can perhaps be seen as an instance of discontinuity in glossing practices. It is very possible that each glossing tradition has had to independently invent grammatical glossing, and that the trigger for their invention can tell us a lot about changes in analytical frameworks and the need to express new categories or configurations. In other words, while the intended audience probably does affect the glossing provided for linguistic data, grammatical glosses can only be used once they have been developed within a certain tradition.

c) The physical appearance of glosses is affected by typesetting decisions (most obviously segmentation and alignment), and as a result, it is not clear who has contributed to giving glossed material the appearance it has when it appears in print. Until the advent of digital submission of manuscripts, data compilers, typesetters, journal editors were certainly responsible for changes to elements of glossed texts in published work. This is a crucial point in working with earlier materials (such as those from the 1850's), when more individuals were involved in the production of journals or books, and when the professional editors had to make sense of hand-writing notes, with more or less readable diacritics. In some cases, the question is raised explicitly in journals: Figure 21 below is taken from a note following the 1857 *Journal of the Asiatic Society* article by Hodgson describing the languages Vayu and Bahing:

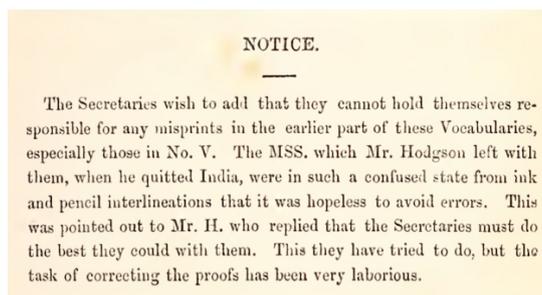


Figure 21. Notice inserted between Table of contents and
page 1 of the 1857 issue of the JASB

If the Secretaries in question were told to "do the best they could" with Hodgson's materials -- materials which, incidentally, are surprisingly inconsistent in their use of segmentation and glossing⁹⁷ -- then who can be said to be responsible for the production of the exact form of the material we are using to examine the evolution of glossing practices? Note of course that contemporary publishing practices put almost the entire burden of preparation of annotated data on the author (only in one case in my experience has an editor suggested changes to my Leipzig-conforming glosses), but that is because instructions are very clear, and because the author provides an almost-fit-to-print article. The question of who has made changes to glossed data is nonetheless very important when seeking to document changes through time: If annotation is to reveal the theoretical stance of the author, then preferably it is the author him/herself annotating the data, and not a typesetter who makes modifications based on the need to have an example or a text fit on the page.

Setting aside these issues which pose some challenges to the study of the history of glossing, there is in fact a great deal to be learned from studying changes over time in the presentation of glossed material. Some studies I have carried out in this direction are described below.

3.2.2 Annotation studies: methodology and materials

One approach I have found useful in attempting to trace the evolution in glossing over time is to start from contemporary practices, which are more or less standardized, and therefore form a relatively consistent set of practices, and work backwards in time, attempting to identify when

⁹⁷ See van Driem (2004) for some interesting suggestions about research on Hodgson's archived manuscripts.

aspects of the practices evolve; when possible, I have also sought to identify the reasons for these changes.

The difficulty with this approach is the massive amount of data that potentially makes up the relevant corpus: theoretically this is all glossed text, in descriptions but also in theoretical work. Trends are only visible with some perspective, and it is only recently that it has become clear to what extent the Leipzig Glossing Rules have been adopted, at least within the fields of typology and language description/documentation: linguistics journals regularly make mention of them when referring to the preparation of examples for the submission of manuscripts.⁹⁸ In order to render the volume of data manageable, the studies on the evolution of glossing I have carried out have been delimited geographically, focusing on descriptions of Tibeto-Burman languages of Nepal and Algonquian languages, occasionally extending to other languages within the language area of the two groups.

The time period covered by the studies has quite simply been determined by the available materials. For Kiranti languages, as well as other Tibeto-Burman languages of Nepal, the earliest grammatical descriptions in English⁹⁹ date back to the 1850's, in the form of Hodgson's work on Vayu and Bahing published in articles of the *Journal of the Asiatic Society of Bengal* from 1857 and 1858. Even though descriptive work on Algonquian languages goes considerably further back in time (the first Algonquian grammar being Eliot (1666)), I have limited my Algonquian corpus so that it matches the Nepal corpus temporally.

The choice of languages was determined by interest and familiarity, on the one hand, and morphology, on the other. Because studying glosses is somewhat painstaking work, and involves trying to second-guess the linguist's annotation decisions, it is useful to have some familiarity with the languages in the corpus. I hope the first section of this document will have

⁹⁸ A few examples--of the many-- are listed here (bolding of 'Leipzig Glossing Rules' is mine):

From *Studies in Language* (Brill) Guidelines for contributors:

"Examples should have numbers in parentheses, and these parenthesized numbers should be used for references in the body of the text. Examples should be glossed following the **Leipzig Glossing Rules**."

From Instructions for authors for *Linguistics* (DeGruyter):

"Authors are advised to adhere as closely as possible to the De Gruyter Mouton journal style sheet, especially regarding the form of Figures, Tables, Citations, and the References. For Examples, authors should attempt to follow the **Leipzig Glossing Rules**."

From Instructions for authors for *Cahiers de linguistique Asie Orientale* (Benjamins): "Linguistic examples must be numbered and follow the **Leipzig Glossing Rules** (eva.mpg.de/lingua/resources/glossing-rules.php) and *The Generic Style Rules for Linguistics* (eva.mpg.de/fileadmin/content_files/staff/haspelmt/pdf/GenericStyleRules.pdf)."

⁹⁹ While such a study could of course be carried out on work written by linguists of any background, it also seemed important to focus on English-speaking linguists who would be potentially be able to read each others' work and be influenced by the same theoretical developments in the field.

suggested that I am relatively at ease with the structure of (some) Kiranti languages. An Algonquian language class taught by Guillaume Jacques and Anton Antonov at INALCO in 2016 first introduced me to the structure of Algonquian languages,¹⁰⁰ via Ojibwe and Cree; the class was followed up by a (now temporarily dormant) Algonquian text reading group. This gave me a strong interest and very passing familiarity with Algonquian languages. The polypersonal indexation found in both Algonquian and Kiranti languages was an additional advantage to using both language groups to study annotation practices: (pre-contemporary) linguists annotating data from these languages were faced with verbal morphology considerably different from that of European languages, and therefore with decisions about how to account for the indexes of multiple roles.

I have undertaken a number of studies on the evolution of glossing (two of which have appeared in published form, with the others presented at conferences), and shall describe them in turn. They are completed by the preparation (currently underway) of a volume (edited by Franck Cinato, Aimée Lahaussais and John Whitman) which will assemble contributions from linguistic traditions and periods (ancient China, medieval and contemporary Japan, Ireland, among others) in an attempt to examine glossing practices from a comparative perspective and determine what features can be attributed to universal linguistic practices, and what features are culturally/linguistically/temporally determined.

Lahaussais & Léon (2015) looks at the emergence of norms governing data annotation, looking at the impact of work by Voegelin and his students in setting up practices of transcription and translation of data on Amerindian languages. The article can be seen as a first foray into looking at the formalization of translation as part of data annotation: while contemporary field linguists are left somewhat adrift by the lack of instructions concerning the translation tier (as opposed to the transcription and glossing tiers, which have been equipped with rules)--to wit, the difficulty typologists have in using translation tiers in order to identify constructions for their research (Cysouw & Wälchli 2007)--there were attempts in the 1950's to formalize the processes governing the production of the translation tier (as well as the glossing tier). These processes, while they resulted in methodological advances, and the possibility of automatization, do not seem to have had a lasting impact on language description.

¹⁰⁰ The Jacques-Antonov method involves diving into transcribed/translated (but un glossed!) texts and working through them with a dictionary until they start to reveal themselves.

A second article (Lahaussois 2016c) focuses on the translation tier in glossed text, in this case without attempting to place it in any kind of theoretical context but instead exploring how the translation tier has evolved over time within the context of materials deposited in oral archives. Looking in particular at texts deposited in the PanGloss Collection,¹⁰¹ I was able to identify a great many instances of the translation tier being used not for free translation (as per EMELD recommendations)¹⁰², but rather for literal translations, a tendency which had been pointed out previously (Bow, Hughes & Bird 2003). The literal aspects of the translations provided were grouped into a taxonomy of deviations from what might constitute a free translation (word order issues; tense concordance issues; morphosyntactic issues), and my hypothesis was that in a context where the glossed textual material is separated from the grammar--as is the case in an oral archive, where the texts are accessed independently of grammatical descriptions--linguists consciously use the translation tier in order to make clear the syntactic (and sometimes morphosyntactic) features of the source language. It seems that until standards are established to impose some systematicity on the treatment the translation tier--on par with what has happened for the transcription and glossing tiers--, it will be used by different linguists to do a number of different things. One of the results is that this makes it difficult for typologists to effectively make use of glossed data to search for features of interest. A set of "Translation rules" could, for instance, propose standard uses for typographical symbols (such as [], ()) often found in translation tiers, which are variously used to indicate a literal translation but also syntactic constituents of relevance; it could also provide clear instructions about negotiating the literal to free continuum, with advice on how to highlight relevant linguistic features.¹⁰³

A conference presentation (Lahaussois & Léon 2017) was another attempt to delve into the influence of the Americanist tradition on the development of glossing practices. The *International Journal of American Linguistics* was selected, by virtue of its being an important venue for language description in the Americas and for theoretical developments (with editors and contributors such as the Voegelins, Boas, Hockett...), as the source for the corpus, which consisted of all articles for the period of 1917-1970 presenting descriptive data. The corpus

¹⁰¹ http://lacito.vjf.cnrs.fr/pangloss/index_en.html

¹⁰² See <http://emeld.org/school/classroom/text/levels.html> for discussion of recommendations for different tiers of interlinear glossed text.

¹⁰³ This would presumably have been part of Enrique Palancar's project on annotation which had been selected for funding by the now-defunct Fédération de typologie et universaux linguistiques.

was mined with respect to 2 questions: the physical arrangement of annotated data on the page and the nature of glossed elements (whether these were words, morphemes, or something else).

The essential trends that were found concerning physical arrangement of annotated data are that data was presented in interlinear fashion from 1917 to 1945, after which point data begins to appear in a sequential presentation (even when fully glossed).¹⁰⁴ The disappearance of interlinearly aligned data corresponds with a trend of glossing morphemes instead of words, as evidenced by an increased use of segmentation markers in the transcription tier. The decade from 1945 to 1955 is a transitional period for units of analysis, with the publication of major theoretical texts (Hockett 1954; Harris 1942) on the definition and delimitation of morpheme and on models of grammar, and it is perhaps not surprising that this should lead to changes in the presentation of glossed materials. Although they were not part of the corpus, a look at *IJAL* articles from the mid-1980's on again reveals a reemergence of an interlinear presentation, possibly corresponding with proposals for interlinear morphemic glossing (Lehmann 1982).¹⁰⁵

While it seems like there is important material in this study, it again reveals the difficulties of trying to pinpoint periods of transition: even within the same journal, and in the same (albeit large) linguistic area of the Americas, the presentation of data varies widely. While trends do emerge, they do so fuzzily, and the significance is difficult to correlate with other synchronic developments.¹⁰⁶

The next study (Lahaussais 2017c), which will be a chapter in the book on comparative glossing practices (Cinato, Lahaussais, Whitman (eds), in prep), looks at a corpus of descriptions and materials on Algonquian and Kiranti languages published in the period from

¹⁰⁴An example of a fragment of sequentially (vs interlinearly) arranged glossed text is Harrington (1946: 237):

'æ'ho·w'æ·'hey'dl nɑ̄ hæ·'tey'dl t'æ'p
how'dlhey'dl. [*They-were-traveling-along and
someone an-antelope killed.*] 'æ'pɛ·y'nɛ·y'
nɔ̄ kyæ'tæy'ki' tsæ'nhey'dl. [*They-were-
butchering-it and the-chief came-up.*] t'æ'p

¹⁰⁵ Lehmann is quite explicit about the connection between morphemic glossing and an item and arrangement approach (see, for example, his website;

https://www.christianlehmann.eu/ling/ling_meth/ling_description/grammaticography/gloss/index.php?open=two_approaches)

¹⁰⁶ One other hypothesis we had for the disappearance of interlinearized data presentation was that it was related to the period of World War II and shortages of materials such as printing paper; a sequential alignment of annotated data saves space on the page (which is why to this day articles often have a combination of both interlinear annotation and sequentially aligned annotation for examples introduced within the text).

1850-1910, and specifically at the glossing of 2nd person pronouns and indexes. The corpus made it possible to explore the period immediately preceding that of the *IJAL* study described above, in order to determine some of the characteristics of glossing in the period in question. The presentation mostly focused on the glossing of 2nd person pronouns and indexes: the neutralization of number and role in modern 2nd person pronouns resulted in 'you' being an imprecise gloss, especially for languages with verbal indexation of both subject and object arguments. To counter this, both Algonquian and Kiranti descriptions of the period covered by the corpus make use of the archaic pronouns of Early Modern English (spoken 1500-1700, on other words already outdated at the time of the corpus) in order to distinguish between features necessary for accurate glossing.

	subject 2SG	subject 2PL	object 2SG	object 2PL
archaic	thou	ye	thee	you
modern	you	you	you	you

Figure 22. Archaic pronouns of English and their modern equivalents

I claimed that their use constituted a step in the development of grammatical glossing (at times when glossing was exclusively lexical): these pronouns are not used as mere lexical glosses, but as a means of encoding additional grammatical information, and as such they represent a step in the development of grammatical glosses cooccurring alongside lexical glosses.

Another study (Lahaussais 2019c) focuses on the use of segmentation in materials describing and comparing Tibeto-Burman languages of Nepal from 1850 to 1935. Because there is not a huge amount of material for the period in question, I resorted not only to grammatical descriptions (often brief sketches) but to compilations of language data (often drawn from work collected by others, but somewhat reannotated). The corpus was the following:

- Hodgson's descriptions of Vayu and Bahing (Hodgson 1857a; Hodgson 1857c; Hodgson 1858)

- Hunter's comparative dictionary, with languages of Nepal mostly based on Hodgson (Hunter 1868)

- Campbell's compilation of specimens of languages of India, made up of data collected by colonial officers using a single lexical and phrase list (George Campbell 1874)

- Linguistic Survey of India vol 3.1, focusing on the Tibeto-Burman family, a compilation of first-hand data collected by colonial officers (Grierson 1909)

-Articles based on first-hand data on Thulung, Kulung, Sampang (Wolfenden 1933; Wolfenden 1934; Wolfenden 1935)

Over the course of the 70 years of the corpus, the data is initially (ie in Hodgson 1857-8; Hunter 1868; Campbell 1874) fairly inconsistent with respect to segmentation: segmentation is only used, in word lists and phrases, for clear inflectional/derivational material matching the Greco-Latin 'accidents' (person, gender, tense, case, number, etc). When texts are presented, they are unsegmented, with divisions by word; translations follow in a separate section.

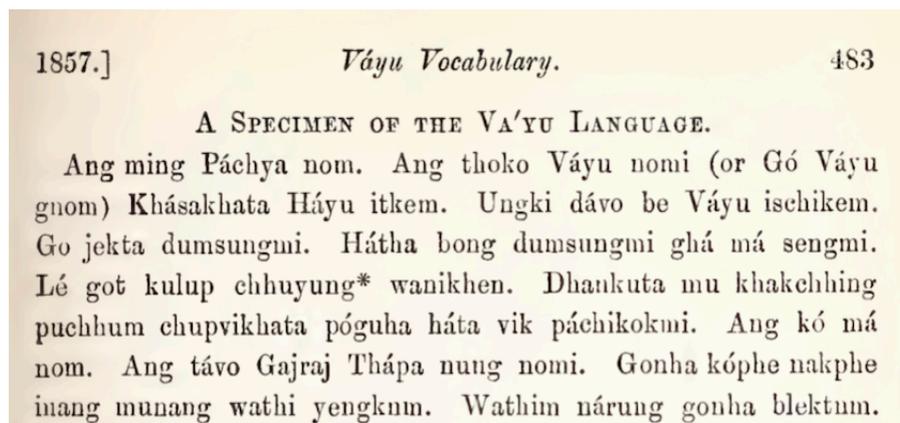


Figure 23. Transcription of text specimen in Vayu,
with no segmentation, (Hodgson 1857a: 483)

In the middle part of the time period (represented by materials in the *LSI*, (Grierson 1909)), examples in word lists and phrases, and even in texts, are much more consistently segmented, suggesting a deepening awareness of the morphology of the languages.¹⁰⁷ We do however find examples of prefixes and suffixes (in those very terms) that are not hyphenated, making it unclear what the criterion is for determining whether an affix gets a hyphen or not¹⁰⁸ and what a segment might represent in examples.

¹⁰⁷ Grierson felt that the *LSI* was a tool that would help colonial officers master the languages in their districts, doing away with interpreters. While this is an extraordinarily ambitious position--the very brief language sketches and comparative word lists are hardly enough to give one even a rudimentary knowledge of the language, let alone the ability to participate in court cases in the indigenous language (Majeed 2019a). Nonetheless, Grierson, in his capacity as editor of the entire project, working through the specimens of the Parable of the Prodigal Son he was sent in a great many of the languages, would presumably have developed a keen sense of the morphology of the languages in his sample.

¹⁰⁸ See for example the following: "Negative particle. The negative particle is a prefixed mā ; thus kā mā khāng-khi-kā, I do not wish ; kā mā hān-kā, I am not going..." (Grierson 1909: 279); "Number--there are two numbers,

Texts, on the other hand, are consistently segmented, even if the interlinear "glossing" tier also includes segmentation marks which are used for several purposes. The hyphens in the glossing tier serve to both group together into a single 'gloss' the English words that make up the translation of the (segmented) unit of transcription that they correspond to, and to segment the glosses for morpheme-like elements. As they are not typographically differentiated, the overall effect is of inconsistency.

At the end of the time period in the corpus (Wolfenden, 1935), segmentation is extensive, and etymological in nature, setting apart any element which can be connected etyma in the proto-language. This is the case with word *pā-p*, 'father', presumably a single unsegmentable word by contemporary speakers (as it is today), but segmented here, and accompanied by a footnote explaining that "-p almost certainly represents a formerly vocalized syllable" (ie. only *pā* is considered a morpheme to be used for reconstruction).

PARABLE OF THE PRODIGAL SON

<i>Kōñ mī-čyō-kā-m nī-čī ū-čyō bā-ī-dā. Mē nī-čī-dā</i>
One man of two his sons were. Them two among
<i>yāk-kē ū-čyō-kā ū-pā-p-kā-m-dā rāk-tā “Pā-ā, ā-mā</i>
younger his son his father to said “Father, my
<i>ā-dē kāū-lūñ hām-kō-bū kō-lē bā-nē”.</i> Ū-pā-p-kā
my share wealth that which is that give me”. His father

Figure 24. Annotation in (Wolfenden 1935) showing segmentation for the purposes of reconstruction and comparison with the proto-language

This use of segmentation as an etymological tool appears to be a particularity of the author of the *Outlines of Tibeto-Burman Linguistic Morphology* (Wolfenden 1929), and is not a use of segmentation I have seen elsewhere. The glossing accompanying the text does not make use of the segments, but rather translates entire phrases (made up of several segmented words) identified by physical layout, with the literal translation centered under groups of words making up phrases in transcription. Wolfenden is thus an outlier with respect to annotation on a number of fronts for his time.

the singular and the plural. The plural can be distinguished by adding suffixes such as *haru* and *pāli* ; thus *ā-pā haru*; fathers ; *chāmai pāli*, daughters” (ibid, 280)

A final study (Lahaussais 2020b) I wish to discuss focuses on glossing in the *LSI*, and on the role of large-scale linguistic surveys as vehicles for change, on account of their wide distribution and the interest they arouse. The study explores an 1897 text, *Specimen translations in various Indian languages* (Grierson 1897), which presents 65 language versions of the Parable of the Prodigal Son. These specimens were intended as a collection of translations of the same text in order to facilitate the collection of the parable for the *LSI*: it was believed that within the 65 versions, there would be at least one linguistically close enough to any language of India to help colonial administrators collect its translation. The interest of the volume from the point of view of glossing is that it is a collection of (mostly) annotated texts that predates the editorial changes made by Grierson in the *LSI*; in many cases, the original texts are the same, and the comparison of how they appear, from the point of view of annotation, in the 1897 volume and the *LSI* is revealing about the process of preparing these texts for publication.

The study was also the opportunity for me to refine an analytical grid which I use in order to track changes in annotation practices, and which is presented below in Figure 25. The three (potential) annotation tiers are treated separately, with consideration of various aspects (and sub-aspects) of the properties of each one. While a grid such as this may seem to be a fairly banal proposal, it has been very helpful to me in making it possible to objectively break down features of the annotation systems, rather than attempting a comparison on the basis of a more holistic reading.

Transcription						
(type) native orthography						
(type) Romanized spelling						
(type) phonetic system						
(type) other						
(segmentation) none						
(segmentation) some						
(segmentation) comprehensive						

Glossing						
(presence) none						
(presence) partial glossing						
(presence) fully glossed text						
(type) numbering						
(type) word-for- word						
(type) literal translation						
(type) morphemic						
(segmentation) none						
(segmentation) present						
(segmentation) order of segments follows metalanguage						
(segmentation) order of segments is mixed						
(segmentation) order of segments follows transcription						
(segmentation) number of segments matches transcription						
(segmentation) number of segments does not match transcription						
Translation						
(type) none						
(type) literal						
(type) free						

(position) set apart from annotation						
(position) interlinear						

Figure 25. Analytical grid for tracking evolution of annotation practices

With respect to the potential for such a grid to reveal anything concrete about changes in glossing practices, consider the simplified,¹⁰⁹ completed grid in Figure 26 below. The grid shows the features of both the 1897 *Specimens* volume (left-hand column) and the *LSI* (right-hand column) for the same languages. For ease of reading, the *Specimens* features are marked in parentheses. Using the grid, we can thus track the changes to various aspects of annotation across the two volumes, and come to some conclusions about annotation practices in the *LSI*.

language	Kashmiri		Marathi		Telugu		Lepcha		Kachari		Singpho	
	<i>Spec.</i>	<i>LSI</i>										
Transcription												
(segmentation) none			(x)								(x)	
(segmentation) some	(x)			x								
(segmentation) comprehensive		x			(x)	x	(x)	x	(x)	x		x
Glossing												
(presence) none			(x)				(x)					
(presence) partial glossing									(x)			
(presence) fully glossed text	(x)	x		x	(x)	x		x		x		x
(type) numbering											(x)	
(type) word- for-word	(x)	x		x		x		x	(x)	x		x

¹⁰⁹ The sections that have been removed are those concerning type of transcription and those concerning the translation tier (as the texts studied are exclusively the Parable of the Prodigal Son, the translation is not provided).

(type) literal translation					(x)							
(type) morphemic												
(segmentation) none					(x)							
(segmentation) present	(x)	x		x		x		x	(x)	x		x
(segmentation) order of segments follows metalanguage	(x)			x								
(segmentation) order of segments is mixed		x										x
(segmentation) order of segments follows transcription						x		x	(x)	x		
(segmentation) number of segments does not match transcription	(x)	x		x		x		x	(x)	x		x
(segmentation) number of segments matches transcription												

Figure 26. Completed grid contrasting 1897 *Specimens* volume and same languages in *LSI*

These conclusions are somewhat abstract in the absence of detailed examples from each of the sources, but the changes that we find applied across the *LSI* are the following:

- All annotation is interlinear, with transcription and glossing tiers
- Segmentation is found in both transcription and glossing tiers
- Glosses are word-for-word glosses

- Segments within glosses are generally in non-English word order (but not considered morphemic because the segmentation marks are used to group rather than divide); the fact that they are in order of identifiable morphemes within transcription is a big step
- Some segments within glosses even add grammatical information, although only lexical glosses are found

The study made it possible to identify some areas of change in annotation practices over the period between the 1897 *Specimens* volume and the publication of the *LSI* (the final volume of which was published in 1921). Considering the impact of the *LSI* at the time ((Majeed 2019a; Majeed 2019b), it is likely that it, like other large-scale linguistic surveys, played a role in propagating changes to annotation practices to other areas of the world.

3.3 Methodological tools

This section presents my involvement in a few projects of a methodological nature: the development of a comparable corpus of myths in three Kiranti languages (3.3.1); a database on linguistic questionnaires, and some reflections on how to create an archive for questionnaires while accounting for their historical and epistemological context (3.3.2); and a Khaling verb dictionary, with incorporated recordings of all example sentences (3.3.3).

3.3.1 Corpora: Kiranti comparable corpus

In this section, I will discuss a comparable corpus project built within the framework of an ANR project entitled HimalCo. The Kiranti Comparable Corpus,¹¹⁰ based on narrative corpora in three Kiranti languages (Thulung, Khaling and Koyi), was devised as a methodological tool with the two principal aims:

a) to find a solution to the problem of using culturally foreign stimuli in order to collect authentic narrative data. The effective use of the 'traditional' stimuli, such as the picture book *Frog where are you* (Mayer 1969) and the Pear Story film (Chafe 1980), have been questioned in this respect (Hellwig 2019; Lüpke 2009), notably when these stimuli are used to build parallel corpora for the purposes of comparison. Because the material is produced through translation, there are a number of issues about its usability, which tends to be affected by a series of biases (Wälchli 2007: 133). These have been discussed at length in the articles making up the STUF (Sprachtypologie und Universalienforschung/Language Typology and Universals) special issue on the use of parallel texts for typological studies (Cysouw & Wälchli 2007). Our intent in the project was to build a multi-language corpus facilitating morphosyntactic comparison that was not subject to problems resulting from translation-sourced materials.

b) to develop a sense of how closely related the three Kiranti languages in questions are, in a context in which the status of these languages as forming a subgroup is not resolved (Ebert 2003; Gerber & Grollman 2018).

Narrative material in Kiranti is ideal for building a comparable corpus: Kiranti languages share a mythological cycle, which has been described by a number of scholars (Gaenszle 1991; Ebert & Gaenszle 2009; Allen 2012). Even before becoming aware of research into the mythological cycle, I had noticed that a number of the traditional stories I had collected

¹¹⁰ <http://himalco.huma-num.fr/corpus/comparable/index.htm>

in the three languages shared narrative elements, even if the names of the protagonists were sometimes different and the number and sequence of events did not always match up across the versions. We determined, with the project, to collect as many of the stories making up the cycle as possible, in as many languages/dialects and with as many speakers as possible, and to align them in the manner of a parallel corpus.

We realized that because of significant differences in the narratives making up our corpus, due to individual differences in story-telling as well as to the fact that the different Kiranti languages had evolved different versions of the stories, that the concept of parallel corpus was not relevant, and that it was more appropriate to label our collection a *comparable* corpus than a *parallel* corpus.¹¹¹ Despite the imperfect match between the versions of the myths, we still felt it important for the purposes of comparison to align the data from different language versions.

The methodology for alignment was devised by trial and error: initially, the texts (which were all annotated with Toolbox) were compared in pairs, charting which sentences described the same content in an excel spreadsheet. This is described in early descriptions of the methodological aspects of the corpus (Lahaussois & Guillaume 2012; Lahaussois 2015). Although the pairwise comparison of versions of stories was viable given the size of the corpus I had assembled, the subjectivity of the process was somewhat troubling, and it soon became obvious that the only sensible way of carrying out the task of alignment was basing comparison not on sentences in individual stories but rather on the events making up the basic story. This again raised the issue of which version of the story to consider the underlying one -- when for example each language ordered events differently -- and forced us to assume that the different language versions derived from a proto-story. Note that there is no reason to believe that there ever was a proto-version: the stories could very well have been transmitted from village to village, changing through contact, as opposed to all sharing a single original source. Nonetheless, anthropologists like Allen (2012) claim that these stories are a common inheritance also shared by Tibetan peoples, suggesting that the Kiranti versions are part of a wider narrative tradition. Because of the uncertainty about the history of these stories, I shall

¹¹¹ "A comparable corpus can thus be defined as a corpus containing components that are collected using the same sampling method, eg. the *same proportions* of the texts of the *same genres* in the *same domains* in a range of *different languages* in the *same sampling period*. [...] By contrast, a parallel corpus can most easily be defined as a corpus that contains native language (L1) source texts and their (L2) translations." (McEnery & Hardie 2012: 20)

hereafter refer to them as *matrix* stories: these are the reconstituted narrative scaffolding, taking into account the various versions that I have access to. Given that the corpus only contains three languages for the time being (ideally of course we would increase it through contributions of annotated and sound-synchronized versions narratives from the same mythological cycle in other Kiranti languages), the question cannot be resolved as to how these stories were transmitted across the different languages.

The steps of the methodology were the following:

- a) Collection of narratives in Koyi, Khaling, Thulung, with full annotation using Toolbox;
- b) Identification of stories that are the same across the languages;
- c1) Pairwise comparison of versions of the same story, identifying sentences describing same narrative content [this was the initial procedure, later changed to c2];
- c2) Identification, on the basis of all the versions collected, of the key events making up each matrix story;
- d) Identification of the sentences in each individual version of the story with key events of the matrix story.

The mapping of individual sentences of each version with key events constituted the alignment of the stories. An engineer on the project, Séverine Guillaume, then developed an interface for the visualization of these alignments and of a concordancer with which to search the resulting corpus.

The Comparable Corpus is made up of two types of files. These are Annotation files, of which there is one per version of each story. The information contained in these files includes the transcription, glossing and translation of the individual sentence, word, and morpheme units that make up the text (Jacobson, Michailovsky & Lowe 2001; Thieberger & Jacobson 2010). There are also Alignment files, of which there is one per story. The alignment files identify the sentences in each version of each story that correspond to the events of the matrix story. A screenshot of the resulting Table of similarities for three versions of one story is shown in Figure 27.

Event	cocoon-c989b53a-e86d-30a8-9b51-cf738ac8e09f	cocoon-a0aed3de-fe98-35f2-9841-31acc163cae8	cocoon-27913aa2-d305-36c0-b454-4c4b3b63c6f5
1	S2 S3	S2 S4	
2	S5 S6	S4 S5	
3	S7		
4	S10 S11	S6	
5	S12 S13	S8	
6	S24 S25	S9	
7	S28 S29	S10	
8	S30 S31	S11	
9			S1 S2
10	S34	S13	S3
11	S39	S15 S16	S4
12	S39 S40 S41 S48 S49	S18 S21	S5
13	S58	S22	S5
14	S61 S62	S23	S6
15	S67 S68	S26	S6 S8
16	S69	S27	S8
17	S73 S74 S75	S28 S29	
18			S9 S10
19	S80	S32	S11
20	S81		S13
21	S83	S31	S13
22			S14
23			S15 S16

Figure 27. Table of similarities: shows which sentence units in each version of a story correspond to events

The table of similarities for each story (which contains all of the alignments, even though the screenshot in Figure 27 only shows the alignment of three versions) is one of the tools used to build the comparable corpus, and is part of the back-office, and is not intended as the means of searching the corpus. Instead, a number of tools are proposed for selecting, viewing and analyzing the data.

Corpus selection: The corpus is currently made up of five proto-stories, with up to six versions of each. The versions are identified by language -- Khaling, Koyi and Thulung (either Mukli or Deusa dialect) -- and by speaker, with twelve different speakers having contributed to the corpus. Because of the corpus design, only versions of a same story are aligned. Users can therefore only select a single story to view at any one time, with a choice of versions (ranging from a minimum of two to the maximum number of versions available, which is six, for the Orphans story), using the corpus selection interface, shown in Figure 28 below. This action yields a sub-corpus; All that follows refers to what can be done with that sub-corpus.

Presentation
Comparable corpus
Dictionaries
Read r

Presentation **Corpus selection**

Kiranti mythology

Corpus access

Proto-story	Language/Dialect	Speaker	
Ogress	Khaling	Dhana	<input type="checkbox"/>
	Khaling	Dhan Maya	<input type="checkbox"/>
	Thulung/Mukli	Bishnu	<input type="checkbox"/>
	Thulung/Mukli	Chandrakala	<input type="checkbox"/>

View selected corpus

Proto-story	Language/Dialect	Speaker	
Orphans	Khaling	Dhana	<input type="checkbox"/>
	Khaling	Lal Bahadur	<input type="checkbox"/>
	Koyi	Mila	<input type="checkbox"/>
	Koyi	Phanindra	<input type="checkbox"/>
	Thulung/Deusa	Lakpa	<input type="checkbox"/>
	Thulung/Mukli	Chandrakala	<input type="checkbox"/>

View selected corpus

Proto-story	Language/Dialect	Speaker	
Ur-Mother	Khaling	Dhana	<input type="checkbox"/>
	Koyi	Mila	<input type="checkbox"/>
	Koyi	Phanindra	<input type="checkbox"/>
	Thulung/Mukli	Chandrakala	<input type="checkbox"/>
	Thulung/Deusa	Lakpa	<input type="checkbox"/>

View selected corpus

Proto-story	Language/Dialect	Speaker	
Leper story	Khaling	Bijjame	<input type="checkbox"/>
	Khaling	Purna	<input type="checkbox"/>
	Thulung	Chandrakala	<input type="checkbox"/>

View selected corpus

Figure 28. Partial view of corpus selection interface

(<http://himalco.huma-num.fr/corpus/comparable/selection.php>)

Corpus viewing: The corpus interface makes it possible to view the selected sub-corpus in two different modes: Integral Text View and Similarity View. In Integral Text view, the full text of the selected versions is visible, arranged in a column. Because a certain proportion of the material in any given story will not have equivalents in the others, and will therefore not appear when only the similarities are viewed, this mode makes it possible to

maintain the narrative and morphosyntactic integrity of each version of the story. When a sentence in a given version is involved in a similarity, it is signaled by a hyperlinked label (“Similarity #”) and identified by color. The color makes it possible, when scrolling through the text, to identify visually which sentences participate in a given similarity and the general physical layout of the correspondences, which do not occur in the same order in the different versions of the stories. The Integral Text View is illustrated in Figure 29 below.

Proto-story Ur-Mother

Khaling Dhana	Koyi Phanindra	Thulung/Deusa Lakpa
<p>S1 lə ʔanammim samma tamra ʔinji ʔa ʔinji ʔa ʔeisiʔe Will you listen to the story from long ago of the beginning of the earth?</p>	<p>S1 asina sumnima salama-bo soma tʰins-a-m de-ki-lo ninambu-tsoptu mu-ka tsuktsu-tso ruwaharj paruharj mo-ni-m tsʰa A long long time ago, when we talk of man's creation, (we say) there were two gods in the sky above, Ruwahang and Paruhang.</p>	<p>Event 10</p> <p>S1 make oram bajara kole mijakma rwama ritse korja baiqa ʔe Long ago on this earth, there was a woman called Miyakma and she was all alone.</p>
<p>S2 ʔanam deunasʔe blettem ʔuŋa ʔeisi ŋimansu hai Deunas told the story a long time ago and I will now tell it to you</p>	<p>S2 jo idə bakʰaju bi pu soma det-ka asu jo ɔ-mo-ni-m tsʰa As for those called men, there were none on the earth.</p>	<p>S2 wanmim sy wo mibamba ʔe There was no one else around.</p>
<p>S3 ʔanam te tephem has mukde baja ʔeɟkim ʔikbrā:l tsəri baja mugo:we Long ago, there was no place for humans to live, what in our language we call Cheri baya</p>	<p>S3 so lo somo jo ɔ-go-m tsʰa There were not even stones or trees.</p>	<p>Event 12</p> <p>S3 mijakmakam dhyksikhrem malmu thari dystam baiqa ʔe The time came for Miyakma to find a mate.</p>
<p>S4 ʔanam tsəri baja mugo:we There was no livable land</p>	<p>Event 5</p> <p>S4 dʰaiʔo uk ŋisi-bi bakʰaju-no sopmu-lonka ribipma pipi tʰins-a tsʰa Then, one day, from the breath of the earth, Ribipma was created.</p>	<p>S4 tala bajara mini rwama sy wo mibamba ʔe But there was nothing on earth called a human.</p>
<p>S5 dhām jo mugo:we tephem ʔa peri ʔeɟki mephem tsəri jo mugo:we peri jo mugo:we There was no sky, which we call heaven, there was no earth, there was no heaven</p>	<p>Event 10</p> <p>S5 dʰano ribipma pipi-ja ja ebo ja ebo mu-to mu-to solome tsʰoʔ tsʰa</p>	<p>S5 sikewamanqewamim korja bamri ʔe Only Sikewa and Mankewa were there [two bird types].</p>
S6		S6

Figure 29. Integral text view for 3 versions of the Ur-Mother story

Notice that in the screenshot in Figure 29, the Thulung version of the story begins with what has been identified as Event 10, which corresponds to Sentence 5 in the Koyi version (itself following Event 5), while in the Khaling version of the story, Event 1 comes considerably later. The arrangement into columns makes it possible for the viewer to scroll down through each story and get a visual sense of which parts are similar, and where they are to be found.

The Similarity View pops up when one selects one of the color-coded similarities in the Integral Text View. The Similarity View for Event 5 (corresponding to the purple-colored event in Figure 29) is shown in Figure 30 below.

Proto-story Ur-Mother

Event 5

Khaling (Dhana)

S41

melo nempħrħgħmē buk-τē ?e ?lko saŋħl buk-τē mēħl ja:thħ khulu dħħħ?e
nempħrħgħmē rēntēlo tsħi ninoomridħm buk-τē ?e

melo nempħrħgħmē buk-τē ?e ?lko saŋħl buk-τē mēħl
at.that.time tree.species give.birth-3sg_pst_hs other vegetation_intens give.birth-3sg_pst do-then
ja:thħ khulu dħħħ?e nempħrħgħmē rēnt-τē-lo tsħi ninoomridħm
later wind.name wind_erg tree.species attack_3sg_pst-when_contr NiwamRidam
buk-τē ?e
give.birth-3sg_pst_hs

And another tree came up, and when the far away khulu wind blew on it, Ninwam Ridam grew.

Koyi (Phanindra)

S4

dħai?ħ uk ŋħsi-bi bakħaju-nō sopmu-lōŋka ribipma pipi tħħħs-a tsħa

dħai?ħ uk ŋħsi-bi bakħaju-nō sopmu-lōŋka ribipma pipi tħħħs-a tsħa
then one daytime-LOC earth-GEN breath-ABL [name] grandmother create-3SG.PST HS

Then, one day, from the breath of the earth, Ribipma was created.

Figure 30. Screenshot of Similarity View for Event 5 of the Ur-mother story

Note that the material in Figure 30 involves only two sentences, one from the Koyi version and one from the Khaling version. They are presented in interlinear format, facilitating linguistic comparison. This raises the important issue of consistency in glossing practices across the different versions that make up the corpus (Figure 30 reveals some issues in this regard, in the form of typographical differences in the grammatical glosses). For ease of comparison, it is preferable for annotation practices to be as similar (or at least transparent) as possible, something which complicates the integration of stories glossed by different linguists.

A few issues emerge: the visualization interface shown in Figure 30 is for comparison of the data. The choice to present the examples horizontally (after initial attempts at keeping the data in columns, as per the Integral Text View seen in Figure 29) was meant to facilitate the viewing of the results, but this is still not an ideal presentation. A second major issue concerns the subjectivity of the alignment. An outside viewer might well wonder what these sentences share that justifies labeling them as similar, and indeed, in isolation from the larger story ark they do not appear to have much in common. They are both connected to "Event 5" of the "Ur-

mother"¹¹² matrix story, which is additionally given a descriptive label (not seen here, but it is "The wind blew on the tree and the goddess came to life").

It is at this juncture that it is important to point out that the similarities found in the corpus can be grouped into three types:

Type a: similarities with shared narrative function only

Type b: similarities with shared narrative content

Type c: similarities with shared morphosyntactic constructions

These types have a certain amount of overlap, with each type also containing the features of the type immediately above it. They can generally be categorized as the following.

For Type a, sentences will refer to a same event in the story, but without sharing lexical or morphosyntactic material. An example is an aligned event referring to three protagonists, but with different references in the stories: in one version, the two sisters might be named and the brother referred to as "brother", while in another version, the sisters may be referred to as "sisters" and the brother named. This is different from the idea of protagonists being given different names in different versions of the story: Khaling Grômme and Lasme are Thulung Dzau and Khleu, a scenario we would label as having shared lexical content (Type b), as a database of lexical correspondences could be set up which would enable automatization of alignment. Type a similarities, on the other hand, are unlikely to be automatically alignable because they do not involve any lexical correspondences.

Type c are those similarities which, in addition to shared narrative function and content, have shared morphosyntactic constructions, and are the ones which are of real interest for linguistic comparison.

A final tool available for corpus users is the concordancer: it runs concordances on the selected sub-corpus, and can search for elements in the text transcription and glossing tiers. In essence, it functions as a trilingual glossary, insofar as a search for a lexical item (in English, via its gloss) yields results in the three languages of the corpus. It is also useful for identifying inconsistencies of glossing or of transcription (depending on how one runs the search).

The Kiranti Comparable Corpus was one of the products of an ANR project (2013-2016), coordinated by Guillaume Jacques, and involving myself and Alexis Michaud as PIs and Séverine Guillaume and a hired engineer as support staff. The following issues are what, from

¹¹² This is the name given to one particular story which is the shared foundation myth.

a distance of 3 years, I now see as missed opportunities and methodological flaws in the corpus we have built (which may yet be corrected): comparability of data and size of corpus.

a) Comparability of data: the sentences that participate in a similarity are difficult to compare from a linguistic perspective, even when they are presented in interlinearized form. Aligned sentences are viewed in Similarity view, such as in Figure 30 above. While removing the sentences from their narrative context favors an examination of their linguistic content, the way in which the different versions are displayed does not facilitate their comparison or the identification of what they might share (beyond the obvious, which is that they are both considered to be descriptions of the same Event).

One way to improve this would be to have the similarities coded to indicate what kind of similarity is at play (recall that the types were the following: shared narrative function; shared narrative content; shared morphosyntactic content). Ideally the type of similarity would be identifiable directly from the Integral Text View mode, so that a linguist interested mostly in morphosyntactic comparison might not have to look through all Similarities to find those relevant to him/her.

One recent partial improvement made to the visualization mode is the highlighting of the search term resulting from a concordance within the glosses in the Similarity view. An example is provided in Figure 31 below for ideophones. As mentioned earlier (see 2.2.1.2), Thulung and Khaling share a type of ideophone which contributes semantics associated with speed and force. These were not given grammatical glosses when the texts making up the corpus were first analyzed (only considerably later did I realize that this was a special grammatical category shared by the languages), and were instead glossed as 'suddenly'. Nonetheless, the concordancer, which is able to access grammatical and lexical glosses, as well as formatives in the transcription tier, can generate a list of all sentences incorporating this term. When one selects a Similarity view, the results are highlighted everywhere they occur within a given similarity, greatly facilitating the task of comparison.

Event/Èvènement : Vulture carries off mother to his nest

Khaling (Dhana)

S34

tunól bî: ho:teɲa me ʔumém ni dzhòŋ kúrteɲa khoottɛ

tunól bî: hɛ-te-nɔ me ʔu-mém ni dzhòŋ
 one.day mythical.vulture -come--3sg_pst-then that -3_sg_poss-mother indeed **suddenly**
 kur-te-nɔ khott-tɛ
 -carry--3sg_pst-then -take.away-3sg_pst

One day the vulture came and carried their mother off suddenly.

Khaling (Dhan Maya)

S13

manɔ ni dhâmtu bî: khɔrentɛsi, me bî:ʔe go tsok sũŋkhatɛ ʔe

manɔ ni dhâm-tu bî: khurt-le-te-si me bî:-ʔe
 and indeed -sky_loc_high mythical.vulture -make.turn--aux--3sg_pst_refl that -mythical.vulture-erg
 go tsok suŋ-khɔ-te ʔe
 _foc **suddenly** -pick.up--aux-3sg_pst _hs

A vulture was turned up in the sky and suddenly picked her up suddenly.

Thulung/Mukli (Bishnu)

S3

mɛ ne umam tsɔi bjuka lɔɖy ʔe

mɛ ne u-mam tsɔi bju-ka lɔk-dy ʔe
 that _foc -3_sg_poss-mother _contr -eagle-erg -take.away-3_sg_pst_a _hs

An eagle carried their mother away.

Thulung/Mukli (Chandrakala)

S2

me umamlai tsɔi kole bjuka lɔɖy reŋsha ʔe djeɔɖyɔma

mɛ u-mam-lai tsɔi ko-le bju-ka lɔk-dy reŋsha ʔe
 that -3_sg_poss-mother-dat _contr -one-cl -eagle-erg -take.away-3_sg_pst_a it.seems _hs
 djɛ-dy ma
 -pick.up-3_sg_pst_a and.then

A vulture picked up their mother and carried her away.

S3

tsokodjeɔɖyɔma lɔɖy reŋsha ʔe lɔɖy reŋsha ʔe memma hala gele lɔɖyɔma dɔkpu dɔksaɖɔla nem bɔnyɔɖyɔma dzyllɔm baiɖa ʔe

tsoko-dje-dy ma lɔk-dy reŋsha ʔe lɔk-dy
suddenly--pick.up-3_sg_pst_a and.then -take.away-3_sg_pst_a it.seems _hs -take.away-3_sg_pst_a
 reŋsha ʔe memma hala gele lɔk-dy ma dɔkpu dɔksa-ɖɔla nem
 it.seems _hs then up up -take.away-3_sg_pst_a and.then big -tree-above house
 bɔne-dy ma dzyɔ-dy-m ba-ɖa ʔe
 -make-3_sg_pst_a and.then -place-3_sg_pst_a_nmlz -be-3_sg_pst _hs

It picked up the mother quickly and carried her away, it carried her up, made a nest up in a big tree and put her in it. (tsoko=quickly, adv with pick up)

Figure 31. Similarity view based on concordance, highlighting the search term in the glosses

While the resulting data is not necessarily informative, it has the potential to be so. See for example that in Khaling (Dhan Maya), we find *tsok suŋ-khɔ-te* (suddenly pick.up-AUX-3SG.PST) and in Thulung/Mukli (Chandrakala), *tsoko-djɛ-dy* (suddenly-pick.up-3SG.PST_A). The comparison of ideophones of this type, whose only semantic contribution is suddenness

and forcefulness, is notoriously difficult, and yet the corpus in this instance has helped identify something interesting: even though the associated lexical verb is different in Khaling and Thulung, the ideophones are clearly cognate, and apparently their selection is triggered by the semantics of the verbs in question. A better-coded comparable corpus of this type would make it possible to successfully develop a much more fine-grained analysis of the semantics and of the usage (as they turn up in the description of the same dramatic narrative event) of this type of ideophone, and would thus contribute significantly to their study.

A useful modification to the corpus would be to code the constructions found in each version of the story. This could be done by adding an annotation level to the source data constituting the corpus, with the coding in that annotation level also available to the concordancer.¹¹³ One could thus perform a concordance on, say, relative clauses across the corpus, and see only Similarities that had that particular feature. This would be very useful as a means of focusing in on the potential that the corpus has for real comparison of the morphosyntax of the languages that are represented.

b) Size of corpus: An additional disappointment is not having a larger corpus, both in terms of number of speakers, languages, and stories (but anyone who has tried to collect traditional narratives in these communities -- and elsewhere I suspect -- will attest to the difficulty of the enterprise. This is also mentioned by Allen for Thulung, pointing out that the stories making up this cycle are only told on very specific occasions, making their casual collection challenging (2012: 13–18). The corpus in its current state is fairly minimalist, and serves more as a proof-of-concept than as a finished product. Additionally, improvements to the annotation of texts already in the corpus cannot be easily integrated: my newly re-segmented/transcribed Thulung corpus does not have sentence units that correspond to the ones in the stories originally used; I cannot therefore update the annotated versions of texts without having to go through the process of realigning them anew.

¹¹³ This is however unlikely to happen in the current format of the corpus. The reason is that the language versions are those which are archived in CNRS's PanGloss collection. While PanGloss could potentially accommodate a 4-tiered annotation of texts, with one of those tiers tracking constructions, this might require changes to the methodology currently used to prepare annotated texts for archiving.

3.3.2 Questionnaires

My first experience of questionnaires, apart from familiarity with the STEDT wordlists,¹¹⁴ was Payne (1997), recommended by Leanne Hinton as I headed off to the field for the first time. Considering the many variables that affect field conditions -- in addition to environmental and personal factors, such as health issues and overstimulation -- I have over the years relied on various field checklists to guide my sessions with consultants,¹¹⁵ and this has given rise to curiosity about this aspect of field methodology.

The occasional difficulty of field conditions in Nepal, alluded to above, made me wonder about the possibility of having untrained native speakers oversee the documentation of their own language, using a data collection guide. This would make it possible to do linguistic salvage work in parts of the world where conditions are too dangerous for outsiders, either temporarily or permanently. These musings led to an attempt to write a 'self-description manual', funded by UNESCO and published in CD format within the framework of a project called Initiative Babel (Lahaussois 2006). While the project is naive in a great many ways, it led to an interest in the tools that can accompany and facilitate a description/documentation project, namely questionnaires.

The Fédération de typologie et universaux linguistiques (TUL) funded a project I submitted on linguistic questionnaires: the goal was to collect as many questionnaires as possible, inventory their diversity and organize them into a taxonomy, and develop an archive where linguists could find and deposit the various tools they used and developed.¹¹⁶ The first step of the project was to define the object of our study, which we did by interpreting the term "questionnaires" very loosely, as "any methodological tool used in the collection of linguistic data for a descriptive or typological activity" (Lahaussois 2019d: 33). In order to signal that this was a project-internal use of the term, we referred to these tools as Questionnaires (by analogy with Haspelmath (2010)'s suggestion that language-internal descriptive categories be capitalized). Thus defined, the term covered, in addition to questionnaires (of the translation or analytical type), a number of tools that are not based on questions, and not traditionally considered to be questionnaires: wordlists, stimuli kits, checklists (such as one might find as

¹¹⁴ The approach is described in Matisoff (1978). The questionnaires can be found at <http://tulquest.huma-num.fr/en/node/40>

¹¹⁵ There are of course enormous gaps in my data collection despite this, but I am still grateful for the metaphorical crutch.

¹¹⁶ <http://tulquest.huma-num.fr>

an appendix to an article describing a phenomenon), but also templates for structuring grammars (such as the Lingua Descriptive Series questionnaire (Comrie & Smith 1977), and even less well-defined models, such as the Greco-Latin grammatical model).

The taxonomy the archive is developed around is the following types:

TUL Questionnaires (highlighting the output of other TUL projects)

Areal Questionnaires

Questionnaires by metalanguage

Questionnaires by linguistic subfield

Questionnaires by data type

Questionnaires by medium

This organization was an attempt to recognize that relevant aspects of a Questionnaire are more varied than is apparent in the tradition division between analytical vs. translation-based questionnaires, on the one hand (this is for example what was used by the EUROTYP project, funded by the European Science Foundation from 1990 to 1994 and directed by E. König; see the General Preface to Dahl (2000) for details), or the difference between questionnaires, on the one hand, and stimuli, on the other. We also wanted to account for the fact that other aspects will shape Questionnaires in various ways, namely the metalanguage used in their development, the linguistic subfield they seek to investigate, the data type they are designed to collect (such as lexical, paradigmatic, narrative, conversational, phrasal/clausal), and the medium through which collection occurs (visual-based vs writing-based). The medium, with a broad division between written and visual Questionnaires, focuses on the material form of the Questionnaires. These two types are generally considered so different that there are, to my knowledge, no published taxonomies that take into account both types; Existing taxonomies tend to focus on one kind or another: Mosel (2012) proposes a taxonomy of written questionnaires, with categories for translational questionnaires, scenario questionnaires, and grammatical structure questionnaires; taxonomies of (visual) stimuli can be quite detailed (Lüpke 2009; Hellwig 2006; Majid 2012) but tend to refer to non-stimuli with expressions such as 'traditional elicitation'.

The archive was designed with two features that seem important from an epistemological and historical perspective: it acknowledges the dynamic nature of Questionnaires, and it makes it possible to associate a number of additional descriptions and documents to the Questionnaires (either when they are uploaded or subsequently). Regarding the first, Questionnaires can be and of course often are used to investigate what they were

designed for, but they are also frequently adapted by users to reflect different linguistic or cultural contexts from that intended for the original questionnaire, theoretical evolutions of the field, and even a different topic of investigation (in cases where a linguist draws inspiration from the questionnaire protocol, methodology, or design in order to study something different). They are thus tools that are far from stable, and the archive was designed to make it possible to mark the connection between revised or adapted Questionnaires and their source Questionnaires, linking them together and documenting the changes they had undergone. This can be done in two ways:

a) If the linguist creating an adaptation of an original Questionnaire feels that the new Questionnaire represents something independent enough to deserve its own unique file in the archive, it can be entered as a new Questionnaire. In this case, instead of appearing within the file for the original Questionnaire, it will have its own independent file – with metadata, categories, goals, development context, and protocol – but it can be linked to the original Questionnaire that inspired it by entering the original's in the metadata formula;

b) If on the other hand the distance between the two Questionnaires is considered to be closer, then one can opt to set up the revised Questionnaire as an adaptation of an original by using the Revision file. The Revision file will allow the inputter to assign a new title to the adaptation, to select the original via a drop-down menu listing all Questionnaires in the archive, to complete a text field listing the changes that were made, and to attach the adapted document, along with a revision date. The Revision to the Questionnaire will not have its own independent file in the archive, but rather will appear on the file for the original Questionnaire, along with the new title, the attached document, and associated with another document listing the changes that were made and the date.

Another important aspect of Questionnaires, and of all methodological tools, is that they are contextually anchored: they reflect the state-of-the-art of linguistic knowledge at the time of their development, and their historical context is crucial to understanding their impact and effectiveness. The archive thus provides a space to associate information of a historical nature with each Questionnaire: the main Questionnaire file contains text boxes with information about the goals the Questionnaire was designed to serve, a summary of its protocol, and the development context. Additionally, reviews and documents retracing the history of the

Questionnaire can be associated with a Questionnaire, so that the tools are accompanied by information of various types that serves to contextualize them.¹¹⁷

Another output of the project was an edited volume (Lahaussais & Vuillermet 2019) resulting from a 2-day workshop on elicitation tools for linguistic description and typology held in Paris in November 2017. The volume, entitled *Methodological Tools for Linguistic Description and Typology*,¹¹⁸ showcased a number of inventive Questionnaires, in addition to an introductory article by Birgit Hellwig on "Language diversity, language documentation and psycholinguistics: the role of stimuli" (Hellwig 2019). The Questionnaires discussed covered the vast methodological and topical range from Paperno & Ryzhova (2019)'s "Automatic construction of lexical typological Questionnaires" to François (2019)'s "Proposal for conversational questionnaires", which describes a Questionnaire for eliciting chunks of dialogue, which can quickly be built up into a conversational parallel corpus, via Lovick & Tuttle (2019)'s "Video elicitation of negative directives in Alaskan Dene languages: reflections on methodology", in which the authors describe using videos showing the violation of social norms which they assembled specifically for the collection of negative imperatives and which additionally yielded a number of off-record responses, including a number of unusual linguistic structures.

While the Fédération TUL has unfortunately been closed, the TULQuest project continues, with the addition of new Questionnaires as they are sent in, and the development of a document describing best practices for Questionnaire use and design (with the core group of "TULQuesters", namely Sylvain Loiseau, Claire Moyse, Stéphane Robert, Alice Vittrant, Sylvie Voisin, and Marine Vuillermet).

3.3.3 Khaling verb dictionary

The last contribution to methodological tools I shall mention here is that Khaling verb dictionary (Jacques et al. 2015), developed with Guillaume Jacques, Dhana Khaling and the

¹¹⁷ The archive was also a fun project in which to involve student interns: few MA students, in my experience, handle or even think about linguistic questionnaires, and having to search for and then categorize new questionnaires was a useful exercise for two past interns. They were also involved in writing to the developers of the questionnaires--often famous linguists--to ask about IP rights, something which not only provided them with contacts, but also brought home the complicated nature of IP and attribution when something is developed as a dynamic tool which is meant to be adapted to different situations.

¹¹⁸ The entire volume is available freely at <http://nflrc.hawaii.edu/ldc/sp16/>

Khaling cultural association, and which exists in a number of different versions: an online version, in which sound recordings are associated with each entry,¹¹⁹ PDF versions (downloadable from the same website as the online version) with and without incorporated sound recordings, and a paper publication (printed in Nepal in 2017). While noun dictionaries do not pose particular problems for Kiranti languages (apart from transcription issues which occur when transcribing into devanagari, with difficulties transcribing certain vowels, vowel length and, in the case of Khaling, a two-way tone distinction), dictionaries of verbs are more complicated: the alternating stems found in Khaling require some lexicographical decisions which are rather technical. In "native" dictionaries, it is often the infinitive form of verbs which is listed as the lexical entry, a form which does not contain sufficient stem information to allow conjugation of the verb. Michailovsky (1975: 187) has a comment to this effect: "Some recent word-lists of languages related to Bahing have suffered from the defect that the verb form chosen for quotation has been one from which the root could not be recovered". In the online Khaling dictionary edited by the Tobas (Toba & Toba 2016) as well, verbs are listed by their infinitive forms (with the 1SG present form sometimes given in parentheses), as in Figure 32 below:

अम्दोना [apdonä] ▸ सुत्तु sleep
 अम्दोना [amdonä] (pres. अम्दोडा) ▸ निदाउनु fall asleep ex. अम्दोनाडा खानुतलो मारि भुङ्क केल्ता! When I was just about to fall asleep the wind blew very hard.

Figure 32. Lexical entries for the verbs 'sleep' and 'fall asleep', with the 1SG present form in parentheses for the second line; (Toba & Toba 2016)

The insufficiency of information about stems associated with a given verb suggest that the dictionary is intended for an audience with high levels of Khaling proficiency.

The Jacques et al. (2015) Khaling dictionary, which has an accompanying Android app developed by Remy Bonnet, also lists verbs by infinitive form, in order that a user may find them, but also includes the root form in parentheses after the infinitive form (the root is based on internal reconstruction; see Jacques et al. (2012)). Examples are provided to illustrate the verb, whenever possible with finite forms and varying person scenarios. The verb forms for 1SG (labeled 1s) and 1DI (labeled 1d) non-past are also given (illustrating

¹¹⁹The dictionary is available at <http://himalco.huma-num.fr/dictionaries/SelectDictionary.php?dict=khaling>

vowel alternations, when present), as they are significant for providing information about the root.

In the online version of the dictionary, sound recordings are provided for the infinitive, the 1SG and 1DI forms, and examples.

बेःन्य *bê:ne* (*bekt*) vt. pour in. हाल्लु. ¶ *sāṅku-bi lēmpɛ bektu*. साङ्कुबि लेम्प्य बेक्तु । I pour sugar in the tea. चियामा चिनी हाल्लु । ¶ *ʔuŋʌ toktse-bi ku bektu*. उड तोक्चेबि कु बेक्त । I poured water in the bamboo pitcher. मैले बाँसको जगमा पानी हाल्ले । 1s: *bektu* बेक्तु 1d: *beki* बेकि
See: 44 158

Figure 33. Sample entry from Khaling verb dictionary

Additionally, the entry is accompanied by a number referring to the relevant verb paradigm (transcribed into both devanagari and IPA). An introduction explains how to use the different verb forms provided and paradigms.

उड इन्	रेःन्य	रेःन्त्यनि
उड एऽचि	रेःन्सु	रेःन्त्यान्सु
उड एःन्	रेःन्नु	रेःन्त्यान्नु

Figure 34. 1SG>2 section of the devanagari-transcribed paradigm for *rê:nɛ* (paradigm 44), used as the sample paradigm for *bê:ne*, showing the case-marked pronouns 'describing' the scenarios 1SG.ERG 2SG; 1SG.ERG 2DU; 1SG.ERG 2PL

The IPA-transcribed paradigm (158) accompanying the entry for *bê:ne* is shown in Figure 35, and is laid out in a 'list-style' format (see section 3.1.2 above), the indications of the person/number/clusivity combinations in the various scenarios being presented in a more technical form than in the devanagari version of the paradigm. The combination of the two paradigms makes the verbal data available to both a Khaling audience and to Western scholars.

Table 158: . Transitive verb *rekt* "write"

	non-past	past	imperative
1s>3s	<i>rektu</i>	<i>rektA</i>	
1s>3d	<i>rektusu</i>	<i>rektAsu</i>	
1s>3p	<i>rektunu</i>	<i>rektAnu</i>	
1di>3	<i>reki</i>	<i>rekiti</i>	
1de>3	<i>reku</i>	<i>reku</i>	
1pi>3	<i>rekki</i>	<i>rektiki</i>	
1pe>3	<i>rekka</i>	<i>rektaka</i>	
2s>3s	<i>?irektu</i>	<i>?irekte</i>	<i>rekte</i>
2s>3d	<i>?ireksu</i>	<i>?irektesu</i>	
2s>3p	<i>?ireknu</i>	<i>?irektenu</i>	
2d>3	<i>?ireki</i>	<i>?irekiti</i>	<i>rekije</i>
2n>3	<i>?iré:ni</i>	<i>?irektenu</i>	<i>reknuje</i>
3s>3s	<i>rektu</i>	<i>rekte</i>	
3>3(d)	<i>reksu</i>	<i>rektesu</i>	
3>3(p)	<i>reknu</i>	<i>rektenu</i>	
2/3s>1s	<i>?iré:ηA</i>	<i>?irekA</i>	<i>rekAje</i>
2/3d>1s	<i>?iré:ηAsu</i>	<i>?irekAsu</i>	<i>rekAsuje</i>
2/3p>1s	<i>?iré:ηAnu</i>	<i>?irekAnu</i>	<i>rekAnuje</i>
2/3>1di	<i>?ireki</i>	<i>?irekiti</i>	
2/3>1de	<i>?ireku</i>	<i>?irekutu</i>	<i>rekuje</i>
2/3>1pi	<i>?irekki</i>	<i>?irektiki</i>	
2/3>1pe	<i>?irekka</i>	<i>?irektaka</i>	<i>rekkaje</i>
3>2s, 1d/pe>2s	<i>?iré:</i>	<i>?irekte</i>	
3>2d, 1d/pe>2d	<i>?ireki</i>	<i>?irekiti</i>	
3>2p, 1d/pe>2p	<i>?iré:ni</i>	<i>?irektenu</i>	
1s>2s	<i>ré:ne</i>	<i>rênteni</i>	
1s>2d	<i>rênsu</i>	<i>rêntensu</i>	
1s>2p	<i>rênnu</i>	<i>rêntennu</i>	

Figure 35. List-style paradigm, transcribed in IPA

Being part of such a dictionary project¹²⁰ has brought many epistemological issues to the fore for me, notably the search for a (delicate) balance between user-friendliness and scientific accuracy (Lahaussais (2018a) discusses this), the importance of including recorded examples in dictionaries of endangered languages, and the methodological challenges of collecting the material for the preparation of such a dictionary (with subtle semantic differences that are difficult to take into account when working on languages which are endangered enough that speakers can sometimes be considered to be speaking an idiolect).

¹²⁰ My participation is in the form of collecting illustrative examples for verbs, recording them, writing the introduction to the paper version, and contributing to our team's understanding of the verbal morphology of Khaling (Jacques et al. 2012).

4 FUTURE DIRECTIONS

The following section describes the large-scale projects I plan to undertake over the next few years: two are writing projects, while the third is a database project.

The first is a grammar of Thulung, but of an unusual type, in that I plan for it to be a *contrastive* grammar. One issue in working on a Kiranti language is that it is not always clear what one is working on: a significant number of constructions are calqued on Nepali, to the point that I sometimes feel that I am working on the description of an Indo-Aryan language with a Kiranti lexicon and verbal inflection. While I have not seen this situation put in writing in colleagues' work on other Kiranti languages, and remains very much subjective for now, it is something I have been very much aware of in working on both Thulung and Koyi (considerably less so in Khaling, perhaps because of its greater vitality). In order to investigate this problem, I intend to write a grammar with face-to-face grammatical descriptions of Nepali and Thulung, matching up the description of the same or similar phenomena. The focus will of course be the production of a grammar of Thulung, but it will be significantly informed through having various topics, where relevant, contrasted with how the same topic manifests in Nepali. Areas where this is significant are, among others, ideophones, compound verbs and derivational predicates, the pronominal system. It seems that this type of grammar may bring to the fore similarities and contrasts that were perhaps not made explicit until now, and help make headway towards the question of the features of Thulung syntax which have been calqued from Nepali, or show other evidence of language contact. The similarities between the contrastive grammar project and other projects described above and based on the idea of *aligning* data will not have escaped the reader: in the same way that aligning data in the Kiranti Comparable Corpus has made it possible to identify some interesting commonalities and dissimilarities between the languages, and in the same way that I have attempted to align the tables of contents of grammars in order to see where the authors treated different topics, I feel like the study of the morphosyntax of an individual language can beneficially be illuminated by aligning it with the expression in another (in the case of Nepali, the main contact language) of certain functional domains.

The second writing project, hinted at by the studies described in 3.2, is a book attempting to retrace the stages of the history of interlinear glossing, in order to understand the steps that have led to the current layout. The starting point will be the studies described in section 3.2.2 on Annotation practices, which must be substantially expanded in order to develop

a perspective which encompasses a greater number of language areas. Considering the differences I have noticed in the treatment of examples within descriptions and text specimens, even within the same descriptive document, I feel that the best type of material for the purposes of tracking the evolution of glossing practices is the text specimen, either those accompanying a description or those which are deposited in online oral archives. Using the analytical grid I have proposed (see Figure 25), and focusing on linguistic surveys which include grammatical descriptions and text specimens, which I consider, as suggested above, to be important vectors for change in annotation practices (in addition to having clear, volume-wide editorial practices, and thus limiting typographical confusion coming from different practices by typesetters in different journals), I hope to carefully retrace the general trends of the last 150 years, a period in which annotation, and significantly, morphological theory, underwent great changes. The study of these evolutions must take into account all aspects (and perhaps additional ones as well) of the analytical grid I have proposed, namely the presence and type of segmentation marks in both the transcription and glossing tiers, as well as questions about the type of glosses (which includes the complicated issue of glossing which can take on the form of literal translation, thus confusing the distinction between the material found in the glossing and the translation tiers).

Another planned project is an offshoot of work described in section 3.1.3 above on the grammaticographical models underlying language descriptions. Although the project described in (Kelly & Lahaussais in press) is based on the methodology and effects of aligning the tables of contents of the grammars in the corpus, there is a very interesting project that stops one step short of alignment: this involves developing a very large database of tables of contents of published descriptive grammars, which, through a specialized interface, can be explored according to biographical and bibliographical features (similar to those described in 3.1.3 above and in Kelly & Lahaussais (in press): date, metalanguage, language area, nationality of linguist, training of linguist, etc. It seems quite clear that there are areal traditions of description which impact not only the structuring of grammars but also terminological choices (see, for example, Vanhove (2020)), and many patterns may emerge from an interface that makes it possible to look at grammars and their organizational schemes with respect to sort criteria of the type described above.

Another advantage of this project is that it will result in a database of tables of contents of grammars large enough to make credible claims about the various organizational schemes that are used, and to create a taxonomy of these schemes. This project will be of significant relevance to typologists making use of grammatical descriptions, as well as to historians of

linguistics curious about tracking various traditional grammatical models through to their adaptations in contemporary description grammars.

One disappointment of mine is that I had hoped that writing this thesis would be an opportunity to take into account work by theoretical morphologists that is relevant to the description of Thulung data (in the vein in if Bickel and associates, who manage to do this very well for Chintang; see Mansfield, Stoll & Bickel (in press) as an example). The rich environment at Université de Paris (thanks to the presence of O. Bonami, B. Fradin, B. Cysmann, and a slew of invited professors), and in the EFL Labex projects (specifically Axe 3's OR "Typology of inflectional systems with non-canonical inflection" led by E. Palancar and S. Fedden) in this respect make Paris a remarkable place in which to gain exposure to current morphological theory. Instead of presenting the data herein within the framework of theoretically informed morphology, I have, as a first step, taken this opportunity to describe my Thulung data in a way that is relevant to morphologists. I have attempted to present data (especially on verb roots and stem alternations, on indexes, and on derivational morphology, including multiple exponence) in a way that I hope makes it possible to begin a dialog with these actors, and to bring the rich verbal morphology of Thulung into the limelight.

In conclusion, although one might be surprised to find the different parts of this thesis brought together into a single document, Chapters 1 and 2 being very much works of description based on fieldwork, and Chapter 3 documenting a series of studies and tools based on the epistemology, and in some cases on the history, of descriptive linguistics, the possibilities opened up for me by working in both these areas over the last decade make me very enthusiastic about the next decade of my career, and very grateful to have been given the change to bring together these two different fields within linguistics.

Bibliography

- Aikhenvald, Alexandra. 2015. *The art of grammar: a practical guide*. Oxford: Oxford University Press.
- Allen, N. J. 1975. *Sketch of Thulung Grammar* (Cornell East Asia Papers). Ithaca, NY: China-Japan Program, Cornell University.
- Allen, N. J. 2012. *Miyapma: Traditional Narratives of the Thulung Rai*. Kathmandu: Vajra Publication.
- Aronoff, Mark. 1994. *Morphology by itself: stems and inflectional classes* (Linguistic Inquiry Monographs 22). Cambridge: MIT Press.
- Auroux, Sylvain. 1992. Introduction. Le processus de grammatisation et ses enjeux. In Sylvain Auroux (ed.), *Histoire des idées linguistiques, tome 2 (Le développement de la grammaire occidentale)*, Liège, Mardaga, p. 11-64., 11–64. Liège, Belgique: Mardaga.
- Auroux, Sylvain. 1994. *La révolution technologique de la grammatisation: introduction à l'histoire des sciences du langage*. Liège, Belgique: Mardaga.
- Aussant, Emilie & Aimée Lahaussais. 2020. “Grammaires étendues” et description de morphologie verbale. *Faits de Langues* 50(2). 9–12.
- Baerman, Matthew & Greville Corbett. 2012. Stem alternations and multiple exponence. In Olivier Bonami (ed.), *Stems (Special issue of Word Structure 5.1)*, 52–68. Edinburgh: Edinburgh University Press.
- Benedict, Paul. 1972. *Sino-Tibetan: a conspectus* (Princeton-Cambridge Studies in Chinese Linguistics II). Cambridge: Cambridge University Press.
- Bickel, Balthasar. 1996. *Aspect, Mood, and Time in Belhare* (Arbeiten Des Seminars Für Allgemeine Sprachwissenschaft 15). Zürich: Universität Zürich.
- Bickel, Balthasar. 1999. Nominalization and focus constructions in some Kiranti languages. In Yogendra P. Yadava & Warren W. Glover (eds.), *Topics in Nepalese Linguistics*, 271–296. Kathmandu: Royal Nepal Academy.
- Bickel, Balthasar. 2003a. Referential density in discourse and syntactic typology. *Language* 79(4). 708–736.
- Bickel, Balthasar. 2003b. Belhare. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan Languages* (Routledge Language Family Series 3), 546–570. London: Routledge.
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar, Goma Banjade, Martin Gaenszle, Elena Lieven, Netra Prasad Paudyal, Ichchha Purna Rai, Manoj Rai, Novel Kishore Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83(1). 43–73.
- Bonami, Olivier. 2012. Introduction. In Olivier Bonami (ed.), *Stems (Special issue of Word Structure 5.1)*, 1–6. Edinburgh: Edinburgh University Press.
- Borchers, Dörte. 2008. *A grammar of Sunwar: descriptive grammar, paradigms, texts and glossary* (Languages of the Greater Himalayan Region 5/7). Leiden: Brill.
- Bouquiaux, Luc & Jacqueline M.C. Thomas (eds.). 1976a. *Enquête et description des langues à tradition orale. L'enquête de terrain et l'analyse grammaticale*. Vol. 1. Paris: SELAF.
- Bouquiaux, Luc & Jacqueline M.C. Thomas (eds.). 1976b. *Enquête et description des langues à tradition orale. Approche linguistique*. Vol. 2. Paris: SELAF.
- Bouquiaux, Luc & Jacqueline M.C. Thomas (eds.). 1976c. *Enquête et description des langues à tradition orale. Approche thématique*. Vol. 3. Paris: SELAF.

- Bow, C, B Hughes & S Bird. 2003. Towards a general model of interlinear text. In *Proceedings of EMELD Workshop*, 1–47. <http://emeld.org/workshop/2003/bowbadenbird-paper.pdf> (21 January, 2015).
- Bowern, Claire. 2008. *Linguistic Fieldwork: A Practical Guide*. New York: Palgrave MacMillan.
- Bradley, David. 1997. Tibeto-Burman languages and classification. In David Bradley (ed.), *Tibeto-Burman languages of the Himalayas* (Papers in Southeast Asian Linguistics), 1–72. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Butt, Miriam. 1995. *The structure of complex predicates in Urdu*. Stanford: Center for the Study of Language and Information.
- Bybee, Joan. 1998. “Irrealis” as a Grammatical Category. *Anthropological Linguistics* 40(2). 257–271.
- 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.
- Campbell, George. 1874. *Specimens of Languages of India, including those of the Aboriginal Tribes of Bengal, the Central provinces, and the Eastern Frontier*. Calcutta: Bengal Secretariat Press.
- Chafe, Wallace L. (ed.). 1980. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Chappell, Hilary. 2002. Language contact and areal diffusion in Sinitic languages. In Alexandra Aikhenvald & R.M.W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 328–357. Oxford: Oxford University Press.
- Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages. *Linguistic Typology* 12. 45–98.
- Chelliah, Shobhanna & Willem de Reuse. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.
- Cinato, Franck. 2017. Quelle linguistique dans les gloses du haut Moyen-Âge ? Les gloses de lecture. In Paris, France.
- Colombat, Bernard. 2019a. Accidens. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours* (Orbis/Supplementa 46), 49–54. Leuven: Peeters.
- Colombat, Bernard. 2019b. Accident. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours* (Orbis/Supplementa 46), 54–56. Leuven: Peeters.
- Colombat, Bernard & Aimée Lahaussais (eds.). 2019a. *Histoire des parties du discours* (Orbis/Supplementa 46). Leuven: Peeters.
- Colombat, Bernard & Aimée Lahaussais. 2019b. Les parties du discours ou classes de mots. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours* (Orbis/Supplementa 46), 24–45. Leuven: Peeters.
- Comrie, Bernard & Norval Smith. 1977. Lingua descriptive studies: questionnaire. *Lingua* 42(1). 1–72.
- Corbett, Greville. 2015. Morphosyntactic complexity: a typology of lexical splits. *Language* 91(1). 145–193.
- Creissels, Denis. 2006. *Syntaxe générale, une introduction typologique*. Paris: Hermès-Lavoisier.
- Creissels, Denis. 2009. Uncommon patterns of core term marking and case terminology. *Lingua* 119. 445–459.
- Cristofaro, Sonia & Fernando Zúñiga. 2018. Synchronic vs. diachronic approaches to typological hierarchies. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological*

- hierarchies in synchrony and diachrony* (Typological Studies in Language 121), 3–27. Amsterdam/Philadelphia: John Benjamins.
- Crowley, Terry. 2007. *Field Linguistics, A Beginner's Guide*. Oxford: Oxford University Press.
- Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF - Language Typology and Universals* 60(2). 95–99.
- Dahal, Ballabh Mani. 1999. Phonesthetic elements in Nepali. In Yogendra P. Yadava & Warren W. Glover (eds.), *Topics in Nepalese Linguistics*, 90–102. Kathmandu: Royal Nepal Academy.
- Dahl, Östen. 2000. *Tense and aspect in the languages of Europe* (Empirical Approaches to Language Typology 20–6). Berlin & New York: Mouton de Gruyter.
- DeLancey, Scott. 1985. The analysis-lexis-synthesis cycle in Tibeto-Burman: a case study in motivated change. In John Haiman (ed.), *Iconicity in Syntax* (Typological Studies in Language 6), 367–389. Amsterdam/Philadelphia: John Benjamins.
- DeLancey, Scott. 1989. Verb Agreement in Proto-Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 52(2). 315–333.
- DeLancey, Scott. 2010. Towards a History of Verb Agreement in Tibeto-Burman. *Himalayan Linguistics* 9(1). 1–39.
- DeLancey, Scott. 2018a. Deictic and sociopragmatic effects in Tibeto-Burman SAP indexation. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony* (Typological Studies in Language 121), 345–375. Amsterdam/Philadelphia: John Benjamins.
- DeLancey, Scott. 2018b. Deictic and sociopragmatic effects in Tibeto-Burman SAP indexation. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony* (Typological Studies in Language 121), 345–375. Amsterdam/Philadelphia: John Benjamins.
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Nijmegen: Radboud University.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Doornenbal, Marius. 2009. *A grammar of Bantawa. Grammar, paradigm tables, glossary and texts of a Rai language of Eastern Nepal*. Utrecht: LOT.
- Driem, George L. van. 1988. Reflexes of the Tibeto-Burman *-t directive suffix in Dumi Rai. In David Bradley, Eugenie J. A. Henderson & Martine Mazaudon (eds.), *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*, 157–167. Canberra: Pacific Linguistics.
- Driem, George L. van. 2001. *Languages of the Himalayas. An Ethnolinguistic Handbook of the Greater Himalayan Region*. 2 volumes vols. Leiden, Boston, Köln: Brill.
- 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.
- Driem, George L. van. 2014. Trans-Himalayan. In Thomas Owen-Smith & Nathan W. Hill (eds.), *Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area*, 11–40. Berlin: de Gruyter.
- Driem, George L. van. 2018. Linguistic history and historical linguistics. *Linguistics of the Tibeto-Burman Area* 41(1). 106–127. <https://doi.org/10.1075/ltba.18005.dri>.
- Driem, George van. 1987. *A grammar of Limbu* (Mouton Grammar Library 4). Berlin, New York, Amsterdam: Mouton de Gruyter.

- Driem, George van. 1990. An exploration of Proto-Kiranti verbal morphology. *Acta Linguistica Hafniensia* 22. 27–48.
- Driem, George van. 1993a. The Proto-Tibeto-Burman verbal agreement system. *Bulletin of the School of Oriental and African Studies* 56(2). 292–334.
- Driem, George van. 1993b. *A grammar of Dumi* (Mouton Grammar Library 10). Berlin & New York: Mouton de Gruyter.
- Dryer, Matthew S. 1986. Primary objects, secondary objects. and antititive. *Language* 62(4). 808–845.
- Ebert, Karen. 1994. *The Structure of Kiranti Languages* (Arbeiten Des Seminars Für Allgemeine Sprachwissenschaft 13). Zürich: Universität Zürich.
- Ebert, Karen. 1997a. *A Grammar of Athpare* (Lincom Studies in Asian Linguistics 1). Munich: Lincom Europa.
- Ebert, Karen. 1997b. *Camling (Chamling)* (Languages of the World/Materials 103). Munich: Lincom Europa.
- Ebert, Karen. 2003. Kiranti languages: an overview. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan Languages*, 505–517. London and New York: Routledge.
- Ebert, Karen & Martin Gaenszle. 2009. *Rai mythology: Kiranti Oral Texts* (Harvard Oriental Series 69). Cambridge, MA: Harvard University Press.
- Eliot, John. 1666. *The Indian grammar begun: or, An essay to bring the Indian language into rules, for the help of such as desire to learn the same, for the furtherance of the Gospel among them*. Cambridge [Mass.]: Printed by Marmaduke Johnson. <http://archive.org/details/indiagrammarbeg00elio> (25 August, 2017).
- Elliott, Jennifer. 2000. Realis and Irrealis: Forme and concepts of the grammaticalisation of reality. *Linguistic Typology* 4. 55–90.
- Evans, Nicholas & Alan Dench. 2006. Introduction: Catching language. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: the Standing Challenge of Grammar Writing* (Trends in Linguistics 167), 1–40. Berlin & New York: Mouton de Gruyter.
- Everett, Daniel. 2009. *Don't sleep, there are snakes: Life and language in the Amazonian Jungle*. New York: Vintage Departures.
- 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.
- Gaenszle, Martin. 1991. *Verwandschaft und Mythologie bei den Mewahang Rai in Ostnepal: eine Ethnographische Studie zum Problem zur "ethnischen Identität."* Stuttgart: Steiner Verlag Wiesbaden.
- Genetti, Carol. 1994. *Aspects of Nepali grammar* (Santa Barbara Papers in Linguistics 6). Santa Barbara: Linguistics Department, University of California, Santa Barbara.
- Genetti, Carol. 2007. *A Grammar of Dolakha Newar*. () Berlin: Mouton de Gruyter. (Mouton Grammar Library 40). Berlin: Mouton de Gruyter.
- Genetti, Carol, Ellen Bartee, Alec Coupe, Kristine Hildebrandt & You-Jing Lin. 2008. Syntactic aspects of nominalization in five Tibeto-Burman languages of the Himalayan area. *Linguistics of the Tibeto-Burman Area* 31(2). 97–144.
- Gerber, Pascal & Selin Grollman. 2018. What is Kiranti? *Bulletin of Chinese Linguistics* 11. 99–152.

- Good, Jeff. 2011. The typology of templates. *Language and Linguistics Compass* 5(10). 731–747.
- Good, Jeff. 2016. *The linguistic typology of templates*. Cambridge: Cambridge University Press.
- Grierson, G.A. 1897. *Specimen Translations in Various Indian Languages*. Calcutta: Superintendent, Government Printing, India.
- Grierson, G.A. (ed.). 1909. *Linguistic Survey of India Vol. III.1 Tibeto-Burman Family: General Introduction, Specimen of the Tibetan Dialects, The Himalayan Dialects, and the North Assam Group*. Vol. 3.1. 11 vols. Delhi, Varanasi, Patna: Motilal Banarsidass.
- Guillaume, Antoine. 2016. Associated Motion in South America: Typological and areal perspectives. *Linguistic Typology* 20(1). 81–177.
- Haiman, John. 2018. *Ideophones and the Evolution of Language*. Cambridge: Cambridge University Press.
- Harrington, John P. 1946. Three Kiowa Texts. *International Journal of American Linguistics* 12(4). 237–242.
- Harris, Zellig. 1942. Morpheme alternants in linguistic analysis. *Language* 18(3). 169–180.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Haspelmath, Martin. 2019. Comparing reflexive constructions in the world’s languages.
- Haude, Katharina & Alena Witzlack-Marakevich. 2016. Referential hierarchies and alignment: An overview. *Linguistics* 54(3). 433–441.
- Hellwig, Birgit. 2006. Field semantics and grammar-writing: Stimuli-based techniques and the study of locative verbs. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: the standing challenge of grammar writing* (Trends in Linguistics 167), 321–358. Berlin & New York: Mouton de Gruyter.
- Hellwig, Birgit. 2019. Language diversity, language documentation and psycholinguistics: the role of stimuli. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological Tools for Linguistic Description and Typology* (Language Documentation & Conservation Special Publications 16), 5–30. Honolulu: University of Hawai’i Press.
- Herce, Borja. 2020. *A typological approach to the morpheme*. University of Surrey & University of the Basque Country.
- Hinton, Leanne, Johanna Nichols & John Ohala (eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Hockett, Charles. 1954. Two models of grammatical description. *Word* 10. 210–234.
- 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. *Journal of the Asiatic Society of Bengal* 26. 317–522.
- Hodgson, Brian H. 1857c. 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 (contined from vol 26) [Grammar of the Bahing tribe]. *Journal of the Asiatic Society of Bengal* 27. 393-456 [393–442].
- Hook, Peter. 1974. *The compound verb in Hindi*. University of Michigan, Center for South and Southeast Asian Studies.
- Hovdhaugen, Even. 1996. Missionary Grammars--an Attempt at Defining a Field of Research. In Even Hovdhaugen (ed.), *--and the Word was God: missionary linguistics*

- and missionary grammar* (Studium Sprachwissenschaft 25), 9–22. Münster: Nodus Publikationen.
- Hunter, W. W. 1868. *A comparative dictionary of the Languages of India and High Asia with a dissertation, based on the Hodgson lists, official records, and MSS*. London: Trübner and Co. <https://archive.org/details/cu31924073426532/page/n3>.
- Jacobson, Michel, Boyd Michailovsky & J.B. Lowe. 2001. Linguistic documents synchronizing sound and text. *Speech communication* 33. 79–96.
- Jacques, Guillaume. 2013. Harmonization and disharmonization of affix ordering and basic word order. *Linguistic Typology* 17(2). 187–215.
- Jacques, Guillaume. 2015. Derivational verbal morphology in Khaling. *Bulletin of Chinese Linguistics* 8. 78–85.
- Jacques, Guillaume. 2017. A reconstruction of Proto-Kiranti verb roots. *Folia Linguistica* 58. 177–215.
- Jacques, Guillaume. 2018. Bipartite verbs in Japhug and other Trans-Himalayan languages. *Linguistics of the Tibeto-Burman Area* 41(2). 175–191.
- Jacques, Guillaume & Anton Antonov. 2014. Direct/Inverse systems. *Language and Linguistics Compass* 8(7). 301–318.
- Jacques, Guillaume & Aimée Lahaussais. 2018. Ambulatives in Kiranti. In. Université Paris Diderot, Paris.
- Jacques, Guillaume, Aimée Lahaussais, Boyd Michailovsky & Dhan Bahadur Rai. 2012. An Overview of Khaling Verbal Morphology. *Language and Linguistics* 13(6). 1095–1170.
- Jacques, Guillaume, Aimée Lahaussais & Dhan Bahadur Rai. 2016. Reflexive paradigms in Khaling. *Linguistics of the Tibeto-Burman Area* 39(1). 33–48. <https://doi.org/10.1075/ltba.39.1.02jac>.
- Jacques, Guillaume, Aimée Lahaussais, Dhan Bahadur Rai & Yadav Kumar Rai. 2015. *Khaling-Nepali-English verb dictionary, Version 1.0*. http://himalco.humanum.fr/dictionaries/khaling/dictionary_mp3.pdf.
- Jacques, Guillaume, Aimée Lahaussais & Shuya Zhang. In press. Associated motion in Sino-Tibetan, with a focus on Gyalrongic and Kiranti. In Antoine Guillaume & Harold Koch (eds.), *Associated Motion as a grammatical category in linguistic typology* (Empirical Approaches to Language Typology 64). Mouton de Gruyter.
- Kari, James. 1989. Affix positions and zones in the Athapaskan verb complex: Ahtna and Navajo. *International Journal of American Linguistics* 55(4). 424–454.
- Keenan, Edward & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.
- Kelly, Barbara & Aimée Lahaussais. in press. Chains of influence in Himalayan grammars: models and interrelations shaping descriptions of Tibeto-Burman languages of Nepal. *Linguistics*.
- Kilian-Hatz, C. 2001. In F. K. E. Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, 165–182. Amsterdam & Philadelphia: Benjamins.
- Klamer, Marian. 2001. Expressives and iconicity in the lexicon. In F. K. E. Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, 165–182. Amsterdam & Philadelphia: Benjamins.
- Koch, Harold. 1984. The category of “Associated Motion” in Kaytej. *Language in Central Australia* 1. 23–34.
- Konnerth, Linda & Scott DeLancey. 2019. Introduction to “Verb agreement in languages of the Eastern Himalayan region.” *Himalayan Linguistics* 18(1). 1–7.
- Lahaussais, Aimée. In press. Inauspicious events in Thulung. In Marine Vuillermet, Eva Schultze-Berndt & Martina Faller (eds.), *Apprehensional constructions in a cross-linguistic perspective* (Studies in Diversity Linguistics). Language Science Press.

- Lahaussais, Aimée. submitted. Predicate derivations in Thulung.
- Lahaussais, Aimée. Submitted. Reflexive constructions in Thulung. In Katarzyna Janic, Martin Haspelmath & Nicoletta Puddu (eds.), *Reflexive constructions*.
- Lahaussais, Aimée. 2002. *Aspects of the grammar of Thulung Rai: an endangered Himalayan language*. Berkeley: University of California, Berkeley. <https://tel.archives-ouvertes.fr/halshs-00004761/document>.
- Lahaussais, Aimée. 2003a. Ergativity in Thulung Rai: a shift in the position of pronominal split. In David Bradley, Randy LaPolla, Boyd Michailovsky & Graham Thurgood (eds.), *Language Variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 101–112. Canberra: Pacific Linguistics.
- Lahaussais, Aimée. 2003b. Nominalization and its various uses in Thulung Rai. *Linguistics of the Tibeto-Burman Area* 26(1). 33–57.
- Lahaussais, Aimée. 2004. Thulung Rai. *Himalayan Linguistics Archive* No. 1. 1–25.
- Lahaussais, Aimée. 2006. *Handbook on language preservation and documentation, Languages of Nepal*. UNESCO.
- Lahaussais, Aimée. 2011. The Thulung Rai verbal system: an account of verb stem alternation. *Cahiers de Linguistique Asie Orientale* 40(2). 189–224.
- Lahaussais, Aimée. 2013. Irrealis in the Kiranti languages. In Paris, France.
- Lahaussais, Aimée. 2015. The Kiranti Comparable corpus: a prototype corpus for the comparison of Kiranti languages and mythology. In Mari Jones (ed.), *Endangered Languages and New Technologies*, 17–34. Cambridge: Cambridge University Press.
- Lahaussais, Aimée. 2016a. Reflexive derivations in Thulung. *Linguistics of the Tibeto-Burman Area* 39(1). 49–66. <https://doi.org/10.1075/ltba.39.1.03lah>.
- Lahaussais, Aimée. 2016b. Where have all the interjections gone? A look into the place of interjections in contemporary descriptions of endangered languages. In Carlos Assunção, Gonçalo Fernandes & Rolf Kemmler (eds.), *Tradition and Innovation in the History of Linguistics*, 186–195. Münster: Nodus Publikationen.
- Lahaussais, Aimée. 2016c. 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. 2017a. Relativization strategies and alignment in Thulung Rai. *Cahiers de Linguistique Asie Orientale* 46(1). 73–100. <https://doi.org/10.1163/19606028-04601003>.
- Lahaussais, Aimée. 2017b. Ideophones in Khaling Rai. *Linguistics of the Tibeto-Burman Area* 40(2). 179–201. <https://doi.org/10.1075/ltba.17005.lah>.
- Lahaussais, Aimée. 2017c. Glosses for argument roles in descriptions of languages with polypersonal verbal indexation. In Paris: Université Sorbonne-Nouvelle.
- Lahaussais, Aimée. 2018a. Examining approaches to describing complex verbal morphology for an unwritten language. In Bernard Colombat, Bernard Combettes, Valérie Raby & Gilles Siouffi (eds.), *Histoire des langues et histoire des représentations linguistiques*. Paris: Honoré Champion.
- Lahaussais, Aimée. 2018b. Ambulative constructions in Khaling Rai and Thulung Rai. In Lucknow, India.
- Lahaussais, Aimée. 2019a. Les parties du discours dans les manuels de description linguistique. In Jean-Marie Fournier, Aimée Lahaussais & Valérie Raby (eds.), *Grammaticalia : Hommage à Bernard Colombat* (Langages), 27–34. Lyon: ENS Editions.

- Lahaussais, Aimée. 2019b. Interjection. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours* (Orbis/Supplementa 46), 455–473. Leuven: Peeters.
- Lahaussais, Aimée. 2019c. The emergence of segmentation in descriptions of Tibeto-Burman languages of Nepal. In. Edinburgh: University of Edinburgh.
- Lahaussais, Aimée. 2019d. The TULQuest linguistic questionnaire archive. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological tools for linguistic description and typology* (Language Documentation & Conservation Special Publications 16), 31–44. Honolulu: University of Hawai'i Press.
- Lahaussais, Aimée. 2019e. The morphology of derivational predicates in Thulung. In. Sydney, Australia.
- Lahaussais, Aimée. 2020a. The Shapes of verbal paradigms in Kiranti languages. *Faits de Langues* 50(2). 71–93.
- Lahaussais, Aimée. 2020b. Glossing in the Linguistic Survey of India: Some insights into early 20th century practices. In. Yonezawa, Japan.
- Lahaussais, Aimée & Séverine Guillaume. 2012. A viewing and processing tool for the analysis of a comparable corpus of Kiranti mythology. In *Proceedings of the 5th International Conference on Language Resources and Evaluations: Workshop on Building and Using Comparable Corpora*, 33–41. Istanbul.
- Lahaussais, Aimée & Jacqueline Léon. 2015. Transcription and translation of unwritten languages in American linguistics (1950s to 2000s). In Emilie Aussant (ed.), *La traduction dans l'histoire des idées linguistiques: représentations et pratiques*, 235–258. Paris: Geuthner.
- Lahaussais, Aimée & Jacqueline Léon. 2017. Morpheme as a unit (of transcription and analysis) in American anthropological linguistics. In. London: University of Westminster.
- Lahaussais, Aimée & Yvonne Treis. 2019. Ideophones and interjections: Workshop introduction. In.
- Lahaussais, Aimée & Marine Vuillermet (eds.). 2019. *Methodological tools for linguistic description and typology* (Language Documentation & Conservation Special Publications 16). Honolulu: University of Hawai'i Press.
- Lallot, Jean. 2019. Parapomenon. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours* (Orbis/Supplementa 46), 45–48. Leuven: Peeters.
- Lambert-Brétière, Renée. in press. The Bloomfieldian Heritage in Algonquian linguistics: The verbal complex in Innu. *Faits de Langues* 52.
- LaPolla, Randy. 1992a. On the dating and nature of verb agreement in Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 55(2). 298–315.
- LaPolla, Randy. 1992b. “Anti-ergative” Marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15(1). 1–9.
- LaPolla, Randy. 2003. Overview of Sino-Tibetan morphosyntax. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan Languages* (Routledge Language Family Series 3), 22–43. London: Routledge.
- LaPolla, Randy. 2016. Once again on methodology and argumentation in linguistics Problems with the arguments for recasting Sino-Tibetan as “Trans-Himalayan.” *Linguistics of the Tibeto-Burman Area* 39(2). 282–297.
- 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). Berlin & New York: W. de Gruyter.

- Lovick, Olga & Siri G. Tuttle. 2019. Video elicitation of negative directives in Alaskan Dene languages: reflections on methodology. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological Tools for Linguistic Description and Typology* (Language Documentation & Conservation Special Publications 16), 125–154. Honolulu: University of Hawai'i Press.
- Lüpke, Friederike. 2009. Data collection methods for field-based language documentation. In Peter Austin (ed.), *Language Documentation and Description*, vol. 6, 53–100. London: SOAS.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2004*, 137–175. Dordrecht: Springer.
- Majeed, Javed. 2019a. *Colonialism and knowledge in Grierson's Linguistic Survey of India*. London and New York: Routledge.
- Majeed, Javed. 2019b. The politics of Grierson's Linguistic Survey of India. In: Edinburgh: University of Edinburgh.
- Majid, Asifa. 2012. A guide to stimulus-based elicitation for semantic categories. In Nicholas Thieberger (ed.), *The Oxford Handbook of Linguistic Fieldwork* (Oxford Handbooks in Linguistics), 54–71. Oxford: Oxford University Press.
- Mansfield, John, Sabine Stoll & Balthasar Bickel. in press. Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language*.
- Matić, Dejan & Brigitte Pakendorf. 2013. Non-canonical SAY in Siberia: Areal and genealogical patterns. *Studies in Language* 37(2). 356–412.
- Matisoff, James. 1972. Lahu nominalization, relativization, and genitivization. In John Kimball (ed.), *Syntax and Semantics*, vol. 1, 237–257. New York: Academic Press.
- Matisoff, James. 1978. *Variational Semantics in Tibeto-Burman: the "organic" approach to linguistic comparison* (Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics 6). Philadelphia: Institute for the Study of Human Issues. https://stedt.berkeley.edu/pubs_and_prods/Matisoff_1978_Variational-Semantics-TB.pdf.
- Matisoff, James. 1993. Sangkong of Yunnan: secondary “verb pronominalization” in Southern Loloish. *Linguistics of the Tibeto-Burman Area* 16(2). 123–142.
- Matisoff, James. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.
- Mayer, Mercer. 1969. *Frog, Where Are You*. New York: Dial Books for Young Readers.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- Meakins, Felicity, Jennifer Green & Myfany Turpin. 2018. *Understanding Linguistic Fieldwork* (Understanding Language). London and New York: Routledge.
- Michailovsky, Boyd. 1975. Notes on the Kiranti Verb <East Nepal>. *Linguistics of the Tibeto-Burman Area* 2(2). 183–218.
- Michailovsky, Boyd. 1985. Tibeto-Burman dental suffixes: evidence from Limbu (Nepal). In Graham Thurgood, James Matisoff & David Bradley (eds.), *Linguistics of the Sino-Tibetan area: the state of the art*, 363–375. Canberra: Pacific Linguistics.
- Michailovsky, Boyd. 1988. *La langue hayu* (Sciences Du Langage). Paris: Editions du CNRS.
- Michailovsky, Boyd. 1994. Manner versus place of articulation in the Kiranti initial stops. In *Current Issues in Sino-Tibetan Linguistics*, 766–781. Osaka: The Organizing Committee of the 26th International Conference on Sino-Tibetan Languages and Linguistics.
- Michailovsky, Boyd. 2010. Preliminaries to the comparative study of the Kiranti subgroup of Tibeto-Burman. In *Proceedings of the International Symposium on Sino-Tibetan*

- Comparative Studies in the 21st Century (June 24-25, 2010)*, 145–170. Taipei: Academia Sinica.
- Michailovsky, Boyd. 2012. Internal reconstruction of the Dumi verb: lexical bases and stem formation. *Linguistics of the Tibeto-Burman Area* 35(2). 49–87.
- 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.
- Michaud, Alexis. 2017. *Tone in Yongning Na: lexical tones and morphotonology* (Studies in Diversity Linguistics 13). Berlin: Language Science Press.
- Mosel, Ulrike. 2012. Morphosyntactic Analysis in the Field: A Guide to the Guides. In Nicholas Thieberger (ed.), *The Oxford Handbook of Linguistic Fieldwork*, 72–89. Oxford: Oxford University Press.
- Nida, Eugene. 1949. *Morphology: the descriptive analysis of words*. Ann Arbor, MI: University of Michigan Press.
- Nishi, Yoshio. 1995. A brief survey of the controversy in verb pronominalization in Tibeto-Burman. In Yoshio Nishi, James Matisoff & Yasuhiko Nagano (eds.), *New horizons in Tibeto-Burman morphosyntax* (Senri Ethnological Studies 41), 1–16. Osaka: National Museum of Ethnology.
- Noonan, Michael. 2006. Direct speech as a rhetorical style in Chantyal. *Himalayan Linguistics* 6. 1–32.
- Noonan, Michael, Thomas Payne & David Weber. 2007. Grammar Writing for a grammar-reading audience. In *Perspectives on grammar writing* (Benjamins Current Topics 11), 113–126. Amsterdam/Philadelphia: John Benjamins.
- Opgenort, Jean Robert. 2004. *A grammar of Wambule: Grammar, lexicon, texts and cultural survey of a Kiranti Tribe of Eastern Nepal* (Languages of the Greater Himalayan Region 5/2). Leiden: Brill.
- Opgenort, Jean Robert. 2005. *A grammar of Jero, with a Historical Comparative Study of the Kiranti Languages* (Languages of the Greater Himalayan Region 5/3). Leiden: Brill.
- Pakendorf, Brigitte & Ewa Schalley. 2007. From possibility to prohibition: A rare grammaticalization pathway. *Linguistic Typology* 11. 515–540.
- Paperno, Denis & Daria Ryzhova. 2019. Automatic construction of lexical typological Questionnaires. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological Tools for Linguistic Description and Typology* (Language Documentation & Conservation Special Publications 16), 45–61. Honolulu: University of Hawai'i Press.
- Payne, Thomas. 1997. *Describing morphosyntax: a guide for field linguists*. Cambridge: Cambridge University Press.
- Pike, K. 1962. Dimensions of grammatical constructions. *Language* 38(3). 221–244.
- Post, Mark. 2010. Predicate derivations in the Tani languages: root, suffix, both or neither? In Mark Post & Stephen Morey (eds.), *North East Indian Linguistics*, 175–197. New Delhi, India: Cambridge University Press India.
- Rai, Novel Kishore, Balthasar Bickel, Martin Gaenszle, Elena Lieven, Netra Prasad Paudyal, Ichchha Purna Rai, Manoj Rai & Sabine Stoll. 2005. Triplication and ideophones in Chintang. In Yogendra P. Yadava (ed.), *Current Issues in Nepalese Linguistics*, 205–209. Kirtipur: Linguistic Society of Nepal.
- Rai, Novel Kishore & Werner Winter. 1997. Triplicated verbal adjuncts in Bantawa. In David Bradley (ed.), *Tibeto-Burman languages of the Himalayas*, 119–134. Canberra: Pacific Linguistics.
- Rose, Françoise. 2018. Are the Tupi-Guarani hierarchical indexing systems really motivated by the person hierarchy? In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological*

- hierarchies in synchrony and diachrony* (Typological Studies in Language 121), 289–307. Amsterdam/Philadelphia: John Benjamins.
- Rutgers, Roland. 1998. *Yamphu: Grammar, texts, lexicon* (Languages of the Greater Himalayan Region 2). Leiden: Research School CNWS.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin Ryder, Valentin Thouzeau, Simon J. Greenhill & Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences of the United States of America* 116(21). 10317–10322.
- Samarin, William J. 1967. *Field Linguistics: A Guide to Linguistic Field Work*. New York: Holt, Rinehart and Winston.
- Schackow, Diana. 2015. *A grammar of Yakkha* (Studies in Diversity Linguistics 7). Berlin: Language Science Press.
- Schikowski, Robert, Balthasar Bickel & Netra Prasad Paudyal. 2015. Flexible valency in Chintang. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, 669–708. Berlin: Mouton de Gruyter.
- Schleicher, August. 1859. *Zur Morphologie der Sprache*. Vol. I.7. St. Petersburg: Mémoires de l'Académie impériale des Sciences de Saint-Pétersbourg.
- Schleicher, August. 1861. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: H. Böhlau.
- Schlemmer, Grégoire. 2019. Following the ancestors and managing the otherness. In Marine Carrin (ed.), *Encyclopedia of the religions of indigenous people of South Asia*. Brill.
- Slade, Benjamin. 2013. The diachrony of light and auxiliary verbs in Indo-Aryan. *Diachronica* 30(4). 531–578.
- Slade, Benjamin. 2016. Compound verbs in Indo-Aryan. In Hans Henrich Hock & Elena Bashir (eds.), *The languages and linguistics of south Asia: a comprehensive guide* (The World of Linguistics 7), 559–567. Berlin/Boston: Walter de Gruyter.
- Smith-Stark, Thomas. 2009. La trilogía catequística: Artes, Vocabularios y Doctrinas en la Nueva España como instrumento de una política lingüística de normalización. In Rebecca Barriga Villanueva & Pedro Martín Butragueño (eds.), *Historia Sociolingüística de México. Vol. I: México prehispánico y colonial*, 451–482. México D.F.: El Colegio de México.
- Spencer, Andrew. 2012. Identifying stems. In Olivier Bonami (ed.), *Stems (Special issue of Word Structure 5.1)*, 88–108. Edinburgh: Edinburgh University Press.
- Sprigg, R.K. 1984. The Limbu s-final and t-final verb roots, after Michailovsky 1979 and Weidert 1982. *Nepalese Linguistics* 3. 11–55.
- Spronck, Stef & Tatiana Nikitina. 2019. Reported speech forms a dedicated syntactic domain. *Linguistics* 23(1). 119–159.
- Steele, Susan. 1975. Past and irrealis: just what does it all mean? *International Journal of American Linguistics* 41(3). 200–217.
- Thieberger, Nicholas & Michel Jacobson. 2010. Sharing data in small and endangered languages. In L.A. Grenoble & L. Furbee (eds.), *Language documentation: practice and values*, 147–158. Amsterdam & Philadelphia: John Benjamins.
- Thomas, Margaret. 2019. The monolingual approach in the history of American linguistic fieldwork.
- Thomas, Margaret. 2020. *Formalism and functionalism in linguistics*. New York and London: Routledge.
- Thornton, Anna M. 2019a. Overabundance: a canonical typology. In Franz Rainer, Francesco Guardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation* (Studies in Morphology 5), 223–258. Cham: Springer.

- Thornton, Anna M. 2019b. Overabundance in morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: the interaction between language contact, change, and inheritance. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan Languages* (Routledge Language Family Series 3), 3–21. London: Routledge.
- Toba, Ingrid. 1973. The Khaling verb. In Austin Hale (ed.), *Nepal Studies in linguistics*, vol. 1, 1–14. Kirtipur: INAS.
- Toba, Ingrid & Sueyoshi Toba (eds.). 2016. *Khaling - Nepali - English Dictionary*. SIL International. <https://khaling.webonary.org/browse/browse-khaling/?lang=en>.
- Toba, Sueyoshi. 1984. *Khaling* (Asian and African Grammatical Manual 13d). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tolsma, Gerard. 2006. *A grammar of Kulung* (Languages of the Greater Himalayan Region 5/4). Leiden: Brill.
- Turin, Mark. 2012. *A Grammar of the Thangmi Language: with an ethnolinguistic introduction to the speakers and their culture*. (Languages of the Greater Himalayan Region 5/6). Leiden & Boston: Brill.
- Van Olmen, Daniël. 2010. Typology meets usage: The case of the prohibitive infinitive in Dutch. *Folia Linguistica* 44(2). 471–508.
- Vanhove, Martine. 2020. Finite verb forms in Beja (Cushitic): Labels vs functions, a historical perspective. *Faits de Langues* 50(2). 31–46.
- Vaux, Bert & Justin Cooper. 2003. *Introduction to Linguistic Field Methods*. 2nd printing. Munich: Lincom Europa.
- Wälchli, Bernhard. 2007. Advantages and disadvantages of using parallel texts in typological investigations. (Ed.) Michael Cysouw & Bernhard Wälchli. *STUF - Language Typology and Universals (Special theme issue: Parallel texts: Using translational equivalents in linguistic typology)* 60(2). 118–134.
- Watters, David. 2009. *A grammar of Kham*. Cambridge: Cambridge University Press.
- Whitman, John. 2011. The ubiquity of the gloss. *Scripta* 3. 95–121.
- Whitman, John & Franck Cinato (eds.). 2014. *Lecture vernaculaire de textes classiques chinois / Reading Chinese Classical texts in the Vernacular* (Dossiers d'HEL 7). Paris: SHESL.
- Wilkins, David P. 1991. The semantics, pragmatics and diachronic development of “Associated Motion” in Mparntwe Arrernte. *Buffalo Papers in Linguistics* 1. 207–257.
- Witzlack-Marakevich, Alena, Taras Zakharko, Lennart Bierkandt, Fernando Zúñiga & Balthasar Bickel. 2016. Decomposing hierarchical alignment: Co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics* 54(3). 531–561.
- 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. <https://doi.org/doi:10.1017/S0035869X00095459>.
- Yap, F.-H., K. Grunow-Hårsta & J. Wrona. 2011. *Nominalization in Asian Languages*. Amsterdam: Benjamins.
- Yliniemi, Juha. 2016. Nepali attention marker ni. *Himalayan Linguistics* 15(2). 50–87.

- Zúñiga, Fernando. 2006. *Deixis and Alignment - Inverse systems in indigenous languages of the Americas*. Amsterdam: Benjamins.
- Zwartjes, Otto. 2011. *Portuguese Missionary Grammars in Asia, Africa and Brazil, 1550-1800*. Amsterdam & Philadelphia: John Benjamins.
- Zwartjes, Otto. 2012. The Historiography of Missionary Linguistics: Present state and further research opportunities. *Historiographia Linguistica* 39(2/3). 185–242.

