



Les corpus comparables en typologie

Aimée Lahaussais

► To cite this version:

Aimée Lahaussais. Les corpus comparables en typologie. Master. Approches en SDL, Université Lyon II, France. 2023. halshs-04068500

HAL Id: halshs-04068500

<https://shs.hal.science/halshs-04068500>

Submitted on 14 Apr 2023

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.



LABORATOIRE D'HISTOIRE DES THÉORIES LINGUISTIQUES

L'utilisation de corpus comparables en typologie : deux exemples

Aimée Lahaussais

Histoire des Théories Linguistiques

CNRS, Université de Paris, F-75013 Paris, France

Séminaire de recherche MA "Approches en SDL"

Université Lyon 2

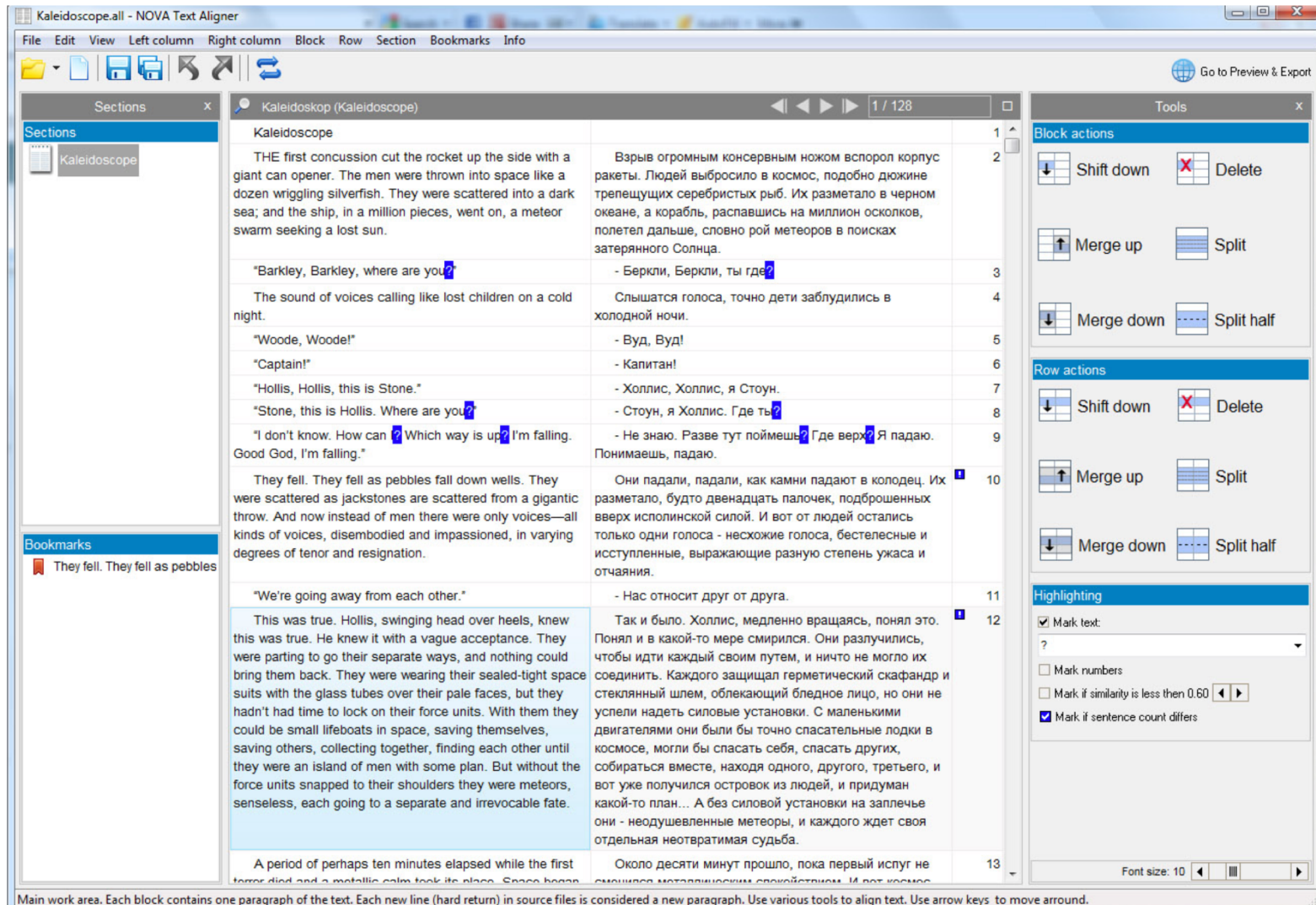
5 avril, 2023

- Les corpus comparables
- Corpus de mythologie kiranti
- Base de données métagrammaticographiques

Corpus comparables vs. corpus parallèles

“A parallel corpus is a collection of texts, each of which is translated into one or more other languages than the original” (Sinclair 1996)

Collections de textes issues de traduction ; contenu “identique” produit à l'aide de logiciels (se basant sur des mémoires de traduction), et corrigé manuellement par la suite



Exemples de corpus parallèles: la Bible, Le Petit Prince, Harry Potter, textes du parlement européen...

^ Sources externes (non révisées)

[...] qui ne peuvent ou ne veulent pas le faire - qu'ils soient grands ou petits - doivent en prendre pour leur grade par l'Union européenne.

↳ europarl.europa.eu

[...] countries that are unwilling or unable to achieve that - whether they be large or small - must get it in the neck from the European Union.

↳ europarl.europa.eu

Toutefois, pour l'application du paragraphe 3 du présent article, l'âge et le grade du fonctionnaire à prendre en compte sont ceux à la date de sa titularisation.

↳ eur-lex.europa.eu

However, for the purposes of paragraph 3 of this Article the official's age and grade to be taken into account shall be those at the time of establishment.

↳ eur-lex.europa.eu

Il pourrait donc prendre une décision indépendante sur le grade à octroyer au produit.

↳ www2.parl.gc.ca

It would be an independent third party separate from the process, so it would be able to make an independent decision as to what the grade on that product would be.

↳ www2.parl.gc.ca

[...] Siramy » a effectué l'essentiel de sa carrière au sein de la DGSE qu'il a quittée pour raison de santé fin 2009 avec le grade de directeur d'administration et en y ayant occupé différentes fonctions.

↳ esisc.net

[...] his career within the DGSE, which he quit for reasons of ill health at the end of 2009 having reached the rank of director of administration after having served in a number of different positions.

↳ esisc.net

Le gouvernement a-t-il nommé une personne de haut grade pour diriger la mise en œuvre de l'ITIE ?

↳ reportingrse.org

Has the government appointed a senior individual to lead on EITI implementation?

↳ reportingrse.org

J'allais m'élever dans la hiérarchie et peut-être allais-je pouvoir prendre ma retraite avec le grade d'adjudant-chef.

↳ www2.parl.gc.ca

I was going to rise up in the ranks and maybe retire as chief warrant officer.

↳ www2.parl.gc.ca

[...] adjoint des Forces canadiennes pendant plus de 31 ans, avant de prendre sa retraite avec le grade de lieutenant-colonel.

↳ ocrt-bctr.gc.ca

He served over 31 years in the Canadian Forces Communication Reserve and AJAG, retiring at the rank of Lieutenant Colonel.

↳ ocrt-bctr.gc.ca

Je vous suis très reconnaissant de me conférer le grade de Docteur en honoris causa en philosophie, que j'accepte non seulement comme une marque d'estime pour ma personne mais comme un hommage à l'UNESCO [...]

↳ unesdoc.unesco.org

I am very grateful to you for honouring me with the degree of Doctor of Philosophical Science honoris causa, which I accept as a tribute not only to myself but also to UNESCO and the ideals it represents.

↳ unesdoc.unesco.org

On a affaibli le principe de solde égale pour grade égal en utilisant la solde de spécialiste pour certains groupes professionnels [...]

↳ oag-bvg.gc.ca

The principle of equal pay for equal rank has been eroded by using specialist pay for some officer classifications, granting [...]

↳ oag-bvg.gc.ca

Utilisés par des typologues, mais contiennent un grand nombre de biais : langue écrite, standardisée, prescriptive, traduite...

Ces biais sont problématiques pour des travaux de typologie linguistique, puisque ceux-ci visent à comparer les langues dans leur état naturel, et les résultats peuvent être faussés si les données ne sont pas authentiques

[Cysouw, M. et B. Wälchli. 2007. Parallel texts: using translation equivalents in linguistic typology. Sprachtypologie und Universalienforschung 60 (2);

Sinclair, J. 1996. Preliminary recommendations on corpus typology. EAGLES document]

Collections de textes semblables en nature mais ne provenant pas de traduction ;

La similarité du contenu permet leur alignement

Exemples : articles de presse traitant, dans différentes langues, d'un même événement politique ou sportif

Les contenus ne sont pas identiques, dont alignement imparfait, mais gain en naturalité linguistique, et quantités massives de données possibles

C'est ce qui dorénavant est utilisé pour entraîner les mémoires de traduction, puisque beaucoup plus important comme volume de données

Corpus de mythologie kiranti

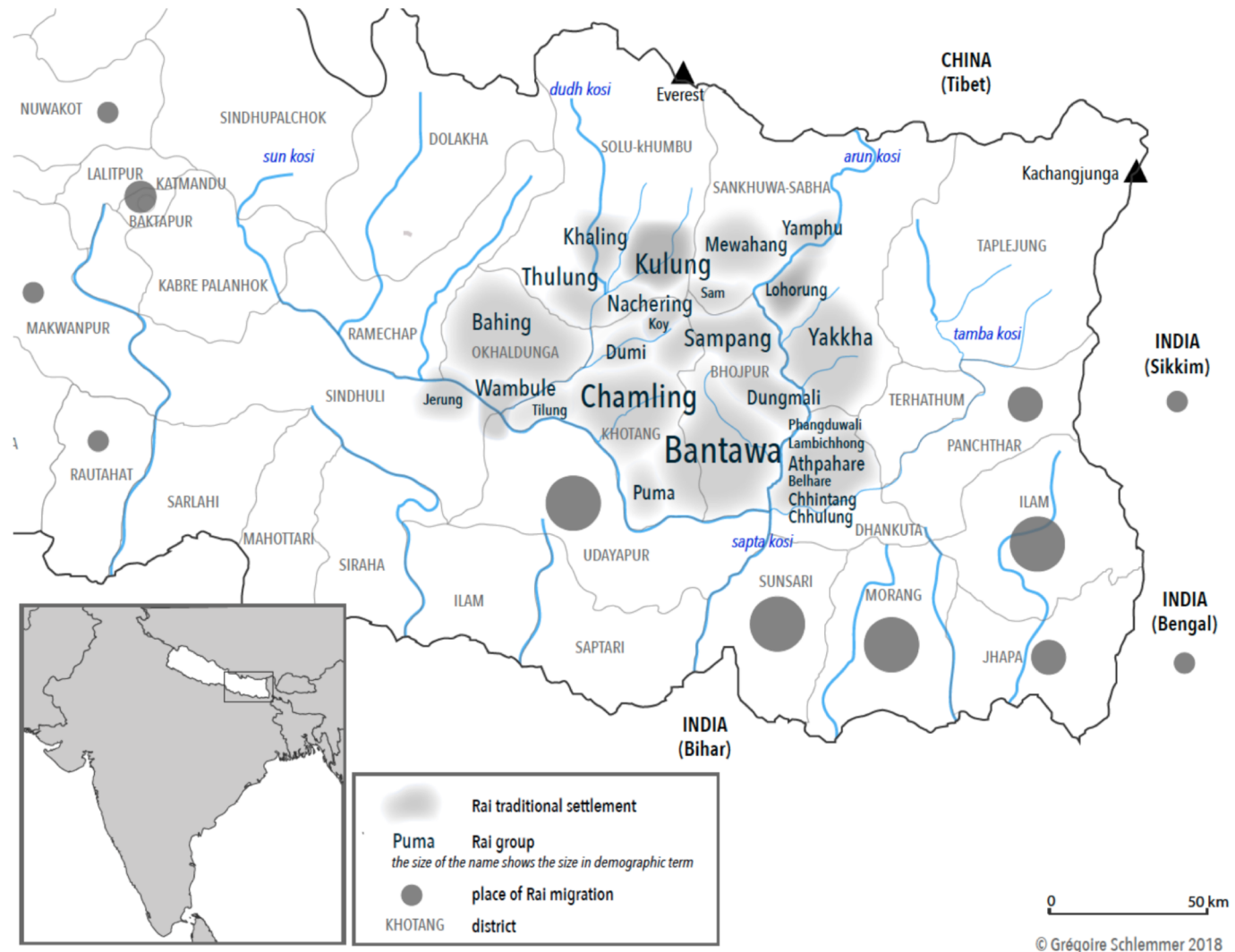
Groupe Kiranti

Langues sino-tibétaines/
trans-himalayennes

Env. 30 langues,
parlées au Népal
oriental

Langues largement
orales

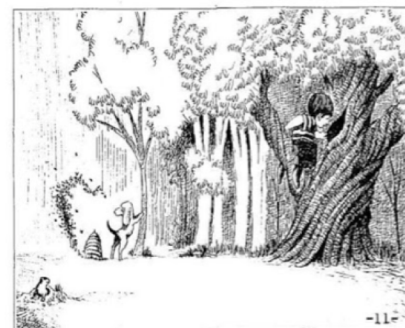
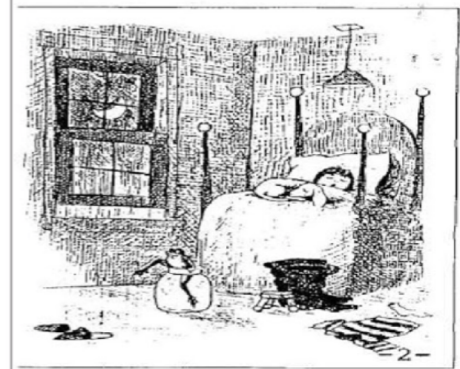
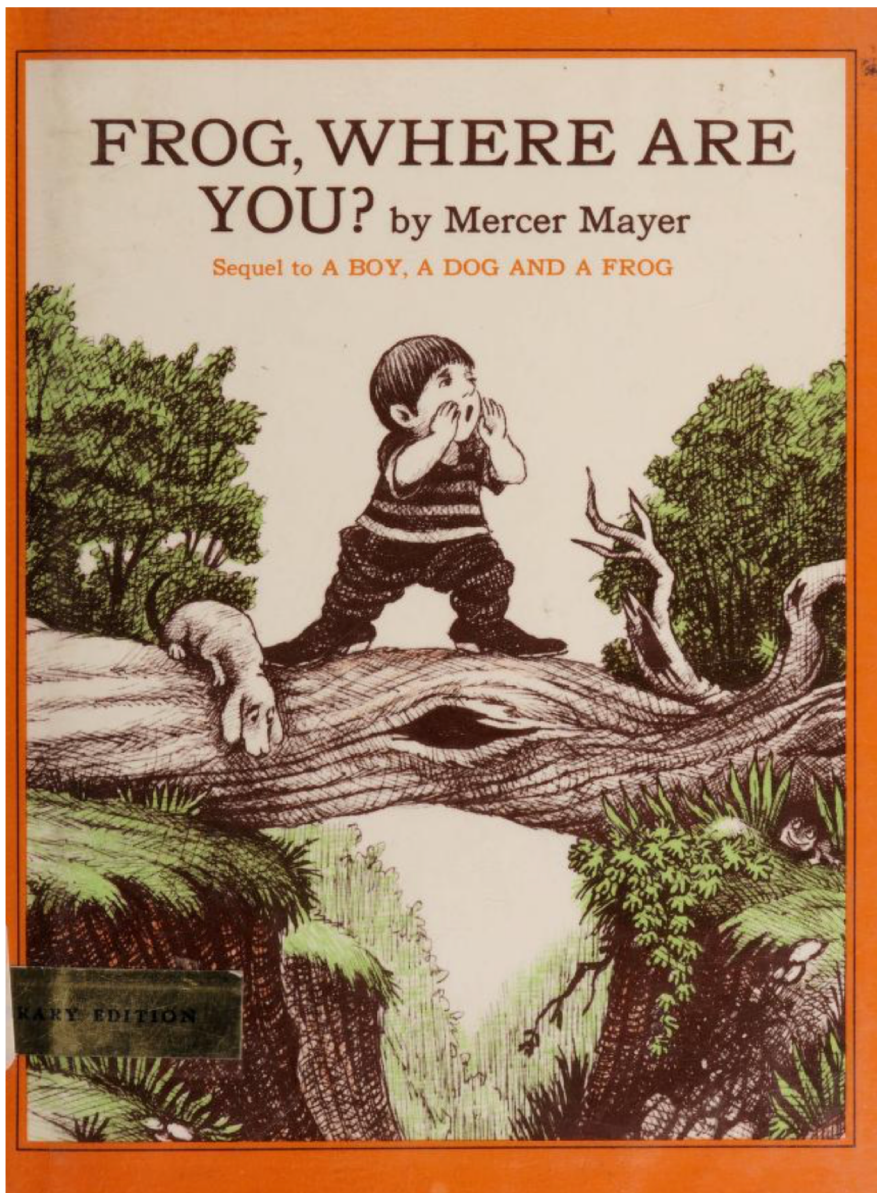
Contact intense
avec le népali
(indo-aryen)



- Questions de recherche :
 - Les langues kiranti sont-elles véritablement apparentées ?
 - Propositions de correspondances régulières et de reconstruction de la proto-langue, mais surtout les initiales
 - Certains spécialistes se demandent s'il s'agit d'un sous-groupe :
« It has never been shown that Kiranti... is a valid genetic unit. » Ebert 2003
 - Peut-on utiliser des matériaux autochtones, plutôt qu'issus de stimuli non adaptés, pour en savoir plus ?
 - Cycle mythologique partagé par les langues kiranti, décrit (entre autres) par Ebert et Gaenszle 2008

[Ebert, K. 2003. Kiranti languages: an overview. In G. Thurgood and R. LaPolla (eds) The Sino-Tibetan languages. London and NY: Routledge

Ebert, K. & M. Gaenszle. 2008. Rai mythology: Kiranti oral texts. Cambridge: Harvard University Press]



Projet HimalCo : corpus parallèles en langues himalayennes

Projet en deux grandes parties : dictionnaires parlants (tous les exemples sont enregistrés) et corpus comparable kiranti

Historique Corpus comparable kiranti :

Project de documentation de 3 langues kiranti (thulung, khaling, et koyi)

Collecte d'histoires très semblables, avec notamment des expressions adverbiales marquées aux même endroits dans les histoires...

→ Histoires appartenant au cycle mythologique

→ 5 histoires racontées dans 3 langues (dont une en 2 dialectes) par plusieurs locuteurs

Proto-story	Language/Dialect	Speaker	
Ogress	Khaling	Dhana	☐
	Khaling	Dhan Maya	☐
	Thulung/Mukli	Bishnu	☐
	Thulung/Mukli	Chandrakala	☐

Proto-story	Language/Dialect	Speaker	
Orphans	Khaling	Dhana	☐
	Khaling	Lal Bahadur	☐
	Koyi	Mila	☐
	Koyi	Phanindra	☐
	Thulung/Deusa	Lakpa	☐
	Thulung/Mukli	Chandrakala	☐

Proto-story	Language/Dialect	Speaker	
Leper story	Khaling	Bijiame	☐
	Khaling	Purna	☐
	Thulung	Chandrakala	☐

Proto-story	Language/Dialect	Speaker	
Ur-Mother	Khaling	Dhana	☐
	Koyi	Mila	☐
	Koyi	Phanindra	☐
	Thulung/Mukli	Chandrakala	☐
	Thulung/Deusa	Lakpa	☐

Proto-story	Language/Dialect	Speaker	
Ghumne Pani story	Thulung	Uttar Kumar	☐
	Thulung	Dal Bahadur	☐

NB : transcription complète de 1 minute d'audio = 1 heure de travail d'annotation !

Sélection du sous-corpus pour visualisation :

Proto-story	Language/Dialect	Speaker	
Orphans	Khaling	Dhana	☒
	Khaling	Lal Bahadur	☐
	Koyi	Mila	☐
	Koyi	Phanindra	☒
	Thulung/Deusa	Lakpa	☐
	Thulung/Mukli	Chandrakala	☒

Event/Èvènement	<u>Solme Lamalit / Dhana</u>	<u>Solme Lamalit / Dhan Maya</u>	<u>Eagle story / Bishnu</u>
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
24			S17
25		S33 S34	S18
26	S87 S88 S89	S36	S19
27	S92		
28		S39	S22
29	S94 S95	S38	S22
30	S96	S40	S22
31			S23
32	S107	S43	S26
33			S27

- Défis pour l'alignement du corpus :

-- Quoi aligner :

Alignement par phrase(s) initialement ; modifié par la suite pour alignement par évènement

-- Faut-il partir d'une version d'une histoire dans une langue choisie au hasard, ou reconstruire une "proto-version" avec les éléments communs à toutes les versions ?

Event: Two were girls, one a boy

Khaling (Dhana)

S2

ʔamsupo sukpu ʔustsəhɛm mo:tnu sakpu melsem ʔutsəsu grômme-kolo lasmɛsu

ʔam-su-po sũ:k-pu ʔus-tse-hɛm mu-tɛnu sâ:k-pu melsem ʔu-tse-su
 3-du-gen three-class 3-du-poss-child-pl be.anim-23-pl-pst two-class girl 3-sg-poss-child-du
 grômme-kolo lasme-su
 Gromme-com Lasme-du

They had three children, two were girls, Gromme and Lasme.

Koyi (Phanindra)

S191

adzi sɔma-tso bigjame-tso made mo-ni-m tsʰɔ kim-bi mo-dza tsʰa

adzi sɔma-tso bigjame-tso made mo-ni-m tsʰɔ kim-bi
 later person-PLU sister-PLU be.anim-3PL.PST-NOM become.3SG.PST house-LOC
 mo-dza tsʰa
 be.anim-DUR.3SG.PST HS

So it is that Human and the sisters stayed, they stayed in the house.

Thulung/Mukli (Chandrakala)

S2

nwale dzaumo kole tarapa memma meɬip memim upap wo siɖa bante lɛsta
 diphu umam wo siɖa

nwa-le dzaumo ko-le tarapa memma me-tɬip mɛ-mim u-pap wo si-ɖa bante
 two-cl daughter one-cl son then that-du that-pl 3-sg-poss-father also die-3-sg-pst where
 lɛks-ɖa diphu u-mam wo si-ɖa
 go-3-sg-pst later 3-sg-poss-mother also die-3-sg-pst

There were two daughters and on son, and then their father died somewhere, and later their mother also died

Concordance ?

Proto-story Orphans

No	Language (speaker)	Sentence	Left context	Word	Right context	Gloss
1	Koyi (Phanindra)	S127	munmuri k ^h okts a k ^h onts asi d ^h anɔ sɔ pu	ts ^h ekuma	nusi ts ^h oʔ si umnusi pu k ^h ur bi buwa	sister
2	Koyi (Phanindra)	S147	ɔɔ mɔ si ɔ ne adzi ne ana	ts ^h ekuma	nusi wɔ oko ɔ k ^h ur bi d ^h oʔo ɔ	sister
3	Koyi (Phanindra)	S187	maʔa kim bi ne ɔɔtsi tsɔ mo ni	bigja	ɔ tsɔ tsɔ mo ni k ^h ɔɔ mo ni	sister
4	Koyi (Phanindra)	S191	luʔ ɔ p ^h in u ts ^h a adzi sɔma tsɔ	bigjame	tsɔ made mɔ ni m ts ^h ɔ kim bi	sister
5	Koyi (Phanindra)	S278	lu ts ^h a naga ne ɔɔ intsi sɔɔ intsi	ts ^h ekuma	intsi tsɔ sɔɔme ts ^h ɔ ɔ kɔkɔ sɔma tsɔ	sister
6	Koyi (Phanindra)	S334	ɔ ne heɔɔ mɔ dza si ts ^h a ɔɔ	ts ^h ekuma	hoʔle d ^h aiʔɔɔ ɔ umbika sɔksu pu ana nuwa	sister
7	Koyi (Phanindra)	S345	ase papa luʔ si heɔɔ d ^h am bi ne	ts ^h ekuma	tsɔ ts ^h ɔm ka ts ^h uʔ mu ts ^h ɔ ts ^h ɔm ka	sister
8	Koyi (Phanindra)	S346	ts ^h ɔm ka ts ^h uʔ mu ts ^h ɔ ts ^h ɔm ka ts ^h ekumɔ	ts ^h ekuma	ts ^h uʔ mu ts ^h ɔʔ ɔ ne kɔpa mo a	sister
9	Koyi (Phanindra)	S347	ts ^h uʔ mu ts ^h ɔʔ ɔ ne kɔpa mo a	ts ^h ekuma	tsɔ dja ɔ tsjuri mu di p ^h in usi	sister
10	Koyi (Phanindra)	S350	se si ts ^h a ts ^h ɔmdam bi ts ^h ɔm ka ts ^h e	ts ^h ekuma	tsɔ ts ^h angara p ^h in usi ts ^h a ts ^h angara jɔ ja	sister
11	Koyi (Phanindra)	S353	a ɔ adzi pi wa pu dja ts ^h a	ts ^h ekuma	ɔ bo hɔbats a ɔ k ^h ɔktsulupa dja ts ^h a	sister
12	Koyi (Phanindra)	S362	si m tselbu ɔ kim hoʔle si d ^h aiʔɔ	bigjame	nusi ts ^h ɔm ka bi pu risi pik uni	sister
13	Khaling (Dhana)	S125	grômme sîl ʔÊndəjo ŋa mÊm tsɛntɛ hÊm ʔɛ	bÊŋmɛ	tsɛ hÊm ʔɛ was tshÊm nu lo ʔatha	sister
14	Thulung/Mukli (Chandrakala)	S112	dʒyl ɖa ŋa ba ɖa memma ko le	ritsɔ	rok mekotima tshimori mal y memlo ka tsai	sister
15	Thulung/Mukli (Chandrakala)	S134	khwaktsilik lai hila be ɖy ʔe ma i	ritsɔ	mim mi ba ni rwak ɖa ma hila	sister
16	Thulung/Mukli (Chandrakala)	S212	be mi lo meʔe lalikundɔ be mi lo	ritsɔ	mim lai gwak mi m tsɔŋra ritsɔ mim	sister
17	Thulung/Mukli (Chandrakala)	S212	lo ritsɔ mim lai gwak mi m tsɔŋra	ritsɔ	mim ka uni nem lanʒa memsaka bhara ne	sister

Idéophones : concordance sur "suddenly"

No	Language (speaker)	Sentence	Left context	Word	Right context	Gloss
1	Thulung/Deusa (Lakpa)	S79	mə ri tʃip ka tʃai mə ser wo	piri	nap tsi ma set tsi ʔe ani medqam	<i>suddenly</i>
2	Khaling (Dhana)	S17	nɛ nʌ ʔusominɛ kʌm ʔətt nɛ mətʃt pɛ	dhawa	dhawa dʒʌkʰʌl kak tɛsu mɛbeːlo khaktsalʌp ʔus dhaː	<i>suddenly</i>
3	Khaling (Dhana)	S17	nʌ ʔusominɛ kʌm ʔətt nɛ mətʃt pɛ dhawa	dhawa	dʒʌkʰʌl kak tɛsu mɛbeːlo khaktsalʌp ʔus dhaː mɛbi	<i>suddenly</i>
4	Khaling (Dhana)	S20	ʔipt do tɛ grɔmmɛ kolo lasmɛ su ʔɛ	dhawa	mɛ dʒʌkʰʌl kak tɛsu lo mɛ lektseːm ʔɛ	<i>suddenly</i>
5	Khaling (Dhana)	S78	mokho ʔuŋ ŋʌ lɛl khət ŋʌ mʌnʌ lasmɛ	phɛrɛ	bher tɛ khət tɛ ʔu kʌm bika ʔʌm	<i>suddenly</i>
6	Thulung/Mukli (Chandrakala)	S13	ŋa ghrok sok dʒ ʔe mə sʌŋ ka	tyry	phɛt sok dʒ retʃha mekotima ne mə tʃyplo	<i>suddenly</i>
7	Khaling (Lal Bahadur)	S19	hɛmba kur tɛ nʌ mɛbelo ni ʔu tʃɔ	sele	phrək tɛ nʌ mɛ khabo khunt de ju	<i>suddenly</i>

Interjections : concordance sur "hey"

38	Koyi (Phanindra)	S331	mama nɔ dosam wɔ njepts a ba tʃʰa	ʔe	ana ne mama wɔ pa na nɔ lɔts	<i>hey</i>
39	Koyi (Phanindra)	S355	u tʃʰa pʰjatpʰjatpʰjat bʰənda dʰaiʔ mu lɔ ne	oho	he intsi kʰoktsulupa de di ma idɔ pɔma	<i>hey</i>
40	Koyi (Phanindra)	S368	di umnusi kɔ sɔmɔ jɔ ɔ gɔ di	lu	idɔ ŋɔ be si luʔ si tʃʰa nɔ	<i>hey</i>
41	Khaling (Dhana)	S91	dʒə tɛ mʌnʌ pɔɔpɔɔp tunəl koɔm ghəl tɛ	ʔe	reːmɛ toːmɛ koɔm mu ghəl wə ʔuŋʌ seseː	<i>hey</i>
42	Khaling (Dhana)	S101	ŋɛn de pɛ mu tɛ mʌnʌ siŋ tɛ	ʔe	rêː ho tɛ bi ʔʌ wêː ʔi thə	<i>hey</i>
43	Khaling (Dhana)	S101	pɛ mu tɛ mʌnʌ siŋ tɛ ʔe rêː	ho	tɛ bi ʔʌ wêː ʔi thə tɛ ho	<i>hey</i>
44	Khaling (Dhana)	S101	ho tɛ bi ʔʌ wêː ʔi thə tɛ	ho	grɔmmɛ ʔi thə tɛ ho siŋ tɛ pɔɔpɔɔp	<i>hey</i>

Ce type de recherche sur un sous-corpus fait apparaître les incohérences d'annotation : gloses lexicales pour des phénomènes mal compris...

- Avantages :
 - le concordancier permet une comparaison lexicale rapide, qui peut ensuite être contextualisée par l'accès aux phrases
 - intérêt ethnographique d'avoir des histoires « alignées », puisque le processus permet de réfléchir aux éléments constitutifs de la proto-histoire (à moins que ces histoires ne soient empruntées ...?!)
 - projet attractif et valorisant pour les communautés locales
- Difficultés :
 - alignement non automatisable (réduisant la taille maximale possible du corpus)
 - les idiosyncrasies dans les versions des histoires sont individuelles et non pas dialectales/linguistiques
 - les constructions ne sont pas glosées (ex. pas moyen de faire une recherche « propositions relatives » ou « verbes dérivés »), et les éléments qui permettraient vraiment de comparer les langues sur le plan morphologique/syntaxique ne sont pas interrogeables facilement

<http://himalco.huma-num.fr>

Base de données métagrammaticographiques

Deux problèmes (parmi tant d'autres) dans l'utilisation des grammaires descriptives pour la typologie :

1) Structuration des tables de matières très hétérogène

Toba 1979 Khaling
40 pages

- 0. Introduction
- 1. Phonology
- 2. Word order
- 3. Nouns
- 4. Pronouns
- 5. Numerals
- 6. Verbs
- 7. Adjectives
- 8. Adverbs
- 9. Conjunctions
- 10. Articles
- 11. Interjections

Michailovsky 1988 Hayu
234 pages

- 1. Les hayu et leur langue
- 2. Phonologie
- 3 Morphologie (1): structure du mot verbal
- 4. Morphologie (2) : structure du mot non verbal ; parties du discours
- 5. Syntaxe
- Texte 1
- Texte 2

Schackow 2015 Yakkha (603p)

- 1 Introduction
2. The Yakkha language and its speakers
- 3 Phonology
- 4 Pronouns, demonstratives, quantifiers, numerals, interrogatives
- 5 The noun phrase
- 6 Adjectives and adverbs
- 7 The geomorphic orientation system
- 8 Verbal inflection
- 9 Noun-verb predicates
- 10 Complex predication
- 11 Transitivity
- 12 Simple clauses
- 13 Nominalization and relativization
- 14 Adverbial clause linkage
- 15 Complementation
- 16 Connectives on the text level
- 17 Discourse particles and interjections
- Appendix A: Texts
- Appendix B: Yakkha kinship terms
- Appendix C: Index of Yakkha formatives

Quels facteurs expliquent ces différences :

Date ?

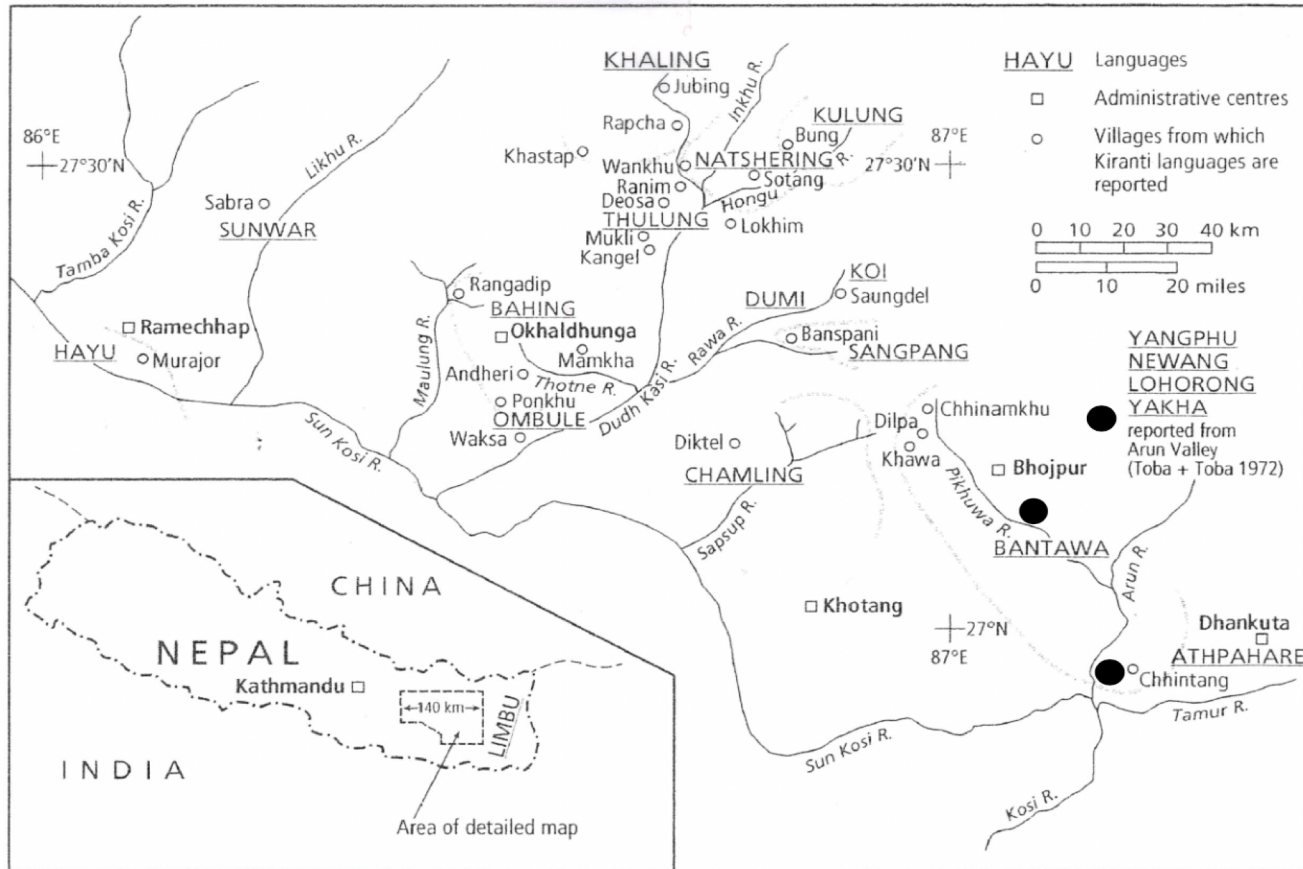
Affiliation de l'auteur ?

Nombre de pages ?

Nombre de chapitres ?

2) Différences terminologiques

Ex. **idéophones** « open lexical class of marked words that depict sensory imagery »
(Dingemanse 2019)



[Carte basée sur Michailovsky 1975]

Métagrammaticographie ?

méta + **grammatico** + **graphie**

= **rédaction de grammaires**

+ **méta**

= **regard évaluatif/critique sur la rédaction de grammaires**

"Grammaticography is an activity related to grammar₂ [...]. At the object level, grammaticography is (the practice - experience or art - of) grammar writing.

At the **meta**-level, it is the investigation of methodological principles that reconcile this practice with linguistic theorizing. [...]"

[Lehmann & Maslova 2004, in *Morphologie: Ein Handbuch zur Flexion und Wortbildung*, vol. 2]

Quelques travaux (+/- descriptifs) sur la *métagrammaticographie*

Ouvrages

- Felix Ameka, Alan Dench & Nicholas Evans (eds.), 2006, *Catching Language: The standing challenge of grammar writing*, 41–68. Berlin & New York: Mouton de Gruyter.
- Thomas Payne & David Weber (eds.), 2007, *Perspectives on grammar writing* (Benjamins Current Topics 11). Amsterdam & Philadelphia: John Benjamins.
- Toshihide Nakayama & Keren Rice (eds.), 2014, *The art and practice of grammar writing* (Language Documentation and Conservation Special Publication 8). Honolulu, HI: University of Hawai'i Press. [open access]

Articles

- Lehmann, Christian & Elena Maslova. 2004. Grammaticography. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie: Ein Handbuch zur Flexion und Wortbildung*, vol. 2, 1857–1882. Berlin & New York: De Gruyter.
<https://christianlehmann.eu/publ/grammaticography.pdf>
- Mosel, Ulrike. 2006a. Sketch Grammar. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 301–310. Berlin & New York: Mouton de Gruyter.
- Mosel, Ulrike. 2006b. Grammaticography: The art and craft of writing grammars. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: The standing challenge of grammar writing*, 41–68. Berlin & New York: Mouton de Gruyter.
- Cristofaro, Sonia. 2006. The organization of reference grammars: a typologist user's point of view. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: The standing challenge of grammar writing*, 137–170. Berlin & New York: Mouton de Gruyter.
- Noonan, M. 2007. Grammar writing for a grammar-reading audience. In Thomas Payne & David Weber (eds.), *Perspectives on grammar writing (Benjamins current topics 11)*, 113–126. Amsterdam & Philadelphia: John Benjamins.

Questions chronologiques

Date d'apparition :

de tables de matières

de gloses interlinéaires

de différentes catégories grammaticales/termes

Quelle influence de la langue native du linguiste ou métalangue choisie sur la description ?

Quelles écoles de description ? Comment les reconnaître ?

Quels schèmes organisationnels pour les grammaires ?

Quelle approche est utilisée par le grammairien ?

Quel cadre théorique de description ? Comment le reconnaître ?

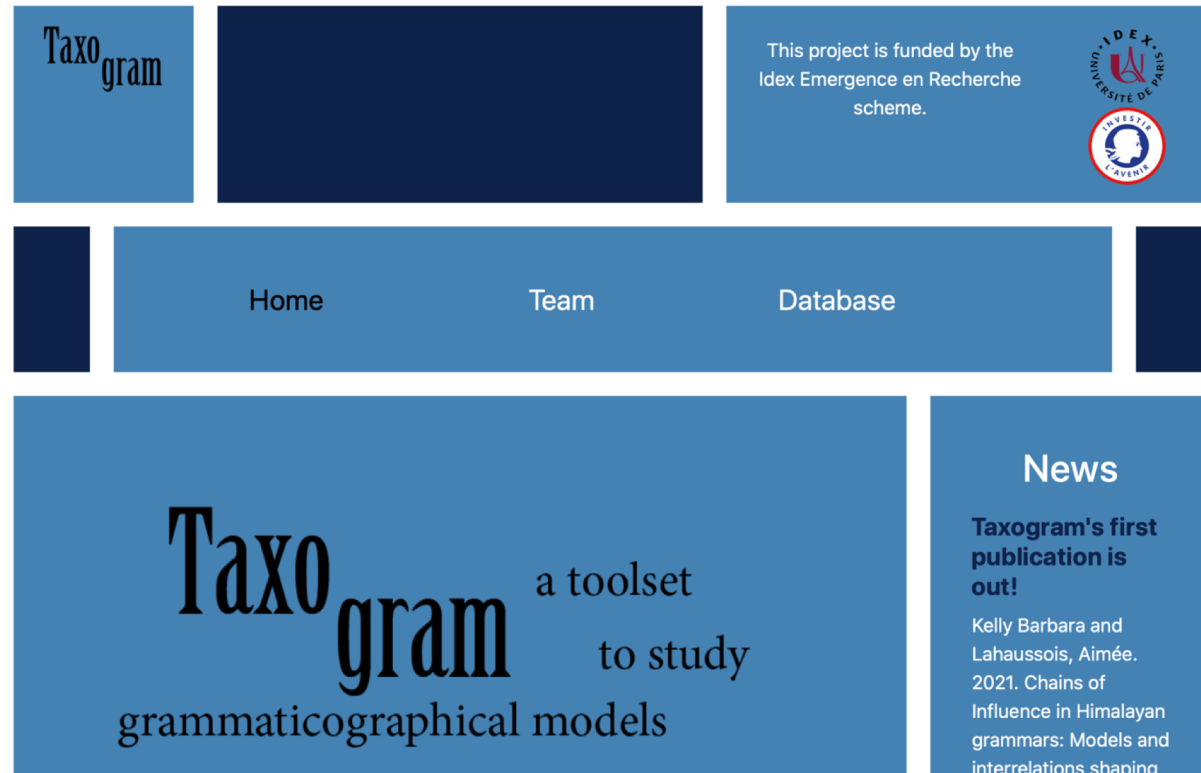
Quels types de grammaires ?

Quels types de publics pour des grammaires ? Peut-on les reconnaître à l'organisation de la grammaire ?

Quelle est la place des glossaires/textes dans les grammaires ?

Les réponses à ces questions sont souvent basées sur une expérience de lecture de grammaires qui est biaisée par les intérêts et expériences du linguiste

Projet Taxogram, <https://taxogram.huma-num.fr/home.html>



Objectif principal :

base de données relationnelle nous permettant de répondre *de manière empirique* à un certain nombre de questions de recherche

- 1) établir une taxinomie des schèmes structurels ou organisationnels de grammaires descriptives
- 2) corrélér ces schèmes avec des critères biographiques et ou bibliographiques
- 3) déterminer des principes d'appartenance à des écoles descriptives sur la base de structuration et de terminologie
- 4) établir une base de données terminologiques 'alignée'

→ Mieux comprendre les choix que font les linguistes écrivant des grammaires et l'impact de ces choix sur l'utilisation faite par les typologues de ces données

Objet d'étude = les grammaires (descriptives)

Pour chaque **grammaire** que nos 'codons' nous collectons :

- a) métadonnées biographiques (linguiste) et bibliographiques (grammaire)
- b) Table des matières
- c) Liste d'abréviations utilisées

A	B	C	D	E	
GlottRef					
Language name					
Glottocode					
Grammar metalanguage					
Author last name					
Author first name					
Scientific editor					
Grammar title					
Publication date					
Author's current affiliation					
Author's affiliation at time of publication					
Author's other affiliation					
Author's other affiliation					
Author's PhD institution					
Publisher					
Collection					
Number of chapters					
Number of pages of appendices					
Number of pages total					
Examples interlinearized (Y N)					
Number of entries in wordlist (estimate)					
Number of texts included					
Number of lines in TOC					
Number of entries in the list of abbreviations					
▶	METADATA	TOC	ABBREVIATIONS	HowTo	

GlottRef
Language name
Glottocode
Grammar metalanguage
Author last name
Author first name
Scientific editor
Grammar title
Publication date
Author's current affiliation
Author's affiliation at time of publication
Author's other affiliation
Author's other affiliation
Author's PhD institution
Publisher
Collection
Number of chapters
Number of pages of appendices
Number of pages total
Examples interlinearized (Y N)
Number of entries in wordlist (estimate)
Number of texts included
Number of lines in TOC
Number of entries in the list of abbreviations
Written as PhD dissertation (Y N)
PhD supervisor
Language family Glottolog
Language subgroup Glottolog
Other relevant genetic grouping
Grammar type
Grammar audience
Approach
Framework
Predominant data type
Availability of data in public archive (Y N)
Other observations
Last name of coder

Fichier métadonnées

Métadonnées portant sur le grammairien et sur la grammaire (biographiques et bibliographiques)

Codes couleur :

orange : données à trouver dans la grammaire

orange pale : un peu de recherche nécessaire pour trouver les données

orange foncé : décisions subjectives

<https://glottolog.org>

Identifiants
'uniques' pour
grammaires
et langues
(grande
variabilité
dans la
nomination !)


Glottolog
Languages
Families
Language Search
References
Reference Search
About

Glottolog 4.3

Welcome to Glottolog 4.3

Comprehensive reference information for the world's languages, especially the lesser known languages.

Information about the different languages, dialects, and families of the world ('languoids') is available in the [Languages](#) and [Families](#) sections. The [References](#) section contains bibliographical information. You can query the bibliographical database by filtering the table view or using [a complex query](#) involving genealogical affiliation, document type, and macro-area.



Catalogue of languages and families

Glottolog provides a [comprehensive catalogue](#) of the world's languages, language families and dialects. It assigns a unique and stable identifier (the Glottocode) to (in principle) all languoids, i.e. all families, languages, and dialects. Any variety that a linguist works on should eventually get its own entry. The languoids are organized via a genealogical classification (the Glottolog tree) that is based on available historical-comparative research (see also the [Languoids information](#) section).

Bibliography

The [References](#) section ('langdoc') provides a comprehensive collection of bibliographical data for the world's lesser known languages. It provides access to 341,507 references of descriptive works such as grammars, dictionaries, word lists, texts etc. Search criteria include author, year, title, country, and genealogical affiliation. References can be downloaded as txt, bib, html, or with the [Zotero](#) Firefox plugin.



Métalangue de la grammaire

corrélations entre métalangue et organisation/terminologie (par nationalité ? mais plus difficile d'obtenir l'information)

	<i>in-form</i>	<i>i-form</i>	<i>i:-form</i>
Vanhove (2014a, 2017a)	<i>Imperfective</i>	<i>Perfective</i>	<i>Aorist</i>
Vanhove (2005)	<i>imperfective</i>	<i>perfective</i>	<i>narrative</i>
Wedekind (2007)	<i>Imperfect</i>	<i>Perfect</i>	<i>Past Continuous / Pluperfect</i>
Appleyard (2007)	<i>present</i>	<i>past</i>	<i>aorist</i>
Appleyard (2004)	<i>present</i>	<i>past</i>	<i>old past</i>
Morin (1995)	<i>inaccompli</i>	<i>accompli</i>	<i>conditionnel</i>
Voigt (1988)	<i>Imperfekt</i>	<i>Aorist / Präteritum</i>	<i>Perfekt</i>
Dahl (1984)	<i>present</i>	<i>perfective past</i>	<i>imperfective past</i>
Hudson (1976)	<i>present</i>	<i>preterit</i>	<i>past</i>
Zaborski (1975)	<i>Present</i>	<i>Old Present</i>	<i>Old Past</i>
Cohen (1972, 1973; 1988)	<i>présent; inaccompli</i>	<i>Passé / parfait; accompli</i>	<i>plus-que-parfait / parfait / conditionnel</i>
Roper (1928)	<i>present</i>	<i>past</i>	<i>conditional</i>
Reinisch (1893-94)	<i>präsens</i>	<i>perfect</i>	<i>plusquamperfect</i>
Almkvist (1881)	<i>Präsens</i>	<i>Perfekt</i>	<i>Ø</i>
Munzinger (1864)	<i>Aorist</i>	<i>Perfect</i>	<i>Plusquamperfect</i>

Vanhove, M. 2020. Finite verb forms in Beja (Cushitic): Labels vs functions, a historical perspective, *Faits de Langues*

Date de publication

permet de révéler des tendances décennales ou séculaires (ou autre) :

années 80-90 organisation phonologie-morphologie-syntaxe-discours

années (20)10-20 organisation par catégories grammaticales

Affiliations, Directeur de thèse (le cas échéant)

Universités où enseigne/a enseigné et où a étudié le grammairien ; autres affiliations

Cercles d'influences (subies et exercées) du grammairien : académique mais aussi non-académique (missionnaire ?)

Possibilité de grouper les grammairiens par 'écoles', généralement autour d'un linguiste et d'une aire linguistique

Faudrait-il (se limiter à ?) diriger des thèses sur des aires linguistiques où on ne travaille pas ?

Données bibliographiques

Editeur, Collection : gabarits imposés ou transmis ?

Nombre de chapitres

Nombre de pages

→ corrélations avec des types d'organisation du contenu ?

(Pawley 2014, dans Nakayama & Rice eds : thèses de plus en plus longues)

Liste d'abréviations + termes correspondants

aperçu rapide de ce qui est grammaticalisé dans la langue

Parfois les abréviations ne sont pas fournies :

Ex. Grammaires missionnaires

Présupposition que les lecteurs et grammairiens partagent le même cadre grammaticographique/terminologique (cadre gréco-latin)

Parallèle contemporain : articles de description/typologie qui ne listent que les abréviations non incluses dans les gloses de Leipzig → cadre partagé

Connexion entre interlinéarisation et utilisation d'abréviations

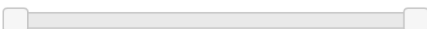
TAXOGRAM

Reset all filters

Keywords

Publication date

0 to 2021



Languages

- ☐ Amharic; (pastor)
- ☐ Amharic-Argobba
- ☐ Amharic; Old Amharic
- ☐ Arandic-Thura-Yura
- ☐ Arapaho
- ☐ Athabaskan

Author's PhD affiliation

- ☐ Addis Ababa University
- ☐ Australian National University
- ☐ Christian-Albrechts University (Kiel, Germany)
- ☐ Collège Royal (Paris)

Sort by

- ☐ Date ☐ Family ☐ Chapters ☐ Pages ☐ Interlinear ☐ Wordlist ☐ Text ☐ TOC ☐ Abbreviation

Abbreviations

Text in TOC

Display word cloud

Display abbr. word cloud

Display TOC

Select all ☐

227 out of 227 grammars displayed

Display all information...

A Grammar and Glossary of the Sherpa Language (100639)

2004

☐

More information...

A grammar of Sunwar: descriptive grammar, paradigms, texts, and glossary (101702)

2008

☐

More information...

Sketch of Thulung grammar with three texts and a glossary

Table of contents

PDF/t

PDF/a

D3

Abbr

KeyW

Foreword

Acknowledgements

Table of contents

Chapter I - Introduction

Chapter II - Phonology

Chapter III - Personal pronouns and verb

Chapter IV - BRIEF COMMENTS ON OTHER PARTS OF SPEECH

Chapter V - FURTHER TOPICS - MAINLY DIACHRONIC

Chapter VI - TEXTS

GLOSSARY

Appendix 1 - Names of trees and wild plants

Appendix 2 - Adverbs of abruptness

ADDENDUM

Khaling

Table of contents

PDF/t

PDF/a

D3

Abbr

KeyW

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

Abbreviations

Bibliography

A descriptive study of Bantawa

Table of contents

PDF/t

D3

KeyW

Acknowledgement

Abbreviations

Map

0. - Introduction

1. - Phonology

2. - Morphology

3. - Pronouns

4. - Verbs

5. - Adverbs

6. - Adjectives

7. - A note on numerals and classifiers

8. - An outline of syntax

9. - Sample texts with translation

10. - Vocabulary

11. - Bibliography

A grammar of Limbu

Table of contents

PDF/t

PDF/a

D3

Abbr

KeyW

Acknowledgements

Table of contents

Abbreviations

Transliteration and transcription

Maps

0 - Introduction

1 - Phonology and phonetics

2 - Nominal morphology

3 - The verbs 'to be'

4 - Morphemic analysis of the simplicia

5 - Aspect and aspectivizers

6 - Mode

7 - Gerunds and periphrastic tenses

8 - Other verbal constructions

9 - Subordination

Sketch of Thulung grammar with three texts and a glossary

Table of contents	PDF/t	PDF/a	D3
Abbr	KeyW		
Chapter I - Introduction			
1. - The Thulung language as an object of research			
1.1. - Name, location, number of speakers			
1.2 - Previous research			
1.21 - References			
1.3 - Relationship to neighbouring languages			
1.31 - References			
1.4 - Sociolinguistic notes			
2 - Aims and methods of present research			
2.1 - Priorities and limitations			
2.2 - Period and location of fieldwork			
2.3 - Informants			
3 - Notation			
3.1 - Nepali words			
3.2 - Citations from previous authorities or varying dialects			
3.3 - Glosses			
3.4 - Cross-references			
Chapter II - Phonology			

Khaling

Table of contents	PDF/t	PDF/a	D3
Abbr	KeyW		
0. - Introduction			
1. - PHONOLOGY			
1.1. - Vowels			
1.2. - Consonants			
1.3. - Syllable patterns			
1.4. - Tone			
2. - WORD ORDER			
2.1. - Declarative			
2.2. - Interrogative			
2.3. - Conditional			
2.4. - Noun phrase			
2.5. - Adjective phrase			
2.6. - Verb phrase			
3. - NOUNS			
3.1. - Number			
3.1.1. - Singular			
3.1.2. - Dual			
3.1.3. - Plural			
3.2. - Gender			

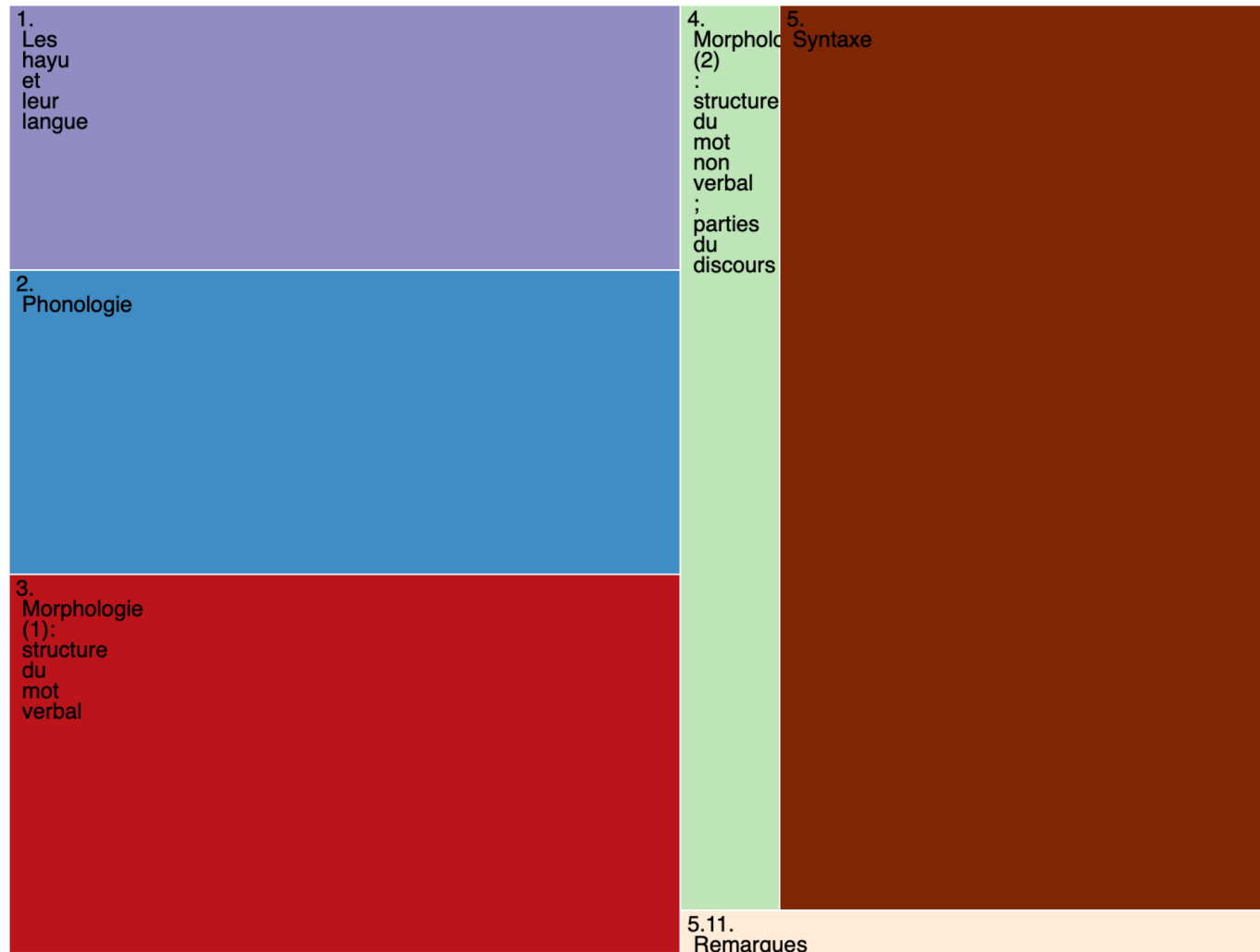
A descriptive study of Bantawa

Table of contents	PDF/t	PDF/a	D3
KeyW			
0. - Introduction			
0.1. - Classification of the language			
0.2. - The term kirat/kirati/kirant			
0.3. - The term khambu (Rai)			
0.4. - Different languages and dialects among the Rais			
0.5. - Some customs and traditions of Rais			
0.6. - The present study of Bantawa			
0.7. - Collection of data			
0.8. - Nature of the study			
1. - Phonology			
1.1. - Inventory of phonemes			
1.2. - Phonemic contrasts			
1.2.1. - Consonants			
1.2.1.1. - Unaspirated vs. aspirated			
1.2.1.2. - Voiced vs. voiceless			
1.2.1.3. - Contrasts of the place of articulation			
1.2.1.4. - Contrast s of the manner of articulation			
1.2.1.5. - Contrast between presence and			

A grammar of Limbu

Table of contents	PDF/t	PDF/a	D3
Abbr	KeyW		
0 - Introduction			
0.1 - The Limbus			
0.2 - The language			
0.3 - This grammar			
1 - Phonology and phonetics			
1.1 - Limbu phonemes and the native phoneme system			
1.2 - Allophones and phonetic description of Limbu phonemes			
1.3 - The Limbu word			
1.3.1 - Hiatus			
1.3.2 - Syllable structure			
1.4 - Assimilation and dissimilation			
1.5 - The orthography			
2 - Nominal morphology			
2.1 - Adjectives			
2.1.1 - Agreement in adjective			
2.1.2 - The nominalizing suffix <i>-tarjba</i>			
2.1.3 - The colour affixes			
2.2 - Pronouns			

Représentation du nombre de pages de chaque item d'un niveau



Sélectionnez un niveau de profondeur hiérarchique :

Représentation du nombre de pages de chaque item d'un niveau

1.1. L'habitat hayu	1.7. Corpus et méthode de recherche	2.5. Combinaisons frontalières de morphèmes	2.6. Autres jonctives	2.7. Phonologie des mots du népali	2.8. Remarques sur la phonologie	3.1. Conséquences morphologiques	3.2. Traditions académiques	3.3. L'indicatif			3.4. L'impératif ; lexhonoratif
1.2. Les Hayu et leurs voisins	2.1. Syllabe	2.2. Phonologie du monosyllabe	3.5. Participes et gérondifs		4.1. Nominaux		4.2. Mots et morphèmes grammaticaux		5.1. La proposition : transitivité et arguments nucléaires		
	3.6. Gérondifs										
	3.7. Verbes à racines alternantes et verbes irréguliers										
1.3. Les traditions hayu	2.3. Phonologie du mot polysyllabique	3.8. Les paradigmes		5.2. Le syntagme verbal				5.4. Arguments périphériques		5.5. Propositions subordonnées	
		3.9. Morphologie ancienne des initiales de racine									
1.4. La situation linguistique au Népal	2.4. Combinaisons à la frontière entre racine verbale et suffixe	3.10. Eléments et structure de la conjugaison verbale		5.3. Le syntagme nominal				5.6. Compléments de phrase		5.9. Modalités de phrase ; particule de phrase	5.10. Particules rhétoriques et ordre des mots
5.7. La proposition relative											
5.8. La coordination de											
1.5. La classification des langues tibéto-birmanes au Népal											
1.6. Etat											

Sélectionnez un niveau de profondeur hiérarchique : ↕

- Comparaison de tables de matières et de structuration de grammaires sur un sous-corpus (selon critères sélectionnées dans les métadonnées)
- Recherche terminologique : première apparition de termes, distribution géographique de termes....
- Visualisation de la proportion de la grammaire accordée à différents sous-domaines ou thématiques : évolution du contenu des grammaires (apparition de sections sur la syntaxe, sur la sémantique, sur les propositions relatives, sur le mouvement associé....) → tendances en typologie
- Evolution de l'appareil descriptif : glosage interlinéaire, abréviations, collections de textes, de glossaires....

Conclusion

Deux projets très différents, mais avec quelques points communs :

- basés sur l'idée de comparaison de données semblables alignées :
 - dans un cas, les expressions linguistiques codant (dans différentes langues) des événements communs dans le cadre d'un narratif ;
 - dans l'autre, les tables de matières des grammaires descriptives
- visée typologique :
 - dans le cas du CCK, pouvoir comparer des matériaux natifs et naturels pour découvrir les différences lexicales, morphosyntaxiques, constructionnelles entre les langues du corpus... ;
 - dans le cas de Taxogram, pouvoir comprendre les influences sur les choix structurants et terminologiques sur les grammairiens, et les prendre en compte lorsqu'on utilise ces grammaires pour faire de la typologie linguistique

MERCI DE VOTRE ATTENTION
