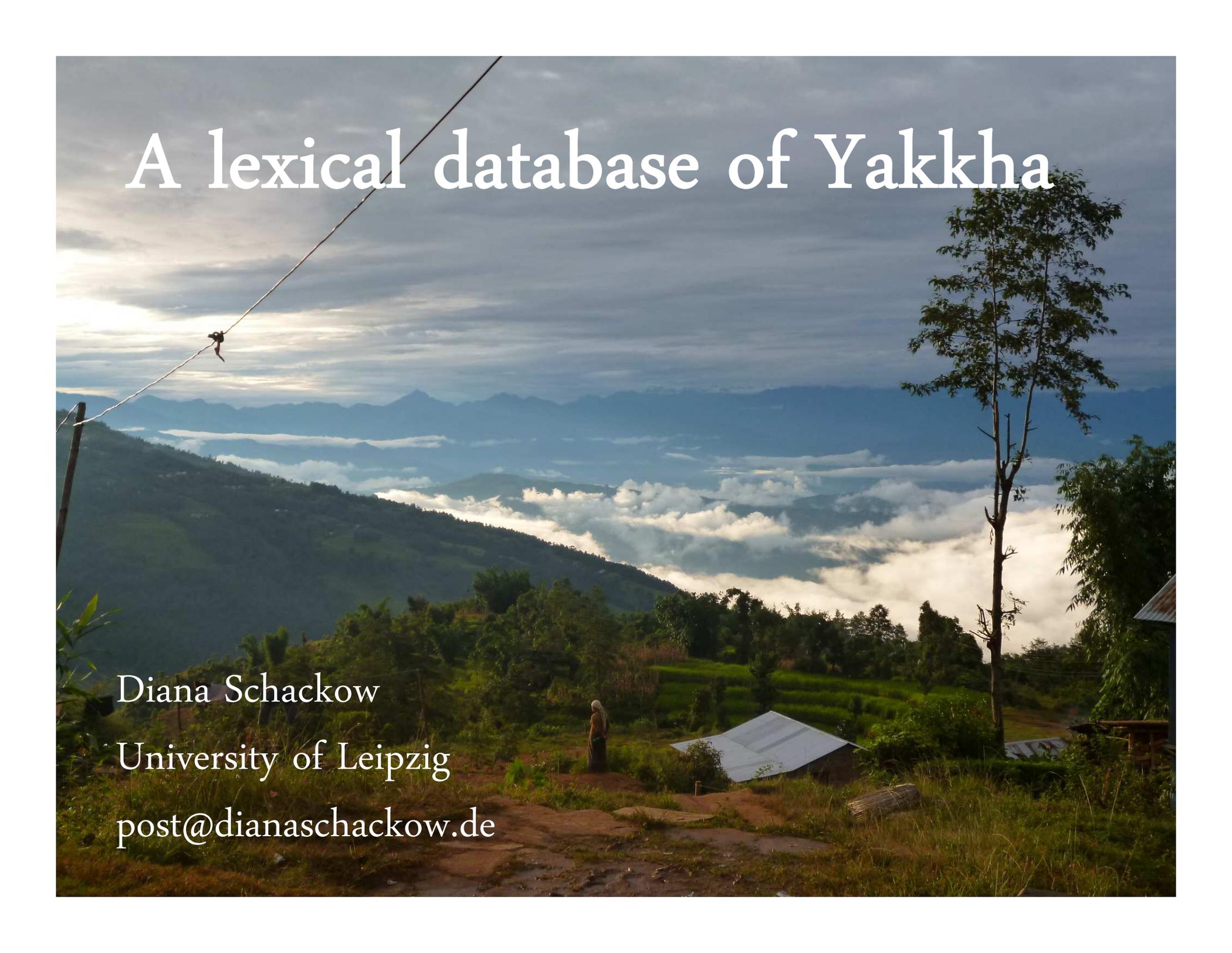


A lexical database of Yakkha

A scenic landscape photograph of a mountainous region. In the foreground, a person is standing on a dirt path. To the right, there is a small building with a corrugated metal roof. In the middle ground, a person is seen walking across a rope bridge that spans across a valley. The background features rolling hills and mountains under a cloudy sky. The overall scene is lush and green, with a mix of natural and man-made elements.

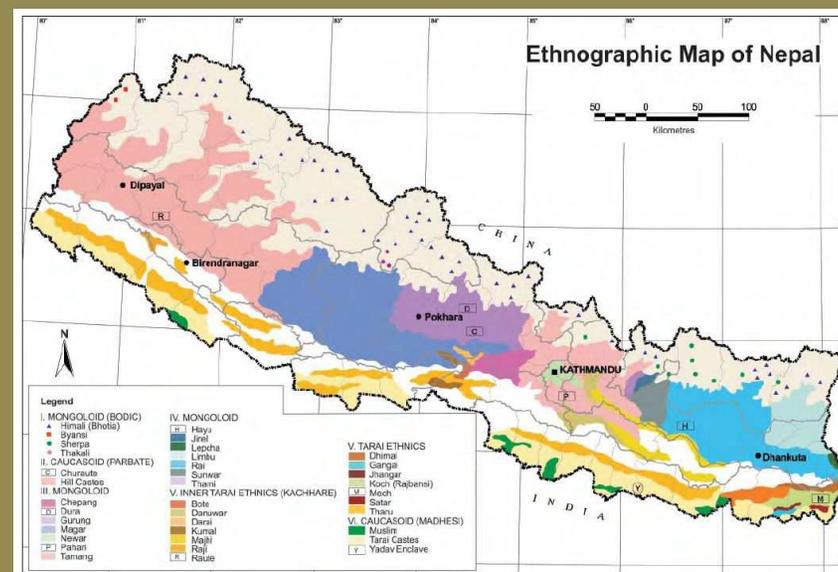
Diana Schackow

University of Leipzig

post@dianaschackow.de

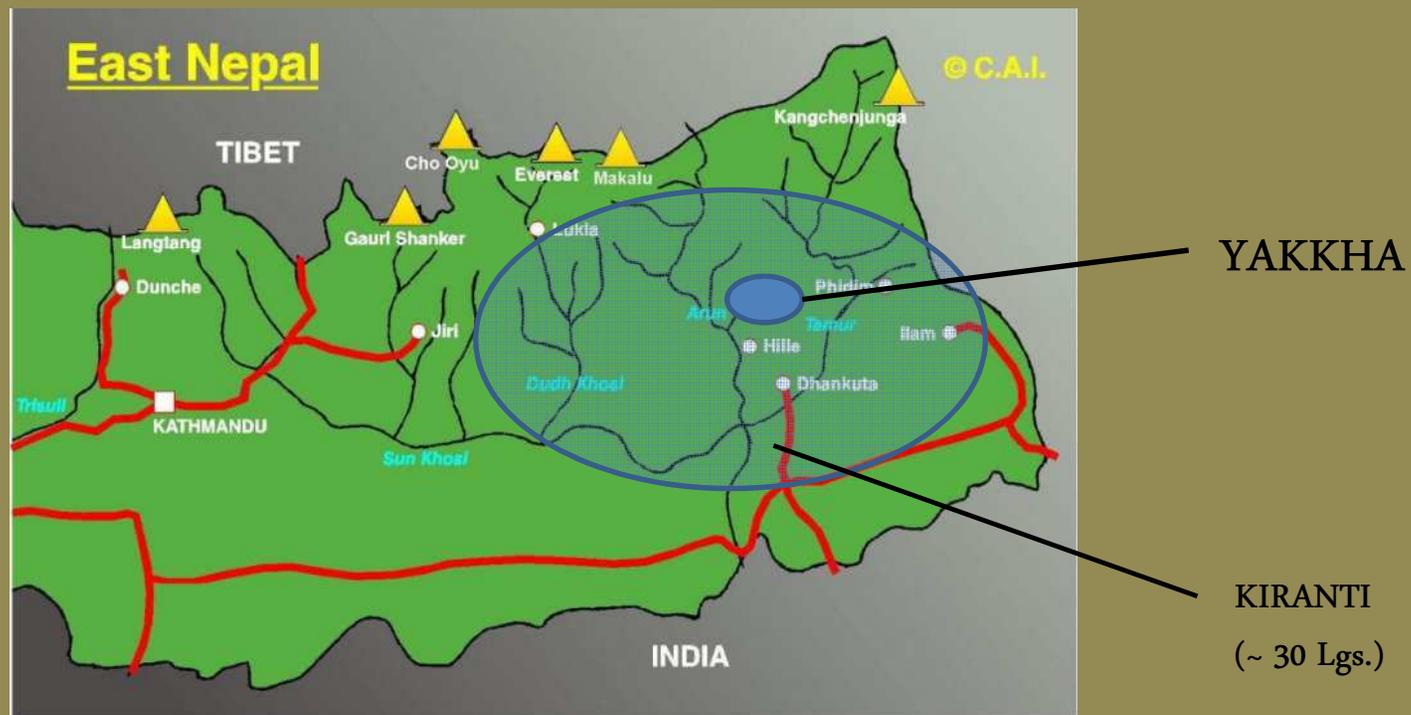
1 Introduction: the language and its speakers

- Ethnic & linguistic diversity in Nepal (~100 lgs. & ethnic groups) dominated by Hindu culture for over 200 years
- Ethnic consciousness on the rise since the end of the monarchy (2008)



Map: Turin (2007). Linguistic Diversity and the Preservation of Endangered Languages

1 Introduction: the language and its speakers



Map: caingram.info

1 Introduction: the language and its speakers

- 2001 census: 14.000 speakers, 17.000 ethnic Yakkha
(numbers seem highly optimistic, language not transferred to the next generation)
- all speakers are at least bilingual (Nepali)
- 4 dialects

1 Introduction: the language and its speakers

- Means of subsistence: agriculture (paddy, millet, corn, cardamom, tea etc.), hunting, fishing
- British army (Gorkha regiments)
- Migration abroad for work/education
 - > many community members outside the homeland, but with internet access



1 Introduction: the language and its speakers

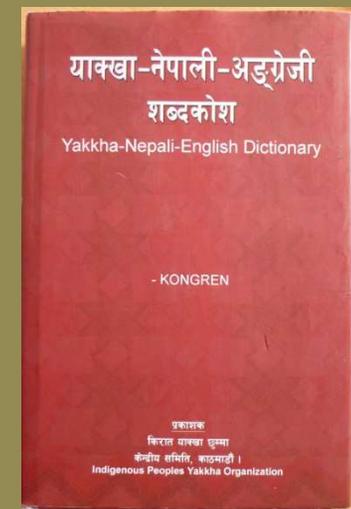
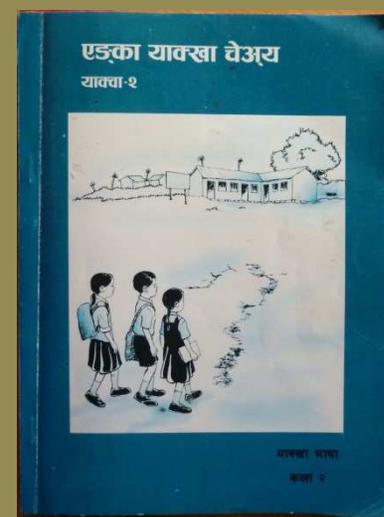
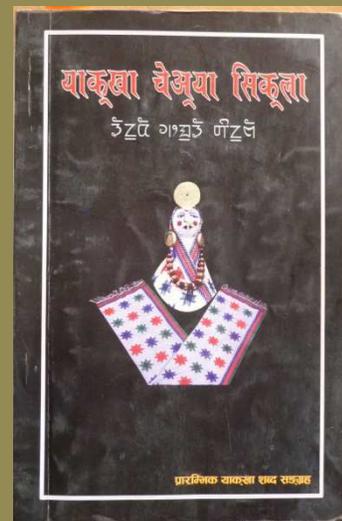
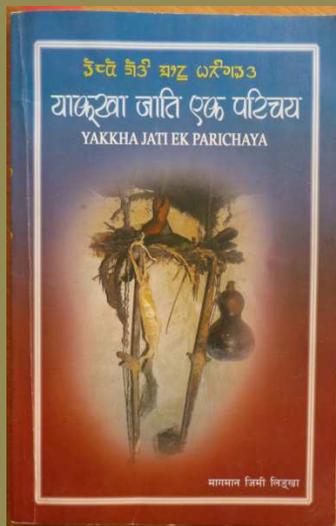
- **Kirant Yakkha Chumma Association:**

social and political activities, workshops, social events

(branch offices in several villages, in Kathmandu and in Ashford, UK)

1 Introduction: the language and its speakers

- school books, a dictionary, ethnographic introductions



2 Structure of the lexicon

- collection and organization of the data with Toolbox & ELAN
- conflicting purposes:
 - (a) user dictionary
 - (b) repository for morpheme-by-morpheme text annotation
- constantly growing, getting reorganized, categories refined or discarded

2 Structure of the lexicon

- `\lex` basic morpheme
- `\lexdev` basic morpheme, Devanagari transcr.
- `\check` own note on progress
- `\exydev` example, Devanagari transcription
- `\exylat` example, Latin transcription
- `\exyeng` example, English translation
- `\ct` citation form
- `\alt` allomorphs
- `\intstr` internal structure of complex words
- `\crossref` crossreferences to complex words, synonyms...
- `\ps` part of speech
- `\ge` English gloss
- `\gn` Nepali gloss
- `\lg` language
- `\ety` etymological notes
- `\gram` grammar notes, e.g. argument structure, case requirements, synt. context
- `\phon` phonology notes
- `\eth` ethnographic notes
- `\sem` semantic details
- `\bzn` botan./zool. name
- `\bzns` source for botan./zool. name
- `\ex` examples with translations
- `\src` person or recording
- `\qst` open questions
- `\pc` picture
- `\sound` sound file
- `\ed` editor
- `\dt` date

2 Structure of the lexicon

Basic information

- `\lex` basic form
- `\ct` citation form (mostly for INF of verbs)
- `\alt` allomorphs
- `\ps` part of speech
- `\ge` English gloss
- `\gn` Nepali gloss
- `\lg` language
- `\ex` examples with translations (mainly for the verbs)

2 Structure of the lexicon

Metadata

- `\check` own note on progress, to filter out completed entries
- `\src` consultant ID or recording ID
- `\qst` open questions
- `\ed` edited by (in case later more people will be working on it)
- `\dt` date last modified

2 Structure of the lexicon

Linguistic and ethnographic details

- `\intstr` internal structure of complex words
- `\crossref` crossreferences to complex words, synonyms, paradigmatically connected words
- `\ety` etymological notes
- `\gram` grammar notes, e.g. argument structure, inalienable possession, case requirements
- `\phon` phonology notes
- `\eth` ethnographic notes
- `\sem` semantic details
- `\bzn` zoological/botanical name
- `\bzns` source for zoological/botanical name

2 Structure of the lexicon

Additions for community dictionary (online/print)

- `\lexdev` lemma, Devanagari transcription
- `\exydev` example, Devanagari transcription
- `\exylat` example, Latin transcription.
- `\exyeng` example, English translation
- `\semcat` semantic category
- `\pc` photo ID

2 Structure of the lexicon

- which form to choose as lemma depends on the purpose of the dictionary (scientific vs. community-usable)
- \lex basic morpheme/ underlying form
- \lexdev basic morpheme, Devanagari transcr.
- \ct citation form

2 Structure of the lexicon

- lemma: citation form vs. underlying form

<i>amum</i>	,grandmother‘	(<i>a-mum</i>)	,1sg.POSS-grandm.‘
<i>cina</i>	,cold‘	(<i>cis-na</i>)	,get.cold-NMLZ‘
<i>phoŋma</i>	,to split‘	(<i>phoks-ma</i>)	,split-INF‘
<i>tiŋdri?ma</i>	,lead to distant place‘	(<i>tikt-tis-ma</i>)	,lead-place-INF‘

3 Examples

Complex verbs:

- `\lex` *tiᅇdra?ma*
- `\exydev` आकाम्निबाक्चि हिप्पाड तित्तुर्वाड्चिड ।
- `\exylat` *akamnibakci hippaᅇ tikturwaᅇciᅇha.*
- `\exyeng` I will bring two friends with me.
- `\alt` *tiᅇ*
- `\intstr` *tikt ra?*
- `\crossref` *tiknamma, tiᅇdri?ma*
- `\ps` vt
- `\ge` bring.along; lead.towards
- `\gn` लिएर आउनु
- `\lg` Y
- `\ety` VV
- `\sem` bring people to deictic center from further away, and from any height level
- `\ex` *ka magman tikturaᅇna – I brought magman with me. akamnibakci hippaᅇ tikturwaᅇciᅇ – I will bring two friends with me. tiktarayagana - He brought you.*

3 Examples

Nouns in composite predicates:

<code>\lex</code>	<i>niṅwa</i>
<code>\lexdev</code>	निङ्वा
<code>\alt</code>	<i>niṅma</i>
<code>\crossref</code>	<i>niṅwaṅacama, niṅwalapma, niṅwakhonḍiṅma, niṅwayuṅma, niṅwahupma, niṅwakaṅma, niṅwalomma, niṅwaṅiṅma, niṅwawama, niṅwaṅaṅapma, niṅwatukma, niṅwapiṅma, niṅwachiṅma</i>
<code>\ps</code>	n
<code>\ge</code>	thought; idea; mind; wish
<code>\gn</code>	मन; इच्छा
<code>\lg</code>	Y

3 Examples

Botanical terms:

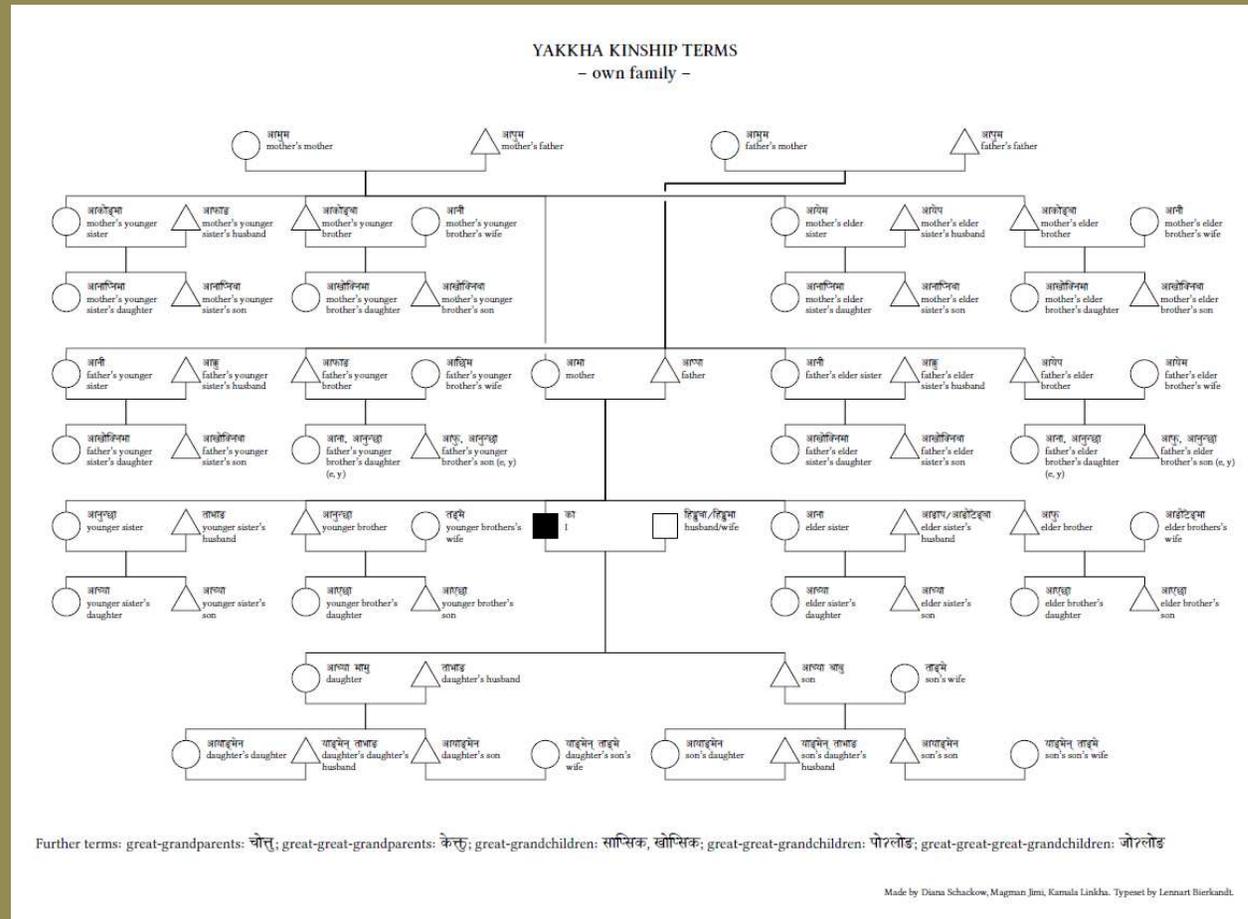
\lex	<i>phimphruwa</i>
\lexdev	फिमफ्रुवा
\ps	n
\ge	chinese.soap.berry
\gn	रिठ्ठा
\lg	Y
\eth	for making soap
\bzn	Sapindus mukerossi
\bzns	Turner: A comparative and etymological dictionary of the Nepali language, N.P. Manandhar: Plants and People of Nepal
\sem	flora

4 Semantic categories

- **body** (body parts, bodily substances, illnesses, wound types)
- **colour**
- **experiential verb**
- **fauna**
- **food** (dishes, plant names, alcoholic beverages)
- **flora** (including fungi)
- **geography** (toponyms, but also geographic, meteorological terms)

4 Semantic categories

- kinship



4 Semantic categories

- **material culture** (household, agriculture, arts and crafts)
- **spatial orientation** (adverbs, motion verbs)
- **supernatural world** (deities, mythological beings and places, types of spells)
- **temporal reference** (adverbs, postpositions)
- **tradition** (festivities, rituals, customs)

5 Outcome - Lexique Pro

The screenshot shows the 'Yakkha - Lexique Pro' application window. The title bar reads 'Yakkha - Lexique Pro'. Below the title bar is a menu bar with 'File', 'Edit', 'View', and 'Help'. A toolbar contains icons for 'Back', 'Forward', 'Home', and 'Pronounce'. The main interface is divided into three sections: a left sidebar with a tree view of categories, a top navigation bar with 'English', 'Nepali', and 'Part of Speech' tabs, and a main content area. The 'Nepali' tab is selected, and the 'fauna' category is highlighted in the sidebar. The main content area displays a list of Nepali words related to fauna, each followed by its English translation and a small icon. The words listed are: आइटे?वान्वाक, आ?वा, छुयुमाइ, चोड्डुवाक, धारावोङ्गि, ढोङ्गुक्माइ, घेङ्चुक, गोगोवा, हेङ्लिक, हिवुम्वा, होकालिक्, होङ्लिङ्वा, होप, इ?वामा, खानुमा, and खेलेक. At the bottom of the window, it shows 'Yakkha: 2402 words' and 'Category: 90 words'.

Yakkha - Lexique Pro

File Edit View Help

Back Forward Home Pronounce

English Nepali Part of Speech

Yakkha Categories Search

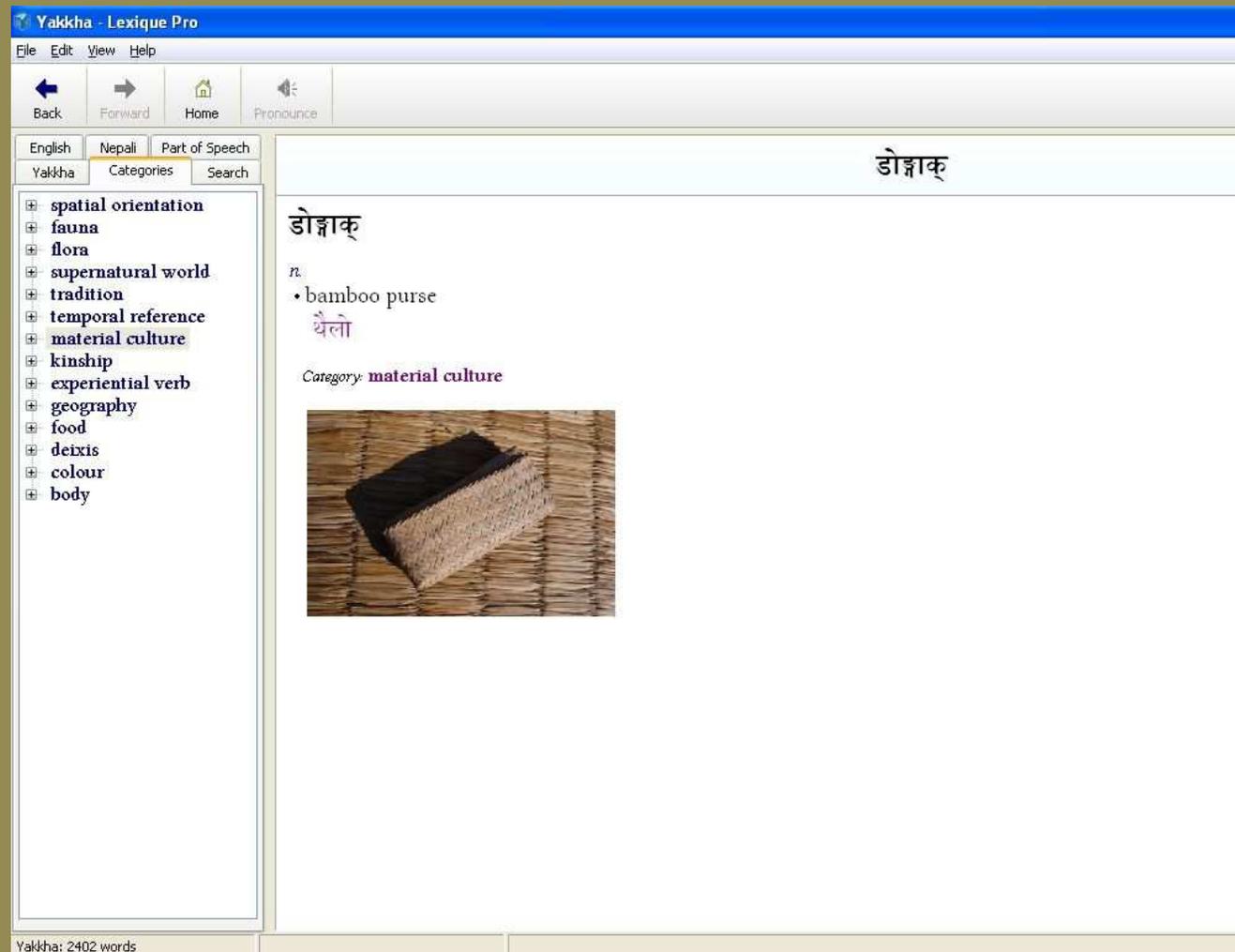
spatial orientation
fauna
flora
supernatural world
tradition
temporal reference
material culture
kinship
experiential verb
geography
food
deixis
colour
body

fauna

आइटे?वान्वाक *n.* a kind of bird; टुपौवा चारा.
आ?वा *n.* crow; काग.
छुयुमाइ *n.* mole; shrew; छुचुन्दो.
चोड्डुवाक *n.* grasshopper; फट्टयाड्डा.
धारावोङ्गि *n.* dragonfly; गाइने किरा.
ढोङ्गुक्माइ *n.* praying mantis; आखाफोर कीरा.
घेङ्चुक *n.* king crow; black bird; चिभे.
गोगोवा *n.* a kind of insect; बिराली.
हेङ्लिक *n.* a kind of insect; रातो सिन्दुरे कीरा.
हिवुम्वा *n.* dung beetle; गुहेकिरा.
होकालिक् *n.* lizard; छेपारो.
होङ्लिङ्वा *n.* a kind of fish; खोलाको पानीमा बस्ने गडेरा.
होप *n.* nest; गुड.
इ?वामा *n.* hornet; कच्छु; अरिहाल.
खानुमा *n.* an insect; सुम्ले किरा.
खेलेक *n.* ant; मसिना; कमिला.

Yakkha: 2402 words Category: 90 words

5 Outcome - Lexique Pro



The screenshot displays the 'Yakkha - Lexique Pro' application window. The title bar reads 'Yakkha - Lexique Pro'. Below the title bar is a menu bar with 'File', 'Edit', 'View', and 'Help'. A toolbar contains icons for 'Back', 'Forward', 'Home', and 'Pronounce'. The main interface is divided into several sections:

- Language Selection:** 'English', 'Nepali', and 'Part of Speech' tabs are visible. Below them are 'Yakkha', 'Categories', and 'Search' buttons.
- Category List:** A vertical list of categories with expandable icons: spatial orientation, fauna, flora, supernatural world, tradition, temporal reference, material culture, kinship, experiential verb, geography, food, deixis, colour, and body.
- Search Results:** The word 'डोङ्गाक्' is displayed in the top right. Below it, the word is repeated in a larger font. The entry includes the part of speech 'n.', the English translation 'bamboo purse', and the Yakkha word 'येलो'. The category 'material culture' is highlighted in purple.
- Image:** A photograph of a rectangular, woven bamboo purse resting on a woven mat.
- Status Bar:** At the bottom left, it says 'Yakkha: 2402 words'.

5 Outcome - Lexique Pro

Yakkha - Lexique Pro

File Edit View Help

Back Forward Home Pronounce

English Nepali Part of Speech
Yakkha Categories Search

- spatial orientation
- fauna
- flora
- supernatural world
- tradition
- temporal reference
- material culture
- kinship
- experiential verb
- geography
- food
- deixis
- colour
- body

छवा

n.
• sugarcane

उखु

Category: [flora](#), [food](#)



Yakkha: 2402 words

5 Outcome - Lexique Pro

- Problems:
 - only 1 font definable for index
 - search will only find the string at the beginning of a lemma, e.g. ‚small knife‘ will not be found with a search term ‚knife‘
 - predefined structure, sometimes difficult to adjust to own requirements

5 Outcome - print dictionary

Yakkha-Nepali-English Dictionary

याक्खा-नेपाली-अंग्रेजी शब्दकोश

Written by

Diana Schackow

With the help of

Kamala Linkha (Jimi)

Manmaya Linkha (Jimi)

Magman Linkha (Jimi)

Mohan Khamyahang (Jimi)

Typeset by

Lennart Bierkandt

Draft, December 2012

5 Outcome - print dictionary

इ॒लो *n* त्यो हौ; के पो watchamacallit
इ॒लो॒ना *ptcl* त्यो के; के पो; what to
say; watchamacallit

इ॒वामा *n* वच्छयुं; अरिङ्गल hornet

इक्मा (इक्) *vt* चिजविज भेला पार्नु;

उठाउनु collect; assemble

फोया चावाक इगुइहा । - I swept up
the dried rice.

इक्मा (इत्त्) *vt* खेद्रु; पछ्याउनु follow;
chase

मिनुमाडा मिमा इत्तुक्सुना । - The cat
chased the mouse.

इक्लाम *n* किल्लिकले; रुद्रघण्टी throat

इक्लाम्साड्मा *vt* स्वर खोलनु clear

throat; harrumph

डिक्लाम साक्सुगार्ई! - Clear your
throat!

इक्नेम्मा (इत्त्-एन्द्) *vt* तल खेदाउनु
chase down

डित्तेन्दुक्सुचिहा । - They have chased
them down.

इखिड *pro* कति how much; how many

chase here

योन्दाड डित्तुभेसुक्सुचिहा । - They have
chased them here from over there.

इङ्मा (इक्स् ~ इङ्) *vt* जांड छपनीले
छत्तु; निचोर्नु; [घाटीमा] निचोर्नु

squeeze; crush; strangle

आमाडा चुवा इक्सुहा । - My mother
filtered the beer (pressing the millet
mash through a sieve).

इछाबा *n* सम्दी son-in-law's father;
daughter in law's father

इछामा *n* सम्दीनी son-in-law's mother;
daughter-in-law's mother

इजाड *adv* किना why

इन *n* अण्डा egg

इन्चामा <इन्; चा> *vt* वेचनु; बिक्री गर्नु;
किन्नु sell; buy and eat

आप्पाडा पिक इनुचोना । - Father sold
a cow.

इन्चामा (इस्-चा) *vi* खेलनु play

फुआमा चाराम्बे इसाचायाना । - Kanchi
played in the yard.

6 Using Devanagari

- Ensuring maximal usability for linguists and community: entries and examples have to be in Latin & in Devanagari
- Note: different alphabetical order in Devanagari
- Grapheme for glottal stop not part of the standard fonts, defined manually (i.e., fonts have to be embedded)

7 Desiderata

- Links to sound files/text corpus
- Occurrence of lexemes in discourse context:
proverbs, songs, jokes, collocations
(current corpus not nearly representative)
- Links to information outside the dictionary:
paradigms, kinship tables, botanical encyclopedic websites (e.g. tropicos.org, sciweb.nybg.org), even youtube videos imaginable
(problem of stability...)