

VARIATION IN LANGUAGE DEVELOPMENT

Elena Lieven & Sabine Stoll
Max Planck Institute for Evolutionary Anthropology



Outline of the course

- Session 1: Introduction, questions, background, data (Sabine Stoll)
- Session 2: Crosslinguistic studies of infant speech perception and production (Elena Lieven)
- Session 3: Composition of early vocabulary: nouns and verbs (Sabine Stoll)
- Session 4: The communicative environment, input and uptake (Elena Lieven)
- Session 5: Crosslinguistic corpus studies (Elena Lieven)
- Session 6: Crosslinguistic experimental studies (Elena Lieven)
- Session 7: Acquisition of tense and aspect (Sabine Stoll)
- Session 8: Acquisition of ergativity (Sabine Stoll)

Outline for Session 1

- Major questions of the field
- Prerequisites for language learning (first year of life)
- Diversity in acquisition tasks:
 - Languages:
 - Grammatical features and tasks of the children
 - Sampling issues
 - Methods of data collection:
 - Diaries
 - Questionnaires
 - Experiments
 - Longitudinal studies
 - Data collection of Chintang, an endangered language of Eastern Nepal

Major questions

- Are there some universal acquisition principles? What is innate?
- What are the factors responsible for the order of acquisition (general cognitive development, language specific factors, cultural environment of child rearing etc.)?
- Are the strategies children use in learning a language more similar within a specific language than across languages?
- If there are different strategies, do they depend on the structure of the language?
- What role does the input and the cultural context play for the acquisition process?

Some observations

- Huge diversity in the languages of the world.
- Different languages pose different challenges to acquisition (Slobin).
- “One cannot study universals without studying particulars.” (Slobin, 1985: 4)
- Crosslinguistic research as a method to reveal both universals and language-specific patterns (Slobin, 1982).

Why is variation an important topic?

- What is universal in language acquisition is one of the major questions, to find out about this, we need to look at variation.
- Predictions of nativist theories: there is no qualitative difference in language acquisition across children.
- Prediction of usage-based theories: qualitative differences are possible and the input has a strong impact on how childrens development looks like.

Variation in Language Acquisition

Within individual languages

- temporal variation (time of acquisition)
- qualitative variation
- context- specific variation

Across languages

- grammatical variation
- cultural variation

Language Comprehension

- Prerequisites for language comprehension:
 - acoustic perception of speech
 - pattern recognition
 - pattern analysis
 - pattern memorization

Prelinguistic development: Prerequisites for language learning

Birth to 6 months

- Recognition of mother's voice (de Casper & Fifer 1980)
- Distinguish native language from other languages (Mehler et al. 1988, Moon et al. 1993)
- Categorical perception of speech sounds (Eimas et al. 1971)
- Recognition of identity of sounds across contexts (Kuhl 1980)
- Segment speech (Saffran et al. 1996)
- Preference for infant-directed speech (Fernald et al. 1989; Cooper & Aslin 1990)

Prelinguistic development

6- 12 months

- Discrimination of phonemic contrasts. Up to approx. 10 months discriminations of all contrasts. Then, only contrast of native language are distinguished. (Werker & Tees 1984)
- By 7 1/2 months children listen longer to familiarized words within longer sentences. (Jusczyk & Aslin 1995)

Prelinguistic development

Abilities around 9 months (e.g. Tomasello 2003)

- Joint attention
- Recognition of symbols
- Imitation (role reversal)
- Intention reading
- Pointing

Development of linguistic abilities: individual variation

- First words (9 months to 1 year of age)
- Strong variation in when children start speaking and how they progress (within a given language)

Table 4.3. Age of acquisition for 10- and 50-word vocabularies in six children

Child	Sex	Lexicon size	
		10 words	50 words
S1	M	1;0	1;5
S2	F	1;1	1;6
S3	M	1;2	1;7
S4	F	1;2	1;8
S5	M	1;4	1;10
S6	F	1;3	1;7

Source: Robb, Bauer, & Tyler 1994:40. Used with permission from Alpha Academic.

Development of linguistic abilities: individual variation

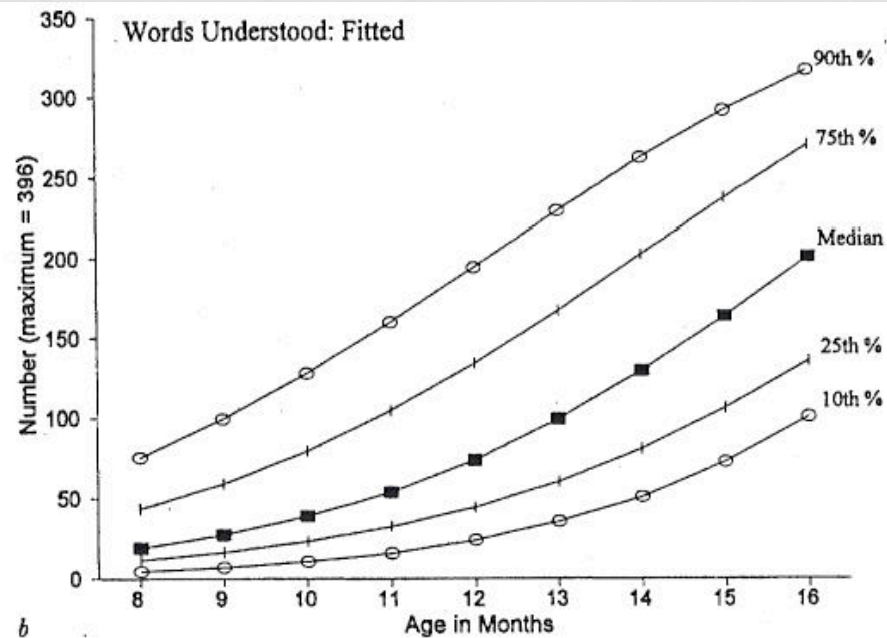


FIG. 2.—Number of words on the Infant form reported to be comprehended children at each month—median values and spread of score distributions. *a*, Observed values. *b*, Fitted values. A portion of this figure is adapted from Fenson et al. (1993, p. 103), with permission of the Singular Publishing Group, Inc.

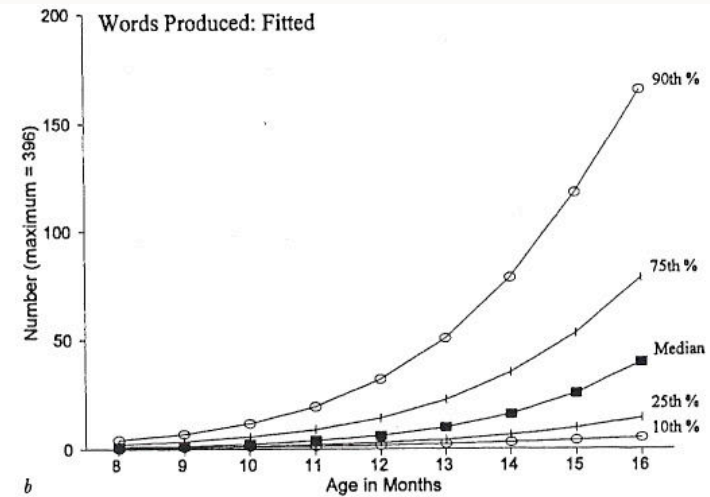
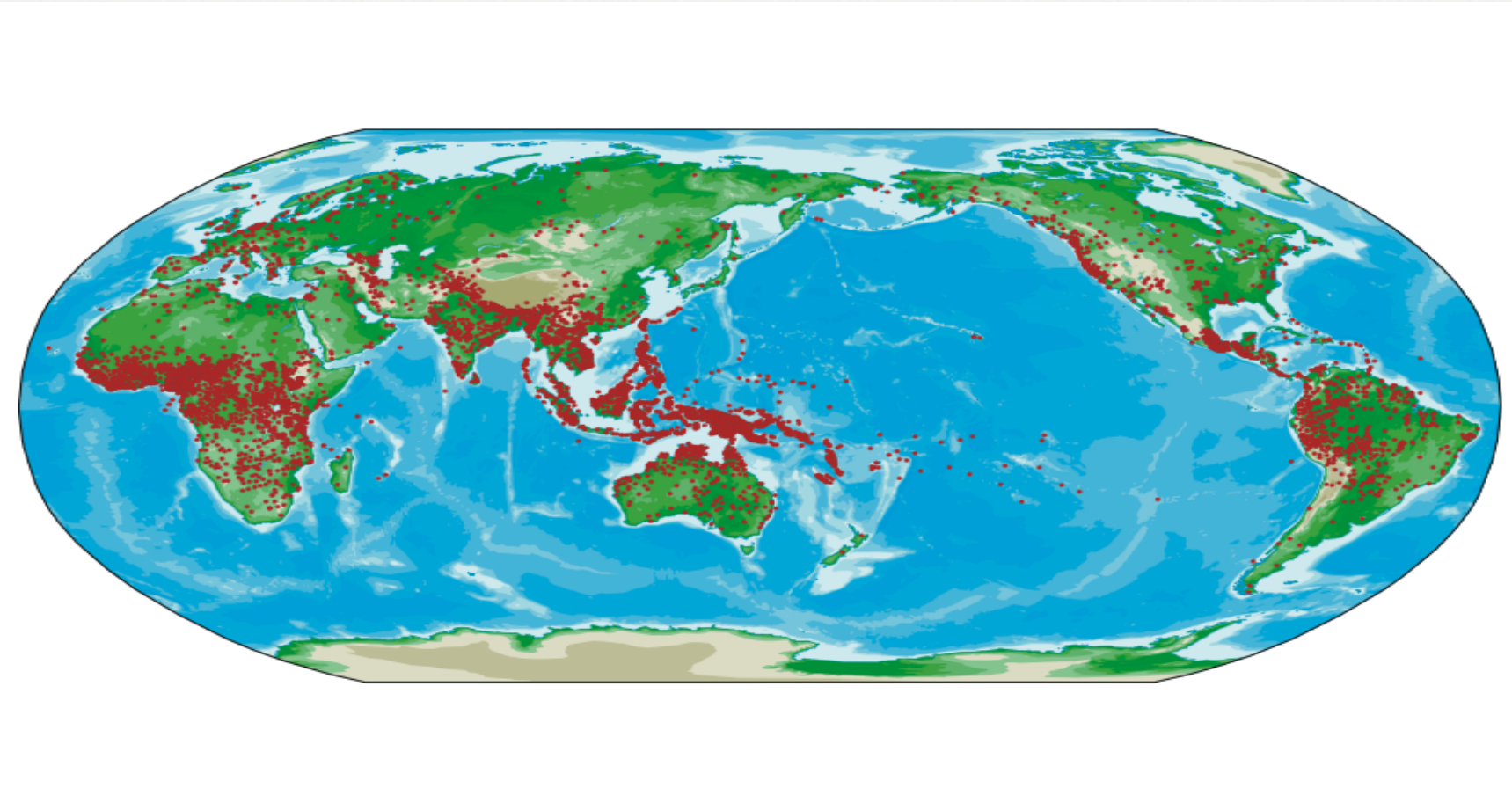


FIG. 3.—Number of words on the Infant form reported to be produced by children at each month—median values and spread of score distributions. *a*, Observed values. *b*, Fitted values. A portion of this figure is adapted from Fenson et al. (1993, p. 104), with permission of the Singular Publishing Group, Inc.

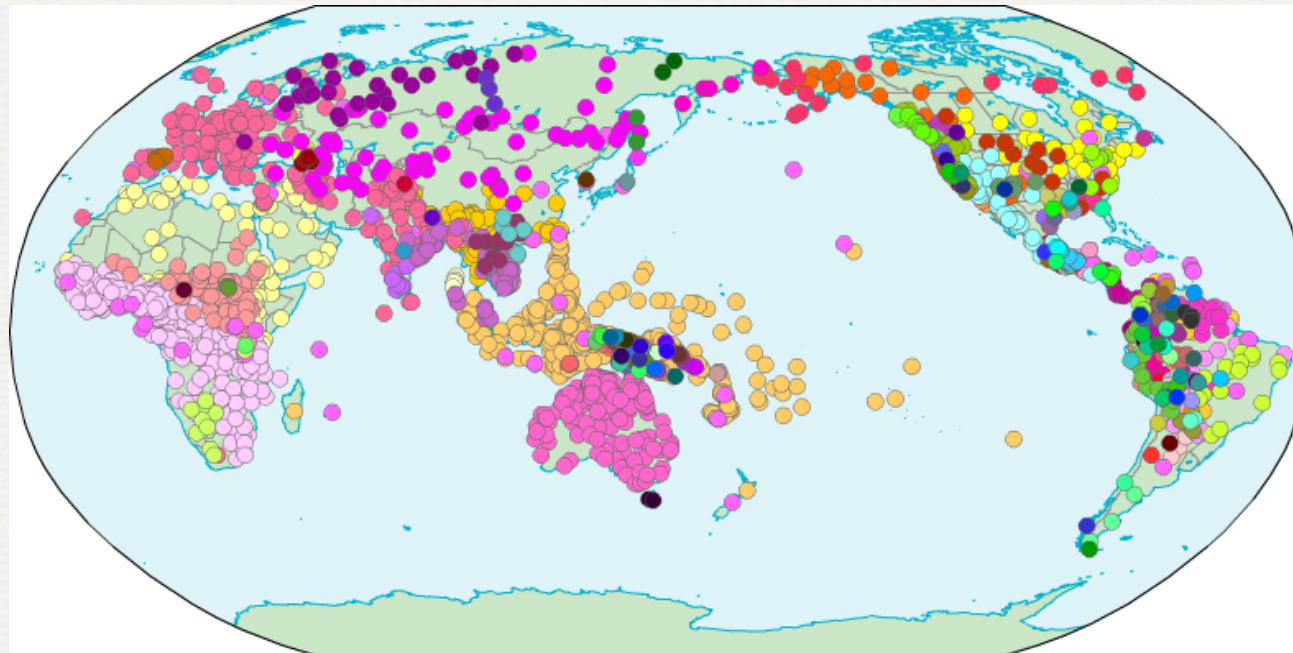
Table 1. *Individual Differences in Language Development: Summary of Claims in the Literature*

Strand 1	Strand 2
Semantics	
High proportion of nouns in first 50 words	Low proportion of nouns in first 50 words
Single words in early speech	Formulae in early speech
Imitates object names	Unselective imitation
Greater variety within lexical categories	Less variety within lexical categories
Meaningful elements only	Use of "dummy" words
High adjective use	Low adjective use
Context-flexible use of names	Context-bound use of names
Rapid vocabulary growth	Slower vocabulary growth
Grammar	
Telegraphic in Stage I	Inflections and function words in Stage I
Refers to self and others by name in Stage I	Refers to self and others by pronoun in Stage I
Noun-phrase expansion	Verb phrase expansion
Morphological overgeneralization	Morphological undergeneralization
Consistent application of rules	Inconsistent application of rules
Novel combinations	Frozen forms
Imitation is behind spontaneous speech	Imitation is ahead of spontaneous speech
Fast learner	Slow learner
Pragmatics	
Object-oriented	Person-oriented
Declarative	Imperative
Low variety in speech acts	High variety in speech acts
Phonology	
Word-oriented	Intonation-oriented
High intelligibility	Low intelligibility
Segmental emphasis	Suprasegmental emphasis
Consistent pronunciation across word tokens	Variable pronunciation across word tokens
Demographic Variables	
Female	Male
Firstborn	Later-born
Higher SES	Lower SES

Potential learning tasks: Languages today (6000-7000)



Language families a subsample of 2560 languages (WALS)



- Andamanese (7)
- Afro-Asiatic (141)
- Sepik (16)
- Northwest Caucasian (7)
- Algic (31)
- Niger-Congo (346)
- Guaicuruan (5)
- West Papuan (13)
- Austronesian (311)
- Arawakan (30)
- Sino-Tibetan (130)
- Tupian (21)
- Mayan (34)
- Nilo-Saharan (87)
- Jivaroan (4)
- Hokan (21)
- Keresan (2)
- other (72)
- Australian (161)
- Indo-European (176)
- Trans-New Guinea (90)
- Nakh-Daghestanian (28)
- Na-Dene (24)
- Ainu (1)
- Cariban (19)
- Muskogean (7)
- Eskimo-Aleut (18)
- Huarpe (1)
- Oregon Coast (3)
- Altaic (65)
- Chukotko-Kamchatkan (5)
- Panoan (11)
- Border (5)
- Harakmbet (1)
- Oto-Manguean (55)
- Andoke (1)
- Anem (1)
- Senagi (2)
- Khoisan (11)
- Macro-Ge (16)
- Zaparoan (3)
- Marind (6)
- Tacanan (4)
- Torricelli (9)
- Caddoan (5)
- Kunza (1)
- Atakapa (1)
- Barbacoan (4)
- Awin-Pare (1)
- Reef Islands - Santa Cruz (2)
- Aymaran (2)

Dryer, 2005

Some examples of variation

Phoneme Inventories (Maddieson):

- Consonant inventories
 - 6 (Rotokas, Papua New Guinea) - 122 (!Xóõ, Southern Khoisan) out of a sample of 562 languages
- Vowel inventories
 - 2 (Yimas (Papua New Guinea) -14 (German)

Some examples of variation

- Inflectional synthesis of the verb (Bickel & Nichols 2005, WALS)
- Degree of synthesis as defined by the number of elements that make up a synthetic verb form
- Large variation from 0 categories per verb form (Vietnamese) to 13 (Koasati)

Some examples of variation

German:

treffen 'to meet'

Chintang

1s	<i>treffe</i> <i>traf</i>
1p	<i>treffen</i> <i>trafen</i>
2s	<i>triffst</i> <i>trafst</i>
2p	<i>trefft</i> <i>traft</i>
3s	<i>trifft</i> <i>traf</i>
3p	<i>treffen</i> <i>trafen</i>
PT	<i>treffend</i> <i>getroffen</i>

	1s	1di	1pi	1de	1pe	2s	2d	2p	3s	3ns	intransitive
1s						<i>tupnaʔā</i> <i>tupnaʔāniŋ</i> <i>tupnehē</i> <i>matupyoknehē</i>	<i>tupnaʔāce</i> <i>tupnaʔāceŋiŋ</i> <i>tupnace</i> <i>matupyoknace</i>	<i>tupnaʔāni</i> <i>tupnaʔāniŋiŋ</i> <i>tupnanihē</i> <i>matupyoknanihē</i>	<i>tubukuj</i> <i>tubukujŋiŋ</i> <i>tubuhē</i> <i>matupyoktuhē</i>	<i>tubukujcuj</i> <i>tubukujcujŋiŋ</i> <i>tubujcihē</i> <i>matupyoktucjihē</i>	<i>tupmaʔā</i> <i>tupmaʔāniŋ</i> <i>tubehē</i> <i>matupyoktehē</i>
1di									<i>tupcoko</i> <i>tupcokoniŋ</i> <i>tubace</i> <i>matupyoktace</i>	<i>tubumcum</i> <i>tubumcumniŋ</i> <i>tubumcumhē</i> <i>matupyoktumcumhē</i>	<i>tupceke</i> <i>tupcekeŋiŋ</i> <i>tubace</i> <i>matupyoktace</i>
1pi									<i>tubukum</i> <i>tubukumniŋ</i> <i>tubumhē</i> <i>matupyoktumhē</i>		<i>tubiki</i> <i>tubikiŋiŋ</i> <i>tubihē</i> <i>matupyoktihē</i>
1de						<i>tupnaʔānciyā</i> <i>tupnaʔānciyāniŋ</i> <i>tupnanciyehē</i> <i>matupyoknanciyehē</i>			<i>tupcokoga</i> <i>tupcokogaŋiŋ</i> <i>tubacehē</i> <i>matupyoktacehē</i>	<i>tubumcumma</i> <i>tubumcummaniŋ</i> <i>tubumcummehe</i> <i>matupyoktumcummehe</i>	<i>tupcekega</i> <i>tupcekegaŋiŋ</i> <i>tubacehē</i> <i>matupyoktacehē</i>
1pe									<i>tubukumma</i> <i>tubukummaniŋ</i> <i>tubummehe</i> <i>matupyoktummehe</i>		<i>tubikiŋa</i> <i>tubikiŋaniŋ</i> <i>tubiehē</i> <i>matupyoktiehē</i>
2s	<i>atupmaʔā</i> <i>atupmaʔāniŋ</i> <i>atubehē</i> <i>{a-ma}tupyoktehē</i>			<i>{a-ma}tupceke</i> <i>{a-ma}tupcekeŋiŋ</i> <i>{a-ma}tubace</i> <i>{a-ma-ma}tupyoktace</i>	<i>{a-ma}tupno</i> <i>{a-ma}tupnikniŋ</i> <i>{a-ma}tube</i> <i>{a-ma-ma}tupyokte</i>				<i>atuboko</i> <i>atubokoniŋ</i> <i>atube</i> <i>amatupyokte</i>	<i>atubukuce</i> <i>atubukuceŋiŋ</i> <i>atubuce</i> <i>{a-ma}tupyoktuce</i>	<i>atupno</i> <i>atupnikniŋ</i> <i>atube</i> <i>{a-ma}tupyokte</i>
2d	<i>atupmaʔānciyā</i> <i>atupmaʔānciyāniŋ</i> <i>atubajcihē</i> <i>{a-ma}tupyoktacijehē</i>								<i>atupcoko</i> <i>atupcokoniŋ</i> <i>atubace</i> <i>amatupyoktace</i>	<i>atubumcum</i> <i>atubumcumniŋ</i> <i>atubumcumhē</i> <i>{a-ma}tupyoktumcumhē</i>	<i>atupceke</i> <i>atupcekeŋiŋ</i> <i>atubace</i> <i>{a-ma}tupyoktace</i>
2p	<i>atupmaʔāniŋ</i> <i>atupmaʔāniŋiŋ</i> <i>atubajnihē</i> <i>{a-ma}tupyoktajnihē</i>								<i>atubukum</i> <i>atubukumniŋ</i> <i>atubumhē</i> <i>amatupyoktumhē</i>		<i>atubiki</i> <i>atubikiŋiŋ</i> <i>atubihē</i> <i>{a-ma}tupyoktihē</i>
3s	<i>utupmaʔā</i> <i>utupmaʔāniŋ</i> <i>utubehē</i> <i>{u-ma}tupyoktehē</i>	<i>maitupceke</i> <i>maitupcekeŋiŋ</i> <i>maitubace</i> <i>{mai-ma}tupyoktace</i>	<i>maitupno</i> <i>maitupnikniŋ</i> <i>maitube</i> <i>{mai-ma}tupyokte</i>	<i>matupceke</i> <i>matupcekeŋiŋ</i> <i>matubace</i> <i>{ma-ma}tupyoktace</i>	<i>matupno</i> <i>matupnikniŋ</i> <i>matube</i> <i>{ma-ma}tupyokte</i>	<i>natupno</i> <i>natupnikniŋ</i> <i>natube</i> <i>{na-ma}tupyokte</i>	<i>natupceke</i> <i>natupcekeŋiŋ</i> <i>natubace</i> <i>{na-ma}tupyoktace</i>	<i>natubiki</i> <i>natubikiŋiŋ</i> <i>natubihē</i> <i>{na-ma}tupyoktihē</i>	<i>tuboko</i> <i>tubokoniŋ</i> <i>tube</i> <i>matupyokte</i>	<i>tubukuce</i> <i>tubukuceŋiŋ</i> <i>tubuce</i> <i>matupyoktuce</i>	<i>tupno</i> <i>tupnikniŋ</i> <i>tube</i> <i>matupyokte</i>
3d	<i>utupmaʔānciyā</i> <i>utupmaʔānciyāniŋ</i> <i>utubajcihē</i> <i>{u-ma}tupyoktacijehē</i>								<i>utupcoko</i> <i>utupcokoniŋ</i> <i>utubace</i> <i>{u-ma}tupyoktace</i>	<i>utubukuce</i> <i>utubukuceŋiŋ</i> <i>utubuce</i> <i>{u-ma}tupyoktuce</i>	<i>utupceke</i> <i>utupcekeŋiŋ</i> <i>utubace</i> <i>{u-ma}tupyoktace</i>
3p	<i>utupmaʔāniŋ</i> <i>utupmaʔāniŋiŋ</i> <i>utubajnihē</i> <i>{u-ma}tupyoktajnihē</i>								<i>utuboko</i> <i>utubokoniŋ</i> <i>utube</i> <i>{u-ma}tupyokte</i>		<i>utupno</i> <i>utupnikniŋ</i> <i>utube</i> <i>{u-ma}tupyokte</i>

Some examples of variation

German

ge-troff-en

an-ge-troff-en

aber:

**ge-an-troff-en*

**ge-troff-an-en*

Chintang

u-ma-tup-yokt-e-hě

3-NEG-meet-NEG-PAST-1

‘He did not meet me’

or:

ma-u-top-yokt-e-hě

ma-top-u-yokt-e-hě

Assumption so far: Free ordering of affixes is impossible

Some examples of variation

- Word order (Subject, Verb, Object)
- All 6 logically possible orders are attested
 - SOV (Japanese)
 - SVO (Mandarin)
 - VSO (Irish)
 - VOS (Nias, Austronesian)
 - OVS (Hixkaryana, Carib, Brazil)
 - OSV (Nadëb, Brazil)
 - No predominant order

Some examples of variation

- Structural diversity: semantic mapping
 - Inclusive vs. exclusive pronouns

German:	<i>ich</i>	<i>wir</i>
Chintang:	<i>akka</i>	<i>ani ananga</i>

inclusive exclusive

Some examples of variation

- Structural diversity: world knowledge
- Kinship and suffixes (Martuthunira, Australia)

Suffix used to mark kinship. Same generation (brother, sister, grandfather) vs. alternating generation (father mother, great-grandfather). Suffix is used to mark same generation set.

Nyinta wiya-nmayi-nha marrkara-ngarli- ku -rla?

2sgNOM see -**COLL**-PAST brother PLURAL-ACC-PST

Did you see your younger brother?s

Languages today

- Approx. 6000-7000 living languages in the world
- Approx. 300 language families
- Approx. 500-700 languages are described or we know at least something about them
- Approx. 300 variables that typologists have extensive knowledge about

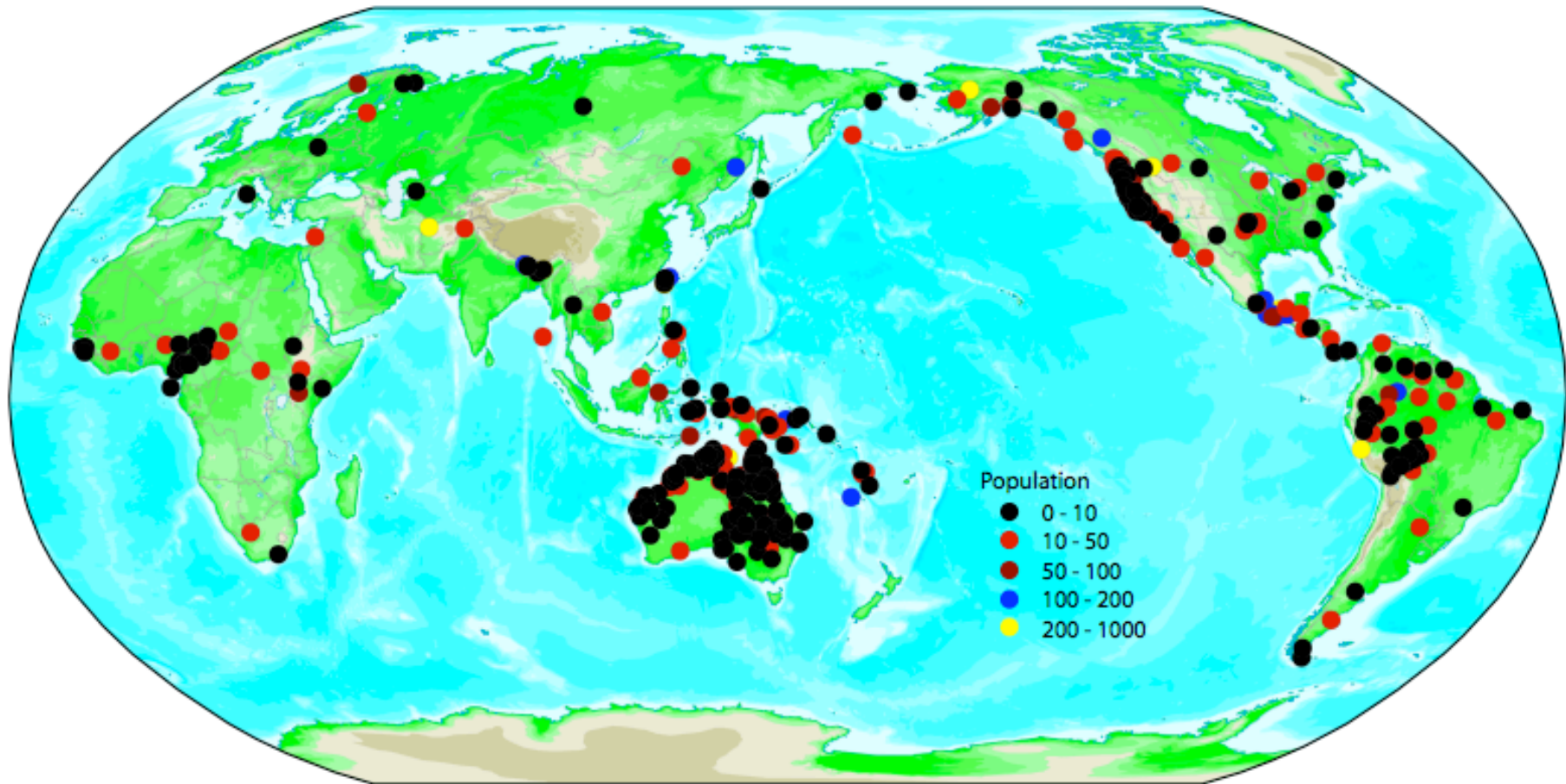
Our knowledge about languages

- Estimation of the minimum number of languages under the following assumptions (Bickel, 2008):
 - on average at every point in time approx. 5000 languages (more realistically: 4000-12000)
 - on average maximally 1000 years per language (criterium of interintelligibility)
 - probable age of modern languages minimum 100'000 years

$$\frac{100,000}{1000} \cdot 5000 = 500,000$$

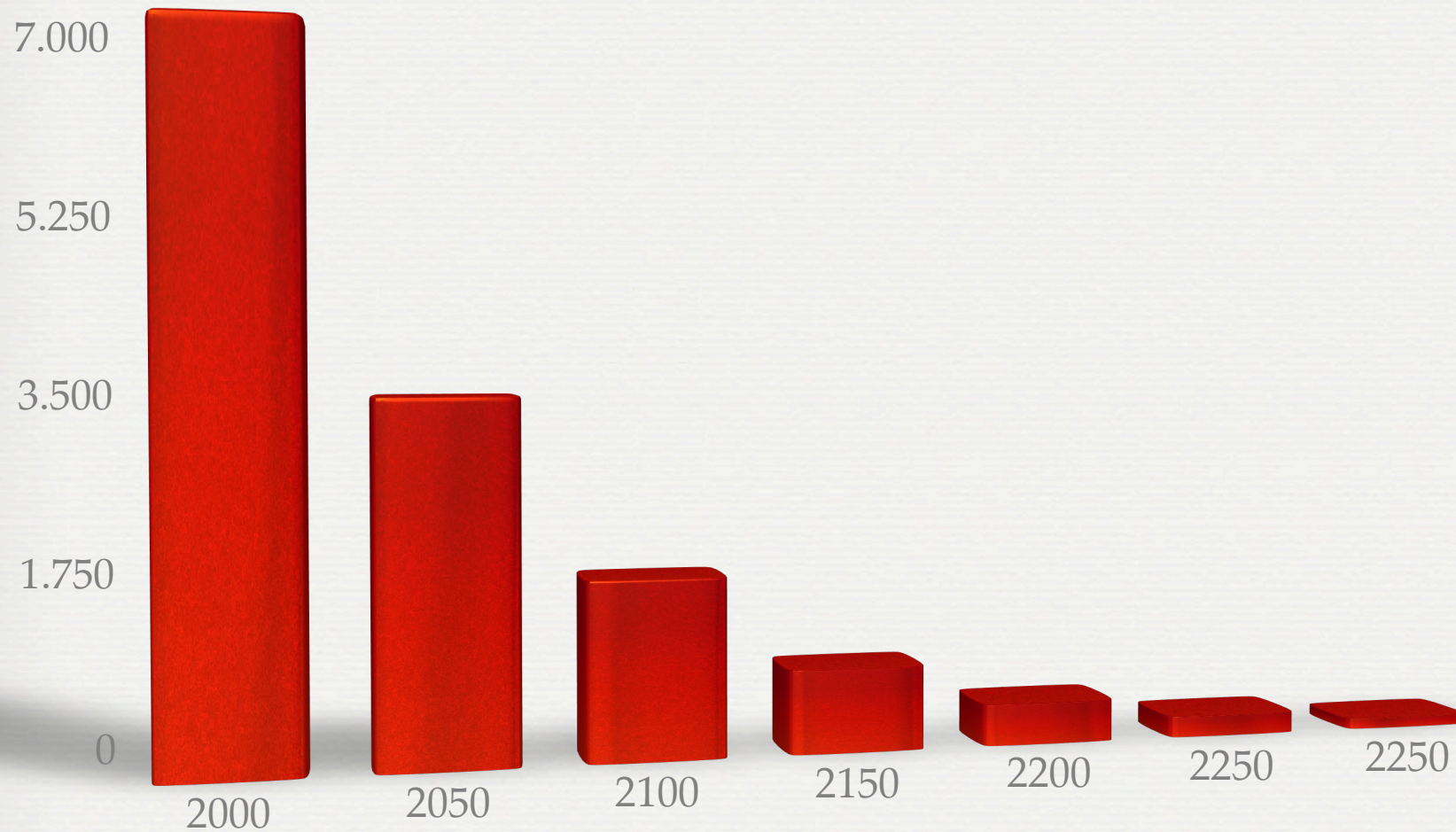
- today's population < 1% of total population (amount of languages)
- of those we know only about 1% (500).
- But soon, we will not be able to learn anything.

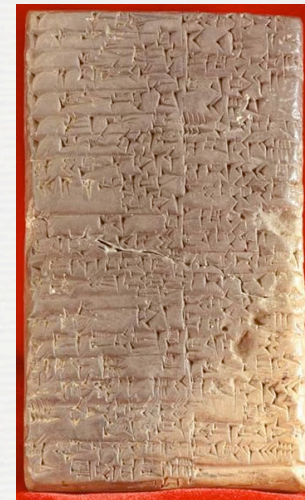
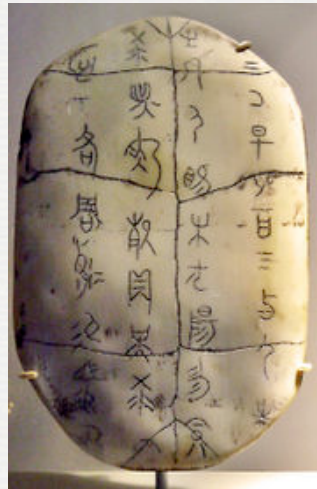
Language death



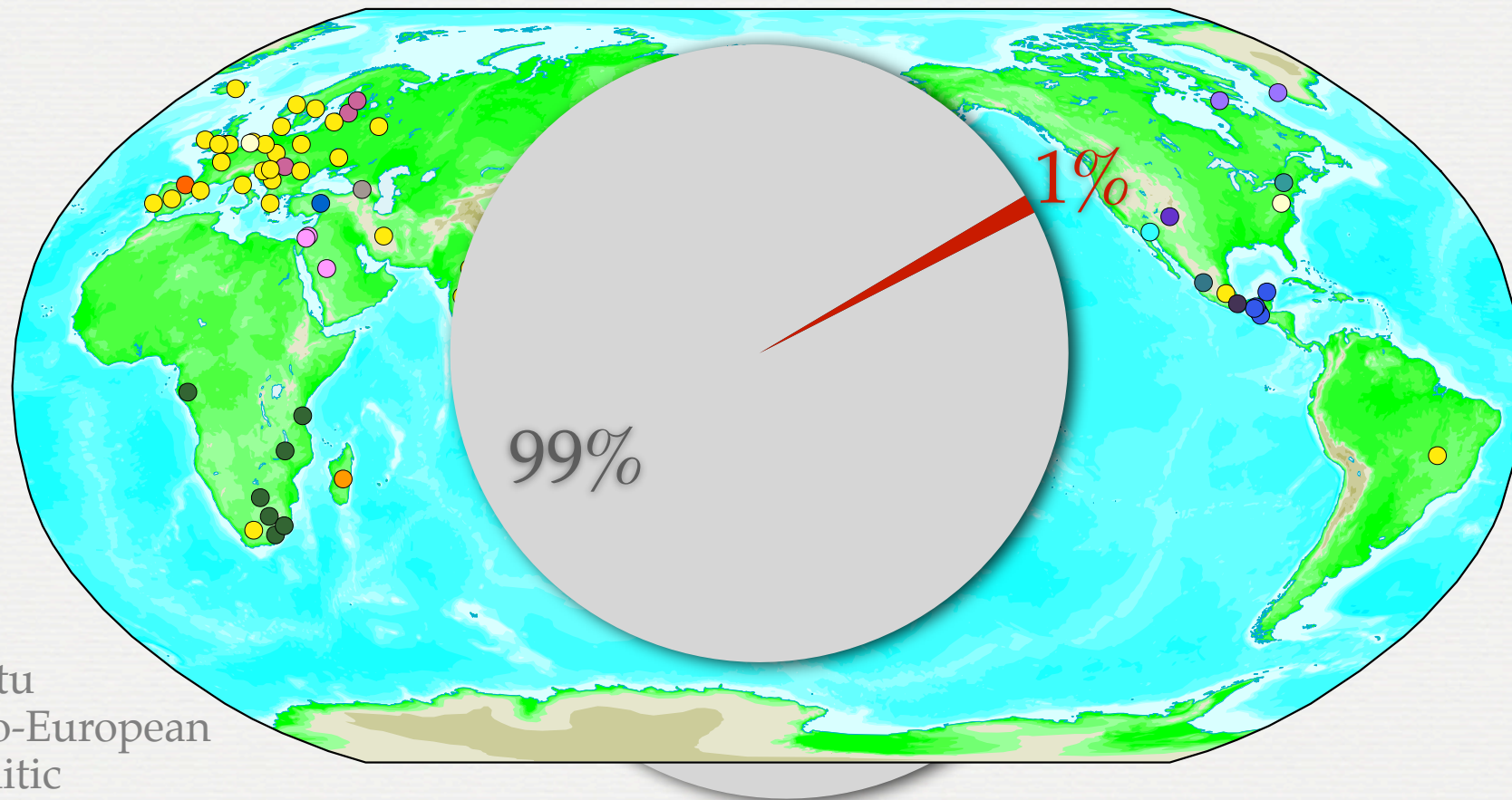
Data from Ethnologue 13, map Bibiko 2006

Language development in the future (estimated death rate of 50% every 50 years):





Survey of languages with acquisition studies



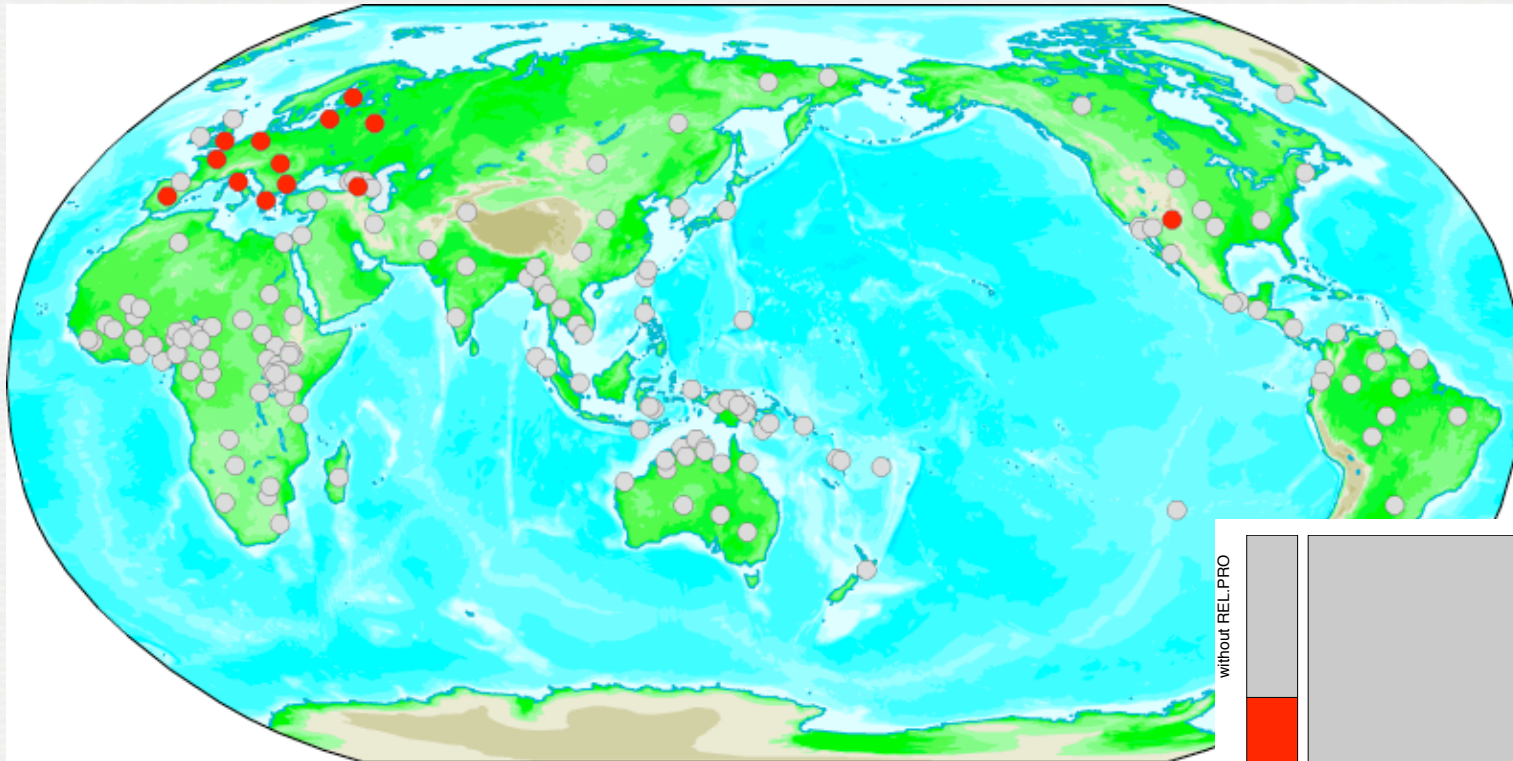
- Bantu
- Indo-European
- Semitic
- Maya
- Sinitic

Lieven & Stoll, DGfS Summer
School, 2010

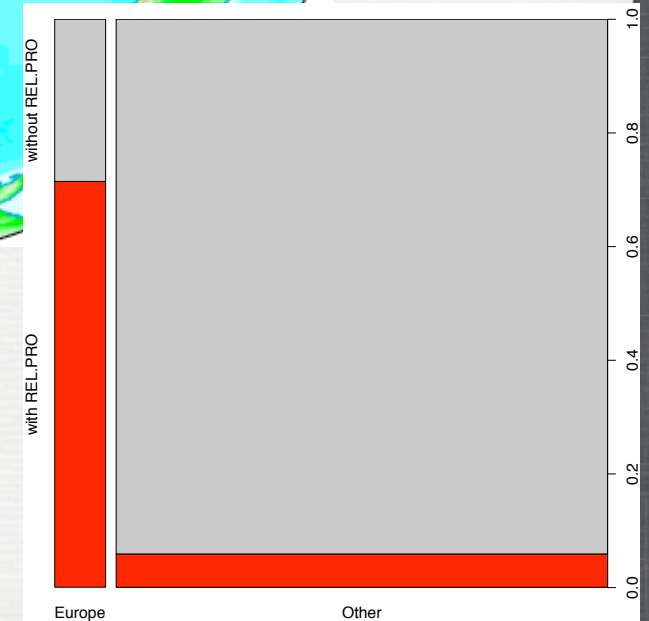
Euroversals (Haspelmath 1998, 2001)

- Definite vs. indefinite article
- Relative pronouns
- 'HAVE'- perfect
- Participle based passive
- External possessor
- 'nobody-came'- constructions
- 'as-big-as'-constructions
- agreement with obligatory NP

Euroversals: Relative pronouns



Data from Comrie & Kuteva 2005,
N=166, EU vs. Rest, $p < .001$; EU vs.
Asia, $p = .05$



Data in language acquisition research

- Diaries (very early in development)
- Questionnaires (MCIDI)
- Experiments
- Corpora
 - Longitudinal corpora
 - Cross-sectional corpora
- Modelling

Diary studies

- German: Preyer (1882), Stern & Stern (1928), Lindner (1885), Leopold (1948)
- French: Gregoire (1937)
- Russian: Gvozdev (1949)
- Polish: Zarebina (1965)
- many modern studies: e.g. Tomasello (1992)

Diary studies

Advantages:

- Every utterance of the child can be noted down in the early phases of acquisition

Disadvantages:

- Only feasible in the very early stages of acquisition
- Difficult to exactly note down the linguistic and extralinguistic context

Questionnaires

Advantages:

- large number of mothers can be tested
- standardized 'test'
- results can be compared easily

Disadvantages:

- cultural issues might apply
- relies on the memory and interpretation of the mother

Experiments

Advantages:

- controlled setting and variables which are the same for all children
- cause and effect can be tested

Disadvantages:

- only one context is tested
- not possible to administer in a lot of cultures

Longitudinal studies

- Audio- or video recording of the child and her caretakers in natural interaction
- Transcription of their speech
- Translation of the data
- Tagging of the corpus (translation, morphological glossing, parts-of-speech glossing)
- Further annotations
- Linking to video- or audio data
- Analysis of the data

Longitudinal studies

- Advantages:
 - spontaneous behaviour of the children is recorded
 - development of individual children can be characterized
 - input of the surrounding adults available for analysis
- Disadvantages:
 - sampling issues
 - potential contextual influences (recording situation)
 - small number of children (what is the norm?)
 - only correlations can be attested


Child Language Data Exchange System (CHILDES)

- <http://chilides.psy.cmu.edu/> (MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.)
- Includes corpora and tools for studying conversational interactions.
 - 130 corpora (30 languages)
 - Tools include:
 - programs for computer analysis of transcripts
 - methods for linguistic coding
 - systems for linking transcripts to digitized audio and video

CHILDES – Child Language Data Exchange System

http://chilides.psy.cmu.edu/

Maps MPI INFO Library MPI Nijmegen cpdp grep help amazon ProBIP Leipzig: E-Z-R Manuscript Central

CHILDES Child Language Data Exchange System 

CHILDES is the child language component of the [TalkBank](#) system. TalkBank is a system for sharing and studying conversational interactions.

<p>System</p> <ul style="list-style-type: none"> Ground rules Guidelines for Contributors Overviews and Introductions Membership list How to subscribe to Mailing Lists 	<p>Programs and Database</p> <ul style="list-style-type: none"> Downloadable Database Browsable Database The CLAN Program WebCLAN Training Videos
<p>Links</p> <ul style="list-style-type: none"> The TalkBank Database IASCL information Other Child Language sites Research based on CHILDES Related Software 	<p>Manuals</p> <ul style="list-style-type: none"> CHAT Transcription CLAN Programs Database Manuals BTS sign transcription system
<p>Phonology and Fonts</p> <ul style="list-style-type: none"> Phon & PhonBank Unicode and IPA for Mac Unicode and IPA for Windows 	<p>Teaching with CHILDES</p> <ul style="list-style-type: none"> Topics in language acquisition. Teaching Tips and Resources. Child Language Bibliographies
<p>Special Procedures</p> <ul style="list-style-type: none"> Procedures and tools for CA analysis Working with digitized video Working with digitized audio The Computerized Comprehension Task 	<p>Morphology and Lexicon</p> <ul style="list-style-type: none"> Part of Speech Analysis by MOR Parental frequency count: view, download, documentation MRC lexical dictionary Syntactic analysis by GRASP
<p>Mirrors</p> <ul style="list-style-type: none"> Antwerp and Chukyo 	<p>Contact</p> <ul style="list-style-type: none"> Brian MacWhinney : homepage

Using local transcripts and local media

1. You need to download the transcripts from "zipped transcripts" below and unzip them. If you want to work with XML versions of the data, rather than CHAT versions, please go to [files](#).
2. If the corpus is linked to [audio or video media](#), you need to download those files and place the media into the transcript folders.
3. You need to download and install the [CLAN](#) program.
4. To open a transcript, you double-click on it. If there is associated media, you can play the media using escape-8 for continuous playback or command-click for playing single utterances.

Zipped Transcripts
English - USA
English - UK
Celtic
East Asian
Germanic
Romance
Slavic
Other Languages
Bilinguals
Clinical
Frog Story Narratives
Other Narratives
PhonBank CHAT Data
PhonBank Phon Data

Using local transcripts with web-served media

This is like the first method, but, if you are on a broadband Internet connection, you can skip step #2, since CLAN will use links in the folder called /media to play the media over the web.

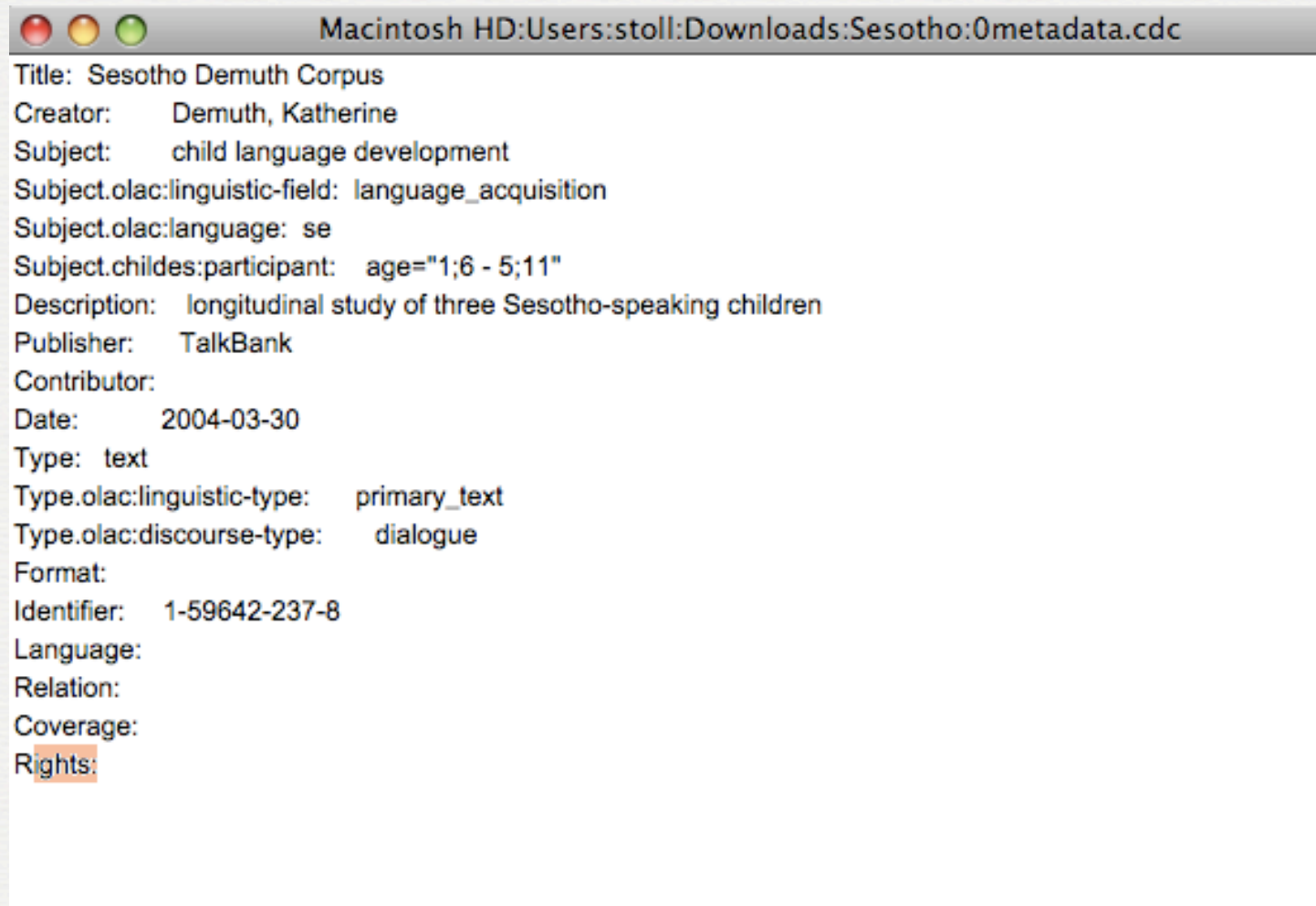
ChilDES: Guilfoyle corpus (Irish)

```
Macintosh HD:Users:stoll:Downloads:Irish:cai.cha
@Font: CAfont:13:0
@Begin
@Languages: ga
@Participants: GEA Investigator, CHI Cai Target_Child, EIT Investigator, MOT Mother
@ID: ga|guilfoyle|GEA|||||Investigator|
@ID: ga|guilfoyle|CHI|1;7.04||||Target_Child|
@ID: ga|guilfoyle|EIT|||||Investigator|
@ID: ga|guilfoyle|MOT|||||Mother|
@Birth of CHI: 01-DEC-1990
@Date: 05-JUL-1992
*MOT: céard é sin ?
%eng: that is that ?
*MOT: ó céard atá faighte agat ?
%eng: oh what do you have ?
*GEA: xx anois babóg .
%eng: xx now a doll .
*CHI: babóg .
%eng: doll .
%xmor: N|doll .
*MOT: nach bh-fuil sí go hálainn ?
%eng: isn't she beautiful ?
*GEA: céard eile atá thíos ann ?
%eng: what else is down there ?
*CHI: xx .
%eng: xx .
*MOT: céard é sin ?
%eng: what is that ?
*CHI: lorraí .
%eng: truck .
%xmor: N|truck .
*MOT: ab eá ?
U1303[E][CHAT] 1
```

Bloom corpus: English

```
Macintosh HD:Users:stoll:Downloads:Bloom70:Peter:peter13.cha
@Begin
@Languages: en
@Participants: CHI Peter Target_Child, LOI Lois Investigator, PAT Patsy
Investigator, MOT Mother
@ID: en|bloom70|CHI|2;5.03|male|normal||Target_Child||
@ID: en|bloom70|LOI||||Investigator||
@ID: en|bloom70|PAT||||Investigator||
@ID: en|bloom70|MOT||||Mother||
@Comment: Mother had invited Patsy and LOI to stay for lunch and lunch
was on Peter's mind from the time they arrived. He had just
had breakfast but asked for food throughout the session.
Mother had tried to put him off in order not to "spoil his
appetite" for lunch, but he persisted and became rather
cranky. He was unable to concentrate on anything for very
long. Patsy and LOI brought no new toys but Peter had just
acquired several new toys from a family about to move out of
the building. He was most excited about a large Shell oil
truck which he proudly showed and enjoyed. He also got a
magnetic alphabet and board and a cash register with wooden
coins inside. Mother was present for most of this session.
Jenny was there for part of it, but was asleep before lunch .
In the middle of his lunch, Peter asked to sit in Jenny's
chair. Mother put him there and Peter played baby for several
minutes, babbling and pretending to drink from a bottle. By
noon, he was very tired and protested when Patsy and LOI went
home.
@Situation: Lunch
*PAT: hi Pete .
%mor: co|hi n:prop|Pete .
%xgra: 1|0|ROOT 2|1|VOC 3|1|PUNCT
%act: <bef> Peter opens apartment door to let Lois and Patsy in
*LOI: <good morning> [>] .
%mor: adj|good n|morning .
%xgra: 1|0|ROOT 2|1|JCT 3|1|PUNCT
*CHI: <l [/] l [/] l xxx door> [<] .
%mor: pro|| unk|xxx n|door .
%xgra: 1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT
*PAT: thank+you .
U1303[E][CHAT] 1
```

Demuth corpus: Sesotho (some metadata)



Demuth corpus: Sesotho

```
Macintosh HD:Users:stoll:Downloads:Sesotho:hvb.cha
@Begin
@Languages: se
@Participants: BOY unidentified_child Playmate, NOI Noi Adult, ELS Elisha Playmate, MEM Mme_Manyili Grandmother, CHI Hlobohang Target_Child, KAT Katherine_Demuth
@ID: se|Sesotho|BOY|||||Playmate|
@ID: se|Sesotho|NOI|||||Adult|
@ID: se|Sesotho|ELS|||||Playmate|
@ID: se|Sesotho|MEM|||||Grandmother|
@ID: se|Sesotho|CHI|2;6.|||||Target_Child|
@ID: se|Sesotho|KAT|||||Investigator|
@ID: se|Sesotho|KHT|||||Playmate|
@ID: se|Sesotho|MHL|||||Mother|
@ID: se|Sesotho|TSN|||||Playmate|
@ID: se|Sesotho|MOL|4;10.|||||Cousin|
*MOL: ya hao e kae Hlobohang ? •
%gls: ya hao e kae Hlobohang ?
%cod: ps9 ij cp9 wh n^name ?
%eng: Where is yours Hlobohang ?
%sit: Rolling stones.
*MOL: e ne e le moo . •
%gls: e-n-e-le moo .
%cod: sm9-t^ne-sm9-cp loc .
%eng: It was here
%sit: H, Mol e
*MOL: jwale a fihla a e nka . •
%gls: jwale u-a-fihl-a u-a-e-nk-a .
%cod: av sm2s-t^p-v^arrive-m^in sm2s-t^p-om9-v^take-m^in .
%eng: Now you arrive you take it
*MOL: nna ka e tswara . •
%gls: nna ke-e-tSwar-a .
%cod: pn1s sm1s-t^p_om9-v^grab-m^in .
%eng: Me I grab
%sit: x2
U1303[E][CHAT] 1
```

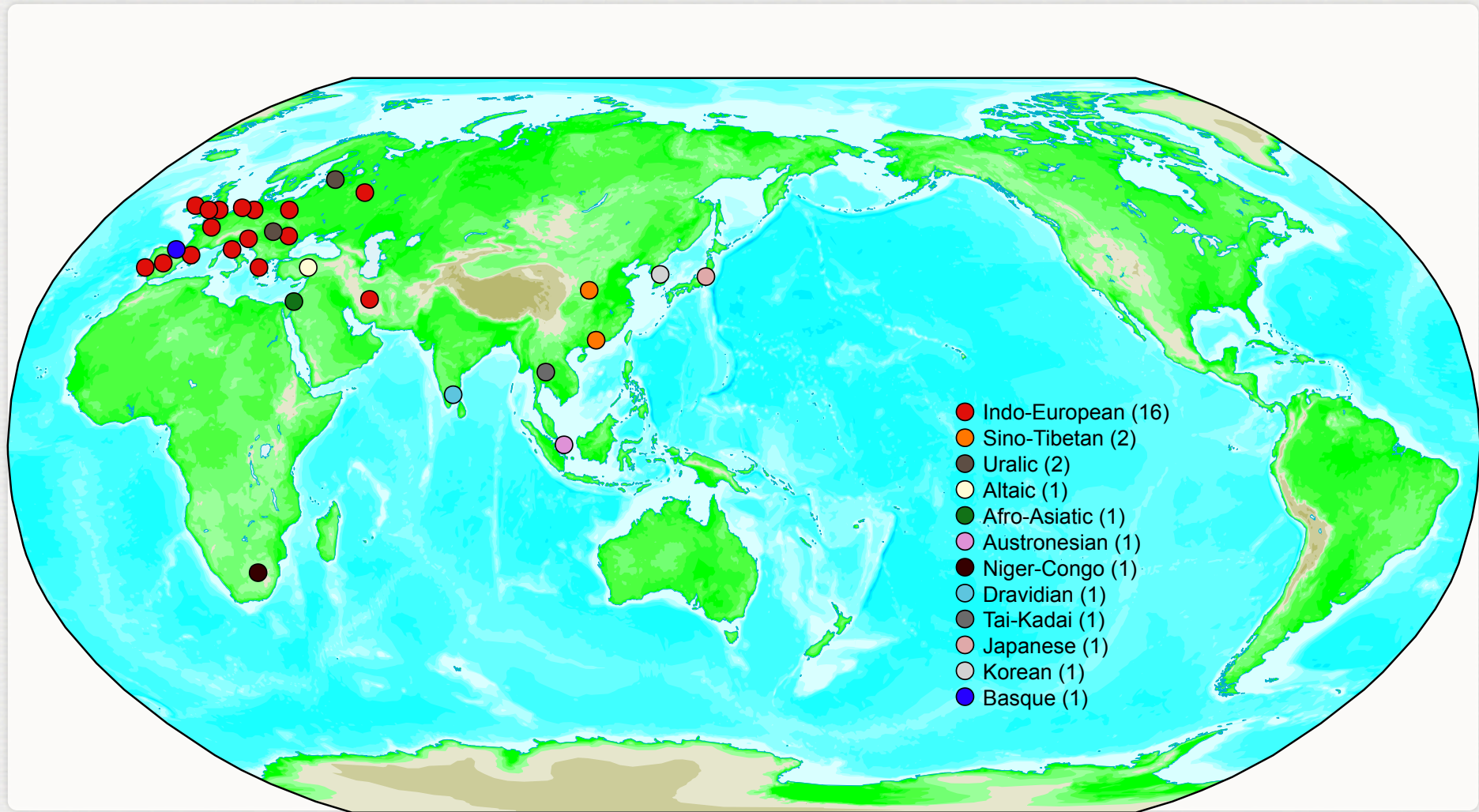
Chang corpus: Chinese

```
Macintosh HD:Users:stoll:Downloads:Chang-1:04.cha
@Begin
@Languages: zh
@Participants: CHI 志强 Child, EXP Observer
@ID: zh|rjp|CHI|5;11.|male|five|middle|Child|
@ID: zh|rjp|EXP|Observer|
@Birth of CHI: 28-MAR-1990
@Date: 07-FEB-1996
@Activities: Replica Play Jungle (RPJ)
@Situation: CHI and EXP create a story with toys
@Comment: Filename is 004RPJ
*EXP: 我们今天来玩森林的游戏好不好?
%mor: pro|wo3-PL=l n|jin1tian1=today v:dirc|lai2=come v|wan2=play n|sen1lin2=forest link|de=possessive v|you2xi4=play adj|hao3=good neg|bu4=not e
*CHI: 好 [^c].
%mor: adj|hao3=good.
*EXP: 你知不知道什么叫森林?
%mor: pro|ni3=you v|zhi1=tell neg|bu4=not v|zhi1dao4=know int|shen2me=what v|jiao4=call n|sen1lin2=forest?
*CHI: 我知道很多动物 [^c].
%mor: pro|wo3=l v|zhi1dao4=know adv|hen3=pretty quant|duo1=more_than n|dong4wu4=animal.
*EXP: 对好这个是森林的草地你帮阿姨把牠打开.
%mor: adj|dui4=correct adj|hao3=good pro|zhe4=this class|ge4 v|shi4=is n|sen1lin2=forest link|de=possessive n|cao3di4=grassland pro|ni3=you v|bang
*EXP: 来这是森林的草地.
%mor: v|lai2=come pro|zhe4=this v|shi4=is n|sen1lin2=forest link|de=possessive n|cao3di4=grassland.
*EXP: 然后呢森林里面是不是一定要有树啊?
%mor: adv|ran2hou4=thereupon sfp|ne=additional n|sen1lin2=forest loc|i3mian4=inside v|shi4=is neg|bu4=not v|shi4=is adv|yi1ding4=definitely v:aux|ya
*CHI: 对 [^c].
%mor: adj|dui4=correct.
*EXP: 对哦好看阿姨有树哦.
%mor: adj|dui4=correct co|o1=chant adj|hao3=good pro|ni3=you v|kan4=look n:relat|a1yi2=aunt v|you3=have n|shu4=tree co|o1=chant.
*EXP: 阿姨有三棵树还有一棵小的.
%mor: n:relat|a1yi2=aunt v|you3=have num|san1=three class|ke1 n|shu4=tree adv|hui2=still v|you3=have num|yi1=one class|ke1 adj|xiao3=small sfp|de
*CHI: 那是 [^c] +...
%mor: pro|na4=that v|shi4=is +...
*EXP: 那是什么?
%mor: pro|na4=that v|shi4=is pro|shen2me=anything?
*CHI: +, 是 (.) 圣诞树 [^c].
%mor: v|shi4=is n|sheng4dan4=Christmas n|shu4=tree.
U1303[E][CHAT] 1
```

Jiwon corpus: Korean

```
Macintosh HD:Users:stoll:Downloads:Jiwon:jw03.cha
@Begin
@Languages: kor
@Participants: CHI Target_Child, MOT Mother
@ID: kor|Jiwon|CHI|2;0.29|female||Target_Child|
@ID: kor|Jiwon|MOT||||Mother|
@Comment: collected on 5-AUG-1992
@Comment: coded on 13-April-2004
*MOT: 지원아.
*MOT: 이거 누구 꺼?
*CHI: 아빠 꺼 엄마 꺼.
*MOT: 아빠 꺼구 엄마 꺼구?
*MOT: 으응, 집 꼬 먹어?
*CHI: 한기, 한기, 한기.
*MOT: 한끼가 뭔데?
*CHI: 한기.
*MOT: 한기?
*CHI: 응.
*MOT: 한개?
*CHI: 응.
*MOT: 어어, 여기 한개 있다구?
*CHI: 네.
*MOT: 한기가 아니라 한개.
*CHI: 한가.
*MOT: 한가?
*MOT: 호호.
*MOT: 한개 그려 봐.
*MOT: 한개 그려 봐.
*CHI: 한기.
*MOT: 한끼?
*CHI: 응.
*MOT: 어이, 지원인 발음도 못하네.
*MOT: 요것도 한입 먹구.
*MOT: 으음 이쁘네.
*MOT: 어이 잘 먹네.
*MOT: 지원이도 야 녹음기 앞에서 말해야지 너.
*MOT: 녹음하고 있어.
*MOT: 지원이 아빠 이빠?
U1303[E][CHAT] 1
```

Languages with accessible corpora (CHILDES)



Language families with accessible corpora

- Basque (1)
- Sino-Tibetan (2)
- Indo-European (16)
- Uralic (2)
- Afro-Asiatic (1)
- Japanese (1)
- Korean (1)
- Niger-Congo (1)
- Dravidian (1)
- Tai-Kadai (1)
- Altaic (1)

Sampling Problems

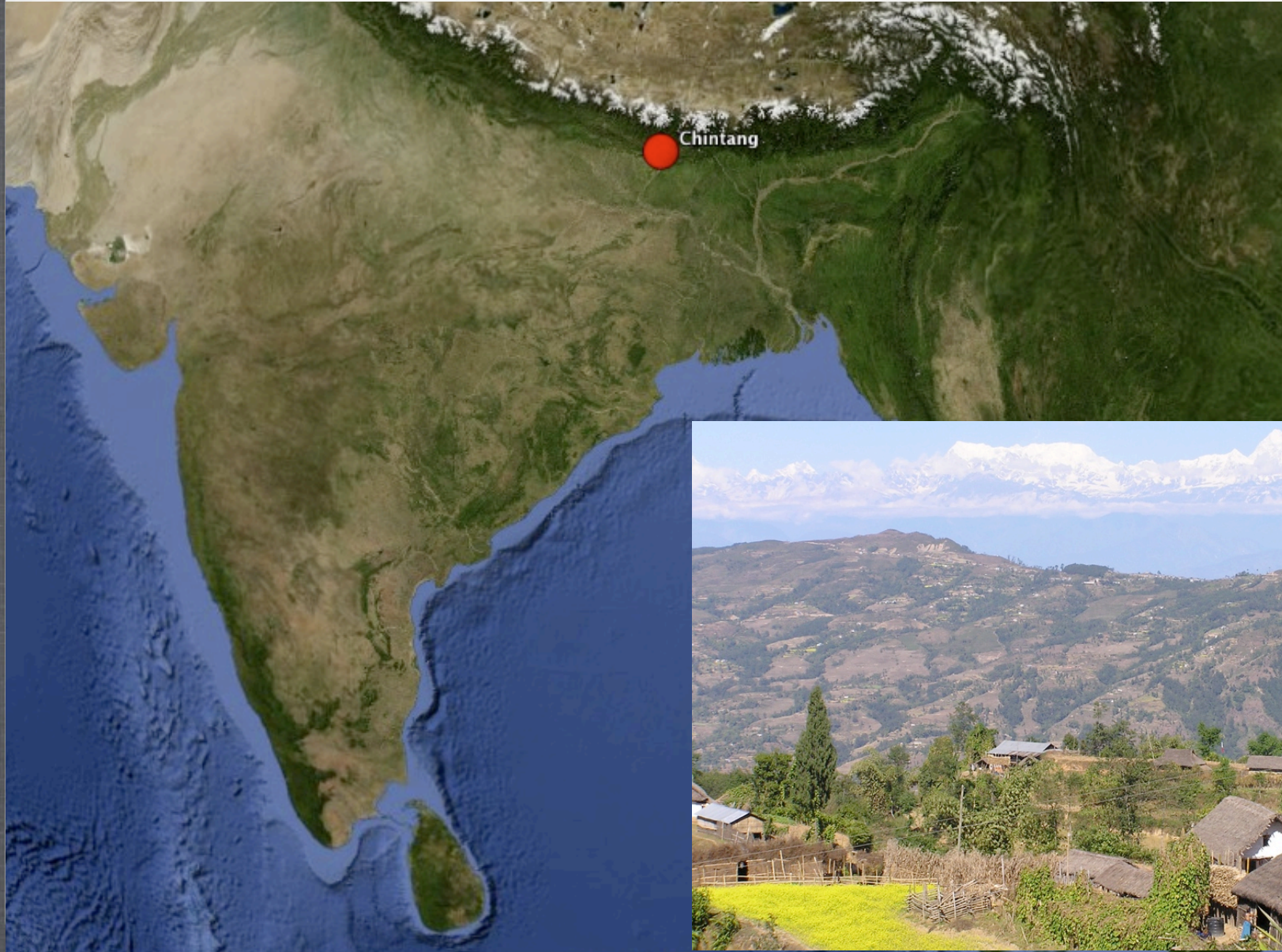
We have only a very limited amount of data available of a very restricted set of languages with a very limited number of the variables that are found in the languages of the world. The sample is skewed towards European languages with a strong bias towards Germanic and Romance.

==> Difficult to make generalizations; mostly our claims can be only language specific or at best probabilistic

Solution: do field work

- Chintang child language corpus (Stoll, Lieven) financed by Volkswagen foundation programs DOBES (PI: Bickel) and Dilthey (Stoll) and by the MPI for Evolutionary Anthropology.
- Fieldwork since 2004.
- Collaboration between linguists, anthropologists and language acquisition researchers

Chintang: a polysynthetic language spoken in eastern Nepal



DOBES: CPDP (2004 - 2009): three components

Linguistics:

Grammar
Discourse
Lexicon

How is the
language
structured?

Balthasar Bickel
Novel Kishore Rai
Vishnu Rai

Ethnography:

Rituals and
religion
Social structure
Everyday life

How is the
language used?

Martin Gaenzsle
Judith Pettigrew

Psycholinguistics:

Acquisition
strategies
Multilingualism

How is the
language learned?

Sabine Stoll
Elena Lieven

Chintang-project members

Linguistics	Anthropology	Language Acquisition	
<p>Bickel B., Rai, N. K. Bhatta, T.N. Paudyal, N.P. Rai, D.K. Dirksmeyer, T. Seeg, J. Zakharko, T.</p>	<p>Gaenzle, M., Pettigrew, J. Rai, I. P. Rai, R.M.</p>	<p>Stoll, S. Lieven, E. Seeg, J. Kuhn, K. Polkau, C. Sauppe, S. Schakow, D. Vissienon, K. Klein, F. Wolters, K. Harbodt, S. Lorenz, N. Poppitz, S.</p>	<p>Rai, J.K Rai, M. Rai, R.M. Rai, L.K. Rai, A. Rai, D.B. Rai, D.K. Rai, D. Rai, G. Rai, S. M. Rai, C. Banjade, G.</p>

Chintang corpus

	words (rounded)
Child language	499.728
Conversations	89.900
Narratives	37.500
Poems and songs	1.500
Myths and ritual texts	37.000
Descriptions	12.300
Sum	677.928

Chintang

- Sino-Tibetan language spoken in eastern Nepal
- approx. 6000 speaker
- trilingual society
- endangered language
- children still learn the language





Chintang corpus

N (children)	Recording	Age			
		1	2	3	4
2	2-3 h per month		■	■	
2	3-4 h per month		■	■	
2	3-4 h per month			■	■

- monthly
- bimonthly
- every 3 month

Chintang Child language project workflow: Managing a large scale project

Building up the Chintang corpus

1. Recording

Chintang natural speech
(different ages & genres)

Location: Nepal

Staff: student assistants

Output: *session.mpg*
session.wav

*add session to session-monitor

*add DVD to media-monitor

2. Metadata

Information about
location, speakers, etc.

Location: Nepal

Staff: student assistants

Software: imdi-editor

Output: *session.imdi*

*check completeness of imdifile

*upload to server

3. Transcription

Text production

Location: Nepal

Staff: student assistants
(Chintang native speakers)

Software: Transcriber

Input: *session.wav*

Output: *session.trs*

*assign the session to a transcriber

* update the monitor
(person, status)

the transcriber also does
the Nepali-translation >>

4. Translation Nepali

Translation into the local lingua franca

Location: Nepal

Staff: student assistants
(Chintang native speakers)

Software: Elan

Input: *session.trs*,
session.wav, *session.mpg*

- *receive .trs and .eaf
- *check completeness
- *check speakercodes (use imdifile and speaker-DB)
- *export as toolboxfile and convert (UTF-8, linebreaks)
- *upload to server
- *update the monitor (person, status, pwords)
- [*calculate workspeed]

5. Translation English

Translation into an international language

Location: Germany

Staff: student assistants

Software: any texteditor

Input: *session.txt*

Output: *session.txt*

- *assign the toolboxfile to a translator
- *update monitor
- *receive toolboxfile
- *check completeness
- *upload to server
- *update monitor
- [*calculate workspeed]

6. Glossing

Linguistic analysis

Location: Germany

Staff: student assistants

Software: Toolbox

Input: *session.txt*

Output: *session.txt*

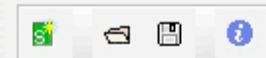
- *assign toolboxfile to a glosser
- *update monitor
- *receive toolboxfile
- *check completeness
- *update monitor
- *add information using scripts (e.g. POS, agegroup of speakers)
- *upload to server

Training of assistants in Nepal





Lieven & Stoll, DGfS Summer
School, 2010



CLLDCh2R10501

- ◆ Project
- ◆ Content
- ◆ Actors
- ◆ Resources
- ◆ References

Editor HTML Links

Project

Name Chintang and Puma Documentation Project

Title Documentation of Chintang and Puma, two Kiranti languages of Eastern Nepal

ID CPDP

Contact Information

Name Prof. Dr. Balthasar Bickel

Address Institut für Linguistik, Beethovenstr. 15, 04107 Leipzig, Germany

E-mail bickel@uni-leipzig.de

Organisation University of Leipzig

Descriptions

Language English



CV

English

Text

The aim of the project is to provide a rich linguistic and ethnographic documentation of two highly endangered but almost totally undocumented languages in eastern Nepal, Chintang and Puma. These languages belong to the Kiranti family of Tibeto-Burman. Chintang is spoken by the Chintang Rai in Chintang Village Development Committee (VDC) of Dhankuta district. Puma is spoken by the Puma Rai in Diplung, Mauwabote, Devasthan and

Add

Transcription...



Lieven & Stoll, DGfS Summer
School, 2010

... into the target language Chintang (in Transcriber)

Transcriber 1.5.1+ (06/25/2005)

(no speaker)
● [...]

LDCh2
● kalpe mo dhara ta

LDCh3
● kamala kham tisakina dhara thano

CLLDCh2R10S01

Resolution

LDCh2	LDCh3
kalpe mo dhara ta	kamala kham tisakina dhara thano

5:22 5:23 5:24 5:25 5:26 5:27 5:28 5:29

Cursor: 05:20.981

Translation into Nepali and English (in ELAN)

Elan - CLLDCh2R10S01.eaf

[Datei](#) [Bearbeiten](#) [Annotation](#) [Zeile](#) [Typ](#) [Suche](#) [Ansicht](#) [Optionen](#) [Fenster](#) [Hilfe](#)

[Tabelle](#) [Text](#) [Untertitel](#) [Metadaten](#) [Steuerung](#)

Lautstärke: 100

Geschwindigkeit: 100

00:05:21.320 Auswahl: 00:05:21.320 - 00:05:22.970 1650

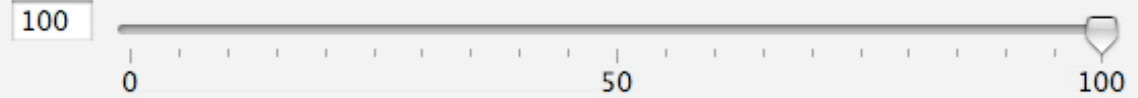
Auswahl-Modus Schleifen-Modus

Layer	00:05:19.000	00:05:20.000	00:05:21.000	00:05:22.000	00:05:23.000	00:05:24.000	00:05:25.000	00:05:26.000	00:05:27.000	00:05:28.000
tx@LDCh2 [188]		khoi		kalpe mo dhara ta				khoi		huī h
nep@LDCh2 [187]		खोई		कल्पना! उ (तल) धारा नै				खोई		उ उ ध
eng@LDCh2 [188]		Let me see		Kalpe,				Let me see!		that I
tx@LDCh3 [201]	kamala huī dhara thano?		huī					kamala kham tisakina dhara thano?		huī dhara thiye?
nep@LDCh3 [201]	कमला! उ धारा झर्दैछ ।		उ ।					कमला!(तभिले) माटो लगाएर धारा झर्दैछ ।		उ धारो झर्‍यो ।
eng@LDCh3 [200]	Kamala, the water-tap is runnin		that					Kamala, after putting soil, the water tap runs.		That water tap is running.
tx@PMR										

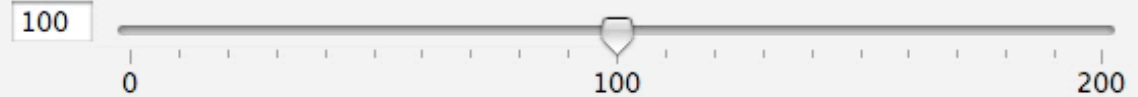


Tabelle Text Untertitel Metadaten Steuerung

Lautstärke:

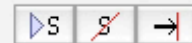


Geschwindigkeit:



00:05:11.010

Auswahl: 00:00:00.000 - 00:00:00.000 0



Auswahl-Modus Schleifen-Modus

	00:05:10.000	00:05:11.000	00:05:12.000	00:05:13.000	00:05:14.000	00:05:15.000	00:05:16.000	00:05:17.000
tx@LDCh2 [188]	a u dura			ai		dhara cattuhē?		
nep@LDCh2 [187]	आ धारो			एई		धारालाई हान्ने ।		
eng@LDCh2 [188]	tap			Hey!		I hit the tap.		
tx@LDCh3 [201]	va			ba		dhara		dha
nep@LDCh3 [201]	रे ।			यी		धारा		धारा
eng@LDCh3 [200]	ll!			this		tap!		tap!
tx@PMR [110]								
eng@PMR [110]								

Glossing (in Toolbox)

The screenshot shows the Toolbox software interface with two windows open:

Window 1: CLLDCh2R10S01_notags.txt

```

\ref CLLDCh2R10S01.213
\ELAN 00:05:21.320
\ELAN 00:05:22.970
\EUDIC LDCh2
\tx kalpe mo dhara ta
\gw kalpe mo dhara ta
\mph kalpana mo dhara ta
\mgf a_person DEM.DOWN tap FOC
\lg N C C/N C
\veng Kalpe, that tap down there.
\vnep कल्पना! उ ( तल) धारा नै।
\dt 10/Mar/2008
  
```

Window 2: Chintang-Lex-SES.db

```

\vex dhara
\ct
\alt dharo
\alt dhala
\ps n
\ge tap
\gn धारा
\g C/N
\ed GB/KK/SES
\dt 25/Jul/2008
  
```

Window 3: Chintang-Lex.db:2

\vex	\alt	\ge	\gn	\ps	\g
dhapt	dhap	hammer; nail; pound	ढोकु	vt	C
dhar	*empty*	edge	धार	n	C/N
dhar	*empty*	blade	धार	n	N
dhara	dharo	tap	धारा	n	C/N
dharan	*empty*	a_place	धरान	n	N
dharati	*empty*	earth	धर्ति	n	N
dharko	*empty*	line	धर्को	n	C/N
dharma	*empty*	virtue	धर्म	n	N
dharman	*empty*	a_person	धरमाने	n	C

Toolbox file with POS-tags

tags

The screenshot shows a software interface with three main windows. The top window, titled 'Toolbox - CLLDCh2R10S01_tags.txt', contains a table of linguistic data. The middle window, titled 'Chintang-Lex-SES.db', shows a list of words with their corresponding POS tags. The bottom window, titled 'Chintang-Lex.db:2', shows a detailed view of the 'dhara' entry in the lexical database.

Window	Content																																																												
Toolbox - CLLDCh2R10S01_tags.txt	<table border="1"> <tr><td>\ref</td><td>CLLDCh2R10S01.213</td></tr> <tr><td>\ELAN</td><td>00:05:21.320</td></tr> <tr><td>\ELAN</td><td>00:05:22.970</td></tr> <tr><td>\EUDIC</td><td>LDCh2</td></tr> <tr><td>\tx</td><td>kalpe mo dhara ta</td></tr> <tr><td>\gw</td><td>kalpe mo dhara ta</td></tr> <tr><td>\mph</td><td>kalpana mo dhara ta</td></tr> <tr><td>\mgl</td><td>a_person.n DEM.DOWN.pro tap.n FOC.gm</td></tr> <tr><td>\lg</td><td>N C C/N C</td></tr> <tr><td>\eng</td><td><i>Kalpe, that tap down there.</i></td></tr> <tr><td>\nep</td><td>कल्पना! उ (तल) धारा नै।</td></tr> <tr><td>\dt</td><td>10/Mar/2008</td></tr> </table>	\ref	CLLDCh2R10S01.213	\ELAN	00:05:21.320	\ELAN	00:05:22.970	\EUDIC	LDCh2	\tx	kalpe mo dhara ta	\gw	kalpe mo dhara ta	\mph	kalpana mo dhara ta	\mgl	a_person.n DEM.DOWN.pro tap.n FOC.gm	\lg	N C C/N C	\eng	<i>Kalpe, that tap down there.</i>	\nep	कल्पना! उ (तल) धारा नै।	\dt	10/Mar/2008																																				
\ref	CLLDCh2R10S01.213																																																												
\ELAN	00:05:21.320																																																												
\ELAN	00:05:22.970																																																												
\EUDIC	LDCh2																																																												
\tx	kalpe mo dhara ta																																																												
\gw	kalpe mo dhara ta																																																												
\mph	kalpana mo dhara ta																																																												
\mgl	a_person.n DEM.DOWN.pro tap.n FOC.gm																																																												
\lg	N C C/N C																																																												
\eng	<i>Kalpe, that tap down there.</i>																																																												
\nep	कल्पना! उ (तल) धारा नै।																																																												
\dt	10/Mar/2008																																																												
Chintang-Lex-SES.db	<table border="1"> <tr><td>\lex</td><td>dhara</td></tr> <tr><td>\ct</td><td></td></tr> <tr><td>\alt</td><td>dharo</td></tr> <tr><td>\alt</td><td>dhala</td></tr> <tr><td>\ps</td><td>n</td></tr> <tr><td>\ge</td><td>tap</td></tr> <tr><td>\gn</td><td>धारा</td></tr> <tr><td>\g</td><td>C/N</td></tr> <tr><td>\ed</td><td>GB/KK/SES</td></tr> <tr><td>\dt</td><td>25/Jul/2008</td></tr> </table>	\lex	dhara	\ct		\alt	dharo	\alt	dhala	\ps	n	\ge	tap	\gn	धारा	\g	C/N	\ed	GB/KK/SES	\dt	25/Jul/2008																																								
\lex	dhara																																																												
\ct																																																													
\alt	dharo																																																												
\alt	dhala																																																												
\ps	n																																																												
\ge	tap																																																												
\gn	धारा																																																												
\g	C/N																																																												
\ed	GB/KK/SES																																																												
\dt	25/Jul/2008																																																												
Chintang-Lex.db:2	<table border="1"> <thead> <tr> <th>\lex</th> <th>\alt</th> <th>\ge</th> <th>\gn</th> <th>\ps</th> <th>\g</th> </tr> </thead> <tbody> <tr><td>dhapt</td><td>dhap</td><td>hammer; nail; pound</td><td>ढीकु</td><td>vt</td><td>C</td></tr> <tr><td>dhar</td><td>*empty*</td><td>edge</td><td>धार</td><td>n</td><td>C/N</td></tr> <tr><td>dhar</td><td>*empty*</td><td>blade</td><td>धार</td><td>n</td><td>N</td></tr> <tr><td>dhara</td><td>dharo</td><td>tap</td><td>धारा</td><td>n</td><td>C/N</td></tr> <tr><td>dharan</td><td>*empty*</td><td>a_place</td><td>धरान</td><td>n</td><td>N</td></tr> <tr><td>dharati</td><td>*empty*</td><td>earth</td><td>धर्ति</td><td>n</td><td>N</td></tr> <tr><td>dharko</td><td>*empty*</td><td>line</td><td>धर्को</td><td>n</td><td>C/N</td></tr> <tr><td>dharma</td><td>*empty*</td><td>virtue</td><td>धर्म</td><td>n</td><td>N</td></tr> <tr><td>dharman</td><td>*empty*</td><td>a_person</td><td>धरमाने</td><td>n</td><td>C</td></tr> </tbody> </table>	\lex	\alt	\ge	\gn	\ps	\g	dhapt	dhap	hammer; nail; pound	ढीकु	vt	C	dhar	*empty*	edge	धार	n	C/N	dhar	*empty*	blade	धार	n	N	dhara	dharo	tap	धारा	n	C/N	dharan	*empty*	a_place	धरान	n	N	dharati	*empty*	earth	धर्ति	n	N	dharko	*empty*	line	धर्को	n	C/N	dharma	*empty*	virtue	धर्म	n	N	dharman	*empty*	a_person	धरमाने	n	C
\lex	\alt	\ge	\gn	\ps	\g																																																								
dhapt	dhap	hammer; nail; pound	ढीकु	vt	C																																																								
dhar	*empty*	edge	धार	n	C/N																																																								
dhar	*empty*	blade	धार	n	N																																																								
dhara	dharo	tap	धारा	n	C/N																																																								
dharan	*empty*	a_place	धरान	n	N																																																								
dharati	*empty*	earth	धर्ति	n	N																																																								
dharko	*empty*	line	धर्को	n	C/N																																																								
dharma	*empty*	virtue	धर्म	n	N																																																								
dharman	*empty*	a_person	धरमाने	n	C																																																								

Toolbox file with agecodes

The screenshot displays a software window titled "Toolbox - CLLDCh2R10S01_agecoded.txt". The main window shows a list of metadata fields for a file named "CLLDCh2R10S01.213". A red circle highlights the fields "\agegroup" (child) and "\age" (1029). Below these are fields for "\tx", "\gw", "\mph", "\mgl", "\lg", "\eng", and "\nep".

\ref	CLLDCh2R10S01.213
\ELANBegin	00:05:21.320
\ELANEnd	00:05:22.970
\EUDICOp	LDCh2
\agegroup	child
\age	1029
\tx	kalpe mo dhara ta
\gw	kalpe mo dhara ta
\mph	kalpana mo dhara ta
\mgl	a_person.n DEM.DOWN.pro tap.n FOC.gm
\lg	N C C/N C
\eng	<i>Kalpe, that tap down there.</i>
\nep	कल्पना! उ (तल) धारा नै।

To the right, a window titled "Chintang-..." shows a list of fields and values for the word "dhara":

\lex	dhara
\ct	
\alt	dharo
\alt	dhala
\ps	n
\ge	tap
\gn	धारा
\g	C/N
\ed	GB/KK/SES
\dt	25/Jul/2008

At the bottom, a window titled "Chintang-Lex.db:2" shows a table of lexical entries:

\lex	\alt	\ge	\gn	\ps	\g
dhapt	dhap	wash	धुनु	vt	C
dhapt	dhap	hammer; nail; pound	डोकु	vt	C
dhār	*empty*	edge	धार	n	C/N
dhār	*empty*	blade	धार	n	N
dhara	dharo	tap	धारा	n	C/N
dhara	dharo	tap	धारा	n	C/N
dharan	*empty*	a_place	धरान	n	N
dhārati	*empty*	earth	धर्ति	n	N
dharko	*emotv*	line	धर्को	n	C/N

Meta-data editor (IMDI)

IMDI Metadata Editor

File View Options Help

CLLDCh2R10501

- Project
- Content
- Actors
- Resources
- References

Local Repository

- Projects
- Content
- Actors
- Languages
- Access

Editor HTML Links

Project

Name: Chintang and Puma Documentation Project

Title: Documentation of Chintang and Puma, two Kiranti languages of Eastern Nepal

ID: CPDP

Contact Information

Name: Prof. Dr. Balthasar Bickel

Address: Institut für Linguistik, Beethovenstr. 15, 04107 Leipzig, Germany

E-mail: bickel@uni-leipzig.de

Organisation: University of Leipzig

Descriptions

Language: English

Text: The aim of the project is to provide a rich linguistic and ethnographic documentation of two highly endangered but almost totally undocumented languages in eastern Nepal, Chintang and Puma. These languages belong to the Kiranti family of Tibeto-Burman. Chintang is spoken by the Chintang Rai in Chintang Village Development Committee (VDC) of Dhankuta district. Puma is spoken by the Puma Rai in Diplung, Mauwabote, Devasthan and

English

Add

Monitor of workflow

chintang_cl_session_monitor (xserve2leipzig.rz.uni-leip)

recor...	session_name	transcription_status	translation_status	interlinear_status	length	child_utterances	adult_utterances	other_child_utte
30	CLLDCh4R01S05							
31	CLLDCh3R01S04							
32	CLLDCh1R01S04							
51	CLLDCh1R02S01	done (GAR)	Nepali done (GAR),		1422			
52	CLLDCh1R02S02	done (AnR)	all done (AnR, DR)	in progress (KK)	244			
53	CLLDCh1R02S03a	done	done	done	142	27; 76	17; 55	98; 142
54	CLLDCh2R02S01	in progress						
55	CLLDCh2R02S02	done (AnR)	all done (AnR, DR)	in progress	169			
56	CLLDCh2R02S03	done (JK)	Nepali done (JK),		478			
57	CLLDCh3R02S01	transcriptions are	done	done	447	177	167	103
58	CLLDCh2R02S04	done	done	done	530	195	289	46
59	CLLDCh2R02S05	done	done	done	135	51	81	3
60	CLLDCh1R02S04a	done	done	done	1153	293	346	514
61	CLLDCh2R02S06	progress (AnR)	progress (AnR)					
62	CLLDCh2R02S07	done (DW)	all done (DW, DR)	progress	723			
63	CLLDCh2R02S08	done (LK)	all done (LK, DS)	done (SES)	894			
64	CLLDCh3R02S02	done	done	done	106	42	35	29
65	CLLDCh3R02S03							
66	CLLDCh3R02S04							
67	CLLDCh3R02S05							
68	CLLDCh3R02S06	done	done	done	221	85	106	30
69	CLLDCh4R02S01	transcriptions are	done	done	264	97	112	55
70	CLLDCh4R02S02a	done	done	done	130	44	36	50
71	CLLDCh4R02S03							
72	CLLDCh4R02S04	done	done	done	513	164	290	59
73	CLDLCh3R01S01							
74	CLDLCh3R01S02	done	done	done	494	11	356	127
75	CLDLCh3R01S03	done	done	done	369	2	367	0
76	CLDLCh3R01S04	done	done	done	284	0	282	2
77	CLDLCh3R01S05							
80	CLLDCh1R02S05	done& being	Nepali done (DKR),		1586			
97	CLLDCh1R03S01	done	\rt done	done	873	125	274	474
98	CLDLCh2R02S01							
99	CLLDCh3R03S01	done	done	done	632	227	69	336
100	CLLDCh2R03S01	done (DKR)	all done (DKR, DS)		936			
101	CLLDCh1R03S02	done	done	done	815	150	237	427
102	CLDLCh2R02S02	done	done	done	723			

The session monitor

chintang_cl_session_monitor (xserve2)

assign rounds 2, 6, 11 first! Enter date of assignment! check progress-files of person x before assigning a new file!

record_ID: 66 responsible: CP
 session_name: CL LDCh3R02S04
 transcription_status: done (AnR)
 translation_status: all done (AnR, DS)
 interlinear_status: in progress since Mar5th
 imdi_status: incomplete
 archive_status:

recording_date: 2004-05-25
 time_duration: 0:56:37
 time_alignment: done in ELAN
 pwords: 3870
 gwords: 0
 TimeCode (as captured):

WHICH FILES ARE ON THE SERVER?

	DRAFT FILES	FINAL FILES
ELAN :	plain ? <input type="checkbox"/> yes	glossed ? <input type="checkbox"/> no
TOOLBOX :	plain ? <input type="checkbox"/> yes tagged ? <input type="checkbox"/> no	agecoded ? <input type="checkbox"/> no
IMDI :	unchecked ? <input type="checkbox"/> no	checked ? <input type="checkbox"/> yes

comments: utterance counting: missing/unclear: DPR 8
 metadata_evaluation: Quality: middle

MONITOR PROBLEMS?
 status_mismatch => mismatch if file is on server AND interlinear_status=done/checked AND file is not glossed (i.e. gwords=0) OR if gwords>0 AND interlinear_status="" OR if pwords>0 AND translation_status="" OR if toolboxfile is on server AND imdfile isn't

UTTERANCE COUNTING
 length (#utterances): 1405
 by targetchild: 329
 by adults: 753
 by other children: 315

of these glossed

OUTPUT OF TRANSCRIPTION + TRANSLATION

	hours	records per hour	done by	month
transcr/transl(Nep)	162	8.6	AnR	2009 Oct-Nov
transl(Eng)		?	DS	2010 Feb
retranscription		?		
glossing	17.5	80.2	CP	2010 March

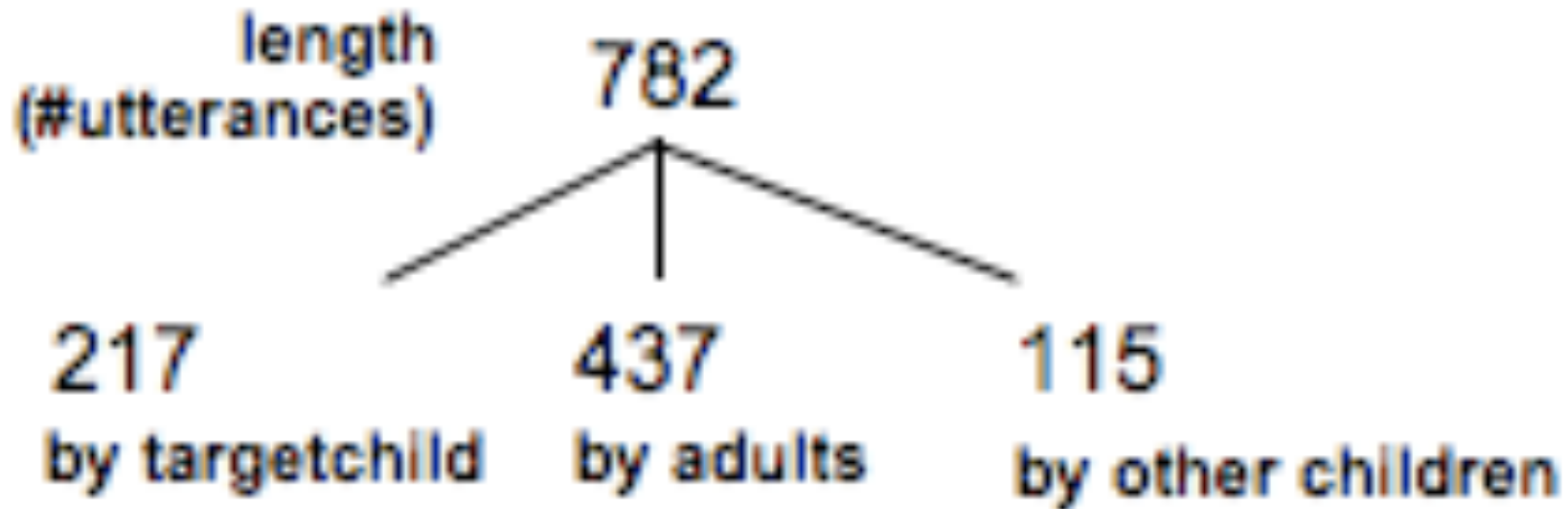
missing_transcriptions_tc

OUTPUT of glossing questions

Record: 1
 Found: 2
 Total: 470
 Unsorted

Adult vs. child language

UTTERANCE COUNTING



217
of these glossed

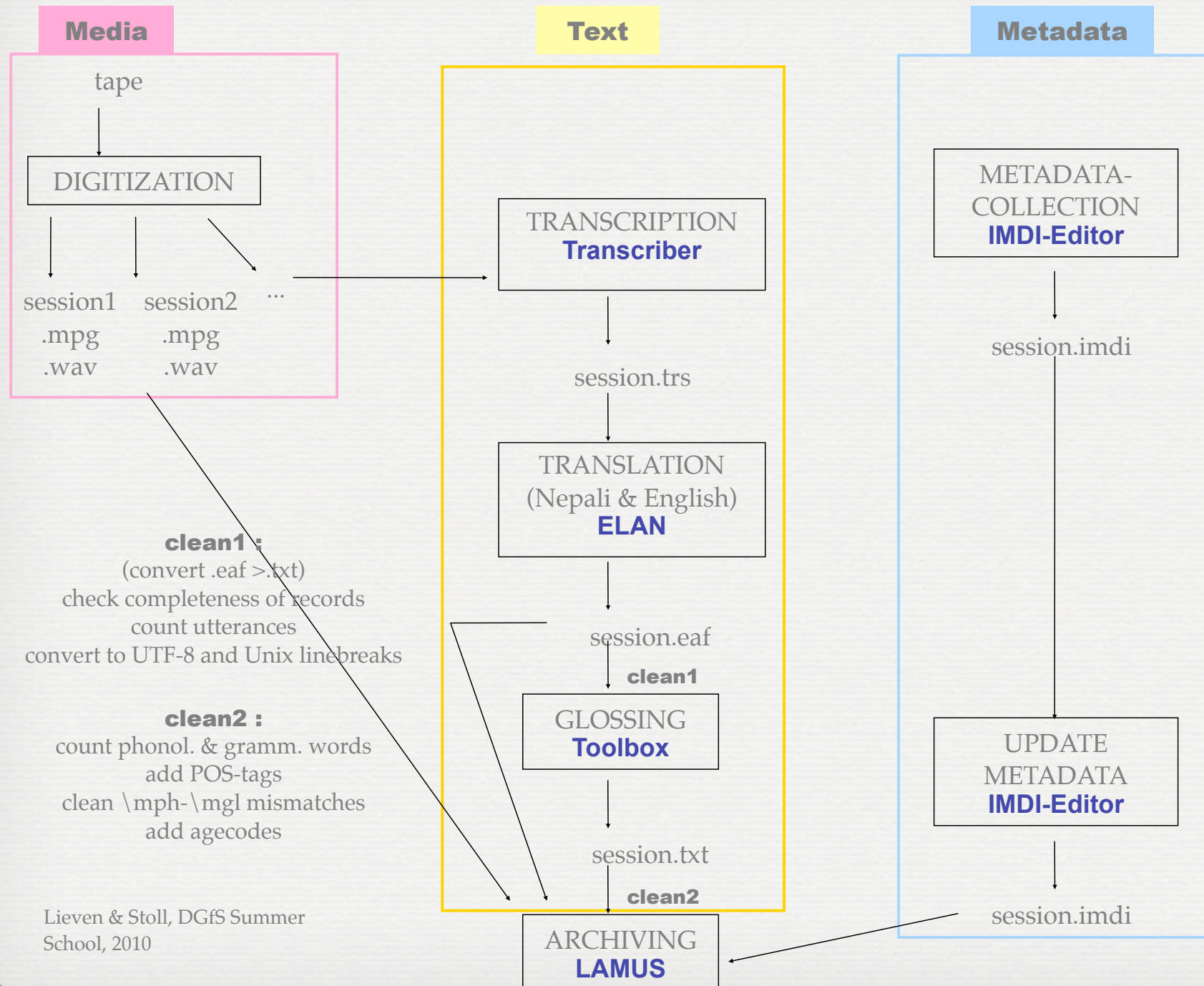
Status of targetchild 1

child CLLDCh1
 total length 47233
 total duration 73:58:57
 total pwords 112769
 total gwords 48062
 child_utterances total 15460
 child_utterances glossed 13202
 adult_utterances total 19319
 other_child_utterances total 114436

ALL CHILDREN			
glossed targetchild utterances : 35512			
	PHONOLOGICAL	GRAMMATICAL	DURATION
total	406277	207229	371:14:24
translated	406277		203:18:16
glossed	313856		152:34:01

TRANSCRIPTION & TRANSLATION
 translated length 47233
 translated duration 55:40:39
 translated pwords 112769

GLOSSING
 glossed length 43142
 glossed duration 51:33:18
 glossed pwords 103512



Thank you for your attention!