WORD DOMAINS

Balthasar Bickel
Kristine A. Hildebrandt
khilde@uni-leipzig.de
I. “Autotyp”: http://www.uni-leipzig.de/~autotyp/

AUTOTYP is not a single individual database project. Rather, it is a large-scale research program involving various thematically specific projects. In each project we develop a series of data and definition files linked together relationally. Currently, the number of files across all projects is about 70. A database file is currently implemented in FileMaker Pro™ for the Macintosh, but will in due time be transferred into a more general, platform independent format.

The following projects are fully implemented and data collection is currently proceeding full steam (click for further information). Lines indicate the most important relationships between projects. (Most projects are interconnected via shared definition files anyway.)

### Thematic Projects
- morphology and phonology of grammatical markers (348)
- synthesis of verbs (222)
- NP structures (417)
- locus (head/dependent marking) (273)
- agreement (339)
- person systems (372)
- morphological alignment (270)
- word domains (29; 273 domains)

### Backbone Projects
- genetic affiliation (676)
- geographical location (521)
  - sampling
  - bibliography
  - statistics

### Service Modules
- language logs
- database logs
- database monitors
II. Words: Phonological & Grammatical

- Definitions
- Examples from database
The Phonological Word ω

- Selkirk (1980); Booij (1983; 1984; 1985); Nespor & Vogel (1986); Peperkamp (1997); Inkelas (1987); McCarthy & Prince (1993); Hall & Kleinhenz (1999); Wiese (1996); Dixon & Aikhenvald (2002)

- A phonological domain that references morphological information = Phonological Word
Phonological processes.rules/constraints that...

• Reference domains ‘beyond the syllable’
• Also reference some morphological or morpho-syntactic domain (e.g. root/stem; stem + affix; stem + stem, etc.)
• ‘morpho-phonology’
Onset: *bl* surfaces word-initial:
1. *blitz* -> [blɪts] ‘lightning’

Coda word-medial: *bl* will *often* resyllabify:
2. *bibl + isch* -> [bi.blɪʃ] ‘biblical’
   = ‘Onset Maximization’

Surfacing coda obstruents devoice
   *Rad* -> [ʁa:t] ‘wheel’ = ‘Final Devoicing’
German P-word

- *lieb + e* love-1SG.PRES -> $[liːbə]$ ‘I love…’
- **BUT:** *lieb + lich* love-ADV -> $[liːpɨç]$ 'dearly'
  - no onset maximization for a *bl*- cluster
  - rather, /b/ devoices to [p]
- **Consonant-initial** suffixes parsed as their own domains for this process (separate p-domains)
  - (libə)$_{ω}$ & (bibliʃ)$_{ω}$ vs. (lib)$_{ω}$ (lІç)$_{ω}$
- Domain of ONSET MAX: stem ± v-initial suffix
- Prefixes excluded: *ab-ändern* [(ʔʌp)$_{ω}$ʔɛn.dəʁn]
- Compounds excluded: *bergabfahrt* [(bɛk)$_{ω}$ʔap.fəʁt]
German Onset Max Word

WTID 309  LID 87  German  UNIT p-domain

WTYPE 210  Coda C to Onset Resyllab

PHON DOMAIN 3  right
RESOLUTION 28  resyllabification

Inputter 24  Kristine
Date 02-02-2005
Reliability 5  Other published sources

Morph Domain
Align ID 3
Align right edge
Domain stem ± suffix
Gram Domain Size 2

Synthesis
Number Formative Slots 2
Bipartite Stem 0
Max number categories in sequence
Number of Available Morphemes 5

Strata
Stratum ID 52
only v-initial suffixes

if source is loanword:
LID1
LID2
LID3
LID4
LID5
LID6

Source native

KNOWN LENDING LANGUAGES

Notes
data: Jochen Geiffuss-Wolfgang (Autotyp Bib); Hall 1999; Booij 1985; Wiese 1996
ONSET MAXIMIZATION applies if 2nd syllable or morph is v-initial. If cv(c), then no ONSET MAX, and rather devoicing of stem/1st syllable coda C:

Notes
Manange (Tibeto-Burman)

• Single Tone Contour Word = stem ± prefix± suffix ± particle (‘clitic’)
• For example, tone /3/: a high, sharp falling tone
• [sΛ₅³] ‘good/tasty/wholesome’
• [α₅⁴-sΛ₄³] NEG-good ‘not good’
• [α₅⁴-sΛ₄³-pΛ₃²/²²] NEG-good-NOM ‘not good one’
• [kjẽ¹¹ α₅⁴-sΛ₄³-pΛ=ko₃²] rice NEG-good-NOM=DEF ‘the bad-tasting rice’
Manange Tone Contour Word

<table>
<thead>
<tr>
<th>WTID</th>
<th>LID</th>
<th>Manange</th>
<th>UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>690</td>
<td>Manange</td>
<td>p-domain</td>
</tr>
</tbody>
</table>

**WTYPe 1** Single Tone Contour/Melody

- **PHON DOMAIN**: 4 span
- **RESOLUTION**: 17 preservative spread/assimilation

**Morph Domain**

- **Align ID**: 7
- **Align**: spans
- **Domain ID**: 64
- **Domain**: stem ± prefix ± suffix ± particle
- **Gram Domain Size**: 4

**Synthesis**

- Number Formative Slots: 1
- Bipartite Stem: 1
- Max number categories in sequence: 1
- Number of Available Morphemes: 4

**Strata**

- Stratum ID: 8
- all but approx 15 disyllabic noun roots

**KNOWN LENDING LANGUAGES**

- LID1
- LID2
- LID3
- LID4
- LID5
- LID6

<table>
<thead>
<tr>
<th>LID</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>Nepali</td>
</tr>
<tr>
<td>375</td>
<td>Tibetan (Standard Spoken)</td>
</tr>
</tbody>
</table>

**Notes** Hildebrandt 2003, 2004, 2005 see WTID 327

Inputter: 24 Kristine
Date: 2-10-2003
Reliability: 3 Grammar Explicit
G-Words

- In **morphology & syntax**: general assumption of a maximum of 2 or 3 domains:
  - $M (\Sigma, \text{Root, Affix})$
  - $X^0$ (Any syntactic head/terminal node)
  - $XP$ (Any syntactic phrase)

- **Other accounts of grammatical units** also very general:
  - Fixed combination or ordering possibilities for elements
  - A certain “conventionalized meaning”
  - A degree of coherence
  - (e.g. Dixon & Aikhenvald 2002)
G-word in Belhare (Tibeto-Burman)

<table>
<thead>
<tr>
<th>WTID</th>
<th>LID</th>
<th>UNIT</th>
<th>PHON DOMAIN</th>
<th>RESOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>234</td>
<td>35</td>
<td>g-domain</td>
<td>5</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**WTYPE 162** Terminal Word (X0)

**Morph Domain**

- **Align ID**: 7
- **Domain**: (object + intrans affxes) ± stem + affxes
- **Gram Domain Size**: 4

**Strata**

<table>
<thead>
<tr>
<th>Source</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>LID1</td>
<td>206</td>
</tr>
<tr>
<td>LID2</td>
<td>169</td>
</tr>
<tr>
<td>LID3</td>
<td></td>
</tr>
<tr>
<td>LID4</td>
<td></td>
</tr>
<tr>
<td>LID5</td>
<td></td>
</tr>
<tr>
<td>LID6</td>
<td></td>
</tr>
</tbody>
</table>

**Notes**

- e.g. phak se'-yu pig 3sS-kill-NPST; phak 'pig' cannot project a phrase of its own in this type of construction
Belhare ‘Terminal Word’

- Domain: (object+intransitive affixes) ± all stem parts +affixes
- \( \text{p}^{\text{hak}} \) se?-yu
  - pig [3sS-]kill-NPST
  - 'S/he kills pigs'
- \( \text{p}^{\text{hak}} \) ‘pig’ cannot project a phrase on its own:
  - *\( \text{eik}^{\text{h}a} \text{ p}^{\text{hak}} \) se?yu ‘S/he kills big pigs’
- Not incorporation: \( \text{p}^{\text{hak}}=\text{to} \) se?yu
III. A comprehensive catalogue of word types

- Description of process/constraint & resolution (if applicable)
- Description of phonological domain & alignment type
- Description of grammatical domain & alignment type
- Exceptions or ‘strata’
- Distribution of types, sizes, alignments
What goes into Word Records?
In addition to Manange...

<table>
<thead>
<tr>
<th>LID</th>
<th>Language</th>
<th>WTID</th>
<th>&lt;File&gt;</th>
<th>Unit</th>
<th>Word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>690</td>
<td>Manange</td>
<td>1</td>
<td>&lt;File&gt;</td>
<td>2</td>
<td>Single Tone Contour/Melody</td>
</tr>
<tr>
<td>691</td>
<td>Tibetan (Lhasa)</td>
<td>84</td>
<td>&lt;File&gt;</td>
<td>1</td>
<td>Single Tone Contour/Melody</td>
</tr>
<tr>
<td>664</td>
<td>Kham</td>
<td>282</td>
<td>&lt;File&gt;</td>
<td>8</td>
<td>Single Tone Contour/Melody</td>
</tr>
</tbody>
</table>
**WordType_1: Single Tone Contour**

<table>
<thead>
<tr>
<th>WTID</th>
<th>word_type</th>
<th>Single Tone Contour/Melody</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>phon_def1</td>
<td>10 Tone</td>
</tr>
<tr>
<td></td>
<td>phon_def2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>phon_def3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phon_Domain</td>
<td>4 span</td>
</tr>
<tr>
<td></td>
<td>Notes_Explanations</td>
<td>1 contour or melody per unit</td>
</tr>
<tr>
<td></td>
<td>Gword_Definition</td>
<td></td>
</tr>
</tbody>
</table>
Some other recurring P-Domain Types

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Number of Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>*CC ‘coda’ position restrictions</td>
<td>5</td>
</tr>
<tr>
<td>*CC ‘onset’ position restrictions</td>
<td>10</td>
</tr>
<tr>
<td>Single Main Stress</td>
<td>5</td>
</tr>
<tr>
<td>*V-initial syllable restrictions</td>
<td>6</td>
</tr>
</tbody>
</table>
A ‘Rare’ P-domain: Yidiny
IV. Traditional Approaches to P & G Domains

- The Prosodic Hierarchy
- Prosodic, Grammatical Alignments, & Cross-Domain Alignments
- What we are finding so far...
Prosodic Hierarchy

Series of nesting domains for phonological rules/processes

- phonological utterance (U)
- intonational phrase (IP)
- phonological phrase (φ)
- phonological word (ω)
- foot (F)
- syllable (σ)
- mora (µ)
Prosodic Hierarchy: Alignment

• UG specifies these prosodic constituents, arranged in a hierarchy
• Hierarchy characterized by principles & constraints (Selkirk 1980; Nespor & Vogel 1986)
  • Strict Layer Hypothesis: “A prosodic constituent of rank n is immediately dominated by a single constituent of rank n + 1”
Alignment: Strict Layer Hypothesis

- *ω: ‘skipping’ of a level
- F
- σ
- σ
- *ω: ‘recursivity’ of levels
- ω
• Recursivity permissable (Ladd 1986)
• Violable constraints (Optimality Theory) (Ito & Mester 1992; Selkirk 1995; Peperkamp 1997):
  • NONRECURSIVITY: No $C_i$ dominates another $C_i$
  • EXHAUSTIVITY: No $C_i$ immediately dominates a $C_k$, $k < i-1$
• Non-Recursivity (Truckenbrodt 1999)=
  • P-phrases in recursive structures must be maximally alike in extension
Grammatical Domains: Alignment

• Correspondence constraints kept very general:
  - ALIGN (PCat, Edge; GCat, Edge), where PCat = \{μ, σ, F, ω, φ\} and GCat = \{Af, Rt, Σ, X^0, XP\} (McCarthy & Prince 1993)
  - Wrap-XP (Truckenbrodt 1999) Syntactic phrases are contained within only 1 phonological phrase
  - Tautomorphemicity (Crowhurst 1994; Bickel 1998; 2003)

\*(G-cat1)(G-cat2)

\[P\text{-CAT}\]
Our Findings: Diversity in Word Domains

• Languages often differentiate more than one P-domain
• Different P-domains reference different morphological information (non-isomorphism)
• P-domains not necessarily contained within each other à la traditional ‘recursion’ expectations
Numbers of P-Word Types:
13 Sino-Tibetan Languages

Kinnauri, Meithei, Lhasa Tibetan

Lahu

Kham, Dege Tibetan
Geographic Distribution: Sino-Tibetan

Kyirong Tib., Dolakha Newar, Belhare
Lhasa, Dege Tibetan
Lahu, Kayah Li
Meithei, Burmese
Manange, Limbu
Kham

△ ≤ 5 P-domains; ○ 6-10; ⊙ 11+ P-domains
Non-Isomorphism: 27 languages

7 languages with 2 non-isomorphic p-words

Kham has 13
Based on Available Morpheme Structure
Ratio: Sino-Tibetan only
Recursion: Lahu (Sino-Tibetan)

PW1: Reduced Stress
prefix-stem-suffix

PW2: Stem Tone Change

\( \sigma \omega \sigma \omega \sigma \)
Recursion: Belhare

- Belhare: 12 P-domains; 6 G-domains
- Bipartite Stems
GD4/PD2, 10: Spans stem, suffixes, particle
GD2: Spans both stem pieces
Prefix-$\Sigma_1.\Sigma_2$-Suffix-Particle
GD1: Spans right piece stem
PD5: Contains at least stem + suffix
GD3, 5: Spans stem and all affixes

PD1: Rt. edge stem (+ suffixes + particle)
PD4: Lft. edge stem (+ suffixes + particle)
PD7: Lft. edge prefix + stem (+ suffixes + particle)

PD3, 6, 8, 9: Lft. edge stem only
V. Future Goals

- Catalogue & Typologize on diversity of domain types
- Focus: Sino-Tibetan & Austro-Asiatic vs. ‘World’ sample
- Enhance database with very detailed information (e.g. Questionnaires)
- Outcome: updating of traditional terminologies: fusion types, affix~clitic~particle
“The Questionnaires”

- [http://www.uni-leipzig.de/~autotyp/download/index.html](http://www.uni-leipzig.de/~autotyp/download/index.html)
- Designed for use in multiple settings, scenarios:
  - Reading/writing of grammars & descriptive phonology
  - Fieldwork/elicitation
  - Transcription & analysis of text data
- P-domains: a ‘bottom-up’ approach:
  - $\sigma \gg$ polysyllabic, monomorphemic $\gg$ polymorphemic $\gg$ phrasal/clausal constructions...
  - Charting phonology that is sensitive to morphological (& syntactic) structure
AUTOTYP: People and funding

- **The AUTOTYP research team (as of October 2004)**
  - Johanna Nichols (Co-Director, Berkeley)
  - Balthasar Bickel (Co-Director, Leipzig)
  - Tracy Hall (Co-Director, Word Domains, Indiana University)
  - Fernando Zúñiga (Research Associate, Eugene/Zürich/Manchester/Santiago)
  - Kristine Hildebrandt (Post-Doc, Leipzig)
  - **RAs in Berkeley**: Gabriela Caballero, Suzanne Wilhite
  - **RAs in Leipzig**: Michael Riessler, Sven Siegmund, Sindy Poppitz, Franziska Crell, Kathi Stutz, Josh Wilbur, Jenny Seeg, Anja Gampe, Sebastian Hellmann
  - **Past team members**: Sandra Biewald, Aimee Lahaussois-Bartosik, Dave Peterson, Keith Sanders, Alena Witzlack-Makarevich, Rebecca Voll