Visualizing morphological patterns in inflectional paradigms

Sebastian Bank, Daniela Henze, Jochen Trommer, Eva Zimmermann

University of Leipzig
DGFs 2013
Visualization Workshop
March 13

Background

Inflectional paradigms: organizing sets of inflectional forms

- arrange forms into tables such that adjacent cells correspond to natural classes

**German verbal inflection**

<table>
<thead>
<tr>
<th>1SG</th>
<th>glaub-e</th>
<th>2SG</th>
<th>glaub-st</th>
</tr>
</thead>
<tbody>
<tr>
<td>2SG</td>
<td>glaub-er</td>
<td>3SG</td>
<td>glaub-t</td>
</tr>
<tr>
<td>3PL</td>
<td>glaub-en</td>
<td>1PL</td>
<td>glaub-en</td>
</tr>
</tbody>
</table>

- identify full and partial syncretism
- correlate with invariant meaning
- more complex patterns: blocking, affix order, hierarchy effects

**Hierarchy Effects in Kiranti and Broader Alcig**
http://proalki.uni-leipzig.de/project/
The internal structure of person portmanteaus
http://portmanteau.uni-leipzig.de/

**Agenda**

- keep things simple, bottom-up
- evolve data structures, formats, conventions, best practices
- assist manual analysis with interactive visualization tools
- implement pattern identification in learning algorithms
- gain theoretical and typological insights

**Organizing sets of inflectional forms**

- maximize neighboring table cell form identity/similarity

**German verbal inflection (present and past)**

<table>
<thead>
<tr>
<th>SG</th>
<th>PL</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e</td>
<td>en</td>
<td>te</td>
</tr>
<tr>
<td>2</td>
<td>st</td>
<td>t</td>
<td>test</td>
</tr>
<tr>
<td>3</td>
<td>t</td>
<td>en</td>
<td>te</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e</td>
</tr>
<tr>
<td>2</td>
<td>st</td>
</tr>
<tr>
<td>3</td>
<td>t</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PRESENT</th>
<th>PAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRS</td>
<td>PST</td>
</tr>
<tr>
<td>PRS</td>
<td>PST</td>
</tr>
</tbody>
</table>
Background

Vizualizing paradigms as form-meaning relations

Our database

Data collection: paradigm representation and sources

- represent paradigms, morphemes, rules, etc. as structured documents
- additional information on relevant syntax, morphology, phonology
- kept as pages under the version control of a MediaWiki
- annotated for aggregation & search (Semantic MediaWiki)

- mostly collected from reference grammars with their segmentation
- underlying and surface forms, features, morpheme kind
- alternative segmentations and analyzes, comparison pages
- command-line tools (e.g. row/column reordering)

Our database

Data: (In)transitive verbal agreement paradigms

Paradigm colorizer: implementation and user interface

- html generated by paradigm and comparison pages is annotated
- ParadigmColoriser.js offers in-browser visualization options

- underlying/surface forms, select/hide related information
- highlight morphemes based on different abstraction levels

Implementation
Visualizing basic patterns with morpheme highlighting

**Kohi** (Kiranti, Nepal, Lahaussois 2009)

**Hypotheses**
- $-ki \leftrightarrow [+1+2+pl]
- -ko $\leftrightarrow [+1-2+pl]
- -to \leftrightarrow [Abs+1+sg]

Analyzing imperfect morpheme distributions

**Kohi**

**Hypothesis**
- $-si \leftrightarrow [+du]

Analyzing marker interactions with morpheme highlighting

**Ainu** (Isolate, Japan, Tamura 2000)

**Intransitive**

<table>
<thead>
<tr>
<th>1s</th>
<th>1p</th>
<th>2s</th>
<th>2p</th>
<th>3s</th>
<th>3p</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>ku-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z-as</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Transitive**

<table>
<thead>
<tr>
<th>1s</th>
<th>1p</th>
<th>2s</th>
<th>2p</th>
<th>3s</th>
<th>3p</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Hypotheses**
- $eci_{-1} \rightarrow [+2+pl]
- eci_{-2} \rightarrow [+1][+2+sg]
- e- $\rightarrow [+2+sg]

Finding hierarchy effects with morpheme highlighting

**Kohi 2 \leftrightarrow 3, 3 \rightarrow 3**

**Morpheme hypotheses**
- $-si \leftrightarrow [+du]
- -ni $\leftrightarrow [+pl]

**Agreement hierarchies**
- $2 > 3
- pl > du$
### Visualizing hierarchy effects with feature highlighting

**Reyesano** (Tacanan, Bolivia, Guillaume 2009)

**Prefix hierarchy**  
2 > 1

**Suffix hypothesis**  
-\(ta\) ↔ [Nom +3]

---

### Finding feature based linearization with feature highlighting

**Older Bahing** (Kiranti, Nepal, Houghton 1858)

**Linearization hierarchy**  
1 > 2 > 3

---

### Determining marker alignment with argument highlighting

**Kulung, agreed with argument** (Kiranti, Nepal, Tolsma 2006)

- single argument marking, not when agreed with argument is singular
- generic SP-markers, special A-markers for X → 3, 1s → 2 portmanteau
Finding hierarchical alignment with argument highlighting

**Hayu, agreed with argument** (Kiranti, Nepal, Michailovsky 1981)

- single argument marking, not when agreed with argument is singular
- $1 > 2$ hierarchy, $1s \rightarrow 2$ portmanteau

**Limitations**

- distributional pattern of pure number markers is still hard to see
- underspecified portmanteau patterns (ambiguous exponents)

**Outlook**

- identify and visualize blocking relations between markers
- visualize linear order relations between markers (slots, template)

**Conclusion**

- restructuring paradigm tables helps with small paradigms
- highlighting of morphemes helps to identify patterns, reduce errors and make aware analysis strategies and options
- different abstraction levels (morphemes, features, agreed with argument) are needed to identify more complex patterns
- visual strategies can be transferred to learning algorithms

Hodgson, Brian Houghton (1858), Comparative vocabulary of the languages of the broken Tribes of Nepal (part II). Journal of the Asiatic Society of Bengal 27:393–456.


