The algebraic structure of morphosyntactic features

Sebastian Bank
sebastian.bank@uni-leipzig.de
Department of Linguistics
University of Leipzig
Features in Phonology, Morphology, Syntax and Semantics
University of Tromsø/CASTL
Oktober 31–November 1, 2013

Introduction

Background: Features in morphological subanalysis

Present and past tense forms of German *spielen* ‘to play’

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>spiel-e</td>
<td>spiel-(e)n</td>
</tr>
<tr>
<td>2</td>
<td>spiel-s-t</td>
<td>spiel-t</td>
</tr>
<tr>
<td>3</td>
<td>spiel-t</td>
<td>spiel-(e)n</td>
</tr>
</tbody>
</table>

PRESENT

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>spiel-te</td>
<td>spiel-te-n</td>
</tr>
<tr>
<td>2</td>
<td>spiel-te-s-t</td>
<td>spiel-te-t</td>
</tr>
<tr>
<td>3</td>
<td>spiel-te</td>
<td>spiel-te-n</td>
</tr>
</tbody>
</table>

PAST

Some underspecified marker hypotheses

/-n/ ↔ [-2 +pl]
/-t/ ↔ [-1]

well-formed feature specification = natural class → **systematic syncretism**

Some feature decomposition for pronouns

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>+1-2-3-pl</td>
<td>-1+2-3-pl</td>
<td>-1-2+3-pl</td>
</tr>
<tr>
<td>PL</td>
<td>+1-2-3+pl</td>
<td>+1+2-3+pl</td>
<td>-1+2-3+pl</td>
</tr>
</tbody>
</table>

Two flavors of feature notations

Given a set of paradigm cells (utterances, contexts)
e.g.

{ 1SG, 1PL, 2SG, 2PL, 3SG.MASC, 3SG.FEM, 3SG.NEUT, 3PL }
or

{ 1SG, 1PL.EXCL, 1PL.INCL 2SG, 2PL, 3SG, 3PL }

Morphosyntactic feature specifications

Give formal representation for the meaning of each individual paradigm cell. Define which sets of paradigm cells correspond to more general meanings.

**Feature-value pairs** (Paradigm Function Morphology, Network Morphology)

{ PER:1, NUM:sg }, ... { PER:3, NUM:sg, GEN:neut }, ... { PER:3, NUM:pl }

**Privative/binary features** (Amorphous Morphology, Distributed Morphology)

[+1 -2 -pl], ...[-1 -2 -pl neut], ...[-1 -2 +pl]
Feature-value pairs

Features as orthogonal categories of mutually exclusive values

- **PER**: 1, 2, 3
- **INCL**: yes, no
- **NUM**: sg, pl
- **GEN**: masc, fem, neut

Cooccurrence restrictions

\[ \{\text{PER:1}\} \subseteq X \lor \{\text{PER:2}\} \subseteq X \rightarrow \{\text{GEN:alpha}\} \neq X \]
\[ \{\text{PER:1, INCL:yes}\} \subseteq X \rightarrow \{\text{NUM:pl}\} \subseteq X \]
\[ \{\text{PER:1, NUM:sg}\} \subseteq X \lor \{\text{PER:3}\} \subseteq X \rightarrow \{\text{INCL:no}\} \subseteq X \]

Ordered attribute paths in DATR

\[ \text{TNS} < \text{PER} < \text{NUM} \]
\[ \langle \text{past 1 sg}>, \langle \text{present 3}>, \ldots \]

Privative/binary features

Feature decomposition

1. **EXCL** = [+1 -2] \quad **SG** = [-pl] \quad **MASC** = [masc] \quad **MASC** = [+m -f]
2. **INCL** = [+1 +2] \quad **PL** = [+pl] \quad **FEM** = [fem] \quad **FEM** = [-m +f]
3. **= [-1 +2] \quad **NEUT** = [neut] \quad **NEUT** = [-m -f]

Feature combinations

<table>
<thead>
<tr>
<th>sg</th>
<th>pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EXCL = [+1 -2 -pl] 1PL.EXCL = [+1 -2 +pl]</td>
</tr>
<tr>
<td>12</td>
<td>1PL.INCL = [+1 +2 +pl]</td>
</tr>
<tr>
<td>2</td>
<td>2SG = [-1 +2 -pl] 2PL = [-1 +2 +pl]</td>
</tr>
<tr>
<td>3</td>
<td>3SG = [-1 -2 -pl] 3PL = [-1 -2 +pl]</td>
</tr>
</tbody>
</table>

Natural classes: syncretism vs. accidental homophony

15 possible assignments to a 4 cell paradigm

<table>
<thead>
<tr>
<th>Natural class syncretism</th>
<th>Elsewhere syncretism</th>
<th>Overlapping distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary features</td>
<td>cells</td>
<td>possible assignments</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>10,480,142,147</td>
</tr>
</tbody>
</table>

Features and their possible combinations

- restrict the sets of paradigm cells that can be part of **systematic syncretism**
- account for the fact that natural class syncretism is **more frequent** than expected if **learners** indistinctively internalized random form-identities

Formal Concept Analysis

Practical application of **order** and **lattice** theory (Birkhoff 1940) introduced by Wille (1982), elaborated in Ganter & Wille (1999).

Rests upon a Galois connection between two sets: a set of **objects** to describe and a set of **attributes** which each object either has or not (boolean flags).

Basic elements of Formal Concept Analysis (FCA)

- The **formal context** \( (O,A,R) \) defines a relation between **objects** and **attributes**.
- The **derivation operator** \( 'r' \) yields **common attributes** for objects and **common objects** for attributes.
- The **concept lattice** \( L(O,A,R) \) defines the **relations** and **operations** on objects-attributes pairs.

Provides precise definitions, terminology, and graphical representations for the way feature notations are used (mostly implicitly) in linguistics.

Has many more practical applications, algorithms, software tools, etc., see [http://www.upriss.org.uk/fca/fca.html](http://www.upriss.org.uk/fca/fca.html)
Formal context: defining a feature system

Context defines the relation between objects and attributes

Drop feature/value distinction: translate all values into privative features

<table>
<thead>
<tr>
<th></th>
<th>+1</th>
<th>2</th>
<th>3</th>
<th>+sg</th>
<th>+pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pe</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pi</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2s</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2p</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3s</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3p</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( O = \{ 1s, 1pe, 1pi, 2s, 2p, 3s, 3p \} \)
\( \mathcal{A} = \{ +1, -1, +2, -2, +3, -3, +sg, +pl \} \)
\( \mathcal{R} \subseteq O \times \mathcal{A} = \{ \{ 1s, +1 \}, \{ 1s, -2 \}, \ldots, \{ 3p, +pl \} \} \)

Dichotomic scale

<table>
<thead>
<tr>
<th>3.masc</th>
<th>3.fem</th>
<th>3.neut</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>excl</td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

Nominal scale

<table>
<thead>
<tr>
<th>very high</th>
<th>high</th>
<th>low</th>
<th>very low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ordinal scale

<table>
<thead>
<tr>
<th>positive</th>
<th>comparative</th>
<th>superlative</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Biordinal scale

| x | x | x | x |

Concept lattice: relations and operations

\( \langle O_1, A_1 \rangle \leq \langle O_2, A_2 \rangle \) when \( O_1 \subseteq O_2 \) (or equivalently \( A_1 \supseteq A_2 \))
Feature systems as context and lattice

Concept lattice, object concepts, attribute concepts

Relations and operations

\[ [+1] \lor [-1] = \top \]
\[ [+\text{sg}] \lor [+p1] = \top \]

\[ [+1] \land [-1] = \bot \]
\[ [+1] \land [+2] = \bot \]


\[ [-1] \land [-3] \neq \bot \quad \text{and} \quad [-1]' \lor [-3]' = \top' \]

\[ [+1 +\text{sg}] \lor [+2 +p1] = [-3] \quad \lor \{ [+1 +\text{sg}], [+2 +\text{sg}], [+2 +p1] \} = [-3] \]

\[ [+1] \land [+\text{sg}] = [+1 +\text{sg}] \quad \land \{ [-2], [-3], [+\text{sg}] \} = [+1 +\text{sg}] \]

tautology, contradiction, implication, subcontrary, intersection, unification

Formal Concept Analysis

 Conjunctive normal form, boolean algebra

trivial, nominal scale

boolean algebra, 2

15 / 24
**Syncretism, underspecification, and insertion competition**

**Present and past tense forms of English ‘to be’**

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>am</td>
<td></td>
<td>1</td>
<td>was</td>
<td>were</td>
</tr>
<tr>
<td>2</td>
<td>are</td>
<td></td>
<td>2</td>
<td>were</td>
<td>were</td>
</tr>
<tr>
<td>3</td>
<td>is</td>
<td></td>
<td>3</td>
<td>was</td>
<td>were</td>
</tr>
</tbody>
</table>

**Fully specified**

\[
\text{am} \leftrightarrow [+1_{sg \text{prs}}] \\
\text{was} \leftrightarrow [-2_{-p1 \text{pst}}] \\
\text{were} \leftrightarrow [+1_{pl \text{pst}}] \\
\]

**Natural class syncretism**

\[
\text{is} \leftrightarrow [+3_{+sg \text{prs}}] \\
\]

**Elsewhere syncretism**

\[
\text{are} \leftrightarrow [\text{prs}] \\
\]

*Insertion with Pāṇinian blocking* (a.k.a. subset principle, elsewhere principle)

Insert the **most specific** marker(s) whose meaning subsume the paradigm cell meaning.

**Insertion of were ↔ [pst]**

\[
\{[-2_{-p1 \text{pst}}] \geq [+1_{+sg \text{prs}}] \}
\]

**Insertion of were ↔ [pst]**

\[
\{[-2_{-p1 \text{pst}}] \geq [+1_{+sg \text{prs}}] \}
\]

**Markedness of extended exponence hypothesis**

The utterance of a subsuming marker does not contribute information. It involves additional formal machinery (feature copying, rule blocks, contextual features, marker sensitivity, enrichment) and correspondingly is harder to learn.

**Contextual feature solution** (insertion as feature discharge, Noyer 1992)

discharged features / non-discharged features

**No masked extended exponence with extensionalism**

<table>
<thead>
<tr>
<th></th>
<th>sg</th>
<th>pl</th>
<th></th>
<th>sg</th>
<th>pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-e</td>
<td>1</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-n</td>
</tr>
<tr>
<td>2</td>
<td>[-1_{+2_{-p1}}]</td>
<td>-n</td>
<td>2</td>
<td>[-1_{+2_{-p1}}]</td>
<td>-n</td>
</tr>
<tr>
<td>3</td>
<td>[-1_{-2_{-p1}}]</td>
<td>-n</td>
<td>3</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-n</td>
</tr>
</tbody>
</table>

Does not interpret [-]insertion in 2SG as extended exponence (but might).

Requires that [-] ↔ [-1] is not a superconcept of [+1] ↔ [+2_{-p1}]. autonomy

But this requires that some paradigm cell is +2 and not -1. extensionalism

**Extensionalist analysis**

\[
\{2s, 2p, 3s, 3p\} \leftrightarrow \{[-1_{1\text{incl.aug}}], [-1_{1\text{incl.min}}] \}
\]

**Contextual features solution**

\[
\{[-1_{1\text{incl.aug}}], [-1_{1\text{incl.min}}] \}
\]

predicts functional pressure to change [-] into [-]

**When markers resist blocking: extended exponence**

**Agreement affixes of Fox animate intransitive verbs** (Bloomfield 1927)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ne-</td>
<td>ne-</td>
<td>11</td>
<td>ke-</td>
<td>ke-</td>
</tr>
<tr>
<td>12</td>
<td>ke-</td>
<td>ke-</td>
<td>12</td>
<td>wa-</td>
<td>wa-</td>
</tr>
<tr>
<td>3</td>
<td>-pl</td>
<td>-pl</td>
<td>3</td>
<td>-pl</td>
<td>-pl</td>
</tr>
</tbody>
</table>

**Extended exponence**

\[
\text{wa} \leftrightarrow [+3] \geq [\text{g}]
\]

**Markedness of extended exponence hypothesis**

The utterance of a subsuming marker does not contribute information. It involves additional formal machinery (feature copying, rule blocks, contextual features, marker sensitivity, enrichment) and correspondingly is harder to learn.

**Contextual feature solution** (insertion as feature discharge, Noyer 1992)

discharged features / non-discharged features

**No masked extended exponence with extensionalism**

<table>
<thead>
<tr>
<th></th>
<th>sg</th>
<th>pl</th>
<th></th>
<th>sg</th>
<th>pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-e</td>
<td>1</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-n</td>
</tr>
<tr>
<td>2</td>
<td>[-1_{+2_{-p1}}]</td>
<td>-n</td>
<td>2</td>
<td>[-1_{+2_{-p1}}]</td>
<td>-n</td>
</tr>
<tr>
<td>3</td>
<td>[-1_{-2_{-p1}}]</td>
<td>-n</td>
<td>3</td>
<td>[+1_{-2_{-p1}}]</td>
<td>-n</td>
</tr>
</tbody>
</table>

\([+2] \neq [-1] \text{ only if there is a } [+1_{1\text{incl.aug}}] \) cell
\([+2_{-p1}] \neq [-1_{+2_{-p1}}] \text{ only if there is a } [+1_{1\text{incl.aug}}] \) cell

However, such an inclusive/augmented reanalysis gives:

1. *Wir spiel-s.
   we play-1INC\_MIN
2. *Wir spiel-e.
   we play-1INC\_AUG
Why to avoid autonomous feature algebra?

- cannot replace extended exponence machinery altogether without undermining natural class restrictivity by adding features
- introduces superficially equivalent options (analytical ambiguity) of exploiting feature autonomy vs. using additional machinery
- results in less specific predictions making analyses harder to test
- why prefer a less restrictive theory when a more restrictive version has not yet been falsified?
- if the choice between [+2] and [-1 +2] is only indirectly observable, how can it be learned?
- is there independent evidence for such ‘morphomic’ features other than the distributional effects they have?

Impoverishment with or without autonomous features

**Autonomy**

\[
A \leftrightarrow [+3] \quad B \leftrightarrow [-1 -2 \text{ pst}]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>PRESENT</td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{pst} \rightarrow \emptyset / [+3 +\text{pl pst}]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>PRESENT</td>
<td></td>
</tr>
</tbody>
</table>

\[
-1 -2 +3 \rightarrow \emptyset / [+3 +\text{pl pst}]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>PRESENT</td>
<td></td>
</tr>
</tbody>
</table>

**Extensionalist**

\[
A \leftrightarrow [+3] \quad B \leftrightarrow [\text{pst}] / [+3]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>PRESENT</td>
<td>PAST</td>
</tr>
</tbody>
</table>

\[
+3 \rightarrow \emptyset / [+3 +\text{pl pst}]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>PRESENT</td>
<td>PAST</td>
</tr>
</tbody>
</table>

Impossible only retreat to the general case

Feature set subtraction in morphological operations

**Feature set subtraction** (Noyer 1992)

\[
\text{impoverishment} \& \text{ fission} (\text{Halle} / \text{Marantz} 1993)
\]

**Impoverishment rule**

\[
[\pm 1] \rightarrow \emptyset / [+\text{pl}]
\]

(Frampton 2002)

\[
[+3 +\text{pl}] - [+1] = [+3 +\text{pl}]
\]

\[
[+1 +\text{pl}] - [-1] = [+1 +\text{pl}]
\]

\[
[+3 +\text{pl}] - [-1] = [-2 +\text{pl}]
\]

\[
[+1 +\text{pl}] - [+1] = ? \quad \text{not} -2 \text{ or} -3
\]

\[
(-2 +\text{pl}, -2, +\text{pl}) \quad (-2 +\text{pl}, -3 +\text{pl}, +\text{pl})
\]

Impoverishment \leftrightarrow feature discharging \emptyset-insertion

(Trommer 1999, 2003)

Subtraction as \emptyset-insertion without autonomous features

\[
[+3 +\text{pl}] - [-1]
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>3</td>
<td>ØA</td>
<td>ØA</td>
</tr>
</tbody>
</table>

\[
+3 +\text{pl}
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>2</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>3</td>
<td>ØB</td>
<td>ØB</td>
</tr>
</tbody>
</table>

\[
-2 +\text{pl}
\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>2</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>3</td>
<td>ØA</td>
<td>ØA</td>
</tr>
</tbody>
</table>

Regarding subtraction as insertion without form-change

- makes various (possibly overly powerful) formalisms more restricted
- allows for a consistent information-based interpretation
Summary

Conclusion

- if features are more than abbreviations for observable distributional facts, even simple formalisms can acquire considerable power
- at least in some cases it is undesirable to use this extra power – not before there is evidence that it is really needed
- Formal Concept Analysis provides the terminology and the tools to spot and disassemble such ‘feature tricks’
- learnability might raise fundamental objections against them
- for the most part feature autonomy can be avoided by always using the most specific notational variant for representing feature sets

References