Underspecification in realisational morphology

Berthold Crysmann and Olivier Bonami

Laboratoire de linguistique formelle — U. Paris Diderot & CNRS

AnaMorphoSys — Lyon, June 2016
Generalisations over exponence

- In many inflection systems, the same exponents may be used in different ways in different contexts.
- We present a formal theory of inflection that is well suited to modeling such situations.
- We highlight 4 types of exponence with variable content:
  1. Parallel exponence
     The same shapes realise related but distinct property sets in different positions in the word.
  2. Polyfunctionality
     The same shapes realise related but distinct property sets depending on part of speech.
  3. Conditioned placement of exponents
     The same shapes realise the same property sets in different positions in different contexts.
  4. Gestalt exponence
     Content is assigned to combinations of exponents rather than individual exponents.
Parallel exponence exemplified

- The paradigms of Swahili subject and object markers are nearly identical.

<table>
<thead>
<tr>
<th>PER GEN</th>
<th>SUBJECT</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SG</td>
<td>PL</td>
</tr>
<tr>
<td>1</td>
<td>ni</td>
<td>tu</td>
</tr>
<tr>
<td>2</td>
<td>u</td>
<td>m</td>
</tr>
<tr>
<td>3</td>
<td>M/WA</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M/MI</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td>KI/VI</td>
<td>ki</td>
</tr>
<tr>
<td></td>
<td>JI/MA</td>
<td>li</td>
</tr>
<tr>
<td></td>
<td>N/N</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td>U/N</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td>KU</td>
<td>ku</td>
</tr>
</tbody>
</table>
Parallel exponence exemplified

- The paradigms of Swahili subject and object markers are nearly identical.
- However, subject and object markers occur in different positions (Stump, 1993).

(1)  

(a) ni-ta-wa-penda  
1SG-FUT-3PL-like  
‘I will like them.’

(b) wa-ta-ni-penda  
3PL-FUT-1SG-like  
‘They will like me.’

→ Position, rather than shape, disambiguates which grammatical function is coded.
Polyfunctionality exemplified

- Tundra Nenets uses the same paradigms of person-number and number-case markers in objective conjugation and possessive declension (Ackerman and Bonami, inpress)

(2)  

a. yemp°q-ŋa-x°yu-da  
dress-FIN-DU-3SG  
‘They two dressed her/him.’

b. ngəno-x°yu-da  
boat-DU-3SG  
‘his/her two boats’
Polyfunctionality exemplified

- Tundra Nenets uses the same paradigms of person-number and number-case markers in objective conjugation and possessive declension (Ackerman and Bonami, in press)

- This holds even in situations of overlapping exponence

\[(2) \quad \begin{array}{l}
\text{a. meə-}m-’ih \\
take-SG.1-DU \\
’We (du.) take it/her/him.’
\end{array}
\]

\[(2) \quad \begin{array}{l}
\text{b. te-}m-’ih \\
reindeer-NOM.SG.1-DU \\
’our (du.) reindeer’
\end{array}
\]

- Thus:

\[
\text{Possessed noun} \sim \text{Objective verb}
\]

\[
\text{possessor} \sim \text{subject}
\]

\[
\text{possessed} \sim \text{object}
\]
Conditioned placement exemplified

- In Moro, object markers occur in different positions in different TMA combinations.

  (3)  
  a.  
  \[
  g-a-\eta\acute{a}-\varleth\varleth-a \\
  \text{SM.CL-RTC-2SG.OM-pull-IPFV} \\
  \text{‘s/he is about to pull you’ (Jenks and Rose, 2015, 271)}
  \]
  b.  
  \[
  g-\acute{a}-\varleth\varleth-\acute{a}-\eta\acute{a} \\
  \text{SM.CL-DIST.IPFW-pull-DIST.IPFW-2SG.OM} \\
  \text{‘s/he is about to pull you from there to here’}
  \]
- Object marker placement predictable from tone pattern
- However, a side effect is that the position of object markers acts as secondary exponents of TMA.
- See Crysmann and Bonami (2016) for many more examples and a typology of variable placement.
Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>nokk</td>
<td>nok-a-d</td>
</tr>
<tr>
<td>GEN</td>
<td>nok-a</td>
<td>nokk-a-de</td>
</tr>
<tr>
<td>PART</td>
<td>nokk-a</td>
<td>nokk-a-sid</td>
</tr>
</tbody>
</table>
Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: \{\text{GEN.SG, NOM.PL}\} vs. all other cells.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>nokk</td>
<td>nok-a-d</td>
</tr>
<tr>
<td>GEN</td>
<td>nok-a</td>
<td>nokk-a-de</td>
</tr>
<tr>
<td>PART</td>
<td>nokk-a</td>
<td>nokk-a-sid</td>
</tr>
</tbody>
</table>
Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: \{\text{GEN.SG, NOM.PL}\} vs. all other cells.
- Theme vowels: \text{NOM.SG} vs. all other cells.

\begin{tabular}{lcc}
 & \textbf{SG} & \textbf{PL} \\
\text{NOM} & nokk & nok-a-d \\
\text{GEN} & nok-a & nokk-a-de \\
\text{PART} & nokk-a & nokk-a-sid \\
\end{tabular}

\textbf{‘beak’}
Gestalt exponence exemplified

- Blevins (2005): while Estonian nouns are easily segmentable, exponents are not associated with stable content.
- Stem alternations: \{\text{GEN.SG}, \text{NOM.PL}\} vs. all other cells.
- Theme vowels: \text{NOM.SG} vs. all other cells.
- Singular forms contrast in shape, although no exponent is dedicated to the expression of a particular case value.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOM</strong></td>
<td>nokk</td>
<td>nok-a-d</td>
</tr>
<tr>
<td><strong>GEN</strong></td>
<td>nok-a</td>
<td>nokk-a-de</td>
</tr>
<tr>
<td><strong>PART</strong></td>
<td>nokk-a</td>
<td>nokk-a-sid</td>
</tr>
</tbody>
</table>

“Case properties are realised by the wordforms [...], and words are characterized by different \textit{combinations} of formatives”.

(Blevins, 2005, 3)
Our goal

- We present aspects of Information-based Morphology, a realisational theory of morphology that embraces the diversity of exponence (Crysmann and Bonami, 2016).
  - In the general case, a realisation rule is a partial generalisation over words linking a set of $m$ morphs with a set of $n$ morphosyntactic properties.
  - Underspecification allows us to state directly generalisations about exponents at various levels of granularity.
- We show how the theory deals with different types of reuse of exponents.
- We treat two crucial examples:
  1. Parallel exponence in Swahili
  2. Gestalt exponence in Estonian
1. **Constructive vs. abstractive** (Blevins, 2006): two modes of description
   - In a **constructive** approach, the shape of words is deduced from other primitives (morphemes, stems, rules, etc.).
   - In an **abstractive** approach, words are primitive; stems, exponents, etc. are abstractions deduced from these primitives.

2. **Exponence vs. Implicative structure**: two empirical questions
   - **Exponence** is the relation between properties expressed by a word and aspects of the word’s shape expressing them.
   - **Implicative relations** are relations between words expressing different property sets.
Important distinctions

- Classical generative morphology is a constructive approach to exponence.
- Blevins (2006); Ackerman et al. (2009) and the following literature adopt an abstractive approach to implicative relations.
- We argue that the two distinctions are orthogonal.
- The present approach:
  - has both constructive and abstractive interpretations;
  - is entirely focused on exponence.
For the purposes of inflection, words can be seen as associations between a phonological shape (PH) and a morphosyntactic property set (MS).

\[
\begin{bmatrix}
\text{PH} & \langle\text{ɪ̆ɪ̆ɪ̆ŋ}\rangle \\
\text{MS} & \left\{\left[\text{LID \ rain}\right],\left[\text{TMA \ prs-ptcp}\right]\right\}
\end{bmatrix}
\]

As a first approximation, rules of exponence can be seen as underspecified descriptions of words.

\[
\begin{bmatrix}
\text{PH} & \langle\ldots\text{iŋ}\rangle \\
\text{MS} & \left\{\left[\text{TMA \ prs-ptcp}\right],\ldots\right\}
\end{bmatrix}
\]
Realisations rules as generalisations over words II

- Because words can consist of more than two bits, we need some way to index position within a word.
  → rule blocks in AMM (Anderson, 1992) and PFM (Stump, 2001)
- Instead we use explicit reference to numbered positions.
  → explicit list of morphs (MPH)

| PH | <reiingen> |
| MPH | { [PH <rei>][PH <en>] } |
| MS | { [LID rain][TMA prs-ptcp] } |

| Rule of exponence: |
| MPH | { [PH <en>],... } |
| MS | { [TMA prs-ptcp],... } |

- Trivial relationship between a word’s phonology (a string) and its morphs (a set of strings indexed for position).
- Easily captures cumulative exponence (1 morph:n properties), extended exponence (m:1) and overlapping exponence (m:n).
Realisations rules as generalisations over words III

- However, this simple view does not allow one to speak of situations where the same association between form and content is used more than once in the same word.
  - Parallel exponence (see above)
  - Exuberant exponence (Harris, 2009)

- We add an extra layer of abstraction:
  1. A word’s representation includes a specification of which realisation rules license the relation between its form and content.
  2. Realisation rules express a relation between a set of morphs of fixed arity and a specific set of morphosyntactactic properties, the morphology under discussion (MUD).

\[
\begin{align*}
\text{MPH} & \left\{ \begin{array}{c}
\text{PH} \\
\text{PC}
\end{array} \right\} \\
\text{MUD} & \left\{ \begin{array}{c}
\text{TMA} \\
\text{prs-ptcp}
\end{array} \right\}
\end{align*}
\]

3. A principle of morphological well-formedness ensures that
   3.1 The properties expressed by rules add up to the word’s property set
   3.2 The morphs introduced by rules add up to the word’s morph list.
Realisations rules as generalisations over words IV

- For the technically inclined:

\[
\text{word} \rightarrow \begin{cases}
\text{MPH} & e_1 \cup \cdots \cup e_n \\
\text{RR} & \left\{ \begin{bmatrix} \text{MPH} & e_1 \\ \text{MUD} & m_1 \end{bmatrix}, \ldots, \begin{bmatrix} \text{MPH} & e_n \\ \text{MUD} & m_n \end{bmatrix} \right\} \\
\text{MS} & m_1 \cup \cdots \cup m_n
\end{cases}
\]

Morphological well-formedness

- In our example:

\[
\begin{cases}
\text{MPH} & \left\{ \begin{bmatrix} \text{PH} & <\text{jein}> \\ \text{PC} & O \end{bmatrix}, \begin{bmatrix} \text{PH} & <\text{yn}> \\ \text{PC} & I \end{bmatrix} \right\} \\
\text{RR} & \left\{ \begin{bmatrix} \text{MPH} & <\text{jein}> \\ \text{PC} & O \end{bmatrix}, \begin{bmatrix} \text{MPH} & <\text{yn}> \\ \text{PC} & I \end{bmatrix} \right\} \\
\text{MUD} & \left\{ \begin{bmatrix} \text{LID} & \text{rain} \end{bmatrix} \right\} \\
\text{MS} & \left\{ \begin{bmatrix} \text{LID} & \text{rain} \\ \text{TMA} & \text{prs-iptc} \end{bmatrix} \right\}
\end{cases}
\]
In short:

- Realisation rules are abstractions over words, stating that some collection of morphs jointly express some collection of properties.
- Morphological well-formedness ensures ‘Total Accountability’ (Hockett, 1947).
- The 1:1 relation of the classical morpheme is one possibility, but the framework accommodates many other situations.
Generalisations over rules

- Back to our initial goal: capturing the variable content of exponents.
- Example: Swahili

(4) a. *ni-ta-wa-penda*
   1SG-FUT-3PL-like
   ‘I will like them.’

b. *wa-ta-ni-penda*
   3PL-FUT-1SG-like
   ‘They will like me.’
Hierarchies of rules

- Strategy familiar from HPSG: organise realisation rules into a (monotonous) multiple inheritance hierarchy

```
realisation-rule

SHAPE

position

MPH  MPH  MPH  MPH
  {PH <ni>}  {PH <wa>}  {PC -3}  {PH <ni>}
MUD  MUD  MUD  MUD
  {PER 1  NUM sg}  {PER 3  NUM pl}  {subj}  {obj}

MPH  MPH  MPH  MPH
  {PH <ni>}  {PH <wa>}  {PC -3}  {PH <ni>}
MUD  MUD  MUD  MUD
  {subj}  {obj}  {PC -1}  {PC -1}

MPH  MPH  MPH  MPH
  {PH <ni>}  {PH <wa>}  {PC -3}  {PH <ni>}
MUD  MUD  MUD  MUD
  {subj}  {obj}  {PC -1}  {PC -1}
```

16
Hierarchies of rules

- Monotonous multiple inheritance hierarchies have a natural abstractive interpretation: nodes in the hierarchy state what some words (or word parts) have in common.
Hierarchies of rules

- A constructive interpretation of the same hierarchies can be given using online type construction (Koenig and Jurafsky, 1994).
- The complete hierarchy is deduced from a reduced hierarchy by expanding all combinations of types.
Hierarchies of rules

- Pre-linking a rule in multiple dimensions blocks overgeneralisation.
Interim conclusion

- We present a view of exponence where:
  - A single rule may link $m$ properties with $n$ exponents
  - Similarities and differences between rules are captured in a monotonous multiple inheritance hierarchy
  - Because it is monotonous and multi-dimensional, the hierarchy can be interpreted abstractively or constructively.
- Allows for a simple account of parallel exponence in Swahili.
- For Swahili, it is crucial that exponents of subject and object marking be introduced separately
  - This allows us to say that rules for subjects and objects have something in common
- We now turn to a system where it is crucial that all exponents be introduced simultaneously.
In Estonian declension, the number of morphs in a word plays a crucial role in exponence.

<table>
<thead>
<tr>
<th></th>
<th>‘beak’</th>
<th>‘workbook’</th>
<th>‘seminar’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SG</strong></td>
<td>nokk</td>
<td>õpik</td>
<td>seminar</td>
</tr>
<tr>
<td><strong>PL</strong></td>
<td>nok-a-d</td>
<td>õpik-u-d</td>
<td>seminar-i-d</td>
</tr>
<tr>
<td><strong>GEN</strong></td>
<td>nok-a</td>
<td>õpik-u</td>
<td>seminar-i-te</td>
</tr>
<tr>
<td><strong>PART</strong></td>
<td>nokk-a</td>
<td>õpik-u-t</td>
<td>seminar-i-id</td>
</tr>
</tbody>
</table>

In these inflection classes:
- The plural is characterised by the presence of 3 distinct morphs
- 1 to 3 morphs in the singular.
- The nominative singular is characterised by a bare stem

This motivates a holistic analysis, where all morphs in a word jointly realize content.

Can be readily captured in the present framework.
Three dimensions controlling:

- **STEM** the choice of a stem alternant
- **THEME** the possible introduction of a theme vowel
- **SFX** the possible introduction of a case-number suffix

```
realisation-rule
  └── stem-rule
      └── wk-st-rule
      └── grl-st-rule
  └── theme-rule
      └── n-sg-rule
          └── g-sg-rule
              └── g-pl-rule
                  └── g-pl-d-rule
                      └── g-pl-t-rule
                      └── grl-p-pl-rule
                          └── spc-p-pl-rule
```
Some rule types in the THEME and SFX dimensions jointly determine the arity of the set of morphs:
Simultaneous introduction in Estonian
Simultaneous introduction in Estonian

realisation-rule

STEM

st-rule

theme-rule

THEME

sg-rule

SFX

n-pl-wk-st-rule

g-sg-wk-st-rule

n-sg-rule

spc-p-sg-rule

grl-sg-rule

pl-rule

n-p-rule

p-p-rule

g-pl-rule

g-pl-d-rule

g-pl-t-rule

grl-p-pl-rule

spc-p-pl-rule

MUD

MPH

PH [nok]

PC 0

LID okk

TV [a]

WEAK-ST [nok]

MUD

PC 1

PC 2

CASE nom

NUM pl

22
Simultaneous introduction in Estonian

realisation-rule

STEM

st-rule

wk-st-rule

girl-st-rule

theme-rule

n-sg-rule

[MPH \{\[]\}]

spc-p-sg-rule

[MPH \{\[]\}]

girl-sg-rule

[MPH \{\[],\[]\}]

SFX

grl-st-rule

ng-sg-wk-st-rule

n-pl-wk-st-rule

pl-rule

[MPH \{\[]\}]

n-p-rule

p-p-rule

g-std-rule

g-pl-t-rule

g-rp-pl-rule

spc-p-pl-rule

MUD

MPH

\{PH <nok>, PC O \}

\{PH <a>, PC 1 \}

LID

\{nokk, TV <a>, weak-st <nokk> \}

NUM

\{case gen sg \}

num

(1)
Simultaneous introduction in Estonian
Simultaneous introduction in Estonian

```
realisation-rule

STEM
  st-rule
  wk-st-rule
  g-sg-wk-st-rule

THEME
  theme-rule
  sg-rule
  grl-st-rule
  [n-sg-rule MPH {[]} [spc-p-sg-rule MPH {[]} [grl-sg-rule MPH {[]}]]]

SFX
  pl-rule
    MPH {}
  n-p-rule
  p-p-rule
  g-pl-rule
  g-pl-d-rule
  g-pl-t-rule
  grl-p-pl-rule
  spc-p-pl-rule

```

```
MPH
  {PH <nokk>}
  {PC O}

MUD
  {LID nokk}
  {TV <a>}
  {ST <nokk>}

```
Simultaneous introduction in Estonian

realisation-rule

STEM

st-rule

theme-rule

wk-st-rule

grl-st-rule

g-sg-wk-st-rule

g-sg-rule

n-sg-rule

spc-p-sg-rule

grl-sg-rule

g-pl-rule

pl-rule

MPH \{[],[],[]\}

n-p-rule

p-p-rule

g-pl-d-rule

g-pl-t-rule

grl-p-pl-rule

spc-p-pl-rule

MUD

MPH

\[
\begin{array}{l}
\text{PH} <\text{nokk}>\\
\text{PC} \text{ O}
\end{array}
\]

\[
\begin{array}{l}
\text{PH} <\text{a}>\\
\text{PC} \text{ 1}
\end{array}
\]

\[
\begin{array}{l}
\text{PH} <\text{sid}>\\
\text{PC} \text{ 1}
\end{array}
\]

\[
\begin{array}{l}
\text{LID} \text{ nokk}\\
\text{TV} <\text{a}>\\
\text{ST} <\text{nokk}>
\end{array}
\]

\[
\begin{array}{l}
\text{CASE part}\\
\text{NUM pl}
\end{array}
\]
Conclusions on Estonian

- This account captures crucial insights of Blevins (2005); Blevins et al. (in press) on the Estonian declension system:
  - Segmentation is clear, but there is no stable association between segments and morphosyntactic content
  - Each dimension captures a series of contrasts, although these contrasts are not strictly tied to positions.
  - Paradigmatic opposition is captured holistically for the word
    - No empty element is needed.

- But:
  - The account can be made sense of both in constructive and in abstractive terms.
  - The account says nothing on implicative relations
    - This is deliberate: we take exponence and implicative structure to be orthogonal questions.
Conclusions

- Exponents with variable content should be a core concern of theories of inflection.
- Information-based Morphology is particularly well-equipped to address such situations:
  - Individual rules express $m:n$ relations between form and content.
  - Underspecification as a single mechanism to capture similarity.
- Two case studies:
  - A proper treatment of Swahili requires individual introduction of exponents
  - A proper treatment of Estonian requires holistic introduction of exponents.
- We provide a formally sound basis for developing a constructional approach to inflection (Gurevich, 2006).
  - Rules of exponence are word-internal constructions
    - organized in a system of paradigmatic oppositions,
    - ranging from the most specific to the most abstract.
  - The combinatorics are very different from that of syntactic constructions.
References


