The polysemy of lexeme formation rules

Delphine Tribout & Olivier Bonami

Workshop Semantics of derivational morphology:
Empirical evidence and theoretical modeling

July 01, 2014
We address the issue of polysemy of lexeme formation processes in French.

The polysemy of particular morphological processes has been studied extensively in the last decade for French.

We propose a formal analysis of empirical studies on polysemy.

We use the HPSG framework.

We work in the theoretical framework of lexemic morphology (Aronoff, 1994; Fradin, 2003).
Polysemous Lexeme Formation Rules

- Deverbal LFRs often produce different types of meaning. Example: -age suffixation rule

<table>
<thead>
<tr>
<th>type</th>
<th>example</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>event N</td>
<td>LAVAGE ‘cleaning’</td>
<td>act of cleaning</td>
</tr>
<tr>
<td>instrument N</td>
<td>AIGUILLAGE ‘railroad switch’</td>
<td>what one uses to switch railroad</td>
</tr>
<tr>
<td>location N</td>
<td>GARAGE ‘parking lot’</td>
<td>place where one parks</td>
</tr>
</tbody>
</table>

- Output lexemes are often ambiguous between two types

<table>
<thead>
<tr>
<th>types</th>
<th>examples</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>event N</td>
<td>CIRAGE ‘polishing’</td>
<td>act of polishing</td>
</tr>
<tr>
<td>instrument N</td>
<td>CIRAGE ‘shoe polish’</td>
<td>what one uses to polish</td>
</tr>
<tr>
<td>event N</td>
<td>PASSAGE ‘passing’</td>
<td>act of going through</td>
</tr>
<tr>
<td>location N</td>
<td>PASSAGE ‘path’</td>
<td>location through which one goes</td>
</tr>
</tbody>
</table>
This situation can lead to posit polysemous LFRs: a single rule with multiple semantics.

\[
\text{age-lfr} : \begin{bmatrix}
\text{PHON} & \phi \\
\text{CAT} & \text{verb} \\
\text{SEM} & \text{‘V’}
\end{bmatrix} \Rightarrow \begin{bmatrix}
\text{PHON} & \phi \oplus a3 \\
\text{CAT} & \text{noun} \\
\text{SEM} & \begin{cases}
\text{‘action of Ving’}, \\
\text{‘place where one Vs’}, \\
\text{‘object used to V’}
\end{cases}
\end{bmatrix}
\]

A polysemous rule gives rise to a single polysemous lexeme

\[
\begin{bmatrix}
\text{PHON} & \text{si\&e} \\
\text{CAT} & \text{verb} \\
\text{SEM} & \text{‘polish’}
\end{bmatrix} \Rightarrow \begin{bmatrix}
\text{PHON} & \text{si\&a3} \\
\text{CAT} & \text{noun} \\
\text{SEM} & \{\text{‘polishing’, ‘shoe polish’}\}
\end{bmatrix}
\]
Semantics of derived lexemes

Derived lexemes are not polysemous

- However, when polysemous lexemes are used, a single interpretation is selected

(1)

J’ ai acheté du cirage noir.
I have bought PART shoe_polish black
‘I bought black polish’

(2)

Le cirage de mes bottes m’ a pris dix minutes
The polishing of my boots me has taken ten minutes
‘It took me ten minutes to shine my boots’

(3)

*Grâce à ce cirage noir, celui de mes bottes sera vite fait.
Thanks to this shoe_polish black that of my boots will be quickly made
(int.) ‘Thanks to this black polish, polishing my boots will be quick’

☞ Problem: if there is a single polysemous lexeme CIRAGE with multiple meanings, why is anaphora impossible in (3)?
Polysemous meaning can be represented as a disjunction of meanings

J’ai acheté du cirage noir ‘I bought black shoe polish’
[ ‘shoe polish’ ∨ ‘polishing’ ]

Le cirage de mes bottes prend 10 min. ‘polishing my boots takes 10 min’
[ ‘shoe polish’ ∨ ‘polishing’ ]

Anaphora does not behave as predicted by polysemous meaning

*Grâce à ce cirage noir, celui de mes bottes sera vite fait.
[ ‘shoe polish’ ∨ ‘polishing’ ]

Instead, specific meaning makes the right prediction

*Grâce à ce cirage noir, celui de mes bottes sera vite fait.
[ ‘shoe polish’ ]
Semantics of derived lexemes

Derived lexemes are not polysemous

- This is not an effect of lexicalization: the same observation holds in the case of nonce formations.
  - example with POMPONNAGE < POMPONNER ‘make up’

(4) Marie a acheté du pomponnage bleu.
Mary has bought blue makeup
‘Mary bought blue makeup’

(5) Le pomponnage de Marie prend 15 minutes tous les matins.
DET of Mary takes 15 minutes every DET mornings
‘It takes Mary 15 minutes to make herself up every mornings.’

(6) *Grâce à ce pomponnage bleu, celui de Marie est vite fait.
Thanks to this bleu that of Mary is quickly made
(int.) ‘Thanks to this blue makeup, Mary makes herself up quickly’
Semantics of derived lexemes

Conclusion

- Derived lexemes have specific meanings
- Thus LFRs need to output multiple specific lexemes, rather than one single lexeme with multiple meanings
- We could postulate multiple specific -age rules

\[
\text{age-evt-lfr : } \begin{bmatrix} \text{PHON } & \phi \\ \text{CAT } & \text{verb} \\ \text{SEM } & 'V' \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON } & \phi \oplus a_3 \\ \text{CAT } & \text{noun} \\ \text{SEM } & '\text{action of Ving}' \end{bmatrix}
\]

\[
\text{age-instr-lfr : } \begin{bmatrix} \text{PHON } & \phi \\ \text{CAT } & \text{verb} \\ \text{SEM } & 'V' \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON } & \phi \oplus a_3 \\ \text{CAT } & \text{noun} \\ \text{SEM } & '\text{object used to V}' \end{bmatrix}
\]

\[
\text{age-loc-lfr : } \begin{bmatrix} \text{PHON } & \phi \\ \text{CAT } & \text{verb} \\ \text{SEM } & 'V' \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON } & \phi \oplus a_3 \\ \text{CAT } & \text{noun} \\ \text{SEM } & '\text{loc where one Vs}' \end{bmatrix}
\]
Semantics of derived lexemes

Conclusion

• But then new problems arise:
  1. How do we account for the similarities among the multiple -age rules?
  2. How do we avoid redundancy between rules?
  3. How do we account for the shared properties of homophonous derived lexemes, like CIRAGE<sub>evt</sub> and CIRAGE<sub>instr</sub>?

☞ We use the HPSG framework and its multiple inheritance hierarchy to model LFRs and to capture both multiple specific rules and shared properties among these rules.
Lexeme formation and the multiple inheritance

• Since Flickinger (1987), there has been a tradition of using inheritance hierarchies to capture some aspects of the structure of the lexicon

• This idea has been extended to account for productive lexeme formation ((Riehemann, 1998),(Koenig, 1999)):
  ▶ Lexicalized derived lexemes are leaf nodes in the hierarchy
  ▶ LFRs are treated as schematic lexical entries for derived lexemes, where the base is not specified.

• Fruitfully applied to French LFRs ((Bonami and Boyé, 2006),(Desmets and Villoing, 2009),(Tribout, 2010))

• We use a variant of this setup
A polysemous LFR can be treated as a collection of specific semantic rules sharing (inheriting) a form schema.
Lexeme formation and the multiple inheritance

- Likewise, a semantic operation shared by different morphological processes can be abstracted away as a rule schema.

\[
\begin{align*}
\text{PHON} & \quad \phi \\
\text{CAT} & \quad \text{verb} \\
\text{SEM} & \quad 'V' \\
\Rightarrow & \\
\text{PHON} & \quad \psi \\
\text{CAT} & \quad \text{noun} \\
\text{SEM} & \quad 'object used to V'
\end{align*}
\]
Lexeme formation and the multiple inheritance

Multiple inheritance hierarchies

- Both types of rule schemas can be combined
Using multiple inheritance hierarchies allows us to account for the polysemy of LFRs.
  ✔ it accounts for similarities among the different cases of -age derivation
  ✔ it avoids redundancy among rules

However we also need to account for the shared properties of homophonous derived lexemes, like \( \text{CIRAGE}_{\text{event}} \) and \( \text{CIRAGE}_{\text{instr}} \)

\( \text{We use the paradigm identifiers worked out by Bonami (2011)} \)
Paradigm Identifiers

- According to Bonami (2011) each lexeme has a Paradigm IDentifier that is shared between multiple lexemes with the same paradigm.
  - $\text{CIRAGE}_{\text{evt}}$ and $\text{CIRAGE}_{\text{instr}}$ inflect in the same way
  - they have the same Paradigm IDentifier
    
    \[
    \begin{array}{|c|}
    \hline
    \text{PID} & \text{cirage} \\
    \text{PHON} & [\text{si}\text{ʁaʒ}] \\
    \text{CAT} & \text{noun} \\
    \text{SEM} & \sigma \\
    \hline
    \end{array}
    \]

- PIDs are complex data structures driving inflection.
- They capture (Fradin and Kerleroux, 2003)’s notion of a flexeme: a family of lexemes with the same inflectional paradigm.
- But they avoid postulating semantically underspecified superlexemes.
Paradigm Identifiers

- All LFRs modify the PID of their input.
- This is stipulated at the MORPHOPHON level

\[
\begin{bmatrix}
\text{age-lfr} \\
\text{PID} : p \\
\text{PHON} : \phi \\
\text{CAT} : \text{verb} \\
\text{SEM} : \sigma
\end{bmatrix} \Rightarrow
\begin{bmatrix}
\text{age-lfr} \\
\text{PID} : \text{age}(p) \\
\text{PHON} : \phi \oplus [a3] \\
\text{CAT} : \text{noun} \\
\text{SEM} : \tau
\end{bmatrix}
\]

- All sub-types of \textit{age-lfr} inherit the PID, and the PID is kept constant by operations that do not affect the inflectional paradigm (e.g. lexicalization, semantic shift)
The multiple inheritance hierarchy allows us to represent every semantic types of output given a morphological process.

However, all semantic types are not productive in the same way for a morphological process.

We also want to account for the productivity of every semantic types of output.
Productivity

- LFRs can be thought of as abstractions of what exists in the lexicon.

- We can have an idea of the productivity of a semantic type by looking at how many lexemes it describes.

  - Koehl (2012) has shown with Fr. -*itude* suffixation that a not productive process can always be used and become productive again.

  - Unlike Baayen (2001), we do not take productivity as the capacity to form hapaxes.

  - Rather, we only look at the frequency of lexemes described by a semantic type in a corpus.

  - We take token frequency, rather than type frequency, into account.

  - We provide an illustration with deverbal -eur suffixation in French.
Productivity
Example: deverbal -eur suffixation in French

- There has been numerous studies on deverbal -eur suffixation in French (among which (Benveniste, 1975), (Fradin and Kerleroux, 2003), (Rosenberg, 2008), (Roy and Soare, 2012))

- -eur suffixation forms agent and instrument nouns, but there also are few result nouns

- We used the French corpus Lexique 3 (http://www.lexique.org/)
  - it contains 55 000 lexemes and 135 000 inflected forms
  - for each inflected form it provides category, phonology, etc. and its frequency in two corpora (a sub-part of Frantext and movie subtitles).
  - 1 591 deverbal nouns extracted from the lexicon manually annotated as agent, instrument or result
Productivity

Example: deverbal -eur suffixation in French

\[
\begin{align*}
\text{PID} & \quad p & \Rightarrow & \quad \text{PID} & \quad \text{eur}(p) \\
\text{PHON} & \quad \phi & \Rightarrow & \quad \text{PHON} & \quad \phi\oplus[\omega\epsilon]\text{K} \\
\text{CAT} & \quad \text{verb} & \Rightarrow & \quad \text{CAT} & \quad \text{noun} \\
\text{SEM} & \quad \sigma & \Rightarrow & \quad \text{SEM} & \quad \tau
\end{align*}
\]

\[
\begin{align*}
\text{eur-agt-lfr} & \quad \left[\text{‘V’}\Rightarrow \text{‘the one who Vs’}\right] \\
\text{eur-instr-lfr} & \quad \left[\text{‘V’}\Rightarrow \text{‘obj. used to V’}\right] \\
\text{eur-res-lfr} & \quad \left[\text{‘V’}\Rightarrow \text{‘result of Ving’}\right]
\end{align*}
\]

4 423.73
DANSEUR ‘dancer’
CHANTEUR ‘singer’

297.9
MINUTEUR ‘timer’
EFFACEUR ‘eraseur’

192.78
SUEUR ‘sweat’
SENTEUR ‘scent’
• We have argued that LFRs have monosemic outputs

• We use multiple inheritance hierarchy in order to model LFRs
  ▶ it allows us to capture multiple specific rules, and
  ▶ to avoid redundancy among them

• We use Paradigm IDentifiers in order to account for properties shared by homophonous derived lexemes

• We include token frequency into the description in order to account for productivity of each semantic type of output within the same morphological process
Thank you for your attention
References


