The polysemy of lexeme formation rules

Delphine Tribout & Olivier Bonami

Workshop Semantics of derivational morphology:
Empirical evidence and theoretical modeling

July 01, 2014

Introduction

- 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

type	example	gloss
event N	LAVAGE ‘cleaning’	act of cleaning
instrument N	AIGUILLAGE ‘railroad switch’	what one uses to switch railroad
location N	GARAGE ‘parking lot’	place where one parks

- Output lexemes are often ambiguous between two types

types	examples	gloss
event N	CIRAGE ‘polishing’	act of polishing
instrument N	CIRAGE ‘shoe polish’	what one uses to polish
event N	PASSAGE ‘passing’	act of going through
location N	PASSAGE ‘path’	location through which one goes

Polysemous Lexeme Formation Rules

- This situation can lead to posit polysemous LFRs: a single rule with multiple semantics.

$$\textit{age-lfr}: \begin{bmatrix} \text{PHON} & \phi \\ \text{CAT} & \text{verb} \\ \text{SEM} & \text{`V'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \phi \oplus \text{aʒ} \\ \text{CAT} & \text{noun} \\ \text{SEM} & \left\{ \begin{array}{l} \text{`action of Ving'}, \\ \text{`place where one Vs'}, \\ \text{`object used to V'} \end{array} \right\} \end{bmatrix}$$

- A polysemous rule gives rise to a single polysemous lexeme

$$\begin{bmatrix} \text{PHON} & \text{siʁ(e)} \\ \text{CAT} & \text{verb} \\ \text{SEM} & \text{`polish'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \text{siʁaʒ} \\ \text{CAT} & \text{noun} \\ \text{SEM} & \left\{ \text{`polishing'}, \text{ `shoe polish'} \right\} \end{bmatrix}$$

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)?

Semantics of derived lexemes

Derived lexemes are not polysemous

- Polysemous meaning can be represented as a disjunction of meanings

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

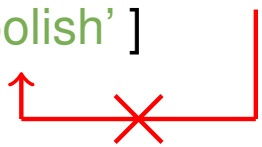
  Le cirage de mes bottes prend 10 min. 'polishing my boots takes 10 min'
  ['shoe polish' ∨ 'polishing']

- 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 PART bleu
‘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

$$\textit{age-evt-lfr}: \begin{bmatrix} \text{PHON} & \phi \\ \text{CAT} & \text{verb} \\ \text{SEM} & \text{`V'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \phi \oplus \text{az} \\ \text{CAT} & \text{noun} \\ \text{SEM} & \text{`action of Ving'} \end{bmatrix}$$

$$\textit{age-instr-lfr}: \begin{bmatrix} \text{PHON} & \phi \\ \text{CAT} & \text{verb} \\ \text{SEM} & \text{`V'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \phi \oplus \text{az} \\ \text{CAT} & \text{noun} \\ \text{SEM} & \text{`object used to V'} \end{bmatrix}$$

$$\textit{age-loc-lfr}: \begin{bmatrix} \text{PHON} & \phi \\ \text{CAT} & \text{verb} \\ \text{SEM} & \text{`V'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{PHON} & \phi \oplus \text{az} \\ \text{CAT} & \text{noun} \\ \text{SEM} & \text{`loc where one Vs'} \end{bmatrix}$$

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$_{evt}$ and CIRAGE$_{instr}$?

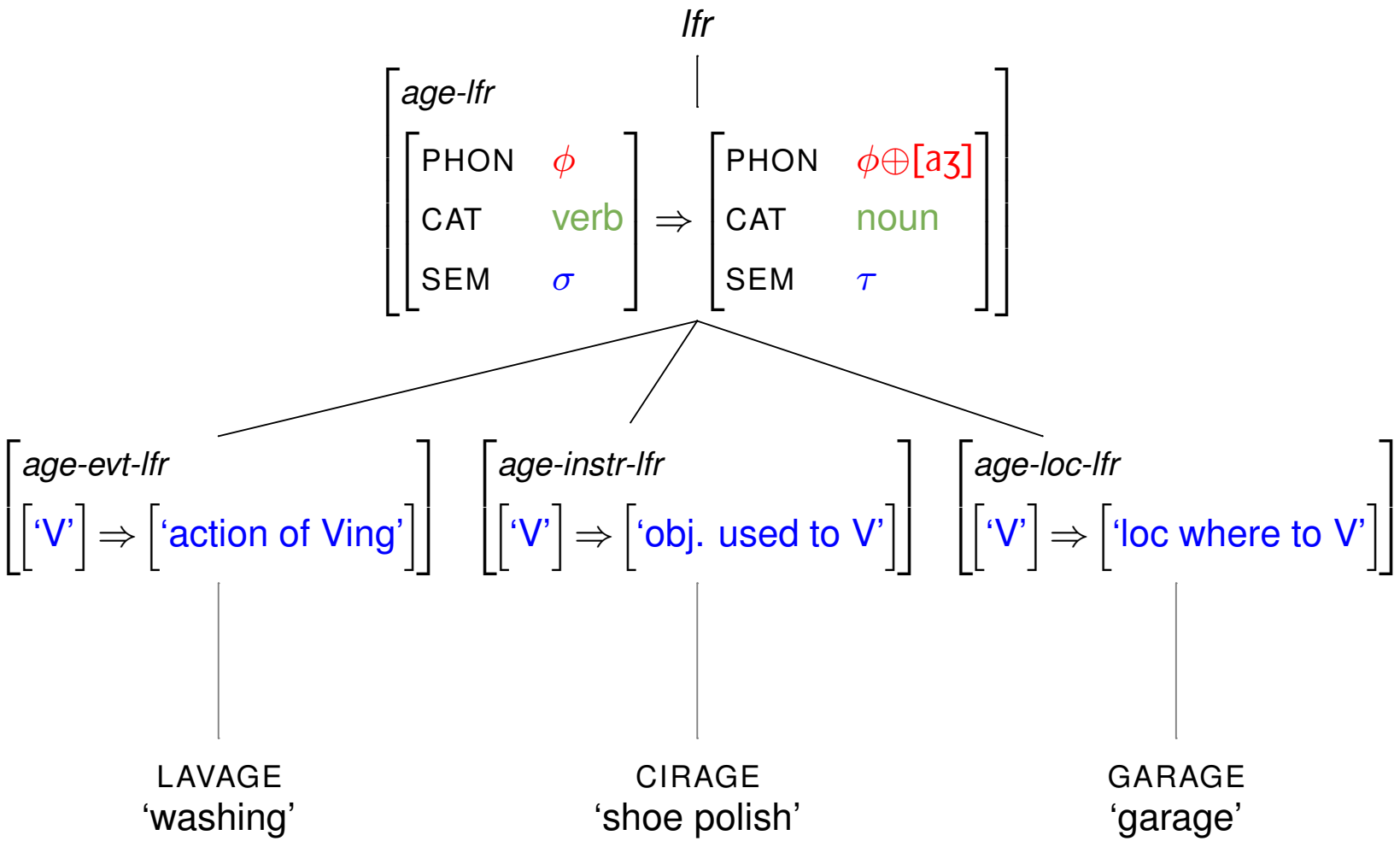
☞ We use the HPSG framework and its multiple inheritance hierarchy to model LFRs and to capture both multiple specific rules and shared properties among these rules

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

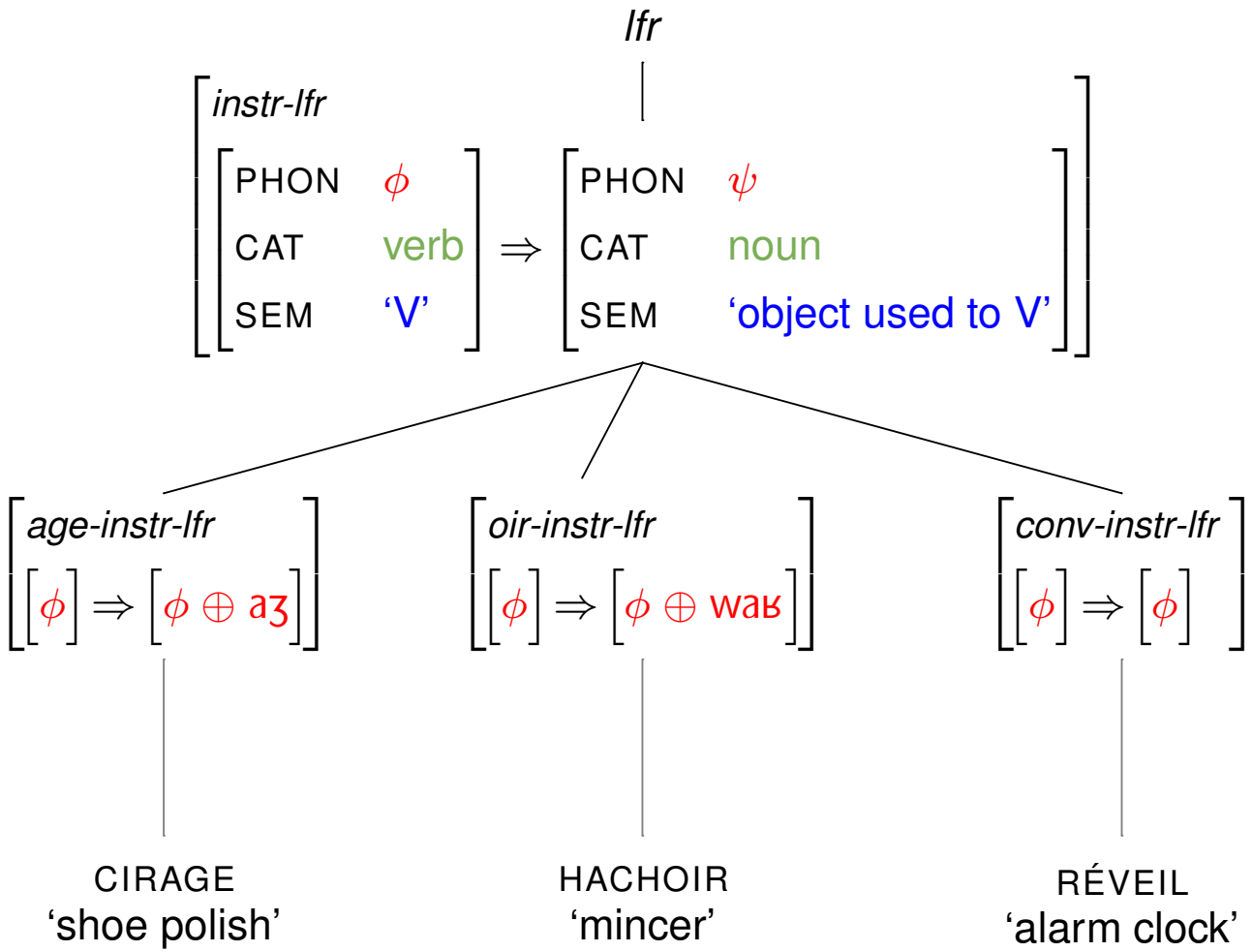
Lexeme formation and the multiple inheritance

- A polysemous LFR can be treated as a collection of specific semantic rules sharing (inheriting) a form schema



Lexeme formation and the multiple inheritance

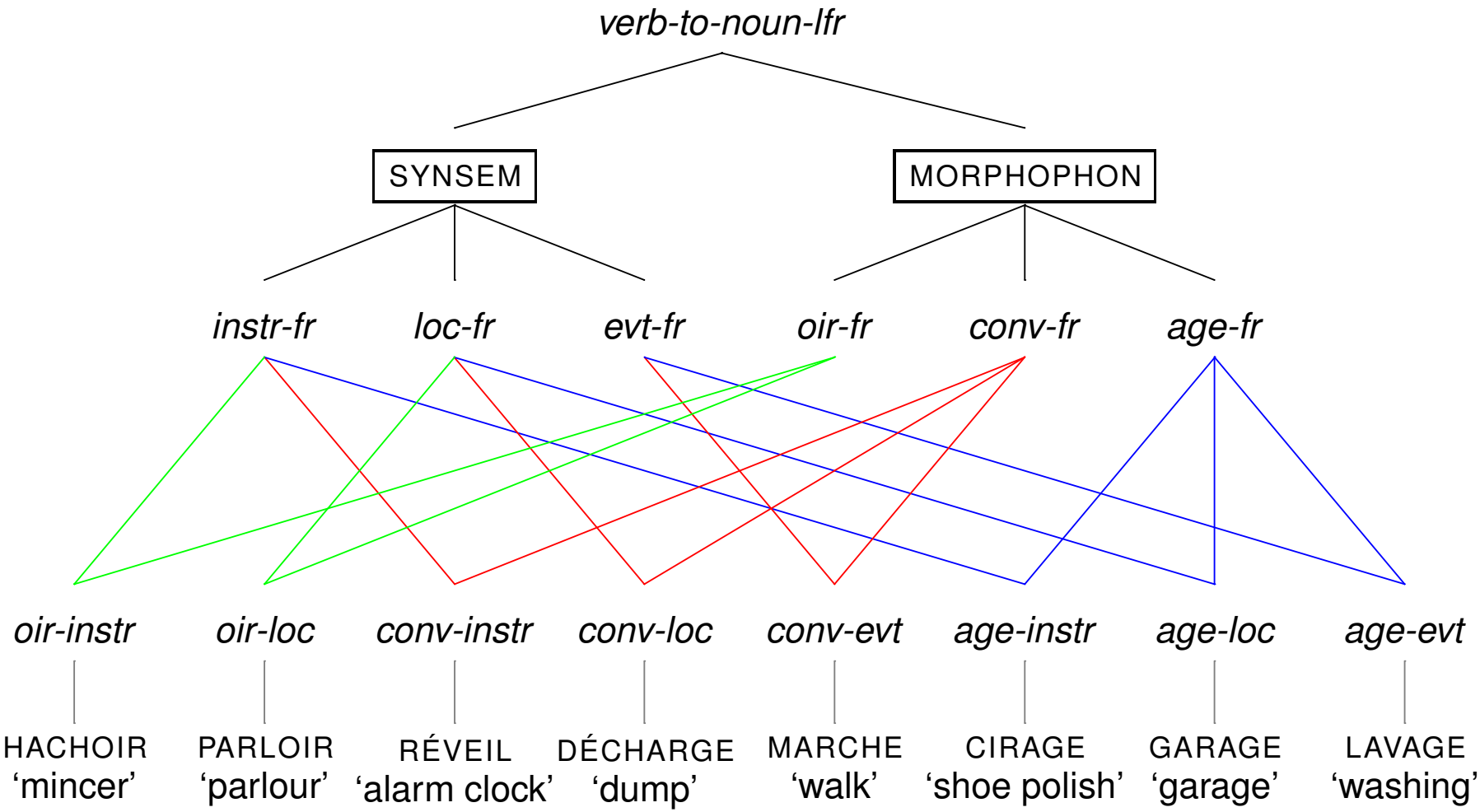
- Likewise, a semantic operation shared by different morphological processes can be abstracted away as a rule schema



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 CIRAGE$_{event}$ and CIRAGE$_{instr}$

☞ We use the paradigm identifiers worked out by Bonami (2011)

Paradigm Identifiers

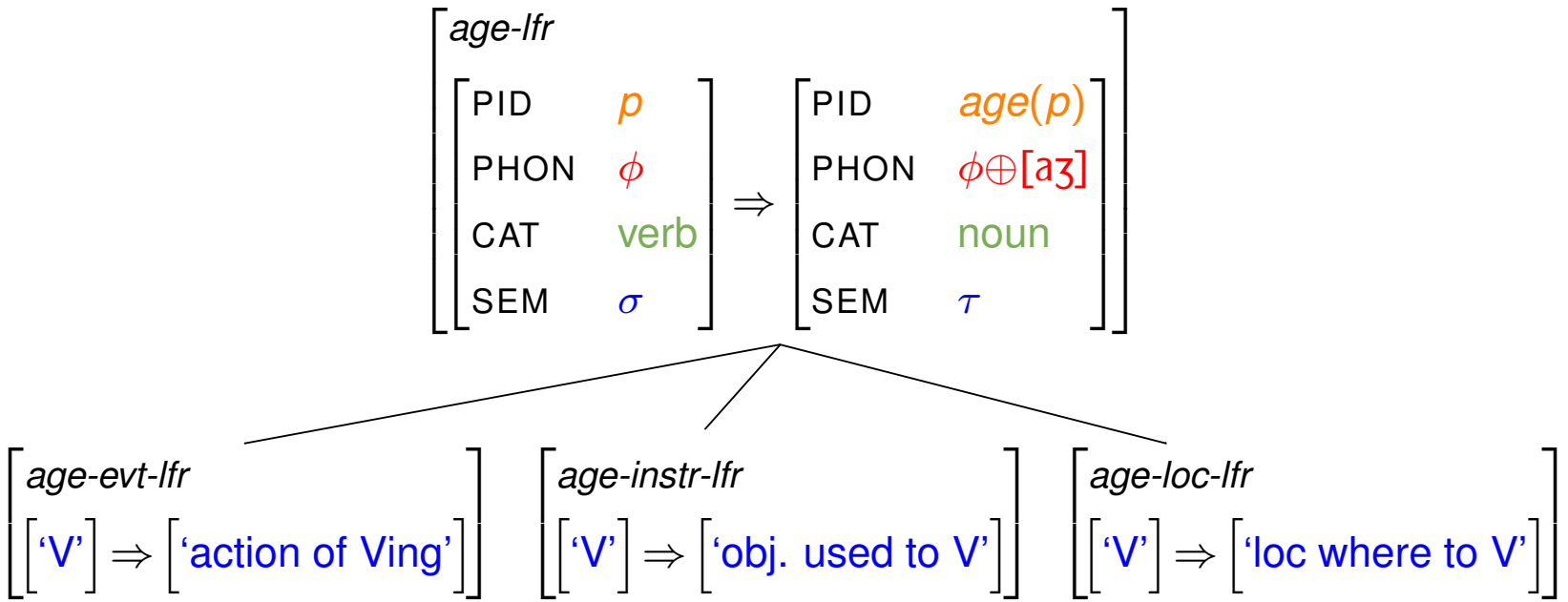
- According to Bonami (2011) each lexeme has a Paradigm IDentifier that is shared between multiple lexemes with the same paradigm
  - CIRAGE$_{evt}$ and CIRAGE$_{instr}$ inflect in the same way
  - ☞ they have the same Paradigm IDentifier

$$\begin{bmatrix} \text{PID} & \textit{cirage} \\ \text{PHON} & [\text{siʁaʒ}] \\ \text{CAT} & \text{noun} \\ \text{SEM} & \sigma \end{bmatrix}$$

- 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



- All sub-types of *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)

Productivity

- 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 tought 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

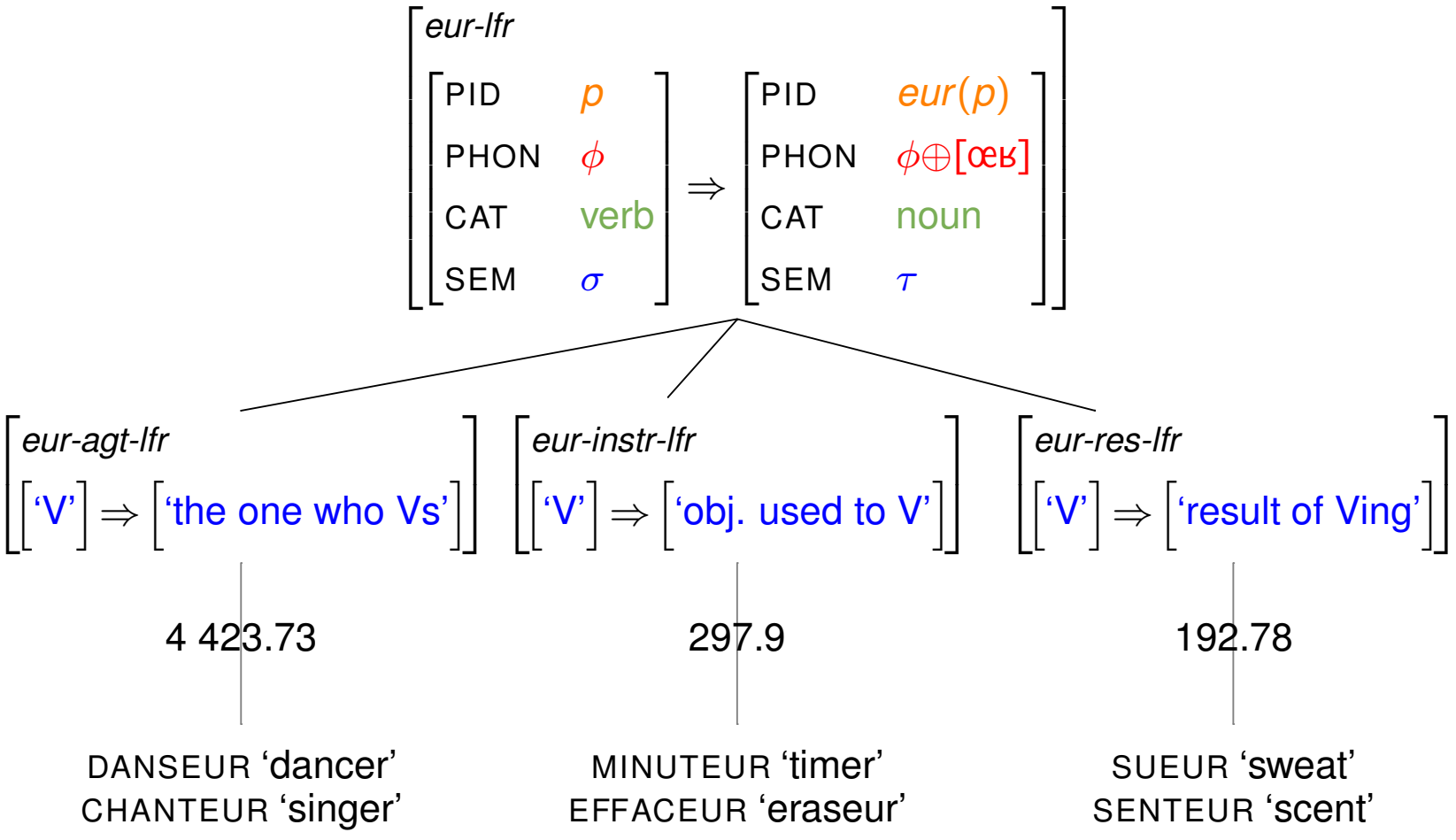
- There has been numerous studies on deverbal -*eur* suffixation in French (among which (Benveniste, 1975), (Fradin and Kerleroux, 2003), (Rosenberg, 2008), (Roy and Soare, 2012))

- -*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



Conclusion

- 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

Conclusion

Thank you for your attention

References

Aronoff, M. (1994). *Morphology by Itself*. The MIT Press, Cambridge.

Baayen, H. (2001). *Word frequency distribution*, volume 18 of *Text, Speech and Language Technology*. Kluwer Academic Publishers, Dordrecht.

Benveniste, E. (1975). *Noms d'agent et noms d'action en indo-européen*. Maisonneuve, Paris.

Bonami, O. (2011). Reconstructing HPSG morphology. In *18th International Conference on HPSG*, Seattle.

Bonami, O. and Boyé, G. (2006). Deriving inflectional irregularity. In Müller, S., editor, *Proceedings of the HPSG'06 Conference*, pages 361–380, Stanford. CSLI publications.

Desmets, M. and Villoing, F. (2009). French VN lexemes: morphological compounding in HPSG. In *Proceedings of the HPSG'09 Conference*, pages 89–109.

Flickinger, D. (1987). *Lexical Rules in the Hierarchical Lexicon*. PhD thesis, Stanford University.

Fradin, B. (2003). *Nouvelles approches en morphologie*. Puf, Paris.

Fradin, B. and Kerleroux, F. (2003). Troubles with lexemes. In Booij, G., de Cesaris, J., Scalise, S., and Ralli, A., editors, *Topics in Morphology. Selected papers from the Third Mediterranean Morphology Meeting*, pages 177–196, Barcelona. ULA-Universitat Pompeu Fabra.

Koehl, A. (2012). *La construction morphologique des noms désadjectivaux suffixés en français*. PhD thesis, Nancy 2.

Koenig, J.-P. (1999). *Lexical Relations*. CSLI Publications, Stanford.

Riehemann, S. (1998). Type-based derivational morphology. *Journal of Comparative Germanic Linguistics*, 2:49–77.

Rosenberg, M. (2008). *La formation agentive en français. Les composés [VN/A/Adv/P]$_{N\,/\,A}$ et les dérivés V-ant, V-eur et V-oir(e).* PhD thesis, Stockholm University.

Roy, I. and Soare, E. (2012). L'enquêteur, le surveillant et le détenu: les noms déverbaux de participants aux événements, lecture événementielle et structure argumentale. *Lexique*, 20:207–231.

Tribout, D. (2010). How many conversions from verb to noun are there in French? In Müller, S., editor, *Proceedings of the 17th International Conference on HPSG*, pages 341–357, Stanford. CSLI Publications.