

# Underspecification and the semantics of lexeme formation

Olivier Bonami & Delphine Tribout

Paris-Sorbonne & IUF, Paris 8  
Laboratoire de Linguistique Formelle

IMM15 — Vienna, Feb. 12, 2012

- The issue we address: the semantic indeterminacy of lexeme formation processes in French
- This domain has been studied extensively in the last decade for French
- Here we focus on the Grammar-LFRs interface and the consequences for LFRs' modeling
- We sketch an explicit model within the general framework of a sign-based approach to lexeme formation
  - ✎ written in HPSG (Pollard and Sag, 1994), broadly compatible with Network Morphology (Corbett and Fraser, 1993; Brown and Hippisley, 2012), and Construction Grammar/Morphology (e.g. Koenig, 1994; Orgun, 1996; Booij, 2010).

The puzzle: general rules, specific outputs

# The semantic indeterminacy of LFRs

- LFRs typically give rise to semantically diverse results

☞ example: nouns in *-age* in French

type	example	gloss
eventuality noun	<i>guidage</i> 'guidance'	act of guiding
instrument noun	<i>maquillage</i> 'make up'	substance used to make sb up
locative noun	<i>garage</i> 'parking lot'	place where one parks

# The semantic indeterminacy of LFRs

- LFRs typically give rise to semantically diverse results

☞ example: nouns in *-age* in French

type	example	gloss
eventuality noun	<i>guidage</i> 'guidance'	act of guiding
instrument noun	<i>maquillage</i> 'make up'	substance used to make sb up
locative noun	<i>garage</i> 'parking lot'	place where one parks

- Output lexemes are often ambiguous between two types

types	examples	gloss of the semantic operations
eventuality noun	<i>cirage</i> 'polishing'	act of polishing
instrument noun	<i>cirage</i> 'shoe polish'	substance used to polish
eventuality noun	<i>passage</i> 'passing'	act of going through
location noun	<i>passage</i> 'path'	location through which one goes

# From indeterminacy to underspecification

- Morphologists are used to assigning to LFRs with apparent polysemous output an abstract and/or underspecified meaning
  - ▶ Aronoff (1980); Plag (1999); Lieber (2004) are notorious examples of this strategy

# From indeterminacy to underspecification

- Morphologists are used to assigning to LFRs with apparent polysemous output an abstract and/or underspecified meaning
  - ▶ Aronoff (1980); Plag (1999); Lieber (2004) are notorious examples of this strategy
- This seems natural from the point of view of morphology and the lexicon

# From indeterminacy to underspecification

- Morphologists are used to assigning to LFRs with apparent polysemous output an abstract and/or underspecified meaning
  - ▶ Aronoff (1980); Plag (1999); Lieber (2004) are notorious examples of this strategy
- This seems natural from the point of view of morphology and the lexicon
- However it might not be optimal for the purposes of syntax, compositional semantics or discourse interpretation.



# From indeterminacy to underspecification

- Morphologists are used to assigning to LFRs with apparent polysemous output an abstract and/or underspecified meaning
  - ▶ Aronoff (1980); Plag (1999); Lieber (2004) are notorious examples of this strategy
- This seems natural from the point of view of morphology and the lexicon
- However it might not be optimal for the purposes of syntax, compositional semantics or discourse interpretation.
- Here we show on the basis of anaphora that we need collections of specific lexemes rather than abstract lexemes.

# Uses of derived lexemes are not indeterminate

- (1) J' avais du cirage noir, mais il m' en fallait du marron.  
I had PART polish black but EXPL me of-it needed PART brown  
'I had black shoe polish, but I needed brown.'
- (2) Le cirage de mes bottes m' a pris trois heures. Celui de mes chaussures  
The polishing of my boots me has taken three hours that of my shoes  
a été plus rapide.  
has been more quick  
'It took me three hours to shine my boots. Shining my shoes was quicker.'

# Uses of derived lexemes are not indeterminate

- (1) J' avais du cirage noir, mais il m' en fallait du marron.  
I had PART polish black but EXPL me of-it needed PART brown  
'I had black shoe polish, but I needed brown.'
- (2) Le cirage de mes bottes m' a pris trois heures. Celui de mes chaussures  
The polishing of my boots me has taken three hours that of my shoes  
a été plus rapide.  
has been more quick  
'It took me three hours to shine my boots. Shining my shoes was quicker.'
- (3) a. \* Grâce à ce cirage noir, celui de mes bottes sera vite fait.  
Thanks to this polish black that of my boots will\_be quickly made  
(int.) 'Thanks to this black polish, polishing my boots will be quick.'
- b. \* Le cirage de mes chaussures m' a pris trois heures. Heureusement que  
The of my shoes me has taken three hours luckily that  
j' en avais du noir.  
I of-it had PART black  
(int.) 'It took me three hours to polish my shoes. Luckily I had black polish.'

# Uses of derived lexemes are not indeterminate

- (1) J' avais du **cirage** noir, mais il m' **en** fallait du marron.  
I had PART polish black but EXPL me of-it needed PART brown  
'I had black shoe polish, but I needed brown.'
- (2) Le **cirage** de mes bottes m' a pris trois heures. **Celui** de mes chaussures  
The polishing of my boots me has taken three hours that of my shoes  
a été plus rapide.  
has been more quick  
'It took me three hours to shine my boots. Shining my shoes was quicker.'
- (3) a. \* Grâce à ce **cirage** noir, **celui** de mes bottes sera vite fait.  
Thanks to this polish black that of my boots will\_be quickly made  
(int.) 'Thanks to this black polish, polishing my boots will be quick.'
- b. \* Le **cirage** de mes chaussures m' a pris trois heures. Heureusement que  
The of my shoes me has taken three hours luckily that  
j' **en** avais du noir.  
I of-it had PART black  
(int.) 'It took me three hours to polish my shoes. Luckily I had black polish.'

☞ If there is a single lexeme **CIRAGE** with an underspecified meaning, why is anaphora impossible in (3)?



# Analysis

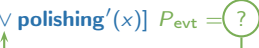
- Suppose we use the crudest possible abstract meaning: a disjunction of the two specific meanings

specific      Grâce à ce cirage noir, le cirage de mes bottes sera vite fait.  
semantics                      polish'                      polishing'

abstract      Grâce à ce cirage noir,      le cirage de mes bottes sera vite fait.  
semantics       $\lambda x. [\text{polish}'(x) \vee \text{polishing}'(x)]$        $\lambda x. [\text{polish}'(x) \vee \text{polishing}'(x)]$

- Property anaphora does not behave in the predicted way:

abstract      Grâce à ce cirage noir,      celui de mes bottes sera vite fait.  
semantics       $\lambda x. [\text{polish}'(x) \vee \text{polishing}'(x)]$        $P_{\text{evt}} = \textcircled{?}$



# Analysis

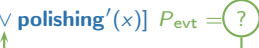
- Suppose we use the crudest possible abstract meaning: a disjunction of the two specific meanings

specific      Grâce à ce **cirage** noir, le **cirage** de mes bottes sera vite fait.  
semantics                      **polish'**                      **polishing'**

abstract      Grâce à ce **cirage** noir,      le **cirage** de mes bottes sera vite fait.  
semantics       $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$        $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$

- Property anaphora does not behave in the predicted way:

abstract      Grâce à ce **cirage** noir,      **celui** de mes bottes sera vite fait.  
semantics       $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$        $P_{\text{evt}} = \textcircled{?}$



- Any other abstract meaning will have the same problem:  
$$\forall x [[\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)] \rightarrow \mathbf{cirage}'(x)]$$

# Analysis

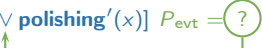
- Suppose we use the crudest possible abstract meaning: a disjunction of the two specific meanings

specific      Grâce à ce **cirage** noir, le **cirage** de mes bottes sera vite fait.  
semantics                      **polish'**                      **polishing'**

abstract      Grâce à ce **cirage** noir,      le **cirage** de mes bottes sera vite fait.  
semantics       $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$        $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$

- Property anaphora does not behave in the predicted way:

abstract      Grâce à ce **cirage** noir,      **celui** de mes bottes sera vite fait.  
semantics       $\lambda x. [\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)]$   $P_{\text{evt}} = \textcircled{?}$




- Any other abstract meaning will have the same problem:

$$\forall x [[\mathbf{polish}'(x) \vee \mathbf{polishing}'(x)] \rightarrow \mathbf{cirage}'(x)]$$

- A specific meaning makes the right prediction:

specific      \*Grâce à ce **cirage** noir, **celui** de mes bottes sera vite fait.  
semantics                      **polish'**                       $P_{\text{evt}} = \textcircled{?}$





## Uses of nonce derived lexemes are not indeterminate

- The data in (3) could be taken to be an effect of lexicalization.
- However the same observation holds in the case of nonce formations.

(4) J' ai acheté du pomponnage bleu.  
I have bought PART bleu  
'I bought blue makeup'

(5) Mon pomponnage m' a pris 15 minutes ce matin.  
My me has taken 15 minutes this morning  
'It took me 15 minutes to make me up this morning.'

(6) \* Grâce à ce nouveau pomponnage, celui de Marie ne prend plus  
Thanks to this new makeup that of Marie NE take only  
que 15 minutes.  
QUE 15 minutes  
(int.) 'Thanks to this new makeup, it takes Marie only 15 minutes to get ready.'

# The problem is general

- Denominal verbs suffixed with *-iser*

(7) Le ministre a **radarisé** la région.

The secretary has the area

'The secretary has installed radars in the area'

(8) Depuis janvier j' ai été **radarisé** deux fois.

Since January I have been two times

'Since January I was caught by a radar twice'

- But:

(9) \* Depuis que le ministre a **radarisé** la région, je l' ai été

Since that the secretary has the area, I it have been

deux fois.

two times

(int.) 'Since radars have been installed in the area, I was caught by a radar twice.'

# The problem is general

- Deverbal adjectives suffixed with *-able* (Hathout et al., 2003)

(10) La truite est **pêchable** dans les rivières de montagne  
The trout is                    in the river of mountain  
'One may fish trouts in mountain rivers'

(11) Cette rivière est **pêchable** de juin à septembre  
This river is                    from June to September  
'One may fish in this river from June to September'

- But:

(12) \* L' été, la truite est **pêchable** dans les rivières qui **le** sont.  
The summer, the trout is                    in the rivers which it are  
(int.) 'During the summer, one may fish trouts in rivers where one may fish.'

# The problem is general

- Deverbal nouns suffixed with *-ette* (Fradin, 2003; Plénat, 2005)

(13) As-tu pris tes *glisettes* pour aller patiner au Rathaus?  
Have-you brought your to go ice-skating at Rathaus?  
'Have you brought your ice skate, so that we can do ice-skating at the Rathaus?'

(14) J' ai fait une *glissette* à la patinoire  
I have done one at the ice rink  
'I did a slithering at the ice rink'

- But:

(15) \* As-tu pris tes *glisettes* pour en faire *une* au Rathaus?  
Have-you brought your to of-it do one at Rathaus?  
(int) 'Have you brought your ice skate, so that we can do ice-skating at the Rathaus?'

# The problem is general

- Denominal adjectives (Fradin and Kerleroux, 2003)

(16) Ce ministre est très populaire au sein du parti  
This secretary is very popular inside of the party  
'This secretary is very popular in the party'

(17) Le mécontentement populaire a conduit à la démission  
The dissatisfaction of people has led to the resignation  
du ministre  
of the secretary  
'The dissatisfaction of people led to the resignation of the secretary'

- But

(18) \*Le mécontentement populaire est tel que le candidat ne l'est  
The dissatisfaction of people is so that the candidate not is  
plus au sein du parti  
anymore inside of the party  
'(int) The dissatisfaction of people is so important that the candidate  
is not popular any more in the party'

# Conclusion on the semantics of derived lexemes

- Derived lexemes have specific meanings

## Conclusion on the semantics of derived lexemes

- Derived lexemes have specific meanings
- Thus LFRs need to output multiple specific lexemes, rather than one single lexeme with abstract or underspecified meaning

## Conclusion on the semantics of derived lexemes

- Derived lexemes have specific meanings
- Thus LFRs need to output multiple specific lexemes, rather than one single lexeme with abstract or underspecified meaning
- We can postulate multiple *-age*, *-iser*, *-able*... rules



# Conclusion on the semantics of derived lexemes

- Derived lexemes have specific meanings
- Thus LFRs need to output multiple specific lexemes, rather than one single lexeme with abstract or underspecified meaning
- We can postulate multiple *-age*, *-iser*, *-able*... rules
- But then new problems arise:
  - ▶ How do we account for the fact that rules come in families?
  - ▶ How do we account for the fact that specific lexemes come in families too?
  - ▶ How do we avoid redundancy between rules?

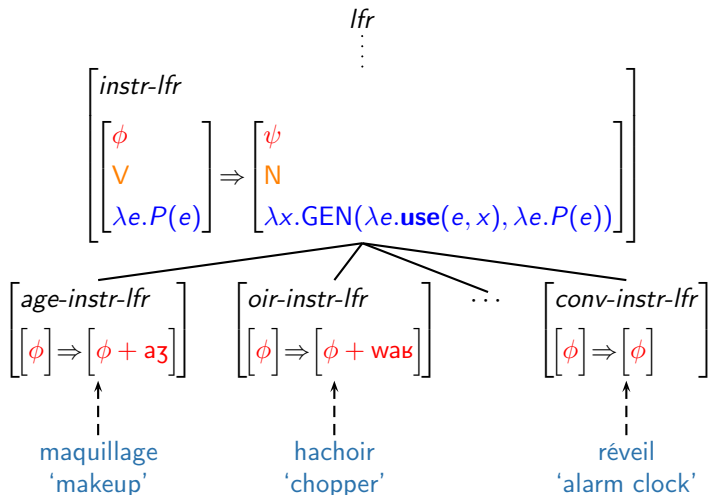
A formal analysis

# Lexeme formation and the multiple inheritance

- Since Flickinger (1987), established 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; Hippisley, 1997):
  - ▶ Institutionalized 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 where:
  - ▶ LFRs form a multiple inheritance hierarchy separate from the stable lexicon
  - ▶ Nonce lexemes are licensed as the output of a LFR
  - ▶ Stable lexemes derive from nonce lexemes through an explicit process of **institutionalization** (Bauer, 1983; Hohenhaus, 2005)

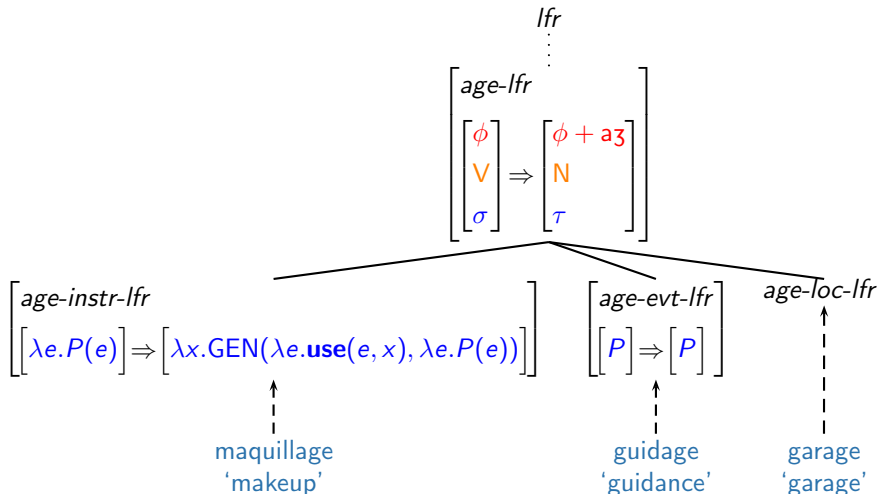
# Abstracting semantic operations from LFRs

- An operation common to a family of morphological processes can be abstracted away as a rule schema
- This reduces the amount of stipulation



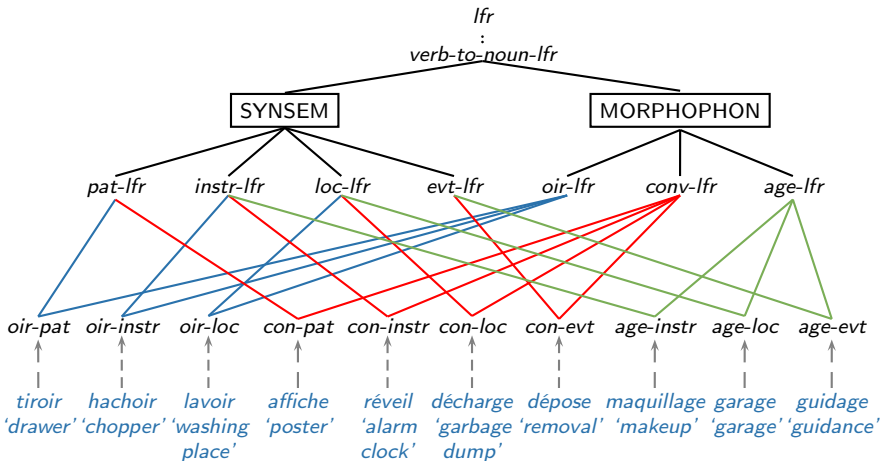
# Distributing information in the hierarchy LFRs

- Likewise, a polysemous LFR can be treated as a collection of fully specified LFRs sharing a form schema



# Multiple inheritance hierarchies

- Both types of rule schemas can be combined



- ☞ Each process/semantics coupling is listed as a separate LFR, without any ensuing loss of generality.

# Paradigm Identifiers

- We assume that each lexeme comes equipped with a **Paradigm Identifier** that is shared between multiple lexemes with the same paradigm.
- PIDs can be seen as arbitrary indices (cf. Spencer, 2005, 's notion of **lexeme identifier**) or as complex data structures driving inflection (Bonami, 2011).

[PID:laver  
laver  
V  
wash']

---

Same PID	cirage 'polish' & cirage 'polishing' devoir 'must' & devoir 'owe' glissette 'ice-skating' & glissette 'ice skate'
Different PIDs	empilage 'piling' & empilement 'piling' cri 'scream' & crier 'a scream' trier 'sort' & triller 'trill'

---

- Captures Fradin and Kerleroux (2003)'s notion of a **flexeme**: a family of lexemes with the same inflectional paradigm.

# LFRs and PID alteration

- All LFRs (including conversion LFRs) alter the PID of their input.
- This is stipulated at the level of MORPHOPHON rule schemas.

$$\left[ \begin{array}{c} \text{oir-lfr} \\ p \\ \phi \\ V \end{array} \right] \Rightarrow \left[ \begin{array}{c} \text{oir}(p) \\ \phi + \text{was} \\ N \end{array} \right]$$

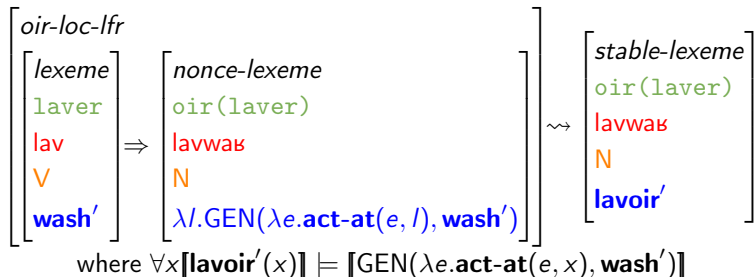
⋮ inheritance  
↓

$$\left[ \begin{array}{c} \text{oir-loc-lfr} \\ p \\ \phi \\ V \\ P \end{array} \right] \Rightarrow \left[ \begin{array}{c} \text{oir}(p) \\ \phi + \text{was} \\ N \\ \lambda l. \text{GEN}(\lambda e. \text{act-at}(e, l), P) \end{array} \right] \xrightarrow{\text{instanciation}} \left[ \begin{array}{c} \text{oir-loc-lfr} \\ \text{laver} \\ \text{lav} \\ V \\ \text{wash}' \end{array} \right] \Rightarrow \left[ \begin{array}{c} \text{oir}(\text{laver}) \\ \text{lavwas} \\ N \\ \lambda l. \text{GEN}(\lambda e. \text{act-at}(e, l), \text{wash}') \end{array} \right]$$



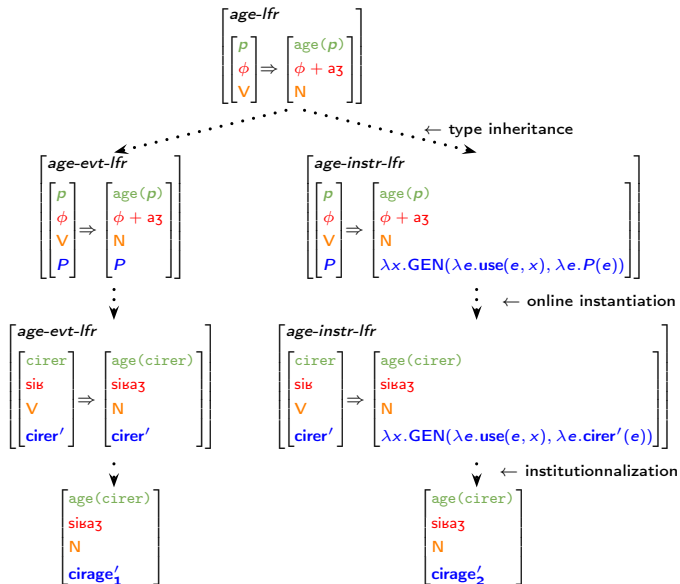
# PIDs and the lexicon

- Nonce lexemes added to the lexicon through an explicit process of **institutionalization**
- This process alters the semantics of the input but not its PID

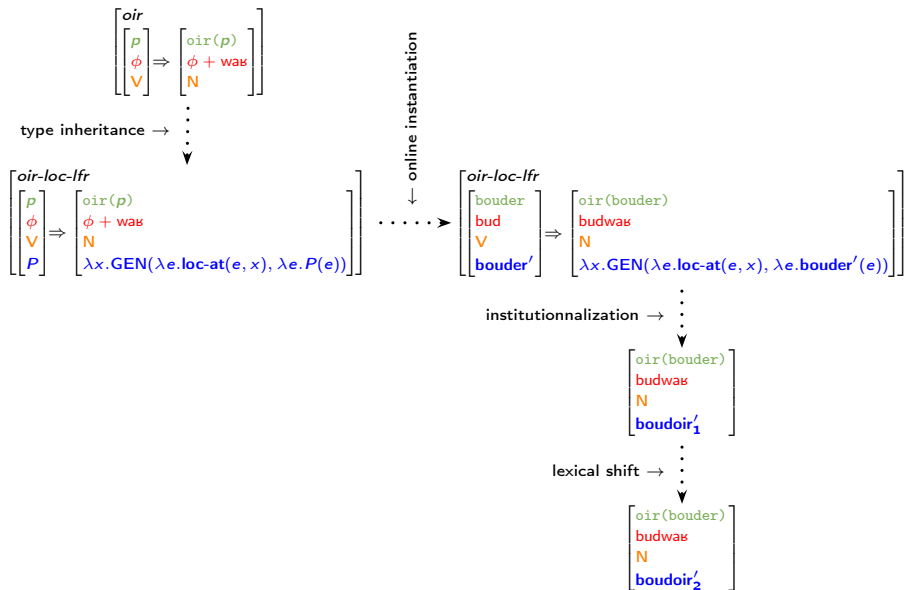


- Lexical meaning shifts also normally leave the PID unchanged

# Spreading PIDs: parallel LFRs



# Spreading PIDs: lexical shifts



## Conclusions

# Conclusions

- Empirical claims:
  - ▶ Individual lexeme formation processes typically have semantically indeterminate outputs.
  - ▶ Yet the interface with syntax, semantics and discourse structure call for semantically determinate derived lexemes.
  - ▶ Not an effect of lexicalization: holds for nonce formations.
- Theoretical proposal:
  - ▶ Model LFRs using a multiple inheritance hierarchy
  - ▶ Distribute lexemic information in separate entries
  - ▶ Relate these entries by a network of Paradigm Identifiers
- Features of this architecture:
  - ▶ Clean distinction between morphology and lexical dynamics
  - ▶ Implements a sign-based version of the **separationist hypothesis** (Beard, 1995)
  - ▶ Interfaces readily with an approach to inflection in the spirit of PFM (Bonami, 2011)

# References I

- Aronoff, M. (1980). Contextuals. In Hoekstra, T., Van der Hulst, H., and Moortgat, M., editors, *Lexical Grammar*, pages 263–285. Foris Publications, Dordrecht.
- Bauer, L. (1983). *English word formation*. Cambridge University Press, Cambridge.
- Beard, R. (1995). *Lexeme-Morpheme Base Morphology*. State University of New York Press, Albany.
- Bonami, O. (2011). Reconstructing HPSG morphology. In *18th International Conference on HPSG*, Seattle.
- Bonami, O. and Boyé, G. (2006). Deriving inflectional irregularity. In *Proceedings of the 13th International Conference on HPSG*, pages 39–59, Stanford. CSLI Publications.
- Booij, G. (2010). *Construction Morphology*. Oxford University Press.
- Brown, D. and Hippisley, A. (2012). *Network Morphology: a defaults-based theory of word structure*. Cambridge University Press, Cambridge.
- Corbett, G. G. and Fraser, N. M. (1993). Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics*, 29:113–142.
- Desmets, M. and Villoing, F. (2009). French VN lexemes: morphological compounding in HPSG. In *Proceedings of the HPSG 2009 Conference*, pages 89–109, Stanford. CSLI Publications.
- Flickinger, D. (1987). *Lexical Rules in the Hierarchical Lexicon*. PhD thesis, Stanford University.
- Fradin, B. (2003). Le traitement de la suffixation en -ET. *Langages*, 152:51–77.
- 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.
- Hathout, N., Plénat, M., and Tanguy, L. (2003). Enquête sur les dérivés en -able. *Cahiers de Grammaire*, 28:49–90.
- Hippisley, A. (1997). *Declarative Derivation: a Network Morphology account of Russian word formation with reference to nouns denoting 'person'*. PhD thesis, University of Surrey.

# References II

- Hohenhaus, P. (2005). Lexicalization and institutionalization. In Štekauer, P. and Lieber, R., editors, *Handbook of Word-Formation*, pages 353–373. Springer.
- Koenig, J.-P. (1994). *Lexical underspecification and the syntax-semantics interface*. PhD thesis, University of California, Berkeley.
- Koenig, J.-P. (1999). *Lexical relations*. CSLI Publications, Stanford.
- Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge University Press, Cambridge.
- Orgun, C. O. (1996). *Sign-Based Morphology and Phonology*. PhD thesis, University of California, Berkeley.
- Plag, I. (1999). *Morphological Productivity. Structural Constraints in English Derivation*. Mouton de Gruyter, Berlin/New York.
- Plénat, M. (2005). Brèves remarques sur les déverbaux en *-ette*. In Lambert, F. and Nølke, H., editors, *La syntaxe au cœur de la grammaire. Recueil offert en hommage pour le 60e anniversaire de Claude Muller*, pages 245–258. Presses Universitaires de Rennes, Rennes.
- Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. CSLI Publications; The University of Chicago Press, Stanford.
- Riehemann, S. (1998). Type-based derivational morphology. *Journal of Comparative Germanic Linguistics*, 2:49–77.
- Spencer, A. (2005). Inflecting clitics in generalized paradigm function morphology. *Lingue e Linguaggio*, 4:179–193.
- Tribout, D. (2010). How many conversions from verb to noun are there in French? In *Proceedings of the HPSG 2010 conference*, pages 341–357, Stanford. CSLI Publications.

# Why property anaphora

- In all the cases we have looked at, the two senses of the ambiguous lexeme correspond to non-overlapping denotations:
  - ☞ The set of shoe polishing events is disjoint from the set of polish portions of matter
  - ☞ Cf. Pinkal's (1995) notion of **h-ambiguity**
- Because of this, the failure of most ambiguity tests can be explained away with an underspecified semantics

(19) \* Le cirage noir a pris 10 minutes  
the polish/polishing black has taken 10 minutes  
 $\text{duration}(\sigma x.((\text{polish}'(x) \vee \text{polishing}'(x)) \wedge \text{black}'(x))) = 10min$

- ▶ Because **polishing'** is a property of events and **black'** is a property of physical objects,

$$\sigma x.((\text{polish}'(x) \vee \text{polishing}'(x)) \wedge \text{black}'(x)) \equiv \sigma x.(\text{polish}'(x) \wedge \text{black}'(x))$$

- ▶ Physical objects don't have durations  $\Rightarrow$  presupposition failure.
- Entity-level anaphora is excluded for the same reason; property anaphora is not.



# Hybrid objects

- Notoriously, some predicates such as *book* allow for co-predication (and anaphora) despite being h-ambiguous

- (20) a. The book on the table is stupid.  
b. This book is much too heavy, and that one is stupid.

- Analyzed in terms of hybrid objects belonging ‘simultaneously’ to two denotation domains (e.g. Godard & Jayez 2003; Pustejovsky 1995; Asher 2011)
- Lexeme formation sometimes outputs such hybrid object predicates


- (21) La présentation de Paul était passionnante. Elle est sur la table si tu veux la lire.  
‘Paul’s presentation was fascinating. In case you are interested in reading it, it is on the table.’

- However the existence of such cases has no bearing on the proper treatment of h-ambiguous lexemes.

# Lexemes and incremental processing

- Proposals treating pairs of homonyms using single entries
  - ▶ Poesio (1996): a single entry with a set of denotations
  - ▶ Asher (2011): a single entry with a dynamic disjunctive type
- Motivated by considerations of incremental processing
- Commendable goal, but

- ▶ Non-homophonous lexemes can have homophonous forms

*je suis*  be.PRS.1SG 'I am'  
follow.PRS.1SG 'I follow'

- ▶ An efficient incremental processor needs to treat these as alternative interpretations
- ▶ For this purpose it might be efficient to compile out a lexicon with unique entries for homophone WORDS
- ▶ No use in also doing so at the level of LEXEMES

☞ Considerations of incremental processing do not motivate unified entries for derived lexemes