

Sous-spécification, construction de lexèmes et structure du lexique

Olivier Bonami & Delphine Tribout

Journée de morphologie formelle
Toulouse, 6 octobre 2011

- Objectif : construire une approche formelle de la relation entre morphologie constructionnelle et lexicque
- Point de départ : règles de construction de lexème (RCL/LFR) intégrées à un lexique hiérarchique
 - ☞ Riehemann (1993, 1998); Koenig (1994, 1999); Hippisley (1997), Hippisley & Brown (sous presse); Bonami & Boyé (2006); etc.

1. Evaluation critique :

- ▶ Vertus d'une classification croisée des RCL
 - ☞ Bonami & Boyé, 2003; Desmets & Villoing, 2009; Tribout, 2010a, 2010b
- ▶ Confusion entre lexique stabilisé et construction en ligne

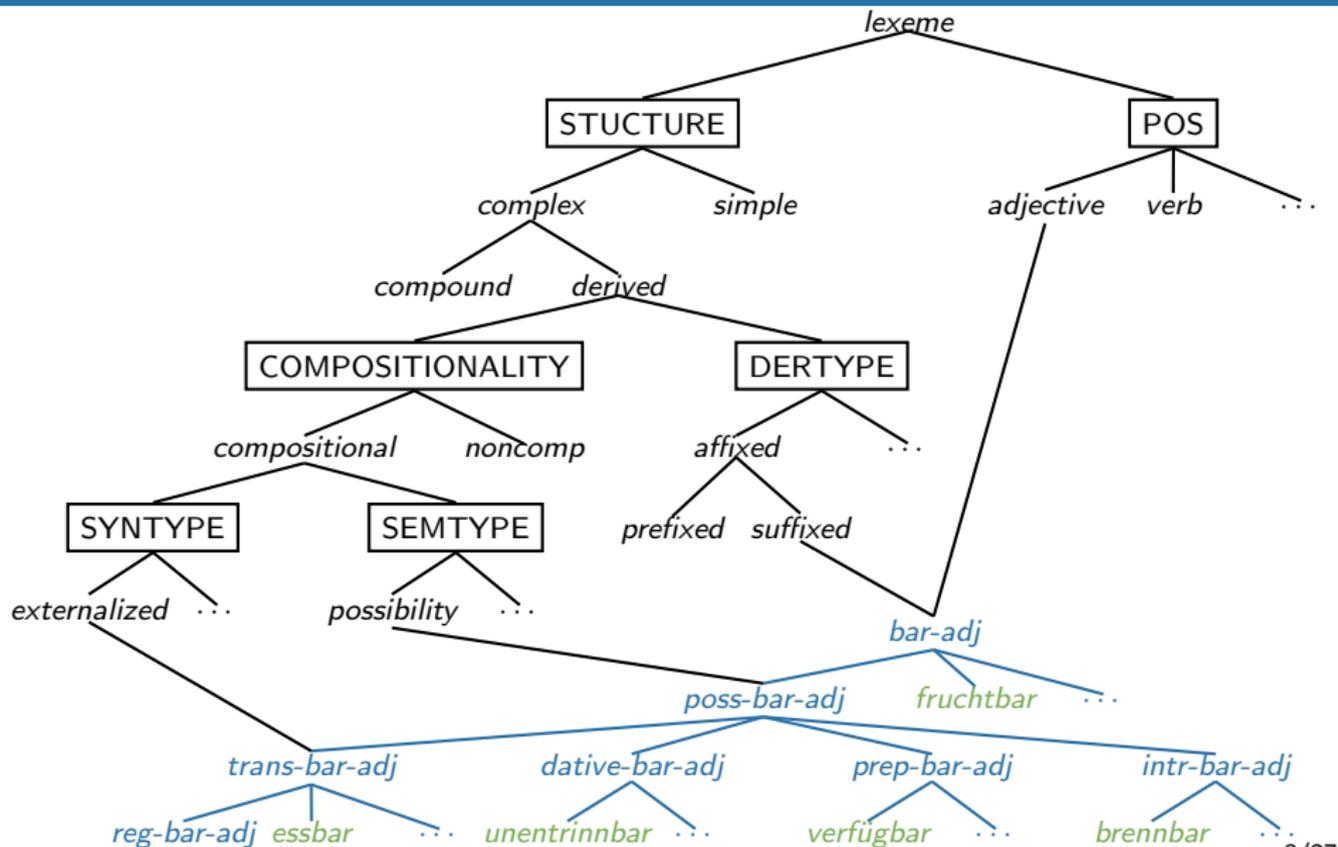
2. Construction partielle d'un modèle explicite de la relation entre lexique et construction

3. Discussion du rôle paradoxal de la sous-spécification en morphologie constructionnelle

Hiérarchie des RCL, hiérarchie des lexèmes

La construction de lexème au cœur d'un lexique hiérarchique

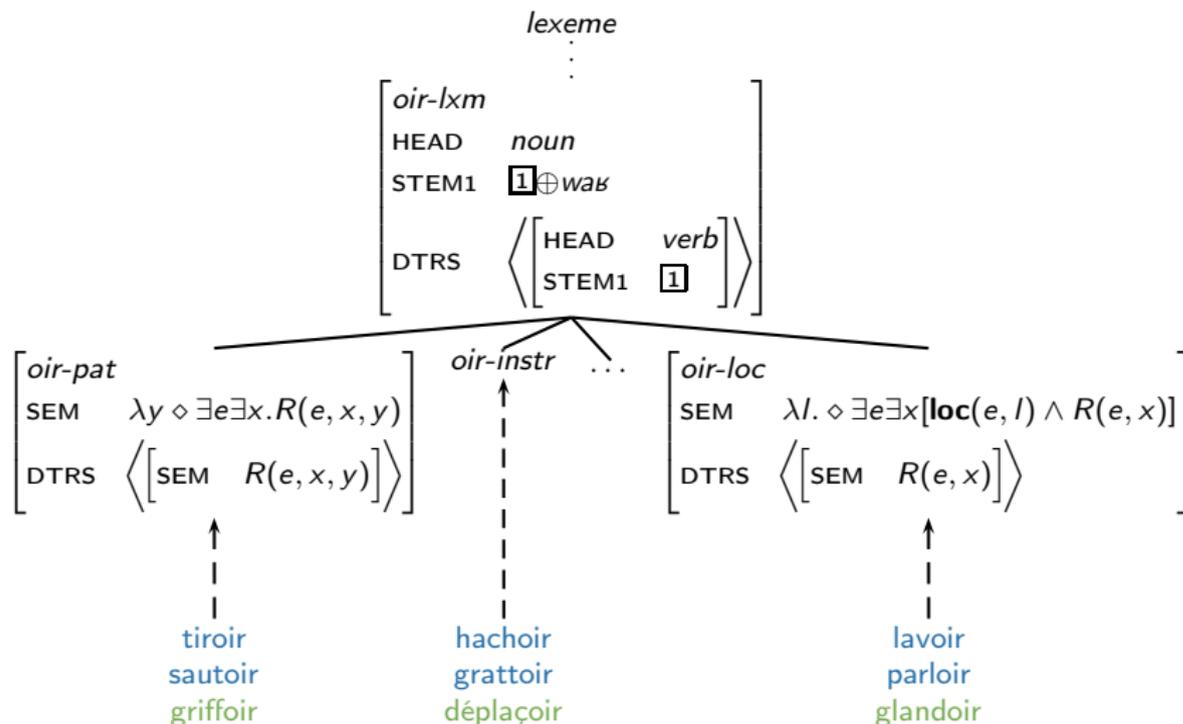
Exemple : Riehemann 1998



Vertus de la classification croisée des RCL

Plusieurs sorties sémantiques pour une même RCL

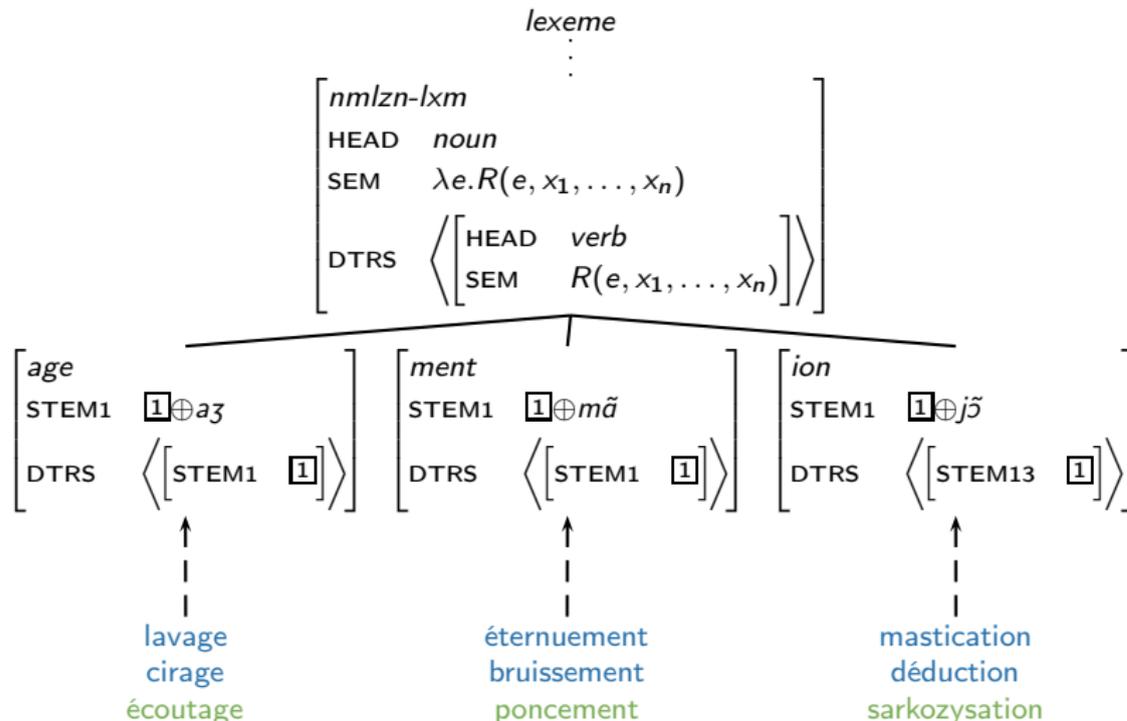
- Exemple : dérivés en *-oir* (données de Namer & Villoing 2008)



Vertus de la classification croisée des RCL

Plusieurs RCL avec la même sémantique

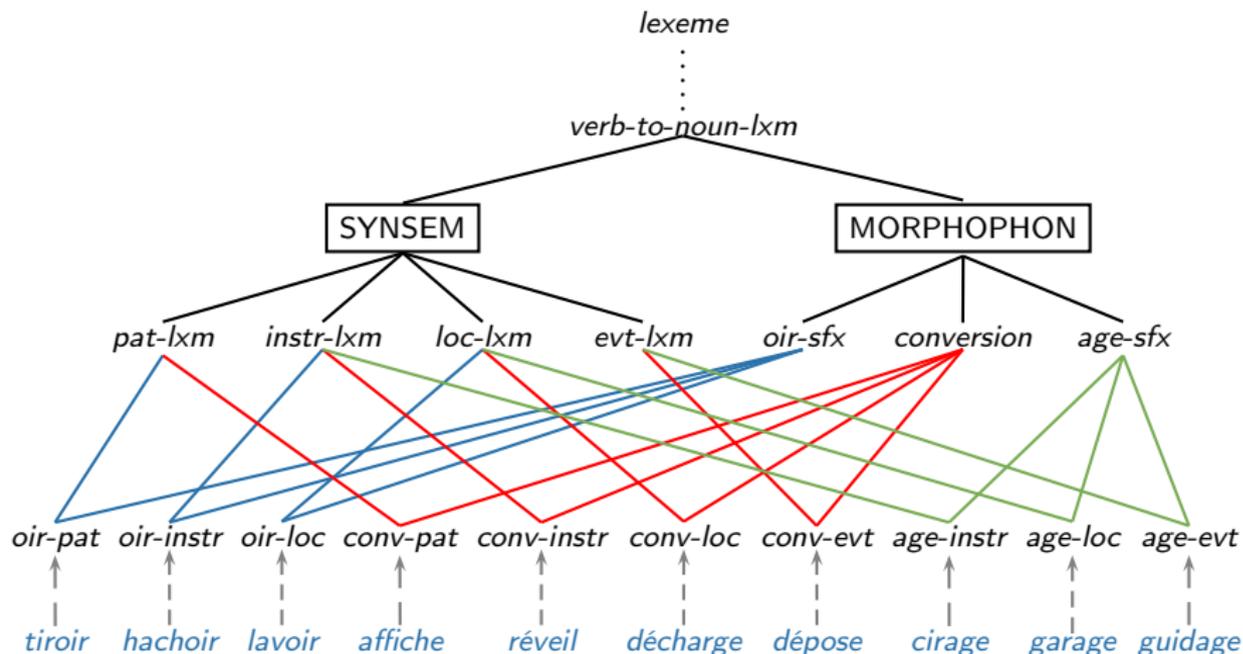
- Exemple : nominalisations



Vertus de la classification croisée des RCL

Exemple : les procédés verbe > nom

- Deux dimensions de classification conceptuellement indépendantes
- Toutes les combinaisons ne sont pas possibles

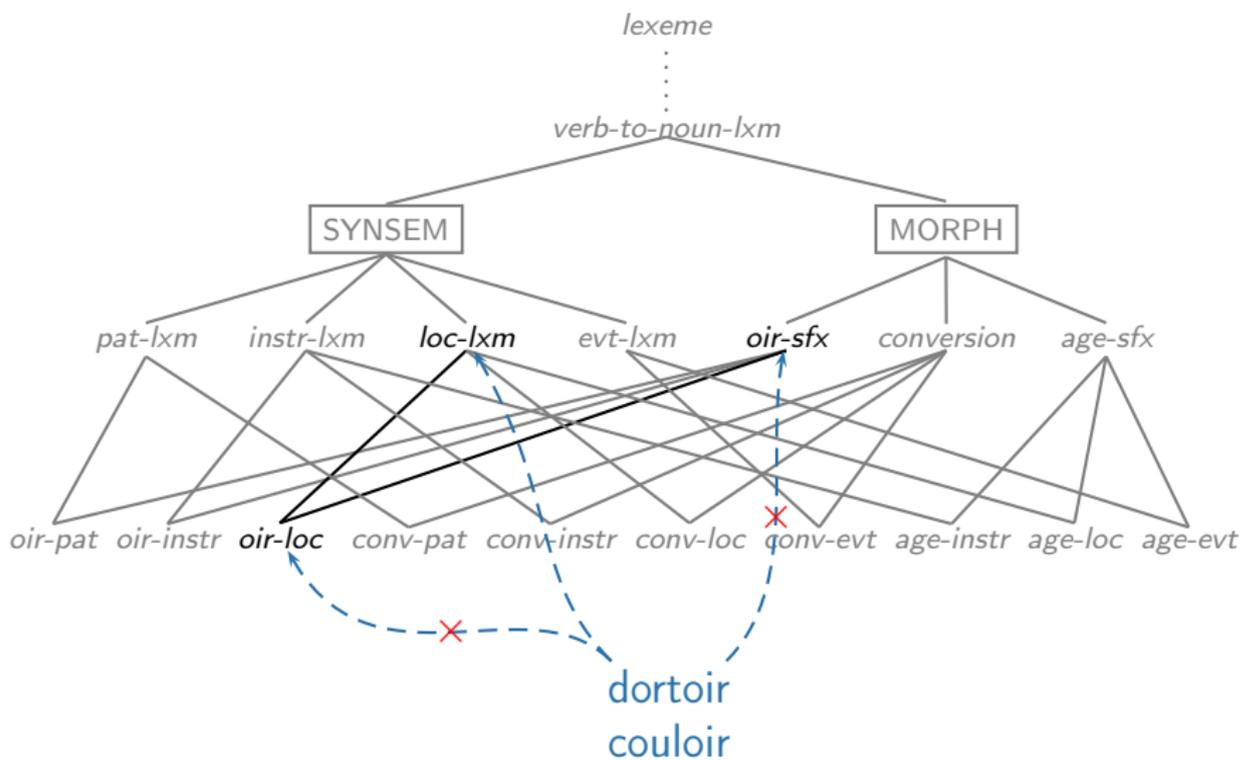


Intégrer deux hiérarchies ?

- Riehemann (et Koenig, et Hippisley) défend une hiérarchie unique mélangeant lexèmes stabilisés et RCL
- Motivation : les RCL émergent comme la généralisation par abstraction d'une classe de lexèmes concrets
- Avec le recul du temps :
 - ▶ Problème conceptuel : la hiérarchie modélise un état synchronique, elle ne peut rien dire directement de la dynamique du lexique.
 - ▶ Les données psycholinguistiques et diachroniques pointent vers une structuration du lexique beaucoup plus dense que ce qui peut être modélisé avec une hiérarchie d'héritage
 - ▶ Rôle des relations de similarité amorphologiques (phonologique, sémantique)
 - ▶ Rôles des relations bidirectionnelles (ex : agent ↔ événement)
 - ▶ Rôle des facteurs de fréquence
 - ▶ Problème d'adéquation descriptive : où placer les lexèmes qui acquièrent dans le lexique des propriétés idiosyncrasiques ?

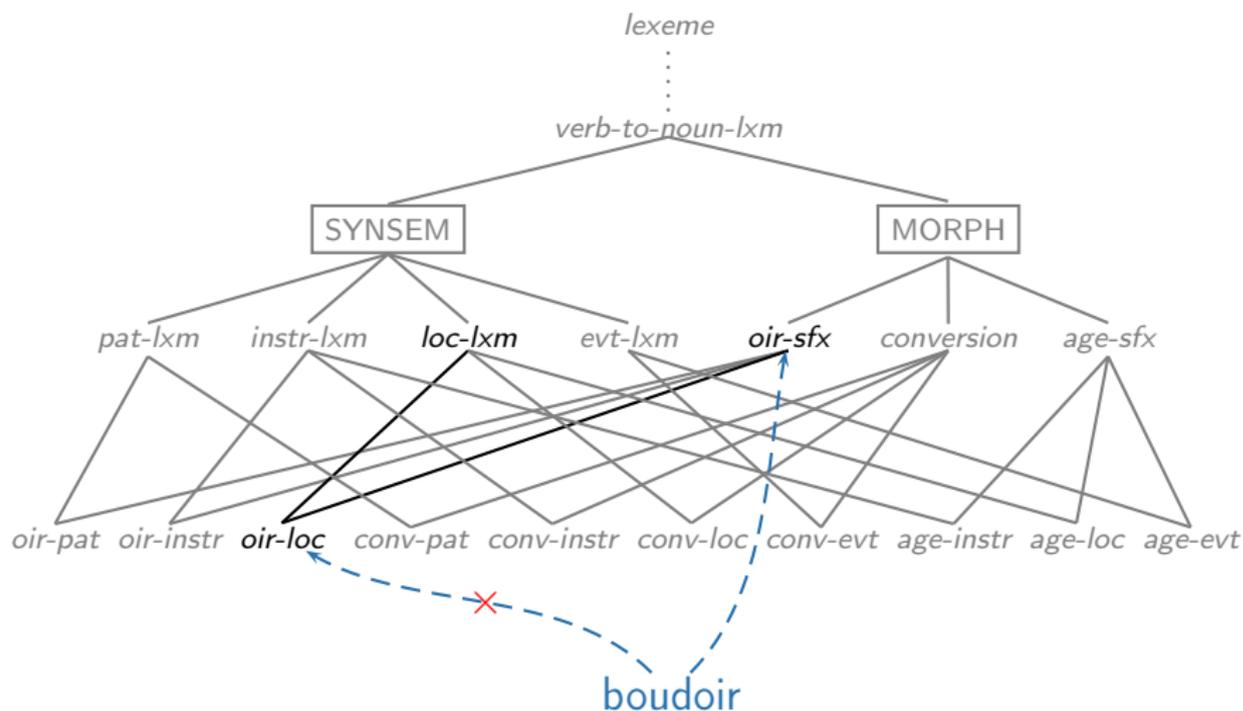
Comment placer les lexèmes dans la hiérarchie ?

1. Forme idiosyncrasique



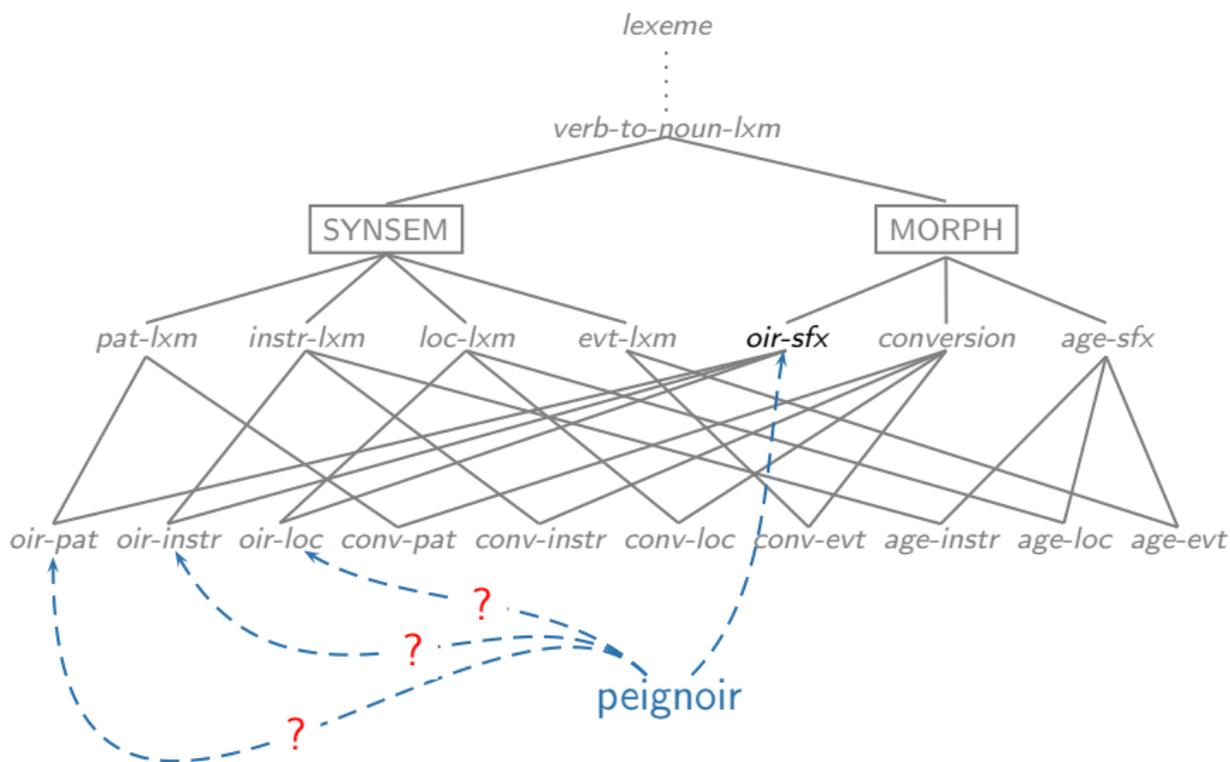
Comment placer les lexèmes dans la hiérarchie?

2. Sens glissé



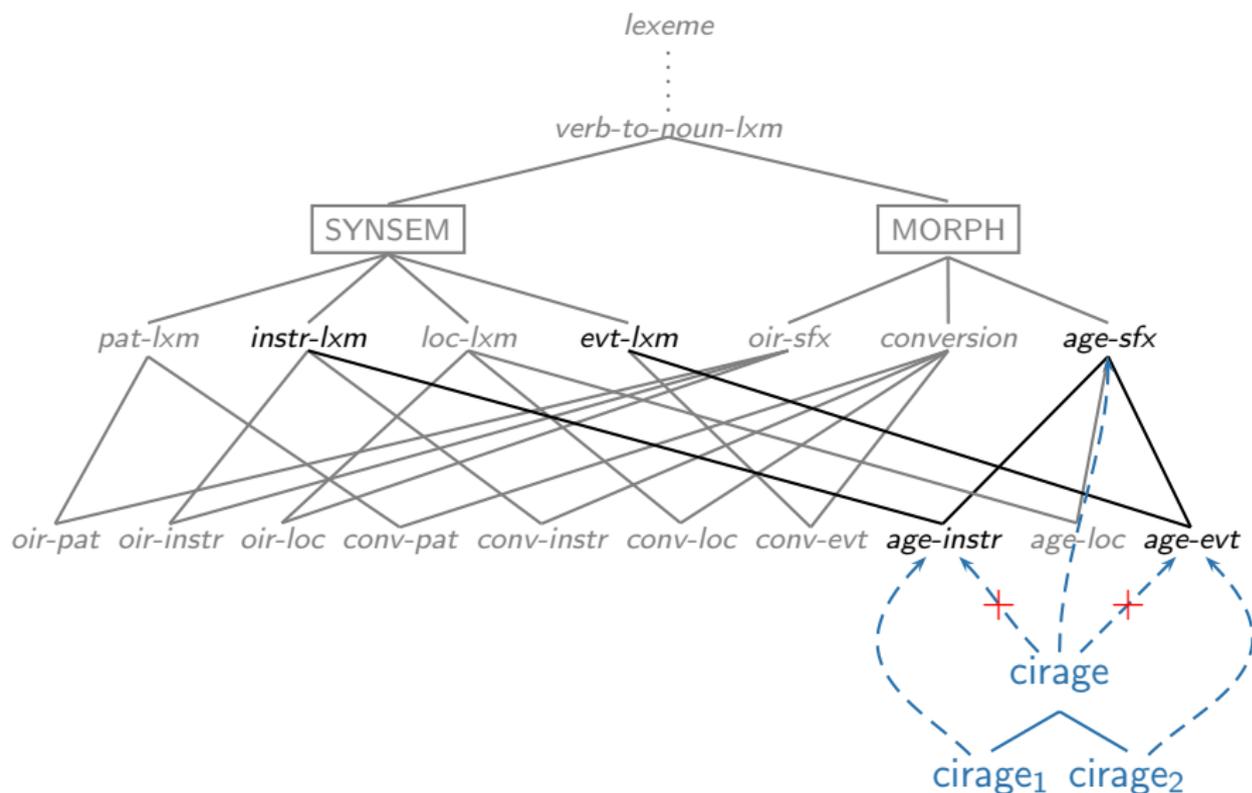
Comment placer les lexèmes dans la hiérarchie ?

3. Sens démotivé



Comment placer les lexèmes dans la hiérarchie ?

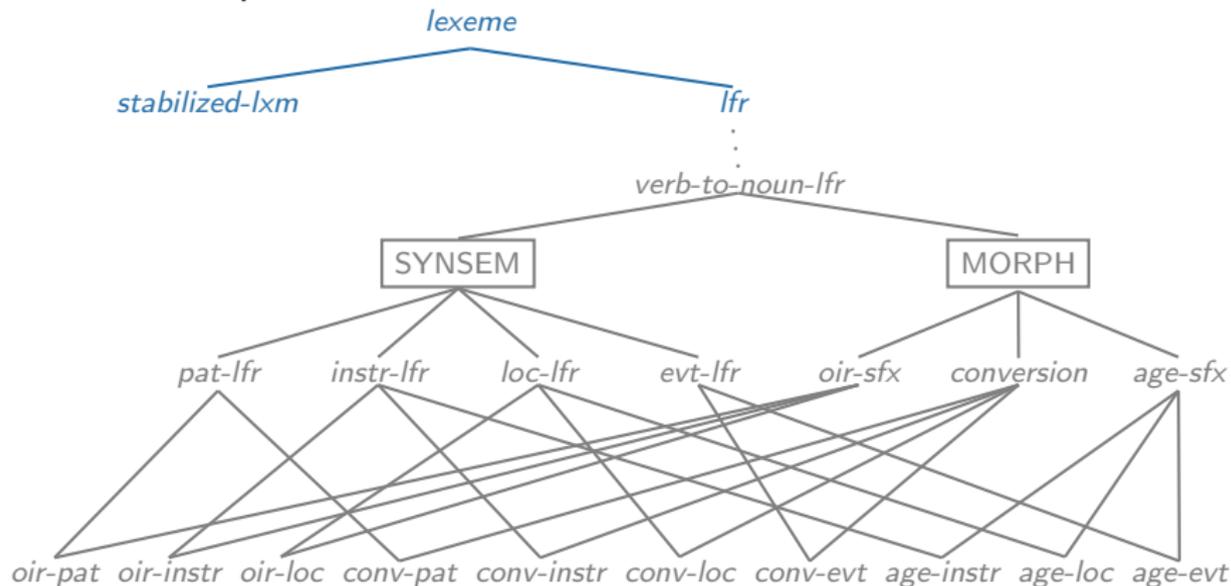
4. Polysémie



Retour à une vision conservatrice du lexique

(librement inspirée de Meurers, 2001 ; Sag et al., 2003 ; Sag, sous presse)

- Le lexique stabilisé est un ensemble d'entrées lexicales



- Un lexème est légitime si :

- Il correspond à une entrée lexicale de type *stabilized-lxm* ; ou
- Il correspond à l'output d'une *lfr* dont l'input est légitime.

Lexique stabilisé et lexique construit en ligne

Typologie des lexèmes

Au delà de la distinction simple/construit

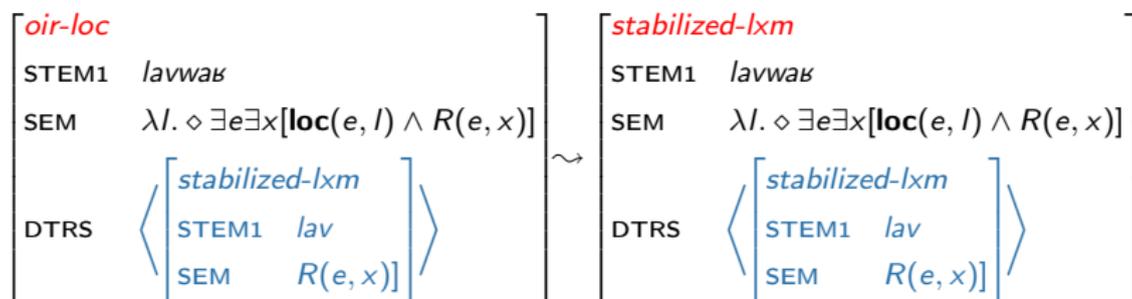
- Les lexèmes stabilisés dans le lexique ne sont pas tous simples
- Tout lexème qui a acquis des propriétés idiosyncrasiques
 - ▶ requiert une entrée lexicale, où ces propriétés sont listées ;
 - ▶ ne peut être de type *lfr*, puisqu'il ne respecte les contraintes d'aucune feuille de la hiérarchie.

lexème. . .	stabilisé	en ligne
simple	$\left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } klwat\grave{e} \\ \text{DTRS } \langle \rangle \end{array} \right]$	—
construit	$\left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lavwa\grave{e} \\ \text{DTRS } \langle \left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lav \end{array} \right] \rangle \end{array} \right]$	$\left[\begin{array}{l} \textit{oir-loc} \\ \text{STEM1 } gl\grave{a}dwa\grave{e} \\ \text{DTRS } \langle \left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } gl\grave{a}d \end{array} \right] \rangle \end{array} \right]$

La dynamique du lexique

L'institutionnalisation (Bauer 1983)

- Opération méta-grammaticale qui ajoute au lexique une entrée équivalente à l'output d'une RCL



La dynamique du lexique

La spécialisation

- Opération méta-grammaticale qui remplace la sémantique d'un lexème stabilisé par une sémantique plus spécifique

$$\left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lavw\grave{a}b \\ \text{SEM } \ll \text{endroit où on lave} \gg \end{array} \right] \rightsquigarrow \left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lavw\grave{a}b \\ \text{SEM } \ll \text{endroit où on lave le linge} \gg \end{array} \right]$$

- Techniquement, introduction d'une constante sémantique

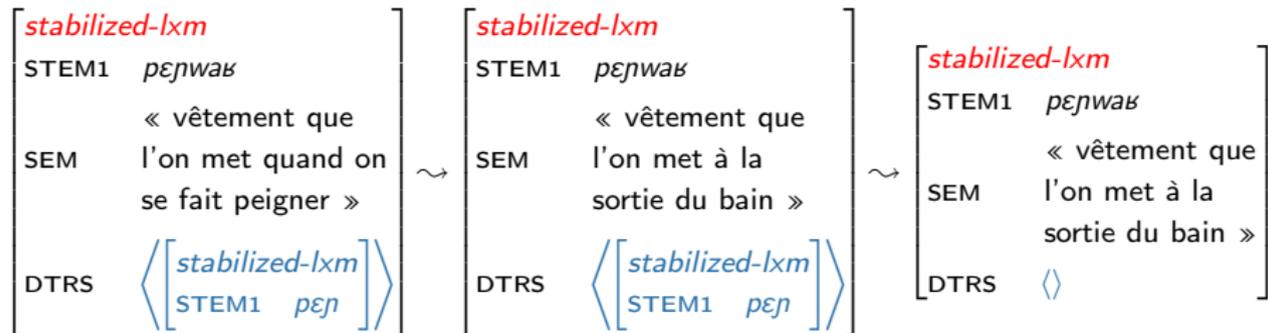
$$\left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lavw\grave{a}b \\ \text{SEM } \lambda l. \diamond \exists e \exists x [\text{loc}(e, l) \wedge R(e, x)] \\ \text{DTRS } \left\langle \begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lav \\ \text{SEM } R(e, x) \end{array} \right\rangle \end{array} \right] \rightsquigarrow \left[\begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lavw\grave{a}b \\ \text{SEM } \lambda l. \text{avoir}'(l) \\ \text{DTRS } \left\langle \begin{array}{l} \textit{stabilized-lxm} \\ \text{STEM1 } lav \\ \text{SEM } R(e, x) \end{array} \right\rangle \end{array} \right]$$

où pour tout l , $\llbracket \text{avoir}'(l) \rrbracket \models \llbracket \diamond \exists e \exists x [\text{loc}(e, l) \wedge R(e, x)] \rrbracket$.

La dynamique du lexique

Glissements et démotivation

- Les **glissements de sens** sont des opérations de modification substantielle de la sémantique associée à un lexème stabilisé
- La **démotivation** est une perte de la relation d'un lexème construit stabilisé à sa base, typiquement pour cause de
 - ▶ glissement de sens ; ou
 - ▶ sortie de la base du lexique



Le rôle de la sous-spécification en morphologie constructionnelle

Les effets de l'institutionnalisation

- Un lexème construit stabilisé a une base déterminée.

<table style="width: 100%; border-collapse: collapse;"> <tr><td colspan="2"><i>oir-loc</i></td></tr> <tr><td>STEM1</td><td>$\boxed{1} \oplus wab$</td></tr> <tr><td>SEM</td><td>$\lambda l. \diamond \exists e \exists x [\mathbf{loc}(e, l) \wedge R(e, x)]$</td></tr> <tr><td>DTRS</td><td>$\left\langle \left[\begin{array}{cc} \text{STEM1} & \boxed{1} \\ \text{SEM} & R(e, x) \end{array} \right] \right\rangle$</td></tr> </table>	<i>oir-loc</i>		STEM1	$\boxed{1} \oplus wab$	SEM	$\lambda l. \diamond \exists e \exists x [\mathbf{loc}(e, l) \wedge R(e, x)]$	DTRS	$\left\langle \left[\begin{array}{cc} \text{STEM1} & \boxed{1} \\ \text{SEM} & R(e, x) \end{array} \right] \right\rangle$	vs.	<table style="width: 100%; border-collapse: collapse;"> <tr><td colspan="2"><i>stabilized-lxm</i></td></tr> <tr><td>STEM1</td><td><i>lavwab</i></td></tr> <tr><td>SEM</td><td>lavoir'</td></tr> <tr><td>DTRS</td><td>$\left\langle \left[\begin{array}{cc} \text{STEM1} & lav \\ \text{SEM} & \mathbf{laver}'(e, x) \end{array} \right] \right\rangle$</td></tr> </table>	<i>stabilized-lxm</i>		STEM1	<i>lavwab</i>	SEM	lavoir'	DTRS	$\left\langle \left[\begin{array}{cc} \text{STEM1} & lav \\ \text{SEM} & \mathbf{laver}'(e, x) \end{array} \right] \right\rangle$
<i>oir-loc</i>																		
STEM1	$\boxed{1} \oplus wab$																	
SEM	$\lambda l. \diamond \exists e \exists x [\mathbf{loc}(e, l) \wedge R(e, x)]$																	
DTRS	$\left\langle \left[\begin{array}{cc} \text{STEM1} & \boxed{1} \\ \text{SEM} & R(e, x) \end{array} \right] \right\rangle$																	
<i>stabilized-lxm</i>																		
STEM1	<i>lavwab</i>																	
SEM	lavoir'																	
DTRS	$\left\langle \left[\begin{array}{cc} \text{STEM1} & lav \\ \text{SEM} & \mathbf{laver}'(e, x) \end{array} \right] \right\rangle$																	

- Cependant, une entrée lexicale est une description sous-spécifiée de l'environnement dans lequel le lexème sera utilisé.

<table style="width: 100%; border-collapse: collapse;"> <tr><td colspan="2">STEM1</td><td><i>lave</i></td></tr> <tr><td colspan="2" rowspan="5">ARG-ST</td><td>$\left\langle \left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right] \right\rangle$</td></tr> <tr><td>SUBJ</td><td>?</td></tr> <tr><td>COMPS</td><td>?</td></tr> <tr><td colspan="2">...</td></tr> </table>	STEM1		<i>lave</i>	ARG-ST		$\left\langle \left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right] \right\rangle$	SUBJ	?	COMPS	?	...		<table style="width: 100%; border-collapse: collapse;"> <tr><td colspan="2">$\left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right]$</td></tr> </table>	$\left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right]$	
STEM1		<i>lave</i>													
ARG-ST		$\left\langle \left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right] \right\rangle$													
		SUBJ	?												
		COMPS	?												
		...													
		$\left[\begin{array}{cc} \text{CAT} & \textit{noun} \\ \text{GEN} & ? \\ \text{NUM} & ? \\ \text{PER} & ? \\ \dots & \dots \end{array} \right]$													

Règle sous-spécifiée, valeur de trait spécifique

- Classe flexionnelle :
 - ▶ *ajuster* vs. **ajustir*
 - ▶ **anobler* vs. *anoblir*
 - ▶ *terrèr* vs. *terrir*

Règle sous-spécifiée, valeur de trait spécifique

- Classe flexionnelle :
 - ▶ *ajuster* vs. **ajustir*
 - ▶ **anobler* vs. *anoblir*
 - ▶ *terrorer* vs. *terrir*

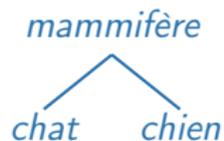
- Genre des inanimés :
 - ▶ *un peignoir* vs. **une peignoire*
 - ▶ **un baignoir* vs. *une baignoire*
 - ▶ *un bouilloir* vs. *une bouilloire*

Règle sous-spécifiée, valeur de trait spécifique

- Classe flexionnelle :
 - ▶ *ajuster* vs. **ajustir*
 - ▶ **anobler* vs. *anoblir*
 - ▶ *terrér* vs. *terrir*
- Genre des inanimés :
 - ▶ *un peignoir* vs. **une peignoire*
 - ▶ **un baignoir* vs. *une baignoire*
 - ▶ *un bouilloir* vs. *une bouilloire*
- Sémantique :
 - ▶ 1. Je me suis servi d'un hachoir pour hacher la viande
 - ▶ 2. #Je me suis servi d'un tiroir pour tirer la voiture
 - ▶ 1. J'ai tiré le tiroir
 - ▶ 2. #J'ai haché le hachoir

La spécificité sémantique

- En matière sémantique, on ne peut pas dire que les lexèmes doivent être « maximalement spécifiques »

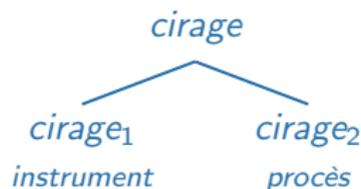
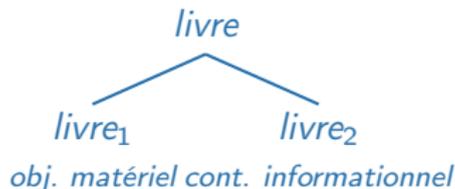


La spécificité sémantique

- En matière sémantique, on ne peut pas dire que les lexèmes doivent être « maximalement spécifiques »



- Il est souvent tentant de faire l'hypothèse d'un lexème à sens sous-spécifié quand on rencontre des emplois différenciés



La spécificité sémantique

- En matière sémantique, on ne peut pas dire que les lexèmes doivent être « maximalement spécifiques »



- Il est souvent tentant de faire l'hypothèse d'un lexème à sens sous-spécifié quand on rencontre des emplois différenciés

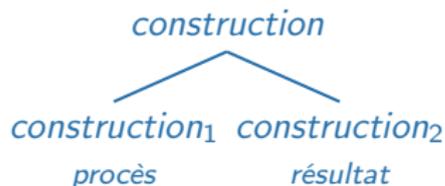


- Mais :

- | | |
|--|--|
| 1. J'ai acheté un livre ₁ rouge. | 1. J'ai acheté du cirage ₁ noir. |
| 2. Un livre ₂ passionnant se lit vite | 2. Un cirage ₂ rapide ne prend qu'une minute. |
| 3. Un livre rouge passionnant | 3. *Un rapide cirage noir |

Spécificité sémantique des nominalisations

- De même, il tentant de faire l'hypothèse d'un lexème *construction* à sens sous-spécifié :

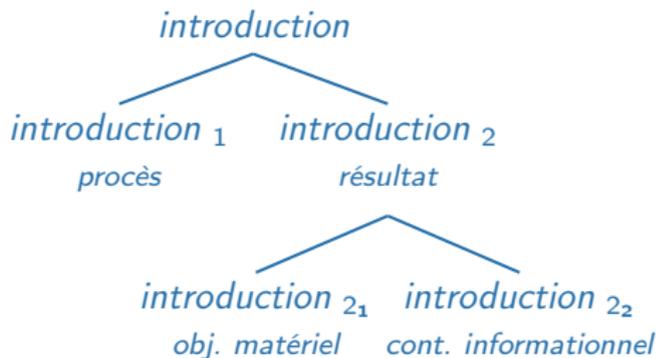


- Mais :

- J'ai assisté à une construction₁ interminable.
 - J'ai vu une magnifique construction₂ en briques
 - *J'ai assisté à une interminable construction en briques
- J'ai vu cette construction. Elle mesure 20 mètres.
 - *J'ai assisté à cette construction. Elle mesure 20 mètres.
- Certaines de ces constructions sont magnifiques. J'en ai visité une l'an dernier.
 - Certaines de ces constructions ont pris des siècles. *J'en ai visité une l'an dernier.

Spécificité sémantique des nominalisations

- Cf. *introduction* :



- (4) a. *L'introduction du problème dans le chapitre 4 est sur la table.
(procès + obj. matériel)
- b. *J'ai lu l'introduction du problème dans le chapitre 4.
(procès + cont. informationnel)
- c. L'introduction que j'ai lue est sur la table.
(cont. informationnel + obj. matériel)

☞ 2 lexèmes *introduction*

- Il reste souhaitable de dire que *cirage*₁ et *cirage*₂, *construction*₁ et *construction*₂ ont quelque chose en commun au delà de l'identité phonologique.
- ☞ C'est le même flexème au sens de Fradin & Kerleroux (2003)
- Bonami (2011) :
 - ▶ Les lexèmes portent un trait PID (Paradigm IDentifier) qui caractérise leur paradigme de flexion
 - ▶ Trait inaccessible à la syntaxe, mais indispensable pour rapprocher les homonymes irréguliers (ex : *devoir*)
 - ▶ Si deux procédés de construction ont le même effet sur la forme, ils introduisent le même PID
- ☞ L'identité de PID est suffisante pour capter la notion de flexème ; ne nécessite pas d'introduire un supertype commun.

Addendum : spécification et construction en ligne

- La spécification des valeurs de certains traits des construits n'est pas un effet de la stabilisation des lexèmes.
- Les lexèmes construits en ligne sont déjà spécifiés
 - ▶ pour la classe flexionnelle :
 1. j'ai stabilisé mes notes en relisant le cours
 2. au coucher du soleil le ciel orangit
 - ▶ pour le genre des inanimés
 1. les freineurs contrôlent et ralentissent le débit de l'eau
 2. cette cloueuse permet de fixer un tableau au mur sans efforts
 3. J'ai acheté une cloueuse hier. ??Ce cloueur permet de construire une charpente en une après-midi.
 - ▶ pour le sens
 1. la truite est pêchable dans les rivières de montagne
 2. ces rivières sont pêchables de juillet à septembre
 3. *la truite est pêchable dans les rivières qui le sont de juillet à septembre

Conclusion

- Architecture proposée pour l'articulation lexicale/RCL
 1. Les RCL sont des lexèmes schématiques organisés en hiérarchie d'héritage.
 2. Les lexèmes stabilisés sont des objets formellement distincts des lexèmes construits en ligne :
 - ▶ les lexèmes stabilisés ne sont pas des instances de RCL, même quand ils sont synchroniquement construits ;
 - ▶ modélisation explicite des différences entre lexème simple, construit stabilisé, historiquement construit mais démotivé, construit en ligne.
 3. Définition d'opérations dynamiques de modification du lexique : institutionnalisation, spécialisation, glissement(s), démotivation.
- Nouveau problème : les lexèmes sont plus spécifiés que le procédé utilisé pour les construire ne l'exige.
 - ▶ Délimitation empirique du problème délicate
 - ▶ Modélisation incertaine : la phénoménologie est-elle limitée au lexique stabilisé ?

Références

- ▶ Laurie Bauer, *English word formation*, Cambridge University Press, Cambridge, 1983.
- ▶ Olivier Bonami, *Reconstructing HPSG morphology*, présentation au colloque HPSG 2011, Seattle, 2011.
- ▶ Olivier Bonami and Gilles Boyé, *La construction des paradigmes*. Présentation à la Journée de morphologie non-conventionnelle, ERSS, Toulouse, 2003.
- ▶ ———, *Deriving inflectional irregularity*, Proceedings of the HPSG'06 Conference (Stanford) (Stefan Müller, ed.), CSLI publications, 2006, pp. 361–380.
- ▶ Marianne Desmets and Florence Villoing, *French vn lexemes : morphological compounding in hpsg*, Proceedings of the HPSG'09 Conference, 2009, pp. 89–109.
- ▶ Bernard Fradin and Françoise Kerleroux, *Quelles bases pour les procédés de la morphologie constructionnelle ?*, Les unités morphologiques. Actes du 3^{ième} forum de morphologie (Bernard Fradin, Georgette Dal, Nabil Hathout, Françoise Kerleroux, Marc Plénat, and Michel Roché, eds.), Presses Universitaires du Septentrion, Lille, 2003, pp. 76–84.
- ▶ Andrew Hippisley and Dunstan Brown, *Network morphology : A defaults-based theory of word structure.*, Cambridge University Press, Cambridge, 2012.
- ▶ Andrew Hippisley, *Declarative derivation : a network morphology account of russian word formation with reference to nouns denoting 'person'*, Ph.D. thesis, University of Surrey, 1997.
- ▶ Jean-Pierre Koenig, *Lexical relations*, CSLI Publications, Stanford, 1999.
- ▶ Susanne Riehemann, *Type-based derivational morphology*, Journal of Comparative Germanic Linguistics 2 (1998), 49–77.
- ▶ Ivan A. Sag, Thomas Wasow, and Emily Bender, *Syntactic theory. a formal introduction*, CSLI Publications, Stanford, 2003.
- ▶ Delphine Tribout, *How many conversions from verb to noun are there in French ?*, Proceedings of the 17th International Conference on HPSG (Stanford) (Stefan Müller, ed.), CSLI Publications, 2010, pp. 341–357.