

Démonette : une architecture ouverte destinée à la description morphologique dérivationnelle

Nabil Hathout(1) & Fiammetta Namer(2)

(1) CLLE/ERSS, CNRS & Université de Toulouse

(2) Université de Lorraine & ATILF-UMR 7118

Thématiques actuelles de la recherche en TAL

17 novembre 2014

Plan

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion

Objectifs

Démonette

Nouvelle ressource lexical du français

- Ressource morphologique dérivationnelle
- Ressource généraliste (à large couverture)

Capacité à inclure des descriptions linguistiques variées

- Démonette est **agnostique** sur le plan théorique et pour les informations décrites.
- Elle dispose d'un **format ouvert** que l'on peut étendre facilement.
- Elle est conçue pour la **préservation** et la **réutilisation** des données et des descriptions / des analyses morphologiques
- L'ambition est de définir un format général pour les descriptions morphologiques dérivationnelles à même de devenir un **standard**.

Caractéristiques remarquables

Démonette sert de support pour la réflexion sur l'organisation des relations morphologiques dans le lexique.

Une structure originale

- Démonette est « orientée relation ».
 - C'est une base de relations morphologiques
 - Les mots sont des jonctions entre les relations.
 - Démonette contient des relations **directes**, **inverses** et **indirectes**.
- Démonette est un réseau non connexe.
 - Elle décrit les relations à l'intérieur des **familles** dérivationnelles.
- Les relations sont munies de définitions sémantiques concrètes et **abstraites**.

Caractéristiques remarquables

Démonette sert de support pour la réflexion sur l'organisation des relations morphologiques dans le lexique.

Une conception cumulative du sens dérivationnel

Chaque relation dérivationnelle induit des propriétés sémantiques sur les mots qu'elle connecte.

- Le sens d'un mot est un **empilement** de ces propriétés sémantiques partielles.
- Les mots sont **typés** sémantiquement.
 - Le typage est cumulatif.
 - Chaque relation dérivationnelle attribue un type à sa source et à sa cible.

- Démonette décrit des relations morphologiques à l'intérieur des familles dérivationnelles.
 - Une extension aux séries dérivationnelles est prévue dans des versions ultérieures.
- Démonette décrit 7 suffixations

exposant formel	trait sémantique	relation catégorielle	exemple
<i>-age</i>	action	V → N	essorer → essorage
<i>-ion</i>	action	V → N	animer → animation
<i>-ment</i>	action	V → N	charger → chargement
<i>-eur</i>	agent	V → N	capter → capteur
<i>-euse</i>	agent	V → N	bercer → berceuse
<i>-rice</i>	agent	V → N	décorer → décoratrice
<i>-if</i>	propriété	N, V → A	incitation → incitatif

Contenu actuel (2)

- Démonette est alimentée par
 - les lemmes de TLFnome,
 - les analyses de **DériF** (Fiammetta Namer) et
 - les filaments de **Morphonette** (Nabil Hathout)
- La version actuelle de Démonette contient (juillet 2014) :
 - 31 204 relations morphologiques
 - 14 800 mots (lexèmes)

Cette version ne sera pas distribuée. Une nouvelle version intégralement reconstruite est en cours de développement.

Plan

- 1 Introduction
- 2 Relations morphologiques**
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion

Dans les familles dérivationnelles

Extrait de la famille de *essorer*

essorage

essorer

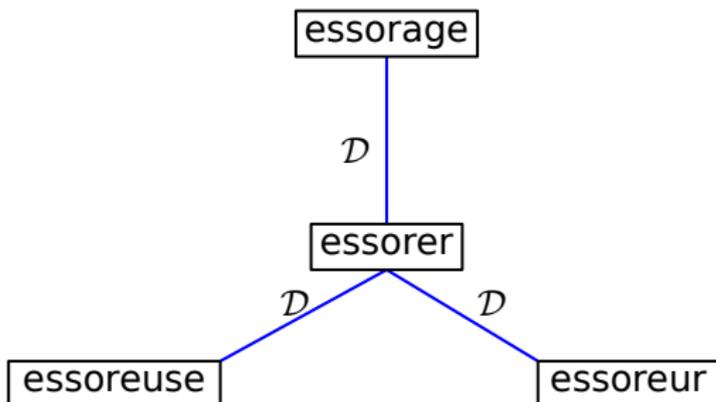
essoreuse

essoreur

Des relations directes ...

Relations entre une base et un dérivé :

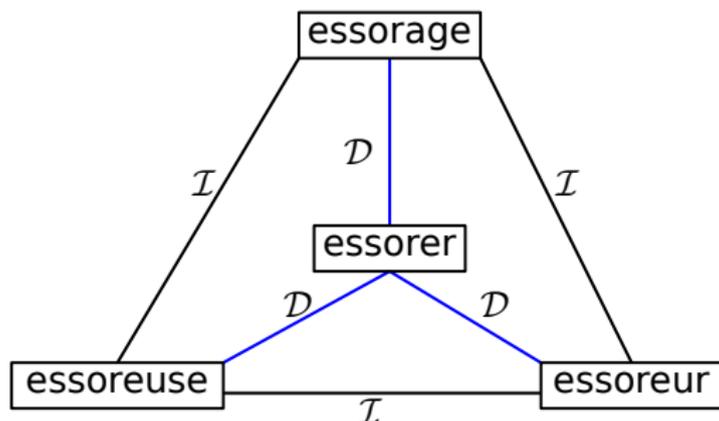
Extrait de la famille de *essorer*



... et indirectes

Il existe d'autres relations entre les membres de la même famille

Extrait de la famille de *essorer*

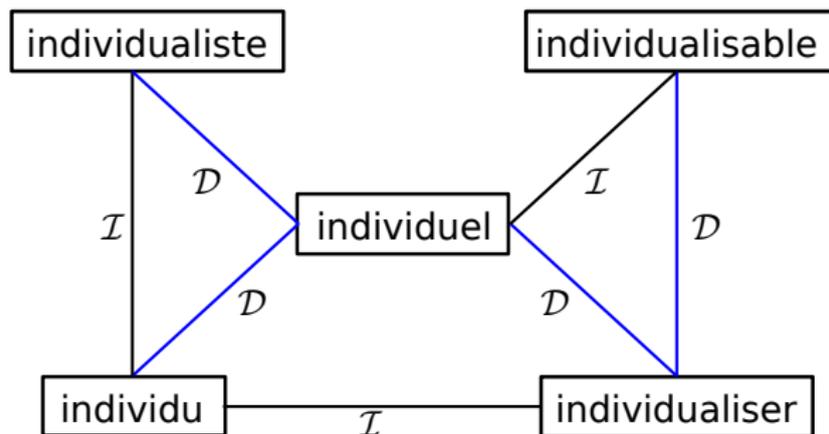


Ces relations sont indirectes.

... et indirectes (2)

Les relations indirectes ne saturent pas obligatoirement la famille.

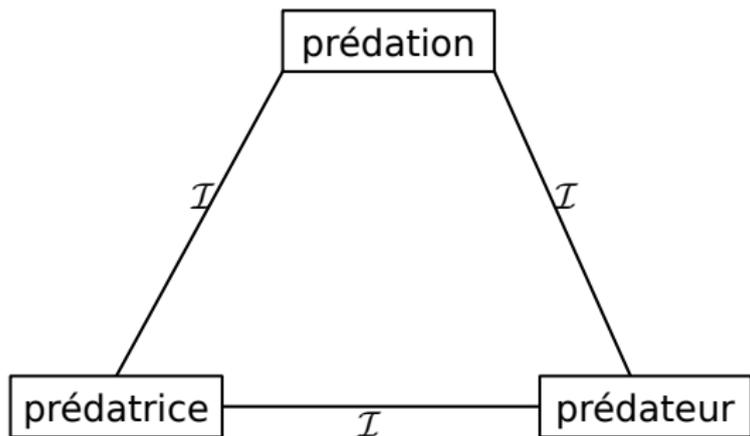
Extrait de la famille de *individuel*



... et indirectes (3)

Dans certaines familles, certaines relations directes peuvent manquer.

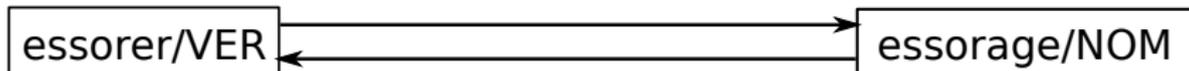
Extrait de la famille de *prédation*



Relations inverses

- Relation morphologique entre un dérivé et sa base.
- C'est la réciproque de la relation directe entre la base et le dérivé.

essorage : *essorer*



essorer est en relation avec *essorage* tout autant que
essorage est en relation avec *essorer* !

Relations dérivationnelles

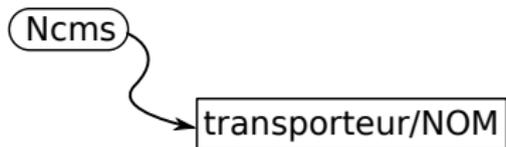
- L'ensemble des relations forment un graphe.
- Il n'y a pas de différence structurelle (fondamentale) entre les relations directes / inverses et les relations indirectes.
- Toutes les relations dans Démonette ont une réciproque.

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques**
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion

Description morphologique des relations directes

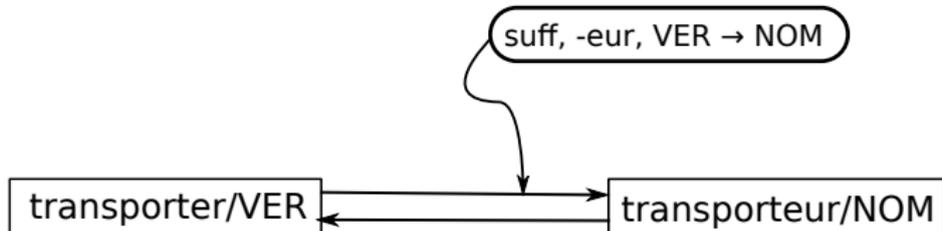
Les mots sont munis de leurs catégories grammaticales et leurs traits morphosyntaxiques.

transporteur



Les relations directes sont munies des procédés constructionnels dont elles découlent.

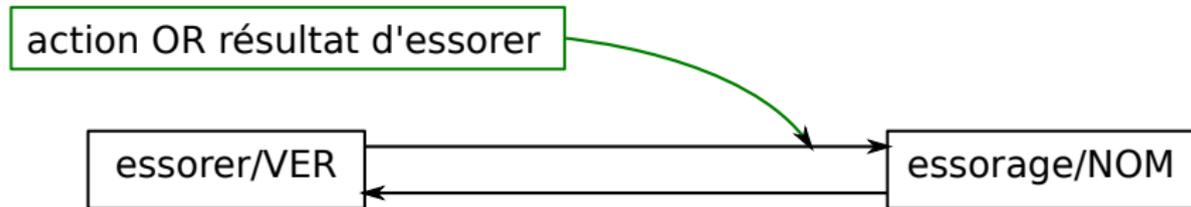
transporter : transporteur



Définitions orientées

Les relations directes sont munies de **définitions orientées**.

essorer : *essorage*

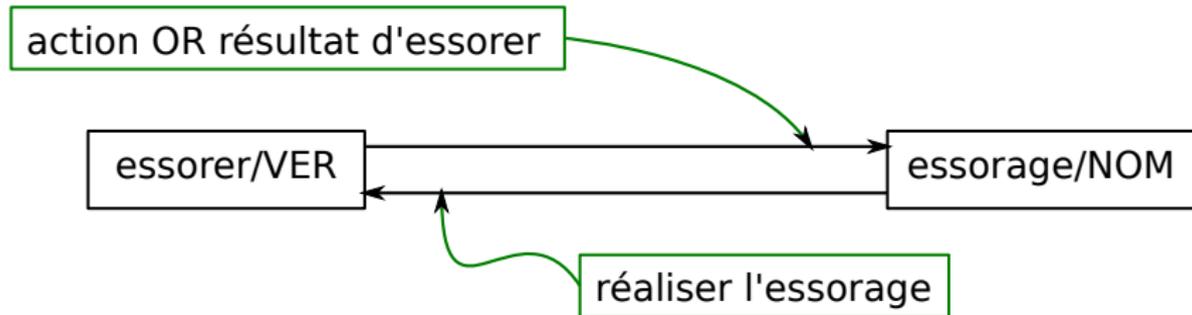


La définition décrit le sens dérivationnel du dérivé relativement à celui de la base.

Définitions opposées

Les relations inverses sont munies de **définitions opposées**.

essorage : *essorer*

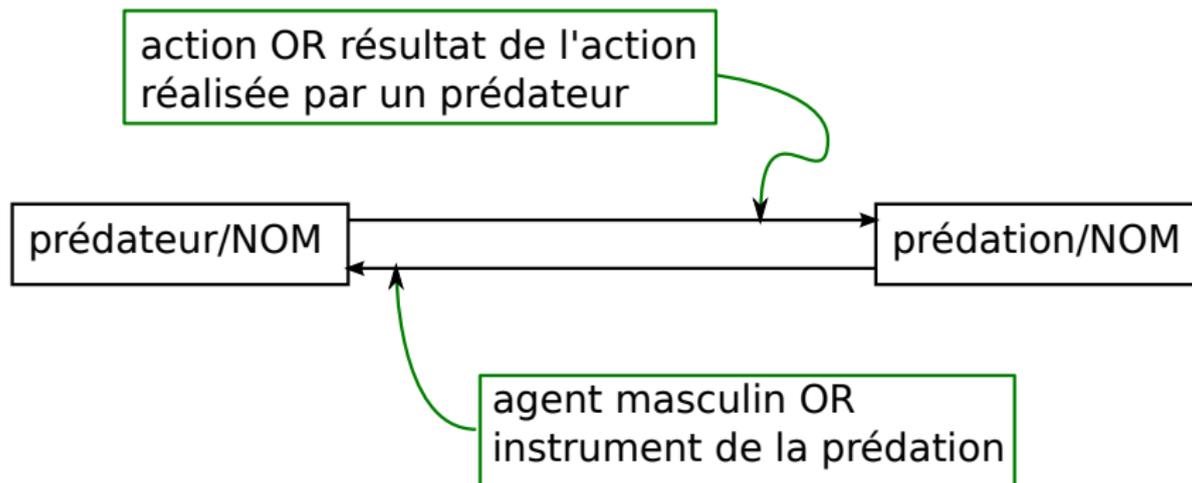


- La définition décrit le sens dérivationnel de la base relativement à celui du dérivé.
- Généralisation de l'idée de **sélection mutuelle** proposée par Michel Roché pour les mots construits en *-isme* et *-iste*.

Définitions transversales

Les relations indirectes sont munies de **définitions transversales**.

prédateur : *prédation*



- La définition décrit le sens dérivationnel du mot cible relativement à celui du mot source.

Définitions abstraites

Tous les noms d'action sont dans la même relation relativement à leur prédicat et réciproquement.

manifestation : action OR résultat de *manifester*

soulèvement : action OR résultat de *(se) soulever*

lavage : action OR résultat de *laver*

manifester : réaliser la *manifestation*

(se) soulever : réaliser le *soulèvement*

laver : réaliser le *lavage*

Définitions abstraites

Tous les noms d'action sont dans la même relation relativement à leur prédicat et réciproquement.

Les **définitions abstraites** rendent comptes de cette régularité.

manifestation : action OR résultat de *manifester*

soulèvement : action OR résultat de *(se) soulever*

lavage : action OR résultat de *laver*

: action OR résultat de **@**

où **@** représente un prédicat.

manifester : réaliser la *manifestation*

(se) soulever : réaliser le *soulèvement*

laver : réaliser le *lavage*

: réaliser le **@ACT**

où **@ACT** représente un nom d'action.

Types morphosémantiques

- Dans les définitions abstraites, les lexèmes sont représentés par des variables typées.
- 5 types morphosémantiques sont définis pour les 7 suffixations actuellement décrites dans Démonette.

<i>étiquette</i>	<i>type</i>	<i>exemple</i>
@ACT	nom d'action	lavage gonflement exclusion
@AGF	nom d'agent féminin	danseuse productrice
@AGM	nom d'agent masculin	danseur
@PROP	adjectif de propriété	combatif
@	prédicat	danser

Types morphosémantiques (2)

Le type morphosémantique représente la classe et le rôle de la cible dans le « scénario » associé au prédicat.

Les types morphosémantiques combinent :

- des catégories grammaticales : nom, adjectif, verbe, adverbe
- le genre des noms : masculin, féminin
- des catégories sémantiques : arbre (*-ier*), plantation (*-aie*), métier (*-ier, -iste*), gentilé (*-ien, -ais*), animal (*-ard*), propriété (*-if, -itude*), action (*-age, -ment, -ion*), etc.
- des rôles thématiques : agent (*-ant, -ard, -eur, -euse, -rice*), instrument (*-eur, -euse, -rice, -oir, -oire*), résultat (*-age, -ment, -ion*), lieu (*-oir, -oire*), etc.
- des relations sémantiques spécifiques : participant antagoniste insuffisant (*-able*), participant antagoniste suffisant (*in-X-able*).

Types morphosémantiques (3)

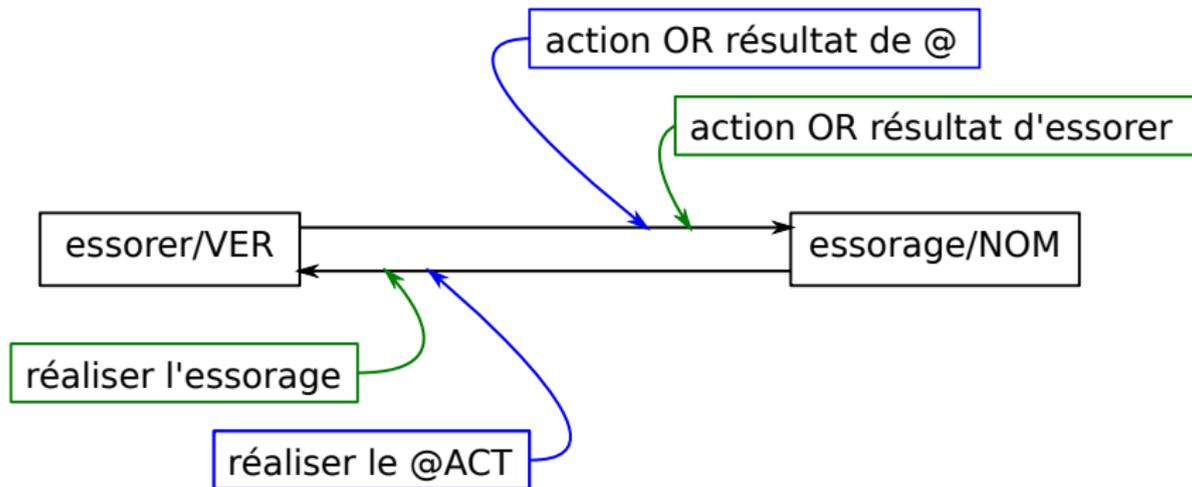
Paradigmes dérivationnels

- Les types morphosémantiques ouvrent la voie vers de **véritables** paradigmes dérivationnels en identifiant une organisation morphosémantique commune pour l'ensemble des familles dérivationnelles.
- Les types morphosémantiques sont à la morphologie dérivationnelle ce que les traits morphosyntaxiques sont à la morphologie flexionnelle.

Définitions abstraites (suite)

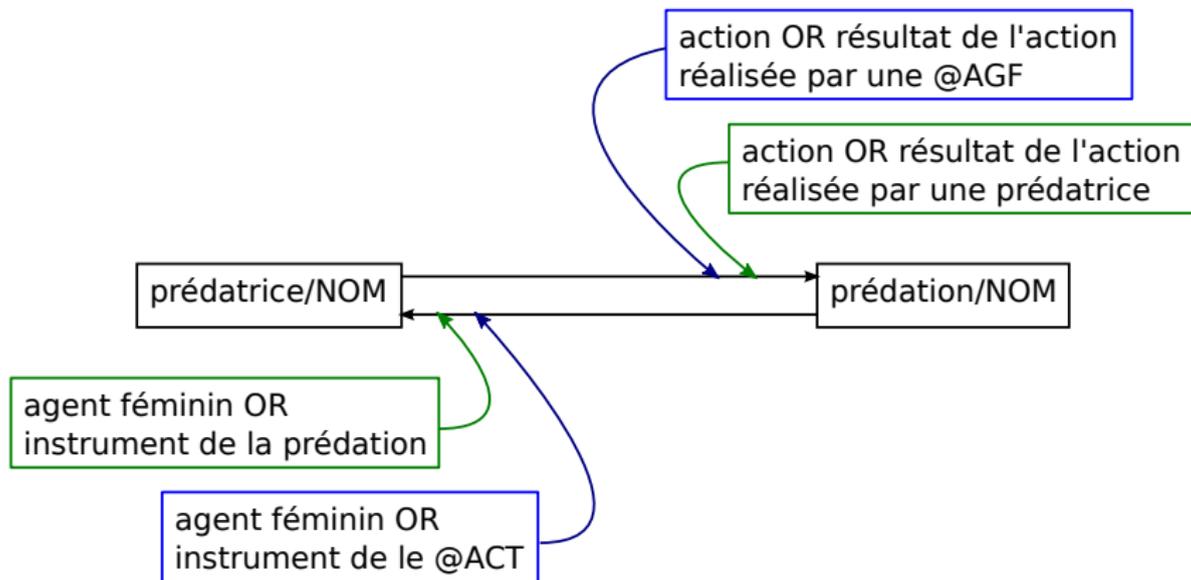
- Les relations sont munies de définitions abstraites obtenues à partir des définitions concrètes.
- Le lexème source de la relation morphologique est remplacée par une variable typée

Définitions orientée et opposée pour *essorer* : *essorage*



Définitions abstraites (suite) (2)

Définitions transversales pour *prédatrice* : *prédation*



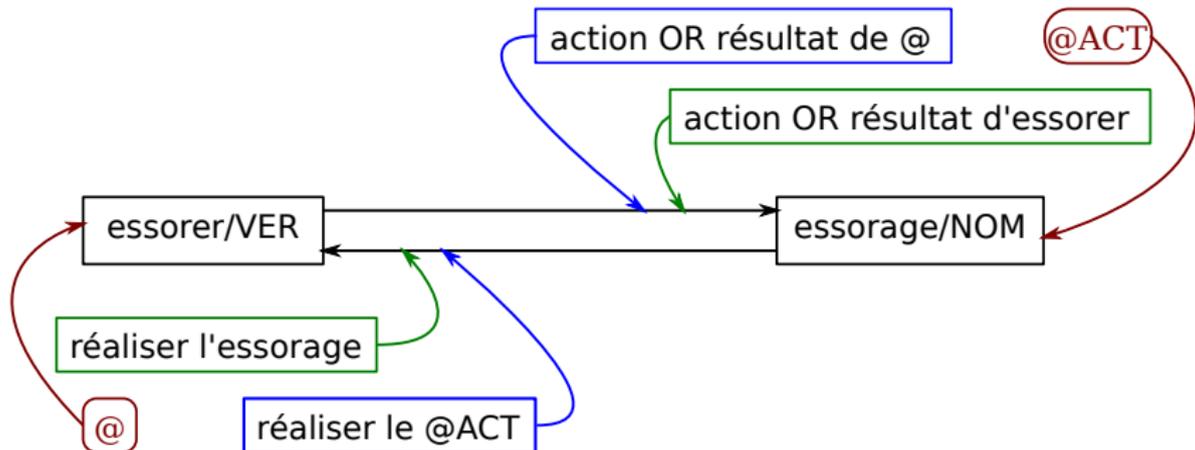
Types morphosémantiques (suite)

Les types morphosémantiques sont attribués par les relations morphologiques à leurs sources et à leurs cibles.

<i>relation</i>	<i>source</i>	<i>cible</i>	<i>exemple</i>
V : (N-age, N-ment, N-ion)	@	@ACT	laver : lavage
V : (N-euse, N-rice)	@	@AGF	danser : danseuse
V : N-eur	@	@AGM	danser : danseur
V : A-if	@	@PROP	penser : pensif
(N-age, N-ment, N-ion) : (N-euse, N-rice)	@ACT	@AGF	prédation : prédatrice
(N-age, N-ment, N-ion) : N-eur	@ACT	@AGM	prédation : prédateur
(N-age, N-ment, N-ion) : A-if	@ACT	@PROP	convulsion : convulsif
(N-euse, N-rice) : N-eur	@AGF	@AGM	danseuse : danseur
(N-euse, N-rice) : A-if	@AGF	@PROP	productrice : productif
N-eur : A-if	@AGM	@PROP	producteur : productif

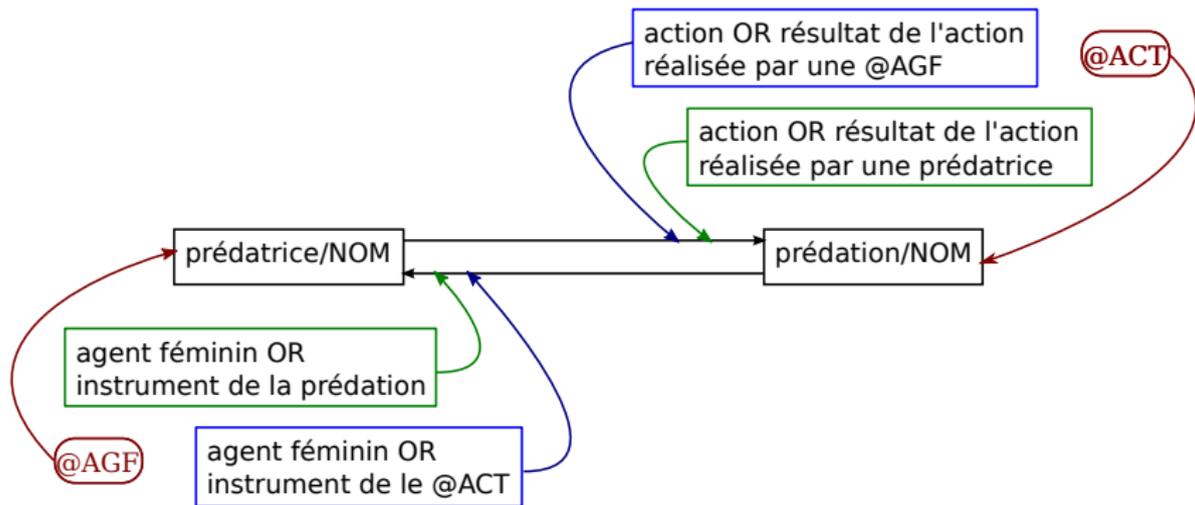
Types morphosémantiques (suite) (2)

Définitions orientée et opposée pour *essorer*: *essorage*



Types morphosémantiques (suite) (3)

Définitions transversales pour *prédatrice* : *prédation*



Le sens construit est cumulatif :

danseuse = agent féminin de *danser*
+ celle qui a pour correspondant masculin *danseur*.

Comment traiter la polysémie des dérivés ?

- Y a-t-il autant de relations qu'il y a de sens possibles ?
- Faut-il représenter les sens différents d'un lexème par des entrées différentes (ex. *production* nom d'action et *production* nom de résultat) ?
- Comment distinguer les sens qui s'ajoutent (AND) et des sens disjoints (OR) ?

Le sens construit est cumulatif :

danseuse = agent féminin de *danser*
+ celle qui a pour correspondant masculin *danseur*.

Comment traiter la polysémie des dérivés ?

- Y a-t-il autant de relations qu'il y a de sens possibles ? → OUI
- Faut-il représenter les sens différents d'un lexème par des entrées différentes (ex. *production* nom d'action et *production* nom de résultat) ? → NON
- Comment distinguer les sens qui s'ajoutent (AND) et des sens disjoints (OR) ? → Pas de distinction.
La sélection et la propagation des sens est déléguée à l'utilisateur.

Plan

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF**
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion

- DériF est un analyseur morphologique
- Il implémente des règles de formation de lexèmes **développées par des linguistes**.
- Les analyses sont contrôlées par des **listes d'exceptions** qui rendent compte des irrégularités accumulées au fil de l'évolution du lexique.
- Les analyses fournissent une *définition* du dérivé relativement à sa base.

enneigement

enneigement, NOM

- DériF est un analyseur morphologique
- Il implémente des règles de formation de lexèmes **développées par des linguistes**.
- Les analyses sont contrôlées par des **listes d'exceptions** qui rendent compte des irrégularités accumulées au fil de l'évolution du lexique.
- Les analyses fournissent une *définition* du dérivé relativement à sa base.

enneigement

enneigement, NOM => [[en [neige NOM] VER] ment NOM]

- DériF est un analyseur morphologique
- Il implémente des règles de formation de lexèmes **développées par des linguistes**.
- Les analyses sont contrôlées par des **listes d'exceptions** qui rendent compte des irrégularités accumulées au fil de l'évolution du lexique.
- Les analyses fournissent une *définition* du dérivé relativement à sa base.

enneigement

enneigement, NOM => [[en [neige NOM] VER] ment NOM]
(enneigement/NOM, enneiger/VER, neige/NOM)

- DériF est un analyseur morphologique
- Il implémente des règles de formation de lexèmes **développées par des linguistes**.
- Les analyses sont contrôlées par des **listes d'exceptions** qui rendent compte des irrégularités accumulées au fil de l'évolution du lexique.
- Les analyses fournissent une *définition* du dérivé relativement à sa base.

enneigement

enneigement, NOM => [[en [neige NOM] VER] ment NOM]
(enneigement/NOM, enneiger/VER, neige/NOM)
:: "(action OR résultat de l'action) de enneiger"

désenneigement

- ✓ désenneigement/NOM : (désenneigement/NOM, désenneiger/VER, enneiger/VER, neige/NOM)
- ✗ désenneigement/NOM : (désenneigement/NOM, enneigement/NOM, enneiger/VER, neige/NOM)

contre-manifestation

- ✓ contre-manifestation/NOM : (contre-manifestation/NOM, manifestation/NOM, manifester/VER)
- ✓ contre-manifestation/NOM : (contre-manifestation/NOM, contre-manifester/VER, manifester/VER)

DériF (3)

schtroumpfement

schtroumpfement/NOM: (*schtroumpfement/NOM*,
schtroumpfer/VER)

pension

✓ *pression/NOM*: (*pression/NOM*, *presser/VER*)

✗ *pension/NOM*: (*pension/NOM*, *penser/VER*)

✓ *pension/NOM*: (*pension/NOM*)

chanteur

chanteur/NOM: (*agent masculin habituel OR auteur masculin
exceptionnel OR instrument*) de chanter

→ *chanter/VER*: [*aspect = dynamique, sous_cat = <NPagent, ...>*]

→ *chanteur/NOM*: [*concret = oui, comptable = oui*]

Contribution de DériF à Démonette

DériF fournit à Démonette :

- des entrées (sommets) ;
- des relations directes (arcs) ;
- des relations indirectes entre mots qui partagent un même radical ;
- des règles de formation de lexèmes pour les relations directes ;
- des définitions concrètes orientées pour les relations directes.

Des modules spécifiques à Démonette permettent d'assigner des types morphosémantiques aux sommets fournis par DériF ...

Typage des lexèmes

```
if (L1.cat = "NOM" and L1.structure = "Xeur"  
    and L2.cat = "VER")  
then L1.type = "@AGM" ; L2.type = "@"
```

Contribution de DériF à Démonette (2)

... et de calculer des définitions opposées.

Définitions concrètes opposées

```
if (L1.cat='NOM' and L1.structure = Xage
    and L2.cat = 'VER'
    and L1.def = "Action OR résultat de l'action de L1.value")
then L2.def = "Réaliser le L2.value"
```

Définitions abstraites opposées

```
if (L1.cat='NOM' and L1.structure = Xage
    and L2.cat = 'VER'
    and L1.def = "Action OR résultat de l'action de L1.value")
then L2.def = "Réaliser le L2.type"
```

Contribution de DériF à Démonette (3)

... et des définitions transversales.

Définitions concrètes transversales

```
if (L1.cat = "NOM" and L1.structure = "Xage"  
    and L2.cat = "NOM" and L2.structure = "Xeur")  
then L1.def = "Action réalisée par L2.value"
```

Définitions abstraites transversales

```
if (L1.cat= "NOM" and L1.structure = "Xage"  
    and L2.cat = "NOM" and L2.structure = "Xeur")  
then L1.def = "Action réalisée par L2.type"
```

Plan

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette**
- 6 Contenu de Démonette
- 7 Conclusion

- Morphonette est un réseau morphologique dérivationnel du français basé sur une conception relationnelle et paradigmatique de la morphologie.
- Les propriétés dérivationnelles des lexèmes sont décrites par leurs familles dérivationnelles et leurs séries dérivationnelles.

modifiable

- famille des mots dont la construction implique *modifier* = *modifier/VER*, *modification/NOM*, *modificateur/NOM*, *modificatif/ADJ*, *modifiant/ADJ*, *modifieur/NOM*, *immodifiable/ADJ*, etc.
- série des mots construits par suffixation en *-able* = *agaçable/ADJ*, *agitabile/ADJ*, *chevauchable/ADJ*, *définissable/ADJ*, *différenciable/ADJ*, *rechargeable/ADJ*, *réconciliable/ADJ*, *soutenable/ADJ*, etc.

Filaments

- Morphonette est composé de filaments.
- Un filament est un triplet $(w, r, s_r(w))$ où :
 - w est une entrée du lexique \mathcal{L} ;
 - $r \in \mathcal{L}$ est un membre de la famille de w ;
 - $s_r(w)$ est l'ensemble des mots qui participent à une analogie avec w et r . $s_r(w) = \{v \in \mathcal{L} / \exists u \in \mathcal{L}, w : r = v : u\}$
- $a : b = c : d$ ssi a est à b ce que c est à d

Filament de *modifiable* relativement à *modificateur*

(*modifiable/ADJ, modificateur/NOM,*
{*amplifiable/ADJ, glorifiable/ADJ, identifiable/ADJ, justifiable/ADJ,*
clarifiable/ADJ, mystifiable/ADJ, rectifiable/ADJ, sanctifiable/ADJ,
simplifiable/ADJ, spécifiable/ADJ, unifiable/ADJ, vérifiable/ADJ})

modifiable:modificateur=amplifiable:amplificateur
modifiable:modificateur= identifier:identificateur
modifiable:modificateur= simplifier:simplificateur

Filaments (2)

- Un mot a autant de filaments qu'il y a de membres dans sa famille dérivationnelle.
- Chaque filament décrit une propriété du mot
- Certains filaments se recouvrent largement, mais d'autres décrivent des propriétés différentes :

(travailleur, travailler, {ravageur, cisailleur, ...}) :
travailleur est un nom en *-eur* dérivé d'un verbe.

(travailleur, travailleuse, {deuilleur, volaille, ravageur, cisailleur, ...}) :
travailleur est un nom masculin qui a un équivalent féminin en /ɔz/.

Morphonette (2)

Morphonette est construit automatiquement en utilisant :

- La mesure de similarité morphologique *Proxinette*;
- L'analogie formelle calculées à partir de signatures.

Morphonette contient :

29 310 entrées ;

96 107 filaments ;

96 107 relations entre membres de la même famille
dérivationnelle ;

1 106 098 relations entre membres des la même série dérivationnelle.

- Morphonette alimente Démonette en relations morphologiques directes, inverses et indirectes.
- Les relations sont obtenues à partir des analogies qui supportent les filaments.
- Morphonette ne permet pas de savoir si une relation est directe, inverse ou indirecte. Il ne fournit pas l'orientation des relations.

Plan

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette**
- 7 Conclusion

Contenu de Démonette

Démonette contient :

	couples
DériF	21 556
Morphonette	9 880
Démonette	23 060

Répartition par type de relation :

relation type	$D \setminus M$	$D \cap M$	$M \setminus D$
direct	8 802	1 145	0
indirect	4 378	6 872	1 504

Contenu de Démonette (2)

Distribution des relations indirectes :

Xteur	158						
Xeuse	3,938	134					
Xrice	338	978	66				
Xment	2,226	98	1,390	40			
Xage	2,946	80	1,932	11	1,656		
Xion	926	1,544	208	976	320	240	
Xif	310	448	36	304	26	24	918
	Xeur	Xteur	Xeuse	Xrice	Xment	Xage	Xion

Les relations indirectes reflètent la préférence de certains suffixes pour les mêmes types de radicaux :

- *-ion*, *-if* et *-rice* préfèrent les radicaux **savants**
- *-age*, *-ment* et *-euse* préfèrent les radicaux **natifs**.
- *-eur* présente les deux comportements :
 - les lexèmes en *-ateur* ont des radicaux savants ;
 - les radicaux des autres dérivés en *-eur* sont natifs.

Plan

- 1 Introduction
- 2 Relations morphologiques
- 3 Descriptions morphosémantiques
- 4 Alimenter Démonette à partir de DériF
- 5 Alimenter Démonette à partir de Morphonette
- 6 Contenu de Démonette
- 7 Conclusion**

- Démonette apporte un grand nombre d'innovations par rapport aux bases de données morphologiques existantes.
 - Relations inverses et indirectes
 - Définitions opposées et transversales
 - Typage des lexèmes
 - Définitions abstraites
- Les informations décrites sont fortement décomposées et très redondantes.
 - Cette caractéristique découle de l'intégration dans Démonette d'informations provenant de deux origines dont la conception est radicalement opposée : DériF et Morphonette.
 - La décomposition des informations permet l'alignement d'informations d'origines différentes.
 - La redondance apporte une grande souplesse (et robustesse) permettant d'y intégrer des descriptions incomplètes ou partielles.
- L'origine de toutes les informations présentes dans Démonette y est indiquée explicitement.

Étendre et compléter Démonette

- Démonette ne contient qu'un petit fragment du lexique du français (7 suffixations ; 15% de TLFnome).
 - Inclure dans Démonette toutes les analyses de DériF et tous les filaments de Morphonette
 - Inclure dans Démonette des bases existantes comme Verbaction ou le *Dictionnaire des mots construits* de (Roché, 2011)
- Inclure les descriptions morphologiques présentes ou acquises de dictionnaires électroniques comme GLAWI (la version XML du Wicktionnaire).
- Compléter les couples par des descriptions phonémiques (GLÀFF).

Décrire les paradigmes dérivationnels

- Compléter l'inventaire des types morphosémantiques
- Caractériser les relations indirectes « utiles » qui seront décrites dans Démonette.
 - Une relation indirecte utile est une relation pour laquelle les locuteurs peuvent formuler une définition transversale

Roché, M. (2011). Quelle morphologie? See Roché et al. (2011), pp. 15–39.

Roché, M., G. Boyé, N. Hathout, S. Lignon, and M. Plénat (2011). *Des unités morphologiques au lexique*. Paris: Hermès Science-Lavoisier.