ParaDis and Démonette
From Theory to Resources for Derivational Paradigms

Fiammetta Namer

Université de Lorraine & ATILF

DeriMo 2019
Prague, 19-20 September 2019

Outline

Introduction

- Démonette: Large-scale derivational database for French
  - Project funded by the French National Research Agency (2018-2021)
  - At the end, at least annotated 366.000 entries
- Source of the data populating the database: a set of derivational lexicons (reliable content)
- Data are reanalysed, annotations are completed, new entries are added.
- Démonette's architecture and content are the adaptation of theoretical principles in derivational morphology.
  - Implements ParaDis: a model of derivational morphology where lexemes, units of analysis, are grouped into families that are organized into paradigms.
- why chosing this theoretical approach?

Outline

At the beginning

- Morpheme-based frameworks.
  - minimal unit of form and meaning
  - syntax-like word structure rewrite rules (e.g. concatenation).
- Production of derivational resources and tools
  - CELEX (Baayen et al., 1995), DerIvaTario (Talamo et al., 2016), CroDeriV (Šojat et al., 2014), Morphological Treebank (Steiner & Ruppenhofer, 2018), WFL (1st version) (Litta et al. 2016)

    (Kyjánek, 2018)

- advantages: simplicity, economy
- drawbacks: whenever there is no one-to-one form-meaning correspondance (zero morpheme, empty morph, polysemous affixes, ..., non-concatenative morphology...)

Improvements

- Two sets of theoretical devices in derivational morphology
  - ternary structure of the lexeme (and lexeme formation rules).(Haspelmath & Sims 2002; Plag 2003)
  - paradigmatic organization for derivation. (Bauer 1997; Blevins 2016; Bochner 1993; Booij 2010; Štekauer 2014; van Marle 1995)
- lexeme- or paradigm-based tools and resources of WF relations: DerivBase (Zeller et al., 2013), DeriNet (Vidra et al. 2019), WFL last version (Litta et al. 2019), DériF (Namer 2013), Morphonette (Hathout 2011).

Lexeme (1)

- the abstract representation of an inflectional paradigm, in the form of a three-dimensional unit: form, part-of-speech, semantic content.

$$\begin{bmatrix} /\text{ri:d}/ \\ \text{V} \\ \text{'read'} \end{bmatrix} \rightarrow \begin{bmatrix} /\text{ri:də}/ \\ \text{N} \\ \text{'person who reads'} \end{bmatrix}$$

- Rules are processes distributed across the three dimensions.

Lexeme (2)

- Each field (formal vs semantic) has an autonomous behaviour in derivation.
  - Several possible formal means for derived words of the same semantic type.
  - (conversely, several possible semantic contents for word structures sharing the same exponent (e.g. Czech *-ka*))

$$\begin{bmatrix} /\text{ˈgʌvənmənt}/ \\ \text{N} \\ \text{'government'} \end{bmatrix} \rightarrow \begin{bmatrix} /\text{ˌgʌvənˈmənt}l/ \\ \text{A} \\ \text{'of the government'} \end{bmatrix}$$

$$\begin{bmatrix} /\text{ˈætəm}/ \\ \text{N} \\ \text{'atom'} \end{bmatrix} \rightarrow \begin{bmatrix} /\text{əˈtɒmɪk}/ \\ \text{A} \\ \text{'of the atom'} \end{bmatrix}$$

Lexeme (3)

- The value at each level results from the application of constraints specific to that level.

$$
\begin{bmatrix} /sit\textscr\~{ɔ}/ \\ Nm \\ \text{'lemon'} \end{bmatrix} \rightarrow \begin{bmatrix} /sit\textscr ɔnje/ \\ Nm \\ \text{'plant that produces lemons'} \end{bmatrix}
$$

Lexeme: limitations

- Lexeme formation rules are binary oriented devices
  - And it may happen that in related wordpairs each word is both (or neither) the base and (nor) the derivative of the other: cross-formation, (Booij & Masini, 2015)

$$
\begin{bmatrix}
\text{/'fæʃɪzəm/} \\
\text{N} \\
\text{'fascism'}
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\text{/'fæʃɪst/} \\
\text{Nm} \\
\text{'person supporting fascism'}
\end{bmatrix}
$$

Lexeme: limitations

- Lexeme formation rules are binary oriented devices
  - And it may happen that in related wordpairs each word is both (or neither) the base and (nor) the derivative of the other: cross-formation, (Booij & Masini, 2015)

$$
\begin{bmatrix} /\text{'fæʃɪzəm}/ \\ \text{N} \\ \text{'ideology supported by fascists'} \end{bmatrix} \leftarrow \begin{bmatrix} /\text{'fæʃɪst}/ \\ \text{Nm} \\ \text{'fascist'} \end{bmatrix}
$$

Lexeme: limitations

- Lexeme formation rules are binary oriented devices
  - And it may happen that the formal base is semantically derived from the formally derived word: back-formation, (Becker, 1993)

$$
\begin{bmatrix} \text{/'vɪvə͵sekt/} \\ \text{V} \\ \text{'practice vivisection'} \end{bmatrix} \; ? \; \begin{bmatrix} \text{/vɪvɪ'sek∫n/} \\ \text{Nm} \\ \text{'vivisection'} \end{bmatrix}
$$

Lexeme: limitations

- Lexeme formation rules are binary oriented devices
  - And it may happen that the formal base is semantically derived from the formally derived word: back-formation, (Becker, 1993)

$$\begin{bmatrix} /\text{ˈvɪvəˌsekt}/ \\ \text{V} \\ \text{'practice vivisection'} \end{bmatrix} \;?\; \begin{bmatrix} /\text{vɪvɪˈsekʃn}/ \\ \text{Nm} \\ \text{'vivisection'} \end{bmatrix}$$

Lexeme: limitations

- Lexeme formation rules are binary oriented devices

- Paradigm-based approaches to derivation overcome orientation and binarity issues
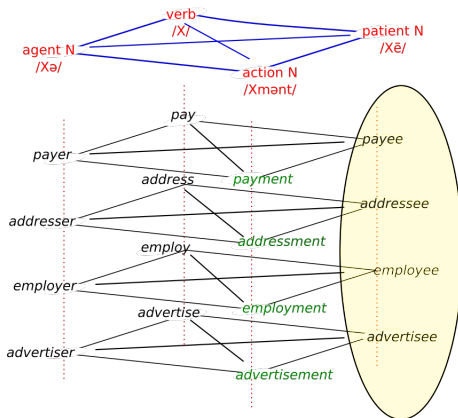
Derivational paradigms: some definitions

- Derivational family: structured set of lexemes, the form and meaning of which depend on each others
- Verb derived set of lexemes
- All of them are (in)directly connected
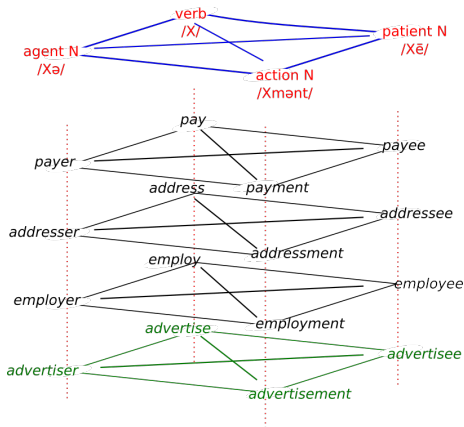
Derivational paradigms: some definitions

- Derivational series: set of aligned lexemes [Bonami & Strnadová, 2018]

- aligned lexemes (same column) : same formal and semantic contrast relations.

- For instance : $N_{action} \leftrightarrow N_{patient}$

Derivational paradigms: some definitions

- Derivational paradigm: arrangement of families whose members have multiple correlations
- Can be represented by the network connecting all the derivational series
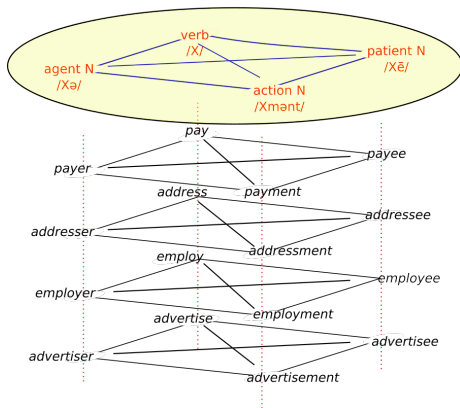
Derivational paradigms: summary

- Paradigm-based frameworks for Word Formation (e.g. Bochner (1993)) are well-equipped for:
  - processing cross- and backformations (vs oriented rules)
  - taking into account of word formation at family level (vs binary rules)
  - (as well as regular rules connecting a derived$_W$ to its base word$_W$)
- but . . .

Outline

Limitations

- to form a paradigm, members of aligned families must have consistent, regular form-meaning relations with each others
- what happens, when formal regularities diverge from semantic regularities ?
    - The case of the so-called "parasynthetic derivation" (Hathout & Namer, 2018a) or "prefix/suffix rule conflation" (Stump, 2019)

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N 'X'	$X\text{suf}_A$ 'of X'	$anti X\text{suf}_A$ 'opposed to X'	$pro X\text{suf}_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
    - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N 'X'	$Xsuf_A$ 'of X'	$anti$Xsuf$_A$ 'opposed to X'	proXsuf$_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N	$Xsuf_A$	$antiXsuf_A$	$proXsuf_A$
'X'	'of X'	'opposed to X'	'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N 'X'	$Xsuf_A$ 'of X'	*anti*$Xsuf_A$ 'opposed to X'	*pro*$Xsuf_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

| X_N | $Xsuf_A$ | *anti*$Xsuf_A$ | *pro*$Xsuf_A$ |
'X'	'of X'	'opposed to X'	'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N 'X'	$X\text{suf}_A$ 'of X'	$anti X\text{suf}_A$ 'opposed to X'	$pro X\text{suf}_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (1)

- Alignment of derivational families
- Semantic relation between an entity and what fights/promotes it. But no full formal interpredictability: fluctuating (and useless?) suffix value
  - (antigovernment/progovernment, antiallergy/proallergy, anticoagulation/procoagulation, antiinfection/proinfection)
- Actually, this value equals that of the relation adjective. Here, formal, not semantic motivation.
- elsewhere in the families: regular paradigmatic relations

X_N 'X'	Xsuf$_A$ 'of X'	*anti*Xsuf$_A$ 'opposed to X'	*pro*Xsuf$_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (2)

- Everything happens as if *anti*Xsuf was formally derived from Xsuf, and semantically from X
- What we need is to have access to the derivational family, but also to be able to retrieve meaning and form values independently of each other.
  - (idem with *pro*Xsuf)

X_N 'X'	Xsuf$_A$ 'of X'	*anti*Xsuf$_A$ 'opposed to X'	*pro*Xsuf$_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Opposition/Promoting adjectives in English (2)

- Everything happens as if *anti*Xsuf was formally derived from Xsuf, and semantically from X
- What we need is to have access to the derivational family, but also to be able to retrieve meaning and form values independently of each other.
  - (idem with *pro*Xsuf)

X_N 'X'	Xsuf$_A$ 'of X'	*anti*Xsuf$_A$ 'opposed to X'	*pro*Xsuf$_A$ 'promoting X'
government	governmental	antigovernmental	progovernmental
allergy	allergic	antiallergic	proallergic
coagulation	coagulative	anticoagulative	procoagulative
infection	infectious	antiinfectious	proinfectious

Other languages, other prefixation patterns

- Similar issue arises with other derivation patterns: the output form requires the knowledge of one member of the derivative's family, whereas its semantic content requires to access another member (Hathout & Namer, 2016)

- This is true in a large number of European languages

mono-/poly- & pluri- Quantitative opposition in French

- In French, *mono*Xsuf$_A$ describes smth containing `"one X"`, opposed to *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ both describing smth containing `"several X"`.
- Here again, *mono/poly/pluri*Xsuf seem formally derived from Xsuf, and semantically from X
  - Moreover, *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ share the same semantic content.
  - overabundance (Thornton, 2012): another kind of form-meaning mismatch

X$_N$ 'X'	Xsuf$_A$ 'of X'	*mono*Xsuf$_A$ 'with one X'	*poly*Xsuf$_A$	*pluri*Xsuf$_A$ 'with several X'
cellule	cellulaire	monocellulaire	polycellulaire	pluricellulaire
atome	atomique	monoatomique	polyatomique	pluriatomique
clone	clonal	monoclonal	polyclonal	pluriclonal
os	osseux	monoosseux	polyosseux	pluriosseux

mono-/poly- & *pluri-* Quantitative opposition in French

- In French, *mono*Xsuf$_A$ describes smth containing "one X", opposed to *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ both describing smth containing "several X".
- Here again, *mono/poly/pluri*Xsuf seem formally derived from Xsuf, and semantically from X
  - Moreover, *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ share the same semantic content.
  - overabundance (Thornton, 2012): another kind of form-meaning mismatch

X$_N$ 'X'	Xsuf$_A$ 'of X'	*mono*Xsuf$_A$ 'with one X'	*poly*Xsuf$_A$	*pluri*Xsuf$_A$
			'with several X'	
cellule	cellulaire	monocellulaire	polycellulaire	pluricellulaire
atome	atomique	monoatomique	polyatomique	pluriatomique
clone	clonal	monoclonal	polyclonal	pluriclonal
os	osseux	monoosseux	polyosseux	pluriosseux

mono-/poly- & pluri- Quantitative opposition in French

- In French, *mono*Xsuf$_A$ describes smth containing "one X", opposed to *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ both describing smth containing "several X".
- Here again, *mono/poly/pluri*Xsuf seem formally derived from Xsuf, and semantically from X
  - Moreover, *poly*Xsuf$_A$ and *pluri*Xsuf$_A$ share the same semantic content.
  - overabundance (Thornton, 2012): another kind of form-meaning mismatch

X$_N$	Xsuf$_A$	*mono*Xsuf$_A$	*poly*Xsuf$_A$	*pluri*Xsuf$_A$
'X'	'of X'	'with one X'	'with several X'	
cellule	cellulaire	monocellulaire	polycellulaire	pluricellulaire
atome	atomique	monoatomique	polyatomique	pluriatomique
clone	clonal	monoclonal	polyclonal	pluriclonal
os	osseux	monoosseux	polyosseux	pluriosseux

Observations

- In each family, whatever prefix and language, form and meaning is predictable for prefixed adjectives
  - formal connection between the two adjectives.
  - semantic prediction of the content of the prefixed adjective, from that of the noun
- A paradigmatic description makes it possible to account for it in a natural way
  - provided that the semantic and formal levels are separated

- classical paradigmatic analysis is too rigid
  - when the relation network between forms does not coincide with the relation network between semantic values
- a new framework must be considered

Outline

Overview (1)

- ParaDis: "Paradigms vs Discrepancies" (Hathout & Namer, 2018b)
- Bring out paradigmatic regularities where they are blurred by meaning-form mismatches
- Combine advantages:
  - lexeme tri-dimensionality at a paradigmatic organization scale

- Hypothesis: the way derivational paradigms work is a sort of spatial projection of the lexeme's ternary organization
  - A paradigmatic system is a 3-level organization
  - The organizational principles of classical paradigmatic derivation models are brought to these three levels
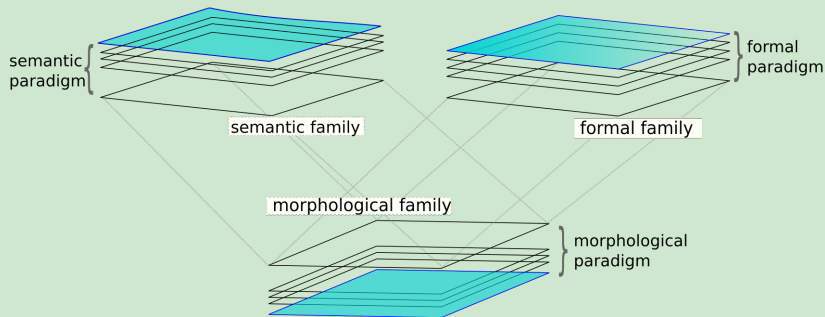
Overview (2)

- The system is made of a formal, a semantic (and a part-of-speech) paradigms. Their meeting point is the morphological paradigm they correspond to.
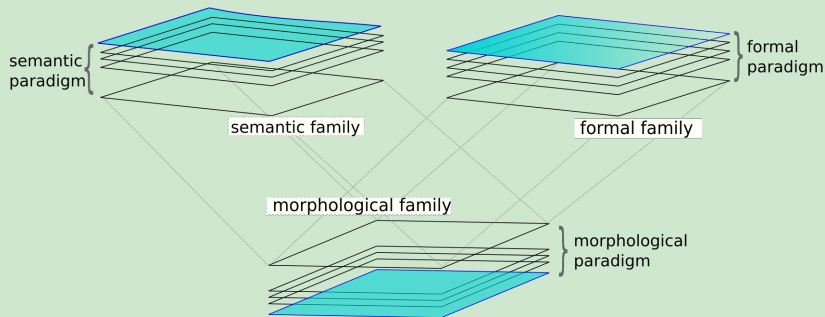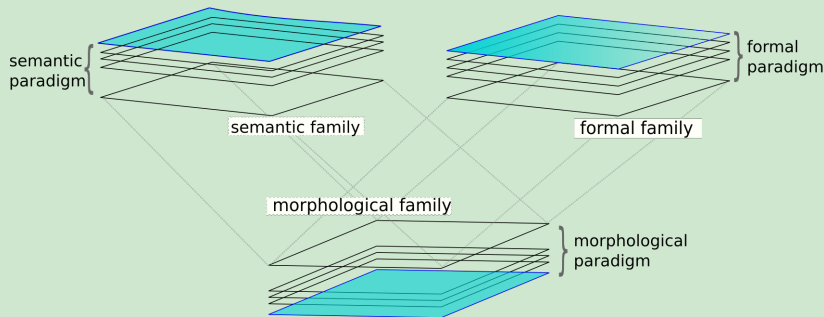
Overview (2)

- Each paradigm is an alignment of families, that is connected networks of items with inter-predictable properties.
- Formal families connect forms, semantic families connect meanings. Morphological families connect lexemes.
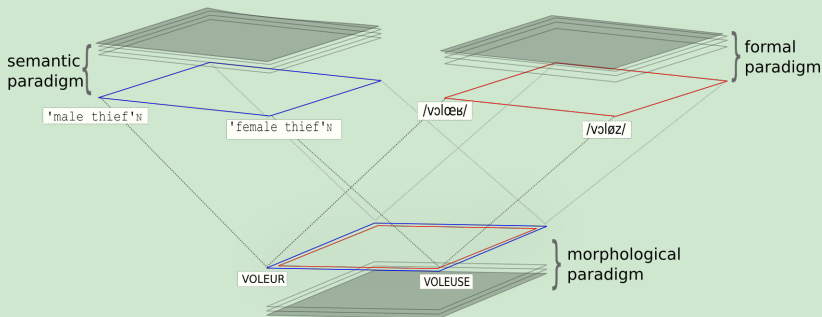
Overview (2)

- Morphological paradigm: the abstract outcome of the independent mechanisms of the formal paradigm and the semantic paradigm
  - Formal and semantic paradigms have autonomous behaviours and have no direct relation with each other
  - Constraints on relations apply locally: formal constraints in the formal paradigm, semantic constraints in the semantic one
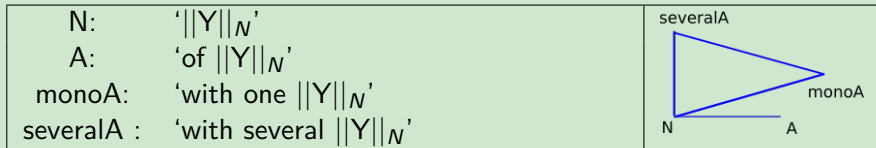
Overview (3)

- When meaning-form relations are regular, paradigms are isomorphic.
- In all the families of the morphological paradigm, any relation between lexemes is regular (both semantically and formally motivated).
  - Gender variation interpredictable for French human agent nouns:
    - ★ Formal /Xœʁ/ ↔ /Xøz/ alternation
    - ★ Semantic male-female correlation between human beings

Mismatched families: *mono*-, *pluri*-, *poly*- prefixed adjs in French (1)

sem netwk	'X'$_N$	'of X'$_N$	'with one X'$_A$	'with several X'$_A$	
form netwk	/X/	/X$_{\text{SUF}}$/	/monoX$_{\text{SUF}}$/	/poliX$_{\text{SUF}}$/	/plyʁiX$_{\text{SUF}}$/
	cellule	cellulaire	monocellulaire	polycellulaire	pluricellulaire

- The network of semantic series ... (i.e. semantic paradigm)
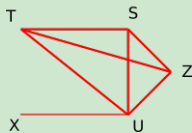- does not match with the network of formal series (i.e. formal paradigm)

N: '||Y||$_N$'
A: 'of ||Y||$_N$'
monoA: 'with one ||Y||$_N$'
severalA :   'with several ||Y||$_N$'

- 4-edge connected graph

Mismatched families: *mono-*, *pluri-*, *poly-* prefixed adjs in French (1)

sem netwk	'X'$_N$	'of X'$_N$	'with one X'$_A$	'with several X'$_A$	
form netwk	/X/	/X$_{SUF}$/	/monoX$_{SUF}$/	/poliX$_{SUF}$/	/plyʁiX$_{SUF}$/
	cellule	cellulaire	monocellulaire	polycellulaire	pluricellulaire

- The network of semantic series ... (i.e. semantic paradigm)
- does not match with the network of formal series (i.e. formal paradigm)

X: $\exists$ W, /X/ =/W/
U: $\exists$ W, /U/= /Wɛʁ/ $\vee$ /Wik/ $\vee$ /Wal/ $\vee$ /Wø/
Z: $\exists$ W, /Z/=/monoV/ $\wedge$ $\exists$ W, (/V/ = /Wɛʁ/ $\vee$ /Wik/ $\vee$ /Wal/ $\vee$ /Wø/)
S: $\exists$ W, /S/=/poliV/ $\wedge$ $\exists$ W, (/V/ = /Wɛʁ/ $\vee$ /Wik/ $\vee$ /Wal/ $\vee$ /Wø/)
T: $\exists$ W, /T/=/plyʁiV/ $\wedge$ $\exists$ W, (/V/ = /Wɛʁ/ $\vee$ /Wik/ $\vee$ /Wal/ $\vee$ /Wø/)


- 5-edge connected graph

Mismatched families: *mono-*, *pluri-*, *poly-* prefixed adjs in French (1)

- The difference between formal and semantic paradigms captures form and meaning mismatch. Nodes and relations in morphological families inherit semantic and formal properties coming from the corresponding nodes and relations in the formal and semantic families.
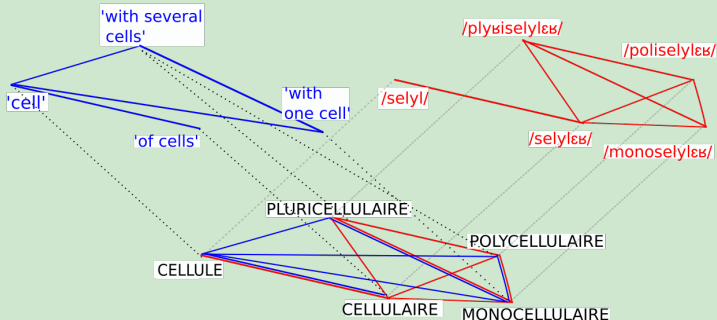
Mismatched families: *mono-*, *pluri-*, *poly-* prefixed adjs in French (1)

- Morphological paradigm: the difference in cohesion within intra-family relations are indicators of discrepancies. Regular relations: double line (double inheritance); formal motivation only (in red) or semantic motivation only (in blue): single line.
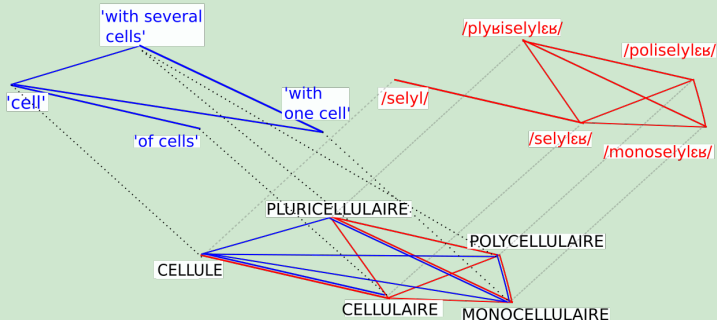
ParaDis: summary

- Framework suitable for:
  - form meaning discrepancies that cannot be solved but at family level: paradigmatic regularities are revealed by the separation of formal, semantic (and part-of-speech) levels of description, expressed each by autonomously structured networks
  - systematic synonymy (overabundance) between two derivational patterns: the semantic paradigm is a network of semantic series with fewer vertices than the formal paradigm
  - such issues ('over-marked' prefixation processes) are frequently observed across languages. Moreover meaning-form asymetry occurs also with other kinds of affixation patterns.
- ParaDis also appropriate for regular paradigms and classical binary oriented base/derived word relations (perfect match between the three paradigms)

Resource implementing ParaDis's principles?

Outline

Translating theoretical principles into a derivational database

- Démonette: derivational database for French. Its goal: represent in a uniform way formal, morphological and semantic properties of morphological relations and the words involved in these relations, regardless of their regularity / canonicity
  - ▶ To achieve this, Démonette implements the three-level paradigmatic system of ParaDis
  - ▶ Démonette: an improved version of a previous prototype: `https://demonette.atilf.fr/`

What do we need?

- Define simple and robust architecture, with an (extendable) feature set capable of covering the complex lexicon (anticipate the description of unexpected cases)
- Bring out (morphological, formal, semantic) families: describe indirect relations, rank relations according to their type
- Deal with infringement to canonicity: meaning-form mismatch, overabundance, affix competition, lexical gap, suppletion, polysemy…
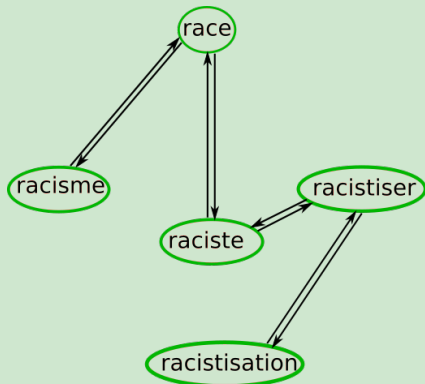
General structure

- An entry: represents a relation between two words of a derivational family
  - if (W_1, W_2) is a relation, then (W_2, W_1) is a relation
  - Each piece of information is a controlled feature/value pairs
- Description grouped in semantic, formal, morphological, phonological (Namer & al. 2017), frequency fields
  - phonological descriptions: allomorphy, suppletion etc.

	Feature	Value	
Entry	$L_1 \rightarrow L_2$	$danser_V$	$\rightarrow$ $danseur_{Nm}$
Lexeme:	Semantic Type	Pred	Person
	Inflect. paradigm	dɑ̃s, dɑ̃sɛ, dɑ̃sɔ̃ ...	dɑ̃sœʁ
Relation:	Structure	ascend2descend, simple	
	Pattern Altern.	X	Xeur
		–	suf
	Semantics	"un danseur danse"	

Typology of Démonette's relations

- Entries may describe classical base/derivative relations

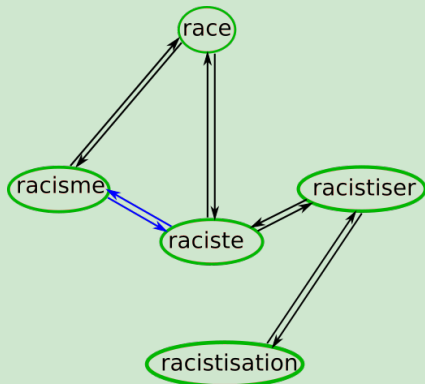| W_1/W_2	Orientation	Complexity	Pattern Alternation
race/raciste	as2de	simple	X/Xiste
raciste/race	de2as	simple	Xiste/X
raciste/racisme	indirect	simple	Xiste/Xisme
racisme/raciste	indirect	simple	Xisme/Xiste
fasciste/fascisme	indirect	simple	Xiste/Xisme
fascisme/fasciste	indirect	simple	Xisme/Xiste
race/racistiser	as2de	complex	X/Xistiser
racistiser/race	de2as	complex	Xistiser/X
racisme/ racistisation	indirect	complex	Xisme/ Xistisation
racistisation/ racisme	indirect	complex	Xistisation /Xisme

Typology of Démonette's relations

- But also relations between siblings (indirect relations, paradigmatic derivation)
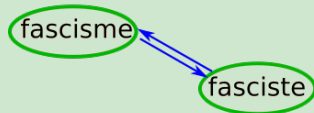
W_1/W_2	Orien-tation	Comple-xity	Pattern Alternation
race/raciste	as2de	simple	X/Xiste
raciste/race	de2as	simple	Xiste/X
raciste/racisme	indirect	simple	Xiste/Xisme
racisme/raciste	indirect	simple	Xisme/Xiste
fasciste/fascisme	indirect	simple	Xiste/Xisme
fascisme/fasciste	indirect	simple	Xisme/Xiste
race/racistiser	as2de	complex	X/Xistiser
racistiser/race	de2as	complex	Xistiser/X
racisme/ racistisation	indirect	complex	Xisme/ Xistisation
racistisation/ racisme	indirect	complex	Xistisation /Xisme

Typology of Démonette's relations

- Regardless of the existence of a common base word (uncomplete families/paradigms)
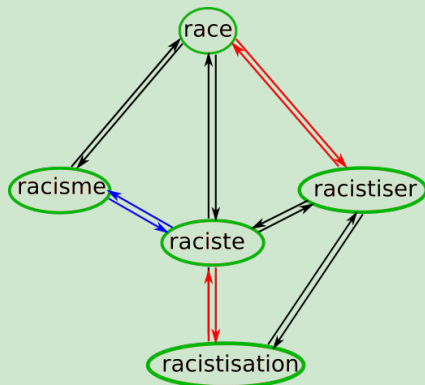
W$_1$/W$_2$	Orien- tation	Comple- xity	Pattern Alternation
race/raciste	as2de	simple	X/Xiste
raciste/race	de2as	simple	Xiste/X
raciste/racisme	indirect	simple	Xiste/Xisme
racisme/raciste	indirect	simple	Xisme/Xiste
fasciste/fascisme	indirect	simple	Xiste/Xisme
fascisme/fasciste	indirect	simple	Xisme/Xiste
race/racistiser	as2de	complex	X/Xistiser
racistiser/race	de2as	complex	Xistiser/X
racisme/ racistisation	indirect	complex	Xisme/ Xistisation
racistisation/ racisme	indirect	complex	Xistisation /Xisme

Typology of Démonette's relations

- But also relations between distant members in a family

W$_1$/W$_2$	Orientation	Complexity	Pattern Alternation
race/raciste	as2de	simple	X/Xiste
raciste/race	de2as	simple	Xiste/X
raciste/racisme	indirect	simple	Xiste/Xisme
racisme/raciste	indirect	simple	Xisme/Xiste
fasciste/fascisme	indirect	simple	Xiste/Xisme
fascisme/fasciste	indirect	simple	Xisme/Xiste
race/racistiser	as2de	complex	X/Xistiser
racistiser/race	de2as	complex	Xistiser/X
racisme/ racistisation	indirect	complex	Xisme/ Xistisation
racistisation/ racisme	indirect	complex	Xistisation /Xisme

Typology of Démonette's relations

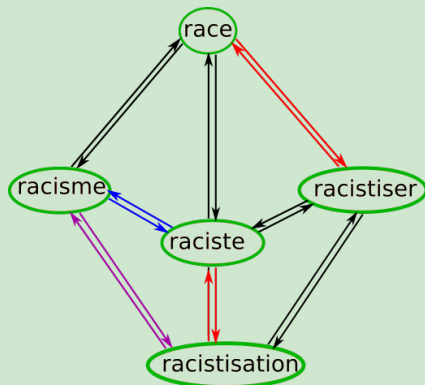- But also relations between distant members in a family

W_1/W_2	Orientation	Complexity	Pattern Alternation
race/raciste	as2de	simple	X/Xiste
raciste/race	de2as	simple	Xiste/X
raciste/racisme	indirect	simple	Xiste/Xisme
racisme/raciste	indirect	simple	Xisme/Xiste
fasciste/fascisme	indirect	simple	Xiste/Xisme
fascisme/fasciste	indirect	simple	Xisme/Xiste
race/racistiser	as2de	complex	X/Xistiser
racistiser/race	de2as	complex	Xistiser/X
racisme/ racistisation	indirect	complex	Xisme/ Xistisation
racistisation/ racisme	indirect	complex	Xistisation /Xisme

Semantic features

- Lexemes are assigned one of the 25 ontological classes belonging to the *Unique Beginners* in WordNet typology
  - Nouns: `Animal, Person, Plant, Artifact, Act, Event, Attribute, Feeling`, …
  - Adjectives: `Modifier`
  - Verbs: `Predicate`

W_1/W_2	$race_N/raciste_N$	$raciste_N/racisme_N$	$raciste_N/racistiser_V$
Class$_1$/Class$_2$	Entity/Person	Person/Cognition	Person/Predicate
Semantic Relation	entity-entity belief	entity-entity partisanship	entity-situation similative
Cross-definition	"A *raciste* promotes the *race*"	"A *raciste* supports *racisme*"	"*Racistiser* smn is to call them a *raciste*"

Semantic features

- The semantic value of the relation is the concatenation of the semantic class hypernym (entity or situation) of the connected lexemes

- Its semantic subtype is derived from this value combined with the lexemes' ontological class, and according to their formal patterns and related information

- Classes + relations $\rightarrow$ cross-definition of each related lexeme

W_1/W_2	$race_N/raciste_N$	$raciste_N/racisme_N$	$raciste_N/racistiser_V$
$Class_1/Class_2$	Entity/Person	Person/Cognition	Person/Predicate
Semantic Relation	entity-entity	entity-entity	entity-situation
	belief	partisanship	similative
Cross-definition	"A *raciste* promotes the *race*"	"A *raciste* supports *racisme*"	"*Racistiser* smn is to call them a *raciste*"

Semantic features

- The semantic value of the relation is the concatenation of the semantic class hypernym (entity or situation) of the connected lexemes

- Its semantic subtype is derived from this value combined with the lexemes' ontological class, and according to their formal patterns and related information

- Classes + relations → cross-definition of each related lexeme

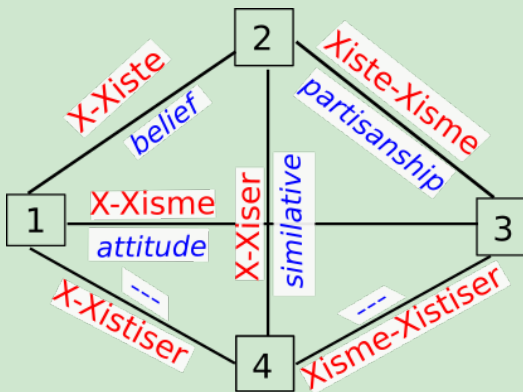| W_1/W_2	$race_N/raciste_N$	$raciste_N/racisme_N$	$raciste_N/racistiser_V$
$Class_1/Class_2$	Entity/Person	Person/Cognition	Person/Predicate
Semantic Relation	entity-entity belief	entity-entity partisanship	entity-situation similative
Cross-definition	"A *raciste* promotes the *race*"	"A *raciste* supports *racisme*"	"*Racistiser* smn is to call them a *raciste*"

Semantic features

- The semantic value of the relation is the concatenation of the semantic class hypernym (entity or situation) of the connected lexemes

- Its semantic subtype is derived from this value combined with the lexemes' ontological class, and according to their formal patterns and related information

- Classes + relations → cross-definition of each related lexeme

W_1/W_2	$race_N/raciste_N$	$raciste_N/racisme_N$	$raciste_N/racistiser_V$
$Class_1/Class_2$	Entity/Person	Person/Cognition	Person/Predicate
Semantic Relation	entity-entity belief	entity-entity partisanship	entity-situation similative
Cross-definition	"A *raciste* promotes the *race*"	"A *raciste* supports *racisme*"	"*Racistiser* smn is to call them a *raciste*"

Grouping families into paradigms

- Retrieving families from relations: displayed by the graph obtained by joining two-by-two the edges (relations) with a node (lexeme) in common
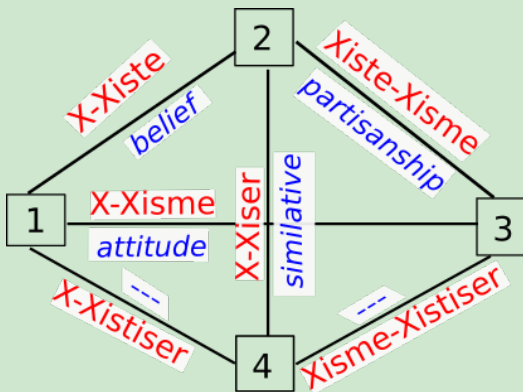

(1,2)	race	raciste
(1,3)	race	racisme
(1,4)	race	racistiser
(2,3)	raciste	racisme
(2,4)	raciste	racistiser
(3,4)	racisme	racistiser
. . .	. . .	. . .

- Retrieving paradigms from families: aligning relations belonging to the same morphological series (sharing the same pattern and semantic value)
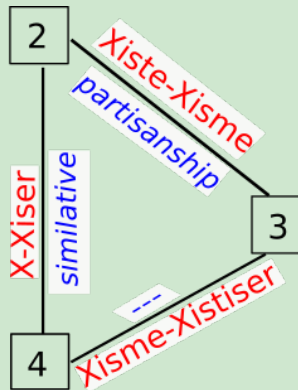
Grouping families into paradigms

- Natural detection of sub-families and sub-paradigms

(1,2)	—	—
(1,3)	—	—
(1,4)	—	—
(2,3)	fasciste	fascisme
(2,4)	fasciste	fascistiser
(3,4)	fascisme	fascistiser
. . .	. . .	. . .

Grouping families into paradigms

- Natural detection of sub-families and sub-paradigms

(1,2)	déclin	décliniste
(1,3)	déclin	déclinisme
(1,4)	—	—
(2,3)	décliniste	déclinisme
(2,4)	—	—
(3,4)	—	—
. . .	. . .	. . .

X-Xiste
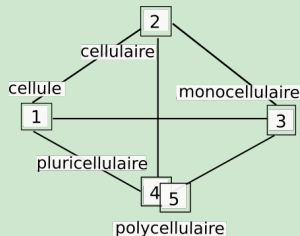belief

Xiste-Xisme
partisanship

X-Xisme
attitude

1 2 3
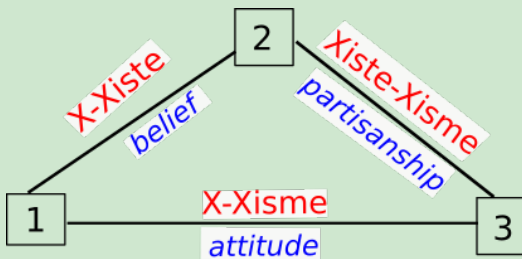
Meaning-form discrepancies: *mono/poly/pluri*Xsuf$_A$

- New value for Complexity: `form-motiv` (fm) and `sem-motiv` (sm)
  - formal paradigms distinguished from semantic paradigms

*mono/pluri/poly*Xsuf$_A$ in French

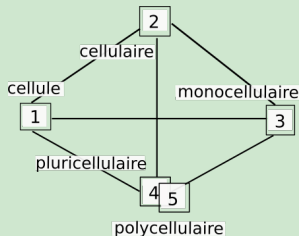|	Patt. Alt	Ori, Compl.	Cross-def
(1,2)	X/ Xaire	a2d, si	"smth cellulaire pertains to the cellule"
(1,3)	X/monoXaire	a2d, sm	"smth monocellulaire contains one cellule"
(1,4-5)	X/pluriXaire X/polyXaire		"smth pluri/polycellulaire contains several cellule"
(2,3-5)	X/monoX X/pluriX X/polyX	a2d, fm	—
(3,4-5)	monoX/polyX monoX/pluriX	ind, si	*polar opposition*
(4,5)	polyX/pluriX	ind sm	*synonymy*

Meaning-form discrepancies: *mono/poly/pluri*Xsuf$_A$

- no semantic annotation on relations tagged with the `form-motiv` (`fm`) value

*mono/pluri/poly*Xsuf$_A$ in French

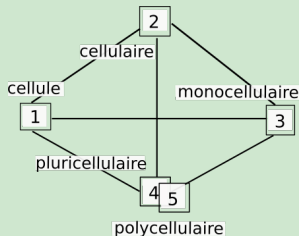|	Patt. Alt	Ori, Compl.	Cross-def
(1,2)	X/ Xaire	a2d, si	"smth cellulaire pertains to the cellule"
(1,3)	X/monoXaire	a2d, sm	"smth monocellulaire contains one cellule"
(1,4-5)	X/pluriXaire X/polyXaire		"smth pluri/polycellulaire contains several cellule"
(2,3-5)	X/monoX X/pluriX X/polyX	a2d, fm	—
(3,4-5)	monoX/polyX monoX/pluriX	ind, si	*polar opposition*
(4,5)	polyX/pluriX	ind sm	*synonymy*

Meaning-form discrepancies: *mono/poly/pluri*Xsuf$_A$

- overabundance: `indirect`, `sem-motiv` (`sm`) relations connecting synonyms words

*mono/pluri/poly*Xsuf$_A$ in French

	Patt. Alt	Ori, Compl.	Cross-def
(1,2)	X/ Xaire	a2d, si	"smth cellulaire pertains to the cellule"
(1,3)	X/monoXaire	a2d, sm	"smth monocellulaire contains one cellule"
(1,4-5)	X/pluriXaire X/polyXaire		"smth pluri/polycellulaire contains several cellule"
(2,3-5)	X/monoX X/pluriX X/polyX	a2d, fm	—
(3,4-5)	monoX/polyX monoX/pluriX	ind, si	*polar opposition*
(4,5)	polyX/pluriX	ind sm	*synonymy*

Outline

Conclusion (1)

- ParaDis: a semantic-driven paradigmatic model for derivation.
- This is a 4-level system, projecting the lexeme's ternary properties at family-wide level.
- Descriptive unit : (formal, semantic, part-of-speech, morphological) family
- Family arrangements form (formal, semantic, part-of-speech, morphological) paradigms
- semantic and formal paradigms are independent of each other without direct correspondence with each other
- They meet in the form of the morphological paradigm (abstract level)
- With this this organisational flexibility, not only classical and regular paradigmatic derivations are easily described, but also constructions involving:
  - ▶ meaning-form discrepancies
  - ▶ systematic synonymy (overabundance)

Conclusion (2)

- The Démonette database implements ParaDis
  - each entry describes a relation between two lexemes of a derivational family: the same lexeme therefore intervenes in as many entries of the base as it has relations within its family,
  - each entry is annotated with respect to the relation and to each of the two related lexemes,
  - relations are defined by three independent sets of properties: structural ones (characterization of the morphological connection itself), formal ones (formal pattern of each lexeme and stem variation, if any) and semantic ones (semantic type of the relation and glosses that mutually defines the two lexemes
- Démonette is intended for researchers in morphology and NLP applications. But its results will also be made available to primary and secondary school French teachers, researchers and speech-language pathologists.
  - The data and results of Démonette will be translated into exercises that will be used to test (possible issues with) children's vocabulary acquisition, or the impact of trauma on comprehension and lexical production

Références (1)

- Aronoff, M. (1976). Word Formation in Generative Grammar. Cambridge, The MIT Press.
- Baayen, R. H., Piepenbrock,R. & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA, Linguistic Data Consortium, University of Pennsylvania.
- Bauer, L. (1997). Derivational Paradigms. Yearbook of Morphology 1996. G. Booij and J. van Marle. Dordrecht, Kluwer: 243-256.
- Bauer, L. (2005). "Productivity: Theories". Handbook of Word-Formation. P. Štekauer and R. Lieber. Dordrecht, Springer: 315-334.
- Becker, T. (1993). "Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology." Yearbook of Morphology 1992: 1-27.
- Blevins, J. P. (2016). Word and Paradigm Morphology. Oxford, Oxford University Press.
- Bochner, H. (1993). *Simplicity in morphology*. Berlin & New York:Mouton de Gruyter.
- Booij, G. (2010). Construction Morphology. Oxford, Oxford University Press.
- Booij, G. & Masini F. (2015). "The role of second order schemas in the construction of complex words. Semantics of complex words". L. Bauer, L. Körtvélyessy and P. Štekauer. Heidelberg, Springer. 47: 47-66.
- Bonami, Olivier & Strnadová, Jana (2018). "Paradigm structure and predictability in derivational morphology". *Morphology*, https://doi.org/10.1007/s11525-018-9322-6.
- Haspelmath, M. & Sims , A. D.(2010). Understanding Morphology, second edition. London/New York, Routledge.
- Hathout, N. (2011a). "Morphonette: a paradigm-based morphological network." Lingue e linguaggio 2011(2): 245-264.
- Hathout, N. (2011b). "Une analyse unifiée de la préfixation en anti-." Des Unités Morphologiques au Lexique. M. Roché. Paris, Hermès: 251-318.
- Hathout, N. & Namer F. (2014). "Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French". Morphology and meaning (Selected papers from the 15th International Morphology Meeting, Vienna, February 2010) F. Rainer, W. U. Dressler, F. Gardani and H. C. Luschützky. Amsterdam, John Benjamins: 177 – 190.
- Hathout, N. & Namer F. (2018a). "La parasynthèse à travers les modèles : des RCL au ParaDis". The lexeme in descriptive and theroretical morphology. O. Bonami, G. Boyé, G. Dal, H. Giraudo and F. Namer. Berlin, Language science Press: 365-399.
- Hathout, N. & Namer F. (2018b). "ParaDis: a Families-and-Paradigms model for derivation. Revisiting Paradigms in Word-Formation" (Workshop organized in the framework of the conference WORD-FORMATION THEORIES III TYPOLOGY AND UNIVERSALS IN WORD-FORMATION IV), Košice, Slovakia.

Références (2)

- Kyjánek, L. (2018). Morphological Resources of Derivational Word-Formation Relations. T. Report. Charles University, ÚFAL. TR-2018-61.
- Litta, E., Passarotti, M., & Culy, C. (2016). "Formatio formosa est. Building a Word Formation Lexicon for Latin". In Proceedings of the third Italian conference on computational linguistics, 185–189.
- Namer, F. (2013) "A Rule-Based Morphosemantic Analyzer for French for a Fine-Grained Semantic Annotation of Texts", Proceedings of the 3d Workshop on Systems and Frameworks for Computational Morphology, C. Mahlow & M Piotrowski eds, Communications in Computer and Information Science, vol. 380, Springer
- Namer, F., Hathout, N. & Lignon, S. (2017). "Adding morpho-phonology into a french morphosemantic resource: Demonette". In Eleonora Litta and Marco Passarotti, eds, Proceedings of the First Workshop in Resources and Tools for Derivational Morphology (DeriMo),. EDUCatt, Milano, Italy:49–60.
- Plag, I. (2003). Word-formation in English. Cambridge, Cambridge University Press.
- Scalise, S. (1984). Generative Morphology. Dordrecht, Foris.
- Šojat, Krešimir; Srebačić, Matea; Pavelić, Tin; Tadić, Marko (2014) "CroDeriV: a New Resource for Processing Croatian Morphology". Proceedings of *LREC'14*. Reykjavik, Iceland: ELRA, 3366-3370.
- Steiner, P. and J. Ruppenhofer (2018). Building a Morphological Treebank for German from a Linguistic Database. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), Miyazaki, Japan, European Languages Resources Association (ELRA).
- Štekauer, P. (2014). "Derivational Paradigms". The Oxford Handbook of Derivational Morphology. R. Lieber and P. Štekauer. Oxford, Oxford University Press: 354-369.
- Stump, G. (2019). "Some sources of apparent gaps in derivational paradigms." Morphology 29(2): 271-292.
- Thornton, Anna M. (2012). "Reduction and maintenance of overabundance. A case study on Italian verb paradigms". In: *Word Structure* 5, pp. 183–207.
- Talamo, L., Celata, C. & Bertinetto, PM (2016) "DerIvaTario: An annotated lexicon of Italian derivatives". In: *Word Structure* 9, pp. 72-102.
- van Marle, J. (1985). On the Paradigmatic Dimension of Morphological Creativity. Dordrecht, Foris Publications.
- Vidra, J., Žabokrtský, Z., Kyjánek, L., Ševčíková, M. & Dohnalová, Š. (2019). DeriNet 2.0, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.
- Zeller, B., Šnajder, J., & Padó, S. (2013). DErivBase: Inducing and evaluating a derivational morphology resource for German. In Proceedings of the 51st annual meeting of the association for computational linguistics (Vol. 1), pp. 1201–1211.