

French corpus data in favor of Construction Morphology

Melanie Uth

University of Cologne

8th Mediterranean Morphology Meeting
September 15, 2011



Outline

Morphological gradience

Construction Morphology

Corpus study

Summary and Conclusion



Gradual morpheme boundaries

Relative frequency of derivatives correlates with degree of t-deletion [Hay(2003)]:

Derivative	Freq (D)	Basis	Freq (B)	Pronunciation
swift-ly	268	> swift	221	⇒ "less of a t"
soft-ly	440	< soft	1464	⇒ "more of a t"

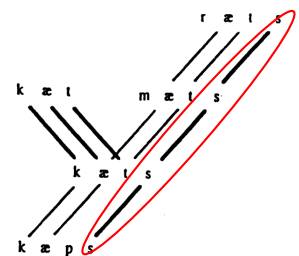
⇒ Morphological 'rules' such as engl. -ly are intimately related to the entirety of corresponding tokens (e.g. -ly-derivatives in concrete usage)



Network Models

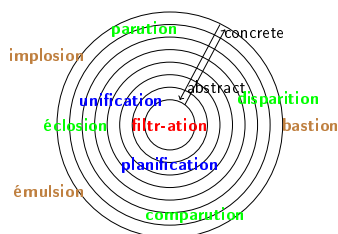
Emergentist modelling of morphological structure in Network Models:

(cf. [Bybee(1995), 429], see also [Bybee(1985), Bybee(1988)] a.o.)



Morphological categories as prototypes

Morphemes = prototype structures emerging from recurrent overlap of phonological and/or semantic features:



CM I: Degrees of schematicity

"Word formation patterns can be seen as abstractions across sets of related words" [Booij(2010a), 2]

"Complex words may have both common and distinctive properties" [Booij(2010b), 53]

⇒ Lexicon contains "different levels of abstraction at which generalizations concerning subsets of complex words can be made" (ibid.)



CM II: Associative patterns and symbolic processing

"[T]here is both analogical word formation ... and word formation based on abstract schemas." [Booij(2010a), 89]

"The use of schemas for word formation patterns presupposes a symbolic approach to representing linguistic knowledge." (ibid.:91)

Symbolic schemas emerge from associative patterns. (ibid.)

⇒ Certain independence but close relation to entirety of concrete derivatives.



Evidence

Affixoid nature of intensifying lexemes in Dutch XA and XN compounds:

- (1)
- ber-e 'bear' ⇒ bere-sterk 'very strong'
⇒ [[bere]_N [x]_{Ai}]_{Aj} ↔ [very SEM]_i_j
 - piep 'squeak' ⇒ piep-jong 'very young'
⇒ [[piep]_V [x]_{Ai}]_{Aj} ↔ [very SEM]_i_j
 - bliksem 'lightening' ⇒ bliksem-bezoek 'fast visit'
⇒ [[bliksem]_N [x]_{Ni}]_{Nj} ↔ [fast SEM]_i_j [Booij(2010b), 57]



Apocope of final schwa of Dutch *einde* ('end') and *aarde* ('earth') in NN compounds:

- (2)
- eind-examen/*einde-examen 'final exam'
eind-station/*einde-station 'final station'
eind-spel/*einde-spiel 'final game'
⇒ [[eind]_N [X]_{Ni}]_{Nj} ↔ [final SEM]_i_j
 - aard-appel/*aarde-appel 'potato'
aard-schok/*aarde-schok 'earthquake'
aard-atmosfeer/*aarde-atmosfeer 'earth atmosphere'
⇒ [[aard]_N [X]_{Ni}]_{Nj} ↔ [SEM]_i related to earth]_j
[Booij(2010b), 86f]



Morphological schemas

- ... develop by abstracting away from concrete lexical items (cf. e.g. engl. *Watergate* → [[x]_{Ni} [gate]_N]_{Nj})
- ... are systematically applied to new bases (cf. *skypable* etc.)
- ... are associated with abstract syntactic and/or semantic features (e.g. [Vtr_i-able]_{Aj} ↔ [[CAN BE SEM_i-ed]PROPERTY]_j, cf. *playable*, *debatable* etc.),
- ... can coerce specific interpretations of derivational bases (cf. *laughable*, *unlistenable*)
- ... cf. [Booij(2010b), Booij(2010a), Booij(2011)]



Corpus analysis

- Quantitative and qualitative analysis of
 - French event nominalizations in *-(at)ion*
 - French agent nominalizations in *-(at)eur*
- 3 Mio word corpus from FRANTEXT (1987-1997)
- Methodology:
 - Search for Latin correspondents (Perseus / CNRTL)
 - Analysis of lexicographic entries concerning mode of formation (CNRTL)



-(at)ion

- Massive borrowing of derivatives from Latin:

repetere	→	repetitus/m	→	repetitio	→	repetition
recipere	→	receptus/m	→	receptio	→	reception
abradere	→	abrasus/m	→	abrasio	→	abrasion
formare	→	formatus/m	→	formatio	→	formation
- Reanalysis, e.g. lat. *dissimulat-us* > lat. *dissimulat-io* > fr. *dissimul-ation* < *dissimuler* (cf. [Kerleroux(2008), 127f], [Schmitt(1988), 193], [Thiele(1981), 34]).
- High productivity of *-ation* and *-ification* ([Kerleroux(2008), 128], [Bonami *et al.*(2009)]).
- Subclasses are potentially productive, too ([Bonami and Boyé(2009)], [Di-Lillo(1983)]).



Share of Latin correspondents (LCs)

Category	Total	+ LC	Example	- LC	Example
regular					
-ation	493	342	<i>exhumation</i>	151	<i>vulgarisation</i> (~ 31%)
subregular					
-ification	29	17	<i>amplification</i>	13	<i>classification</i> (~ 45%)
irregular					
-ition	26	26	<i>punition</i>	0	–
-tion	17	17	<i>distribution</i>	0	–
-ion	58	57	<i>édition</i>	1	<i>implosion</i>
'3rd stem'	257	253	<i>prétention</i>	4	<i>parution</i>
No base	119	116	<i>friction</i>	3	<i>émulsion</i>
Total	477	469		8	(~ 2%)

Regular derivatives

Derivative	CNRTL entry
<i>accentuation</i>	Dér. de <i>accentuer*</i> ; suff. -ation*
<i>alphabétisation</i>	Dér. de <i>alphabétiser*</i> ; suff. -ation(-tion*)
<i>climatisation</i>	Dér. de <i>climatiser*</i> ; suff. -(a)tion*
<i>concertation</i>	Dér. de <i>concerter (se)*</i> ; suff. -(at)ion*
<i>épellation</i>	Dér. du rad. de <i>épeler*</i> ; suff. -(a)tion*
<i>filtration</i>	Dér. de <i>filtrer*</i> ; suff. -(a)tion*
etc. . . .	

Subregular derivatives: Variation

Derivative	CNRTL entry
<i>classification</i>	Dér. du rad. de <i>classifier*</i> ; suff. -ation*
<i>densification</i>	Dér. de <i>densifier</i> ; suff. -ication*
<i>intensification</i>	Dér. de <i>intensifier</i> ; suff. -tion**
<i>massification</i>	Dér. sav. de <i>massifier</i> ; suff. -(a)tion*
<i>planification</i>	Dér. de <i>planifier*</i> ; suff. -(ica)tion*
<i>unification</i>	Dér. de <i>unifier*</i> ; suffixe d'apr. les mots en -ification
etc. . . .	

Irregular derivatives_1: Analogy with Latin pattern

Derivative	CNRTL entry
<i>émulsion</i>	dér. du rad. du part. passé emulsus du lat. class. <i>emulgere</i>
<i>éclosion</i>	dér. de éclos , part. passé de <i>éclore*</i> ; suff. -ion*
<i>parution</i>	dér. de paru , part. passé de <i>paraître*</i> ; suff. -tion*
<i>comparution</i>	dér. de <i>comparaître*</i> d'apr. le part. passé com-paru ; suff. -tion*
<i>disparition</i>	dér. du rad. de <i>disparaître*</i> ; suff. -(i)tion* ; 1611 <i>disparution</i> (Cotgr.), à nouv. En 1725; <i>disparution</i> d'apr. disparu

Irregular derivatives_2: No -(at)ion nominalization

Derivative	CNRTL entry
<i>bastion</i>	<i>bastion</i> = dér. de <i>bastia</i> (suff. augm. -one)
<i>contraception</i>	<i>contraception</i> = angl. <i>contraception</i> , formé du préf. contra- et de -ception de <i>conception</i>
<i>implosion</i>	<i>implosion</i> = formé sur <i>explosion*</i> par substitution du préf. im- à ex- (<i>explosion</i> = Empr. au lat. class. <i>explosio</i> "action de rejeter en battant des mains . . .")

-(at)eur

- Massive borrowing of derivatives from Latin:
 - repetere* → *repetitus/m* → *repetitor* → *repetiteur*
 - recipere* → *receptus/m* → *receptor* → *recepteur*
 - revidere* → *revisus/m* → *revisor* → *reviseur*
 - formare* → *formatus/m* → *formator* → *formateur*
- Latinate derivatives in -(a)teur/trice (*générateur/génératrice*) next to vernacular forms in -eur/euse (*serveur/serveuse*), [Dubois and Dubois-Charlier(1999), 53f].
- Occurrence of scattered derivatives in -ateur from vernacular bases, cf. e.g. *accompagnateur* ([Schpak-Dolt(2010), 98]).
- -eur/euse also in borrowed/latinate derivatives, cf. *pêcheur* < *piscator*, *sauveur* < *salvatore* etc., (ibid.: 110).

Share of Latin correspondents (LCs)

Category	Total	+ LC	Example	- LC	Example
regular					
-eur	280	60	<i>procureur</i>	220	<i>bricoleur</i> (~ 80%)
subregular					
part.pres.stem	13	0	-	13	<i>buveur</i>
-ateur	76	59	<i>créateur</i>	17	<i>respirateur</i>
Total	89	59		30	(~ 34%)
irregular					
-eteur	1	1	<i>sauveteur</i>	0	-
-iteur	5	5	<i>répétiteur</i>	0	-
-uteur	1	1	<i>instituteur</i>	0	-
3rd stem	52	49	<i>acteur</i>	3	<i>projecteur</i>
No base	29	28	<i>spectateur</i>	1	<i>aviateur</i>
Total	88	84		4	(< 5%)

Regular derivatives

Derivative	CNRTL entry
<i>chasseur</i>	Dér. de <i>chasser*</i> ; suff. -eur2*
<i>enregistreur</i>	Dér. du rad. de <i>enregistrer*</i> ; suff. -eur2*
<i>loueur</i>	Dér. de <i>louer2*</i> ; suff. -eur2*
<i>mesureur</i>	Dér. de <i>mesurer*</i> ; suff. -eur2*
<i>payeur</i>	Dér. de <i>payer*</i> ; suff. -eur2*
<i>répondeur</i>	Dér. de <i>répondre*</i> ; suff. -eur2*
etc. ...	

Subregular derivatives_I: Variation

Der. in -eur CNRTL entry

<i>nourrisseur</i>	Dér. de nourrir* ; suff. -eur2* .
<i>buveur</i>	Dér. des rad. bev-, buv- de <i>boire1*</i> ; suff. -eur2* .
<i>régisseur</i>	Dér. du rad. du part. prés. de <i>régir*</i> ; suff. -eur2* .

Der. in -ateur CNRTL entry

<i>respirateur</i>	Dér. de <i>respirer</i> , suff. -eur2* .
<i>accélérateur</i>	Dér. de <i>accélérer*1</i> ; suff. -ateur* .
<i>radiateur</i>	Dér. sav. du lat. radiare ; suff. -ateur* ...
<i>excavateur</i>	Empr. à l'anglo-amér. excavator (1843 ds DAE), dér. de to <i>excavate</i> , v. <i>excaver</i> .

etc. ...

Subregular derivatives_II: Replacement of regulars

Derivatives CNRTL

<i>cultivateur</i>	Dér. de <i>cultiver*</i> ; suff. -(at)eur2* ; a supplanté l'a. fr. <i>cultivoire</i> ...
<i>dénonciateur</i>	Empr. au b. lat. <i>denuntiatio</i> proprement "celui qui annonce"; ... a remplacé l'a. fr. <i>denunceer</i> ...
<i>libérateur</i>	Empr. au lat. <i>liberator</i> "celui qui délivre" ...; a remplacé <i>délivreur*</i> dans le même sens.
<i>fondeur</i>	Formation savante d'apr. le lat. class. <i>fundator</i> "id."; a supplanté l'a. fr. et m. fr. <i>fondeur</i> ...
part.pres.stem	<i>prendreor</i> → <i>preneur</i> , <i>vëor</i> → <i>voyeur</i> , ...

Irregular derivatives

- **projecteur**: Dér. du rad. du supin *projectum* de *projicere* «jeter en avant, projeter»; **suff. -eur2***.
⇒ Sporadic application of **Latinatè 3rd stem pattern**.
- **réflecteur**: 1804 subst. masc., dér. sav. de *réfléchir**; **suff. -eur2***. L'angl. *reflector* ... est att. dep. 1767 ds NED.
⇒ Sporadic **borrowing of English derivatives**.
- **rédacteur**: Dér. sav. de *redactum*, supin de *redigere*, v. *rediger*; **suff. -eur2***. Cf. le m. fr. *redigeur* "rédacteur" 1796 (...).
⇒ Sporadic **replacement of regular derivatives**.
- **aviateur**: unclear etymology.

Summary of results

⇒ Quantitative:

- High percentage of genuinely French regular and subregular patterns
- Hardly any genuinely French irregular derivatives.

⇒ Qualitative:

- Uniform treatment of regular derivatives
- Sporadic analogies, replacement of specific regular derivatives and borrowing of isolated lexemes in the subregular and irregular domain.

Conclusion

- “Morphemes” = Complex morphological categories composed of concrete associative patterns and abstract symbolic part (⇒ CM: Analogical word formation next to abstract schemas).
- Highly regular patterns (with high type frequency) involve high degree of abstraction triggering switch from analogical inferring to symbolic processing (⇒ CM: Use of schemas presupposes symbolic approach).
- Abstract representation with close relation to entirety of concrete derivatives (CM: ⇒ Word formation patterns as abstractions across sets of related words).



Corpora and references

- CNRTL:** *Portail lexical* of the *Centre Nationale de Ressources Nationales et Textuelles*, *Laboratoire ATILF* and CNRS. (http://www.cnrtl.fr/definition_17.01.2010)
- FRA:** Combettes, Bernard & Pierrel, Jean-Marie (2009). *FRANTEXT*: ATILF database of texts, 16th to 21st century. Nancy: CNRS/Université Nancy2. (http://www.atilf.fr/frantext.htm_17.01.2010); extracted corpus fragment = 44 texts, 3.015.325 tokens, dating from 1987 to 1997.
- PER:** Crane, Grebory R., ed. (2009). *Perseus Digital Library. Online Latin Text Corpus*. Medford, MA: Tufts University.
- LS:** Lewis, Charlton T. & Short, Charles (1963). *A Latin Dictionary: Founded on Andrew's ed. of Freund's Latin Dictionary*. Oxford: Oxford University Press.

- Bonami, Olivier and Boyé, Gilles (2009). *Stems in French inflection and lexeme formation*. Paper presented at the *Surrey Morphology Meeting*, 24 April 2009.
- Bonami, Olivier; Boyé, Gilles; Kerleroux, Françoise (2009). *L'allomorphie radicale et la relation flexion-construction*. In B. Fradin, F. Kerleroux, and M. Plénat (eds.), *Aperçus de morphologie*, pages 103–125. Presses Universitaires de Vincennes, Saint-Denis.
- Booij, Geert (2010a). *Construction Morphology*. *Language and Linguistics Compass*, 3(1), 1–13.
- Booij, Geert (2010b). *Construction Morphology*. Oxford University Press, Oxford.
- Booij, Geert (2011).

- Morphology in Construction Grammar.
In T. Hoffman and G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford.
- Bybee, Joan L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. John Benjamins, Amsterdam [u.a.].
- Bybee, Joan L. (1988). *Morphology as Lexical Organization*. In M. Hammond and M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*, pages 119–142. Academia Press, San Diego.
- Bybee, Joan L. (1995). *Regular Morphology and the Lexicon*. *Language and Cognitive Processes*, 10(5), 425–455.
- Di-Lillo, Antoine (1983).

Morphologie des noms en “(-t)-ion” du français.
Cahiers de Lexicologie, 43, 117–135.

- Dubois, Jean and Dubois-Charlier, Françoise (1999). *La Dérivation Suffixale en Français*. Nathan, Paris.
- Hay, Jennifer B. (2003). *Causes and Consequences of Word Structure*. Routledge, London [u.a.].
- Kerleroux, Françoise (2008). *Des Noms Indistincts*. In B. Fradin (ed.), *La Raison Morphologique. Hommage à la Mémoire de Danielle Corbin*, pages 113–132. John Benjamins, Amsterdam [u.a.].
- Schmitt, Christian (1988).

Funktionale Variation und Sprachwandel: Zum Verhältnis von Ererbter und Gelehrter Wortbildung im Spanischen und Französischen.
In J. Albrecht (ed.), *Energie und Ergon - Sprachliche Variation, Sprachgeschichte, Sprachtypologie*. Band 2, pages 183–203. Narr, Tübingen.

- Schpak-Dolt, Nikolaus (2010). *Einführung in die Französische Morphologie*. de Gruyter, Berlin/New York, 3 edition.
- Thiele, Johannes (1981). *Wortbildung der Französischen Gegenwartssprache*. Langenscheidt, Leipzig.