

Morphologie dérivationnelle et polysémie

Vers un lexique morpho-sémantique pour le TAL

Lucie Barque

Université Paris 13 & Alpage

En collaboration avec L. Danlos, P. Haas, R. Huyghe, D. Tribout

Séminaire Alpage

Premier constat

- Analyse automatique : l'ambiguïté lexicale perturbe l'accès au sens
- Les méthodes de désambiguïsation ne donnent pas d'assez bons résultats. Celles qui s'appuient sur des ressources lexicales seraient les plus prometteuses (Navigli, 2009)
 - ▶ Granularité des inventaires de sens

Second constat

- Pas (encore) de lexique sémantique à large couverture pour le français
- Des méthodes de classification sémantique souvent peu explicites (traitement de la polysémie et caractérisation des sens)

Objectif

- Développer une ressource pour le français constituée
 - ▶ D'un lexique morpho-sémantique à gros grain
 - ▶ **verbes** + **noms** déverbaux
 - ▶ D'un corpus annoté (cf. SemCor, pour l'anglais)

Dans cette présentation

- Définition d'un inventaire de sens
 - ▶ Pour les noms
 - ▶ Pour les verbes
- Liens de dérivation morphologique et traitement de la polysémie

Pourquoi un lexique à gros grain ?

- 1 Les lexiques à grain fin ne conviennent pas à la tâche d'annotation (manuelle et automatique)
 - ▶ Accord inter-annotateur avec WordNet : env. 70% (Chklovski and Mihalcea 2002, Snyder and Palmer 2004)
 - ▶ Production d'inventaires réduits (Buitelaar 1998, Navigli 2006, Palmer *et al.* 2007)
 - ▶ Ex. Le verbe *develop* : 21 synsets regroupés en 7 "types"
- 2 Suffisant pour plusieurs tâches
- 3 Mieux adaptés aux techniques de désambiguïsation supervisées

Introduction

Group	Type	Sense No.	Description-Example	Hypernym
1-agent	new, abstract	WN1	products, or mental or artistic creations	CREATE
		WN 2	mental creations - "new theory of evolution"	CREATE
4-agent	improve item	WN6	resources - "natural resources"	IMPROVE
		WN7	ideas - "ideas in your thesis"	THEORIZE
		WN8	train animate beings - "violinists"	TEACH
		WN11	civilize - "countries are developing"	CHANGE
		WN12	make grow, ex. plants- "climate develops the grain"	CHANGE
		WN13	business, grow the market - "develop more customers"	GENERATE
		WN19	music, make more elaborate - "develop the melody"	COMPLICATE
2	new, property	WN3	personal attribute - "a passion for painting"	CHANGE
		WN4	physical characteristic - "beard"	CHANGE
3	new, self	WN5	originate - "new religious movement"	BECOME
		WN9	gradually unfold - "the plot developed slowly"	OCCUR
		WN10	grow - "a flower developed on the branch"	GROW
		WN14	mature - "The child developed beautifully. ..."	CHANGE
		WN20	happen - "Report the news as it develops"	OCCUR
5	chess	WN17	strengthening the position - "Spassky developed."	PLAY
		WN18	a better position for a piece - "develop the rook"	PLAY

Comment distinguer les différents sens d'un mot ?

- Critères pour l'identification des sens (Cruse 1986, Mel'čuk *et al.* 1995)
 - ▶ Critère des **arguments** distincts
 - ▶ FORCER (*fracturer*) : qqn force **qqch**
 - ▶ FORCER (*obliger*) : qqn force **qqn à qqch**
 - ▶ Critère des **co-occurents** distincts
 - ▶ FAMILLE (*parents*) : **fonder une famille**
 - ▶ FAMILLE (*ensemble*) : * **fonder une famille (de qqch)**
 - ▶ Critère des **dérivés morphologiques** distincts
 - ▶ CÉDER (*donner*) → **cession**
 - ▶ CÉDER (*accepter*) → * **cession**
- Jugés insuffisants pour conduire à un consensus (Geeraerts 1993, Pustejovsky 1995, Kilgarriff 1997)

Pourquoi un lexique morpho-sémantique ?

- 1 Les règles de dérivation morphologique s'appliquent à des unités sémantiquement spécifiées, non à des lemmes
 - ▶ Le nom PERDANT n'est pas dérivé de tous les sens de PERDRE
- 2 Les dérivés morphologiques ne sont pas exploités pour la désambiguïsation automatique
 - ▶ Au contraire des arguments et des co-occurents
 - ▶ Probablement pertinents : ex. système de QA
 - ▶ Rappel (Bernhard *et al.* 2011)
 - ▶ Précision
 - Q : perdre + match R : perdant
 - Q : perdre + clés R : *perdant
- 3 Le pouvoir discrétisant du "critère des dérivés distincts" n'a pas été mesuré

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

Classification nominale

Classification à très gros grain [CMLF 2014]

- Tests permettant de classer les N en 3 classes
 - ▶ Objet, Action, Propriété

Classification à gros grain [en prep.]

- Tests permettant de classer les N en 14 classes
 - ▶ Animé, Lieu, Artefact, Objet naturel,
 - ▶ Événement, Domaine,
 - ▶ Propriété, État,
 - ▶ Contenu, Temps, Maladie, Mesure, Monnaie, Phénomène

Remarque

- Classification élaborée dans le cadre de l'annotation de N morphologiquement simples

- **Noms d'objet** : N dénotant des entités, animées ou non, discrètes ou continues, dotées d'une étendue spatiale
 - ▶ N *se trouver* + complément de localisation spatiale
Mon frère/le chargement se trouve dans la voiture
 - ▶ N + complément de constitution matérielle
un manteau en laine/une construction en marbre
 - ▶ N + adjectif de couleur
une soie jaune d'or/un plumage turquoise
 - ▶ N + complément de dimension
une tente de 2 mètres de haut/ un vignoble de plusieurs hectares

- **Noms d'action** : N dénotant des situations dynamiques, duratives ou ponctuelles, téliques ou atéliques.
 - ▶ N avoir lieu à tel moment (à tel endroit)
Le crime/la réunion a eu lieu hier ici-même.
 - ▶ N se produire à tel moment (à tel endroit)
l'exode/le chamboulement s'est produit en 2006.
 - ▶ Effectuer N
Effectuer un travelling/un saut/changement
 - ▶ Procéder à N
Procéder à un baptême/à une restructuration

- **Noms de propriété** : N dénotant des situations d'aspect statif, transitoires ou structurelles
 - ▶ *Un état de N*
un état de misère/tristesse/joie
 - ▶ *Ressentir / éprouver du N*
Ressentir/éprouver de la haine/un mécontentement
 - ▶ *Faire preuve de N*
Faire preuve de flegme/de fougue/d'audace
 - ▶ *Être d'un grand N (= avoir du N)*
Être d'une grande intelligence (= avoir de l'intelligence)

Classification nominale à très gros grain

- N considérés comme dotés de sens multiples uniquement s'ils passent des tests de plus d'une des catégories définies
 - ▶ REPAS : Action / Objet
Le repas a lieu à 13h. / Le repas se trouve sur la table.
 - ▶ INITIATIVE : Action / Propriété
Cette initiative a eu lieu cette année. / Il fait preuve d'initiative.
 - ▶ AMOUR : Propriété / Objet
Il éprouve de l'amour pour elle. / Son amour se trouve dans le hall.
- Pas de prise en compte des distinctions de sens à l'intérieur d'une catégorie
 - ▶ BUREAU : Objet
Un bureau en bois / Le bureau se trouve au premier étage.

- Classification à très gros grain
 - ▶ 3489 N simples annotés (3696 sens)
 - ▶ N ayant un sens **Objet** : 2807
 - ▶ N ayant un sens **Action** : 300
 - ▶ N ayant un sens **Propriété** : 77
 - ▶ N ayant un sens **Autre** : 512 (ex. PONCIF, SOLFÈGE, SAMEDI, AMPÈRE, ...)
 - ▶ N à sens multiples
 - ▶ Env. 200 N "multisémiqiques"
 - ▶ Principale alternance **Action-Objet** (ex. REPAS, CIRQUE, ARRÊT)
 - ▶ Classes clairement discrètes

Classification nominale : WordNet

entity [1]
---- **abstract_entity [5]**
----- **event [2092]**
----- **state [983]**
----- **psychological_feature [747]**
----- **attribute [793]**
----- **relation [334]**
----- **measure [511]**
----- **communication [908]**
---- **physical_entity [22]**
----- **physical_process [319]**
----- **physical_object [2742]**
----- **somebody [2270]**
----- **animal [255]**
----- **grouping [175]**
----- **thing [283]**
----- **matter [841]**

Classification nominale à gros grain

- Définitions :

Animé (ex. ÂNE, PÈRE, CHANTEUR)

- ▶ *Le/ce N décide/choisit de V*

Artefact (ex. HORLOGE, COCKTAIL, AUTOCAR)

- ▶ *Il a fabriqué / confectionné ce N à tel endroit / à tel moment*

Lieu (ex. SALON, GRÈVE, VÉRANDA)

- ▶ *Ce N est un lieu magnifique / pittoresque / sauvage / superbe / ancien*

Temps (ex. SOIRÉE, JUIN, SAMEDI)

- ▶ *Ils se sont rencontrés N / courant N / pendant Dét N / Dét N durant / à N*

Domaine (ex. COMMERCE, JUDO, SOLFÈGE)

- ▶ *Pierre fait du N (faire n'est synonyme ni de fabriquer ni de avoir)*

Événement etc.

- Première annotation

- ▶ 100 N simples, 2 annotateurs
- ▶ Accord inter-annotateur (observé) : 0,69

AQUEDUC - lieu :1 artefact :1

ANTENNE - lieu :1 artefact :2 obj-nat :2

- 👉 Reconduire l'expérience avec le découpage du *Petit Robert*

- N à sens multiples

- ▶ Alternances productives
 - ▶ **Artefact-Domaine** (LUGE, GUITARE)
- ▶ Classes moins intuitivement discrètes
 - ▶ **Artefact-Lieu** (ÉGLISE)
 - ▶ **Événement-temps** (ÉTÉ, JOURNÉE)

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

Une difficulté accrue

- Verbes plus polysémiques que les noms
- Classification (hyponymique) moins naturelle
- Peu de tests linguistiques disponibles dans la littérature
- Alternances syntaxiques (diathèses)

Des ressources disponibles

- **Dicovalence** : ~ 3700 verbes + restriction de sélection
- **LVF** : ~ 12 310 verbes + classification sémantique + nominalisations
- **VerbeNet** : ~ 2600 verbes + classification sémantique + rôles thématiques
- **RLF** : ~ 2300 verbes + classification sémantique + polysémie décrite

Dicovallence (Mertens, 2010)

- Entrées : cadres de valence (paradigmes pronominaux)
 - ▶ ADOPTER
 - 1 qqn adopter qqn (*Ils ont adopté deux orphelins*)
 - 2 qqn adopter qqch (*Le tribunal a adopté cette ordonnance à l'unanimité*)
 - ▶ RECEVOIR
 - 1 qqn recevoir qqch de qqn (*Il a reçu une lettre urgente de sa mère*)
 - 2 qqn recevoir qqch (de qqpart) (*Elle a reçu de beaux cadeaux de mariage*)
 - 3 ...
- Dicovallence et polysémie verbale
 - ▶ Deux cadres = polysémie ? (cf. recevoir 1 et 2, agglomérer)
 - ▶ qqn agglomère qqch
 - ▶ qqn agglomère qqch à qqch
 - ▶ Un cadre = monosémie ?
 - ▶ qqn détester qqn ou qqch : sous-spécification
 - ▶ qqn abattre qqn ou qqch : polysémie
 - ▶ qqn offrir qqch à qqn : syn. de *donner* et syn. de *proposer*
 - ▶ Couverture sémantique ?

VerbeNet (Danlos *et al.*, 2014)

- Ressource sémantico-syntaxique des verbes français
 - ▶ Adaptation du VerbNet anglais (Levin 1993, Kipper-Schuler 2005)
 - ▶ Développée semi-automatiquement à partir de ressources construites manuellement (LVF et LG)
 - ▶ Adaptation des frames sémantico-syntaxiques de VerbNet au français
- Les classes sémantiques
 - ▶ 109 super-classes sémantiques (non-définies)
 - ▶ 270 classes sémantiques décrites

VerbeNet : les classes sémantiques

- **Image Creation** : quelqu'un (**Agent**) cause que quelque chose (**Theme**) soit *inscrit* quelque part (**Destination**)

image_impression-25.1 (broder, graver, tatouer)

- ▶ Agent V **Theme** {{+loc}} **Destination**
- ▶ Agent V **Destination** {de} **Theme**

scribble-25.2 (dessiner, ébaucher, découper)

- ▶ Agent V **Theme** {{+loc}} **Destination**

illustrate-25.3 (illustrer, orner, habiller)

- ▶ Agent V **Destination** avec/de **Theme**

- Quelle classification pour les **verbes** ?
 - ▶ Les super-classes de VerbeNet ?
 - ▶ Les classes sémantiques du RLF ?
 - 📖 Inventaire réduit, linguistiquement motivé
 - ▶ **État** : * *Il est en train de V* (EXISTER, AIMER, ADMIRER)
 - ▶ **Action** : *Il est en train de V* (ÉCRIRE, SIGNER, ADMIRER)
- S'appuyer sur la description des **noms dérivés de verbes**
 - ▶ Inventaire défini (à adapter aux noms dérivés)
 - ▶ Polysémie limitée
 - ▶ ADMIRATION : **état**

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

En **synchronie** :

- La structure polysémique des unités morphologiquement simples est le résultat d'**extensions sémantiques**
 - ▶ Métonymie
 - REPAS-1 (*nourriture*)
 - REPAS-2 (*action de se nourrir*)
 - ▶ Métaphore
 - PLONGER-1 (*sauter dans l'eau la tête la première*)
 - PLONGER-2 (*être condamné*)
- La structure polysémique des unités dérivées est le résultat d'**extensions sémantiques et/ou morphologiques**

1 Polysémie de la **forme dérivée**

- ▶ **ALLUMETTE** : allumette (au soufre) / allumette (au fromage)

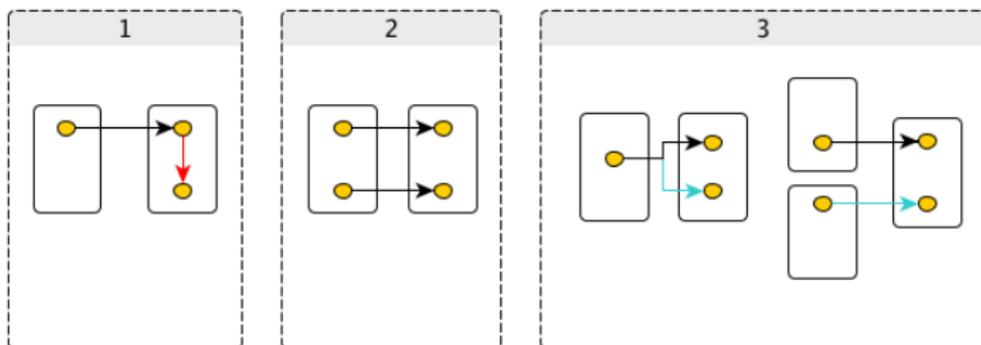
2 Polysémie de la (ou des) **forme(s) de base**

- ▶ **ACIDITÉ** : acidité (d'un fruit) / acidité (d'une remarque)

3 Polysémie / homonymie / sous-spécification de l'**affixe**

- ▶ **RAMEUR** : rameur (musclé) / rameur (d'appartement)
- ▶ **PLUMAGE** : plumage (coloré) / plumage (difficile)

Polysémie des unités dérivées : origines



1 ALLUMETTE

2 ACIDITÉ

3 RAMEUR / PLUMAGE

Polysémie des unités dérivées

- Les unités dérivées en moyenne moins polysémiques que leur base
- Les règles morphologiques permettent de faire des prédictions sur le sens des unités liées
 - ▶ $V_{action} \rightarrow N\text{-}EUR_{anime}$ (88%)
 - ▶ $V_{etat} \rightarrow N\text{-}EUR_{anime}$ (12%)

Polysémie des **noms déverbaux**

- Utiles pour le TAL
 - ▶ Détection d'événements (ambiguïté evt-objet)
 - ▶ Recherche d'information
- Efforts engagés
 - ▶ **Nomage** (Balvet *et al.*, 2012)
 - ▶ **VerbeNet** (Danlos *et al.*, 2014)
 - ▶ **Asfalda** (Candito *et al.*, 2014)

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

Phase "conception linguistique"

- Définition des classes sémantiques nominales et verbales
- Élaboration des règles de dérivation $V \rightarrow N$

Phase "description linguistique"

- **Nomenclature** : Noms déverbaux d'un corpus
 - ▶ Séquoïa (Candito & Seddah 2012)
- **Lexique** :
 - ▶ Description des N et de leur V source avec les inventaires définis
 - ▶ Encodage des liens de dérivation morphologique
- **Corpus** : Annotation des N (et des V, si présents dans le corpus)

Phase "extension semi-automatique du lexique : ajout d'information"

- Information sémantico-syntaxique (Dicovalece-VerbeNet)
 - ▶ Intersection : 2159 verbes de VerbeNet dans Dicovalece
 - ▶ BRODER
 - BRODER-1 (→ BRODERIE)
 - qqn brode (qqch) (*elle brodait des draps*)
 - qqn brode (qqch) qqpart (*elle a brodé son nom sur son mouchoir*)
 - 25(ImageCreation)-1 : Agent V Theme {{+loc}} Destination / Agent V Destination {de} Theme
 - BRODER-2
 - qqn brode sur qqch (*les femmes brodaient sur le thème du mari infidèle*)
 - 37(Communication)-11 : Agent V {sur} Topic

Phase "extension semi-automatique du lexique : ajout d'information"

- Information morpho-sémantique

Démonette (Hathout & Namer 2014)

- ▶ suffixe, sens du suffixe
- ▶ pas de prise en compte de la polysémie

RLF (Polguère 2014)

- ▶ sens de la base, sens du dérivé
- ▶ pas de description des liens de dérivation morphologiques

- ▶ CHASSER-CHASSEUR

- ▶ suffixe : **eur** / sens : **agent**
- ▶ sens de la base : **action** / sens du dérivé : **individu**

- ▶ Règles morphologiques

- ▶ $V_{action} \rightarrow N\text{-}EUR_{anime}$ (88%)
- ▶ $V_{etat} \rightarrow N\text{-}EUR_{anime}$ (12%)

Phase "extension semi-automatique du lexique : ajout d'entrées"

- Ajout d'entrées nominales
 - ▶ Identification automatique de la polysémie (exploitation des définitions de [Wiktionnaire](#) et du [TLFi](#))
- Ajout d'entrées verbales
 - ▶ Exploitation des connaissances syntaxiques ([Dicovale](#))
 - ▶ Exploitation des connaissances morphologiques

Phase "évaluation"

- Évaluation intrinsèque
 - ▶ Comparaison de l'accord inter-annotateur (Polymorph vs Wikipedia)
- Évaluation extrinsèque
 - ▶ [Semantic Role Labelling](#) (Asfalda)
 - ▶ [Question Answering](#)

Introduction

Inventaire des classes sémantiques

Classification nominale (avec P. Haas, R. Huyghe, D. Tribout)

Classification verbale (avec L. Danlos)

Dérivation morphologique et traitement de la polysémie

Polysémie des unités dérivées

Le projet PolyMorph

Conclusion et perspectives

En cours

- Inventaire de classes sémantiques, linguistiquement motivé, pour les N
- Couplage de ressources sémantico-syntaxique pour les V et de ressources morpho-sémantiques pour les paires V-N

À faire

- Développer l'inventaire pour la description des N dérivés
- Inventaire de classes sémantiques, linguistiquement motivé, pour les V
- Annotation d'occurrences de N et de V en corpus

Questions sur la polysémie des unités dérivées

- Quelle proportion de liens de polysémie importés de la base ?
- Quelle proportion de liens de polysémie véhiculés par le suffixe (ou propre à la catégorie du dérivé) ?
- Quelle proportion de liens de polysémie autres ?

Questions sur la polysémie des unités de base

- Quelle proportion de sens discriminés par des dérivés distincts ?