

Apports de la sémantique distributionnelle dans les recherches en linguistique

Cécile Fabre et Nabil Hathout

« Thématiques actuelles de la recherche en TAL »

7 novembre 2016

Objectif et plan de la séance

- Première partie (Cécile Fabre) : Place des modèles de sémantique distributionnelle (SD) dans les recherches en linguistique
 - La SD en quelques mots
 - Apports et limites pour la linguistique
 - Illustration sur deux questions :
 - Etude de relations sémantiques spécifiques
 - Traitement des unités au-delà du mot
- Deuxième partie (Nabil Hathout) : SD et recherches en morphologie

La SD en quelques mots

- Champ de recherche en TAL qui fonde le calcul sémantique sur une représentation vectorielle des contextes des mots
- Hypothèse selon laquelle 2 mots sont sémantiquement proches s'ils apparaissent dans des contextes similaires
- Une sémantique fondée sur l'usage, considéré à partir des emplois des mots dans des corpus

Deux familles de méthodes

- **Méthode classique** : modèles explicites fondés sur le décompte des contextes
 - 4 étapes :
 - Extraction des contextes
 - Décompte et pondération
 - (éventuellement) Réduction de dimensions
 - Calcul de similarité
- **Modèles prédictifs**
 - Procédure d'apprentissage de la représentation vectorielle : le système apprend à assigner des vecteurs similaires à des mots similaires
 - Utilisation d'un réseau de neurones
 - Représentations compactes, de dimension réduite, construites à partir de traits latents
 - Outil word2vec (Mikolov et al. 2013) (Goldberg and Levy 2014)

Une approche devenue centrale en TAL

- Maturité des méthodes
- Testées sur une large gamme de tâches
 - Tâches appliquées
 - Nombreux jeux de données en anglais
 - Détection de synonymes (TOEFL). (Landauer et Dumais 1997)
 - Jugement de proximité sémantique (WordSim 353) (Finkelstein et al. 2002)...
 - Catégorisation de concepts (AB set) (Almuhareb 2006)
- Mise au jour des meilleures configurations selon les tâches
 - Lapesa et Evert 2014, Kiela et Clark 2014, Tanguy et al. 2015, Vulić et al. 2016...
 - Taille des contextes, mesure d'association, mesure de similarité, nombre de dimensions...
 - Différentes configurations selon les catégories grammaticales (Schwartz et al. 2015).
- Des extensions de la méthode
 - Traiter des unités plus larges que le mot (Baroni et al. 2014)
 - Agréger des informations de différentes natures (données multimodales : Bruni et al. 2014)

Et pour la linguistique ?

- Réflexions sur la jonction SD / linguistique
 - Workshop ESSLI : DSALT Distributional Semantics and Linguistic Theory, 2016
 - Meaning in Context (MIC) Symposium, Munich, September, 2015
 - 2015 Toulouse workshop on formal and distributional semantics

Et pour la linguistique ?

- Possibilité de tester à très grande échelle et sur des données variées l'hypothèse distributionnelle
- De nombreuses questions peuvent être abordées en utilisant le critère de proximité distributionnelle
- On peut chercher à articuler ce modèle à d'autres modèles sémantiques :
 - Réflexion sur la complémentarité avec les modèles de sémantique formelle

De nombreux phénomènes sémantiques peuvent être abordés avec la SD

- Catégorisation sémantique
 - Burgess and Lund 1997, Boleda et al. 2004, Schulte im Walde 2006, Zarcone and Lenci 2008, Verhoeven et al. 2012
- Restrictions de sélection
 - Prédire la typicalité d'un constituant relativement à une position syntaxique
 - Pado et al. 2007, Lenci 2011, Zarcone et al. 2015
- Segmentation de composés
 - Riedl and Biemann 2016, Ziering et al. 2016
- Diachronie
 - Changement de sens, évolution de la productivité Gulordava et Baroni 2011, Perek 2016, Hamilton et al. 2016
- Lexique et discours
 - Cohésion lexicale : Adam 2012, Marathe et Hirst 2010
- Comparaison d'espaces sémantiques
 - Hamilton et al. 2016
 - Modelling linguistic individuality : Herbelot et QasemiZadeh 2016

Des freins identifiés

- Prise en compte d'une information linguistique généralement très pauvre (cooccurrences)
- Une notion très floue de la proximité sémantique
 - Agrège les différents types de relation
 - Similarité / association (*relatedness*)
- Des informations difficiles à utiliser, à interpréter
 - Aggravé par l'effet boîte noire des méthodes fondées sur la réduction de dimensions et les réseaux neuronaux
 - “What does a point in vector space, where the dimensions are typically uninterpretable symbols, stand for?” (Erk 2013)
- Difficulté à traiter la polysémie
 - Regroupement de formes → pas d'accès au sens sans procédure de désambiguïsation
(McCarthy et al. 2007), (Boleda et al. 2012)...

Des questions ouvertes

- La sémantique peut-elle se réduire à l'information distributionnelle ?
 - Quels aspects de la sémantique lexicale peuvent être déduits de leur distribution, quels aspects nécessitent l'apport d'informations complémentaires ?

“The distributional paradigm raises the question of the extent to which semantic properties can be reduced to combinatorial relations. Kamp in (Kamp et al. 2014)
 - Aspects pragmatiques et référentiels (Herbelot 2013)
- Utile pour la sémantique lexicale, mais ensuite ?

“Distributional models have been successful in modelling the meanings of content words, but logical semantics is necessary to adequately represent many function words.” (Lewis & Steedman 2013, cité par Koller 2015)
- A quelles questions fondamentales la SD répond-elle ?
 - Koller 2015 : nécessaires « top-down questions »
 - Traiter la compositionnalité, l'inférence, la référence (Lenci 2008)

« Compositionality (...) is an issue that any theory of lexical meaning must adress » (Lenci 2008)
 - Lenci 2008, Erk 2012 : Est-ce un modèle valide sur le plan cognitif ?

Deux exemples de questions linguistiques

- Cibler les relations sémantiques
- Traiter des séquences au-delà du mot

Cibler les relations sémantiques

- Préalable : améliorer les jeux d'évaluation
 - Permettre d'identifier les différents types de relations
 - A « typed » semantic space (Lenci)
- Illustration : antonymie et hyperonymie

Un préalable : améliorer les jeux d'évaluation

- Limite des étalons habituellement utilisés :
 - Ressources génériques (WordNet) ou liées à une tâche spécifique (TOEFL)
 - Ciblent une relation ou au contraire mêlant des relations très hétérogènes
 - Ex : WordSim 353 : pas de distinction similarité (*tasse/bol*) et association (*tasse/café*)
 - Même score pour des relations de type très différent
 - Ex : valeurs similaires pour *monk/brother*, *skin/eye*, *bread/butter*
 - Vocabulaire très majoritairement concret
 - Performances maximales obtenues par les systèmes sur plusieurs jeux (ex Bullinaria & Levy 2012 pour le TOEFL)
 - Sur l'anglais quasi exclusivement

Création de jeux de données dédiés

- Tanguy et al. 2015 :
 - Jeu de données adapté au corpus spécialisé traité
 - Notion générale de proximité
 - Ex : *trait* → *attribut, feature, étiquette, qualia*
- Bless (Baroni et Lenci 2011)
 - 200 noms concrets, 17 classes sémantiques, 5 relations sémantiques, 1 relation random par POS
 - coyote-hyper-animal, alligator-coord-lizard, alligator-mero-mouth, alligator-attri-aquatic, alligator-event-swim, alligator-random-n-addition
- SimLex-999 (Hill et al. 2014)
 - Focalisation sur la similarité (*coast-shore* 9.00, *clothes-closet* 1.96)
 - Mots concrets et abstraits (*the full range of concepts*)

- EVALution 1.0 (Santus et al. 2014)
 - Relations sémantiques de Bless + synonymie, antonymie, possession, inférence
 - Combiner jugement humain et analyse de corpus (fréquences)
 - Focalisation sur les relations prototypiques (score élevé)

• lunch	IsA	meal
• animal	Antonym	vegetable
• bite	Synonym	piece
• battle	PartOf	war
• letter	MemberOf	alphabet
• brick	MadeOf	clay
• machine_gun	Entails	shoot
• banana	HasA	peel
• anger	HasProperty	unpleasant

Cibler les relations sémantiques

- Deux relations spécifiques :
 - Antonymie
 - Hyperonymie
- Plusieurs méthodes :
 - Combinaison d'indices distributionnels et d'autres méthodes (filtre)
 - Sélection de contextes
 - Mesures de similarité spécifiques

Hyperonymie

- Une relation asymétrique
 - Morlane-Hondère 2013, Kotlerman et al. 2010, Lenci et Benotto 2012, Roller et al. 2014
 - Sur le plan extensionnel : $\text{Hyper} \subset \text{Hypo}$
 - Sur le plan intensionnel : $\text{Hypo} \subset \text{Hyper}$
- Hypothèse d'inclusion distributionnelle
 - *Distributional Inclusion Hypothesis* (Geffet & Dagan 2005)
 - “if u is a semantically narrower term than v , then a significant number of salient distributional features of u is included in the feature vector of v as well” (Lenci et Benotto 2012)

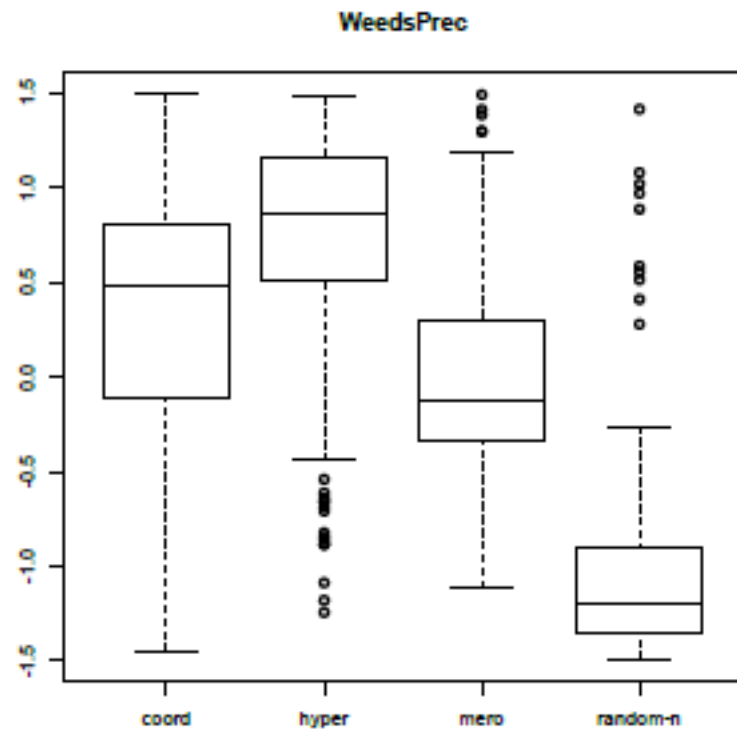
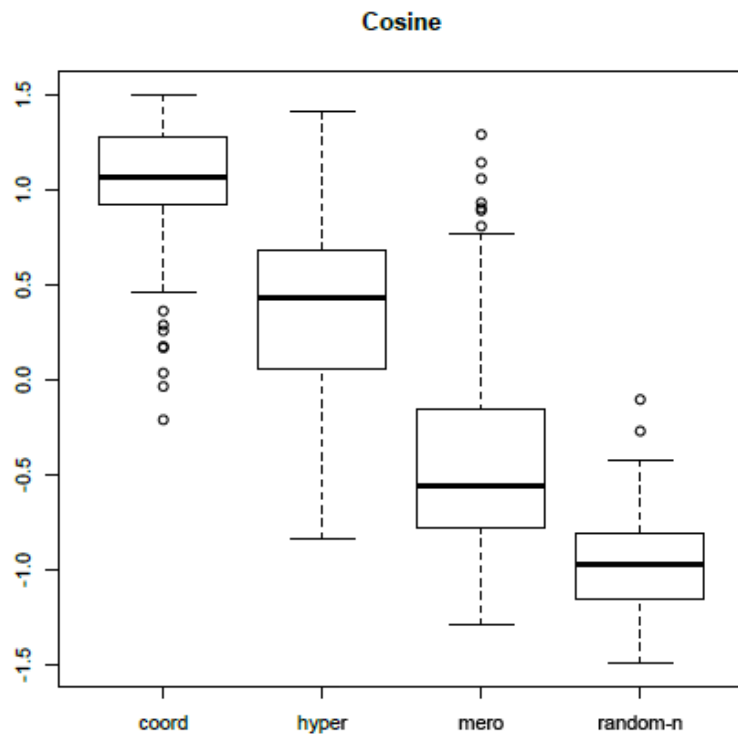
Hyperonymie

- Utilisation d'une mesure de similarité directionnelle
 - quantifie le recouvrement des traits du terme + spécifique u par ceux du terme +général v

$$WeedsPrec(u, v) = \frac{\sum_{f \in F_u \cap F_v} w_u(f)}{\sum_{f \in F_u} w_u(f)}$$

- Lenci & Benotto 2012
 - Données d'entraînement :
 - Exemples positifs du type *big boy* / *boy* (A N1 / N1)
 - Exemples négatifs du type A N1 / N2
 - But : généraliser l'inférence de A N $\rightarrow$ N à N1 $\rightarrow$ N2
 - Données de test : paires de mots issues de Wordnet

- Lenci & Benotto 2012



Antonymie

- Les mesures de SD ne permettent pas de distinguer synonymes et antonymes
 - (Lin 1998) : en moyenne score plus élevé des antonymes
- Exclure l'antonymie :
 - Filtrage par patrons
 - Lin et al. 2003, Turney 2008
 - Utilisation d'un thésaurus pour l'amorçage
 - Mohammad et al. 2013

Distinguer antonymes et synonymes

- Santos et al. 2014
 - les antonymes partagent des contextes moins saillants que les synonymes
- Scheible et al. 2013
 - les différences contextuelles entre antonymes et synonymes peuvent être capturées dans un modèle distributionnel à condition d'utiliser les bons traits
 - construction de 5 espaces distincts (V, ADV, N, ADJ, tout)
 - meilleur modèle : les verbes (66,5% et 70,6% exactitude)

Au-delà du mot

- Deux questions (Snajder 2016)
 - Représenter la sémantique des expressions
 - Détecter le degré de figement des expressions

Représenter la sémantique des expressions

- Jugement de similarité ?

Adjective/noun pairs			Noun/noun pairs			Verbs and objects		
Phrase 1	Phrase 2	Average sim	Phrase 1	Phrase 2	Average sim	Phrase 1	Phrase 2	Average sim
vast amount	large quantity	6.72	state control	government intervention	5.83	reduce amount	cut cost	6.56
old person	elderly lady	5.72	health minister	government leader	4.94	share interest	express view	5.72
small house	little room	3.83	business unit	development plant	3.72	present problem	face difficulty	3.56
social event	low cost	1.56	tv set	bedroom window	1.61	encourage child	leave company	128

Source : Erk 2012

Représenter la sémantique des expressions

- Les vecteurs distributionnels des séquences de mots peuvent être :
 - Récoltés directement à partir du corpus
 - Calculés à partir des représentations de leurs composants
 - Prend en compte la productivité et le manque de données (*sparsity*)

Représenter la sémantique des expressions

- Nombreux travaux impliquant des séquences de 2 mots : VN et AN

Mitchell and Lapata 2008, Baroni and Zamparelli 2010, Guevara 2010...

- Non prise en compte de l'ordre des mots et de la structure
- Opérations sur les vecteurs

- Addition
- Multiplication

	music	solution	economy	craft	reasonable
practical	0	6	2	10	4
difficulty	1	8	4	4	0
practical + difficulty	1	14	6	14	4
practical $\odot$ difficulty	0	48	8	40	0

Source : Snajder 2015

- Passage à la phrase

Baroni et al. 2014, Erk 2012

- « mixture models » ne sont pas adaptés :

- Mots non lexicaux
 - Relations non symétriques
 - Récursivité

- Fonctions distributionnelles

Détecter la compositionnalité des expressions polylexicales (MWE)

- Comparer le vecteur d'une MWE avec les vecteurs combinés de ses constituants
 - Schone and Jurafsky, 2001; Salehi et al., 2014; Hermann and Blunsom, 2014
 - Krcmár et al. 2013, 4 mesures :
 - Mesures de substituabilité (*hot dog*, *hot terrier*)
 - Mesure d'endocentricité (*hot dog*, *dog*)
 - Mesure de compositionnalité (*hot dog*, *hot + dog*)
 - Recouvrement des voisins distributionnels (*hot dog* → *food*, *chips* vs *dog* → *cat*, *bark*)
 - Riedel et Biemann 2015 :
 - Ordonner les MWE selon leur “uniqueness”
 - Probabilité pour un terme d'être remplacé par un mot simple
 - Ex : *hot dog* → *burger*, *sandwich*, *food*...

- Extraction des 200 ngrams les plus similaires :
 - Overlap du contexte (= voisin à droite et à gauche)
- Score d'uniqueness (uq)
- Complété par un score de completeness (c)

<i>red blood cell</i>		<i>red blood</i>	
Sim. term	Sc.	Sim. term	Sc.
erythrocyte	133	red	148
red cell	129	white blood	111
RBC	95	Sertoli	93
platelet	70	Leydig	92
red-cell	37	NK	86
reticulocyte	34	mast	85
white blood	33	granulosa	81
leukocyte	29	endothelial	81
granulocyte	28	hematopoietic stem	79
the erythrocyte	28	peripheral blood monon	78

Context term	Position	Count
<i>red blood cell</i>		
transfusions	right	48
(	right	42
transfusion	right	33
<i>red blood</i>		
cells	right	557
cell	right	344
corpuscles	right	13

- $DRUID = uq(t) - ic(t)$

Transition

- Baroni, M., Bernardi, R., Do, N. Q., & Shan, C. C. (2012). Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 23-32). Association for Computational Linguistics.
- Baroni, M., Bernardi, R., & Zamparelli, R. (2014). Frege in space: A program of compositional distributional semantics. *LiLT (Linguistic Issues in Language Technology)*, 9.
- Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal Distributional Semantics. *J. Artif. Intell. Res.(JAIR)*, 49(1-47).
- Erk, K. (2012), Vector Space Models of Word Meaning and Phrase Meaning: A Survey. *Language and Linguistics Compass*, 6: 635–653. doi:10.1002/Inco.362
- Erk, K. (2016), What do you know about an alligator when you know the company it keeps ? *Semantics and Pragmatics* Vol. 9, 1-63.
- Erk, K. (2013). Towards a semantics for distributional representations. In *Proceedings, 10th International Conference on Computational Semantics (IWCS-2013)*.
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. arXiv preprint arXiv:1606.02821.
- Kamp, H.; Lenci, A. & Pustejovsky, J. (2014), 'Computational Models of Language Meaning in Context (Dagstuhl Seminar 13462)', *Dagstuhl Reports* 3(11).

- Krcmár, L., Jezek, K., & Pecina, P. (2013). Determining Compositionality of Word Expressions Using Word Space Models. *NAACL HLT 2013*, 13, 42.
- Lenci, A. (2008) , Distributional semantics in linguistic and cognitive research, *Italian Journal of Linguistics*, 20(1), 1-31.
- Levy, O., & Goldberg, Y. (2014). Dependency-Based Word Embeddings. In *ACL* (pp. 302-308).
- Marathe, M., & Hirst, G. (2010, March). Lexical chains using distributional measures of concept distance. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 291-302). Springer Berlin Heidelberg.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119). Mitchell, J. and Lapata, M. 2010. *Composition in Distributional Models of Semantics*. *Cognitive Science*, 34:8, 1388--1439.
- Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1): 149–188.
- Riedl, M., & Biemann, C. (2015). A single word is not enough: ranking multiword expressions using distributional semantics. In *Proceedings of EMNLP* (pp. 2430-4440).
- Riedl, M., & Biemann, C. (2016). Unsupervised Compound Splitting With Distributional Semantics Rivals Supervised Methods. In *Proceedings of NAACL-HLT* (pp. 617-622).

- Santus, E. and Lu, Q. and Lenci, A. and Huang, C. (2014) Unsupervised Antonym-Synonym Discrimination in Vector Space. In: *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014*, 9-11 December 2014, Pisa.
- Scheible, S., Im Walde, S. S., & Springorum, S. (2013). Uncovering Distributional Differences between Synonyms and Antonyms in a Word Space Model. In *IJCNLP* (pp. 489-497).
- Verhoeven, B., Daelemans, W., & Van Huyssteen, G. B. (2012). Classification of noun-noun compound semantics in Dutch and Afrikaans. In *Proceedings of the Twenty-Third Annual Symposium of the Pattern Recognition Association of South Africa* (PRASA 2012) (pp. 121-125).
- Vulić, I., Schwartz, R., Rappoport, A., Reichart, R., & Korhonen, A. (2016). Automatic selection of context configurations for improved (and fast) class-specific word representations. arXiv preprint arXiv:1608.05528.
- Weeds, J., Clarke, D., Reffin, J., Weir, D. J., & Keller, B. (2014). Learning to Distinguish Hypernyms and Co-Hyponyms. In *COLING*.
- Zarcone, A., Padó, S., & Lenci, A. (2015). Same same but different: type and typicality in a distributional model of complement coercion. In *NetWordS* (pp. 91-94).
- Ziering, P., Müller, S., & van der Plas, L. (2016). Top a Splitter: Using Distributional Semantics for Improving Compound Splitting. *ACL 2016*.

Apports de la sémantique distributionnelle dans les recherches en linguistique

Cécile Fabre & Nabil Hathout

CLLE/ERSS, CNRS & Université de Toulouse

Partie 2. Morphologie

Morphologie

- Les travaux de morphologie flexionnelle ne font pas intervenir de sémantique à proprement parler.
- La morphologie dérivationnelle est essentiellement lexicale :
 - L'objet d'étude est le lexème construit et sa relation avec un lexème de base
 - Les études portent sur des procédés particuliers (sous-ensembles du lexique)
 - La sémantique décrite est l'**instruction sémantique** du procédé. Le sens est habituellement représenté par une glose grossière ('qui peut être @' pour *-able*)
 - Le lexique a accumulé au fil de son évolution des irrégularités : des dérives sémantiques, des décalages entre forme et sens, etc.
 - Les études morphologiques sont souvent guidées par la forme des lexèmes, ex. étude des lexèmes dont les lemmes finissent en *-able*.

Morphologie (2)

- La morphologie utilise les corpus et le Web pour trouver des lexèmes construits non enregistrés dans les dictionnaires et pour rechercher des sens particuliers de ces dérivés (morphologie extensive).
 - Les caractéristiques particulières des corpus ne sont pas considérées.
- La morphologie utilise des corpus pour estimer la productivité des procédés morphologiques.
- Quelques études décrivent les procédés morphologiques à partir des propriétés des contextes dans lesquels apparaissent les lexèmes construits (ex. thèse de Marine Lasserre (2016)).

Morphologie et sémantique distributionnelle

- Les relations dérivationnelles s'établissent entre des mots qui présentent des variations de forme et sens régulières.
- Les mots d'une famille dérivationnelle présentent des similarités formelles et sémantiques importantes.
- L'AD mesure la proximité sémantique entre les mots d'un corpus (Hypothèse distributionnelle), mais ...
 - L'estimation de la similarité entre les mots tient à la fois au partage de propriétés sémantiques et à leur appartenance à un même topique.
 - Difficile à utiliser pour décrire le sens d'un lexème en particulier:
 - L'AD ne permet pas de distinguer les différents sens des mots.
 - La similarité sémantique distributionnelle est **opaque** car elle est calculée « globalement » pour tous les mots du corpus.
 - Il n'existe pas de connexion entre les voisinages distributionnels et les instructions sémantiques des procédés dérivationnels.

→ Le sens construit dérivationnel a peu été étudié en sémantique distributionnelle.

Vue d'ensemble

- ❶ La sémantique distributionnelle comme instrument au service de la morphologie
- ❷ La sémantique distributionnelle pour la création de ressources morphologiques
- ❸ Représentation vectorielle de l'instruction sémantique des procédés morphologiques

Vue d'ensemble

- ❶ La sémantique distributionnelle comme instrument au service de la morphologie
- ❷ La sémantique distributionnelle pour la création de ressources morphologiques
- ❸ Représentation vectorielle de l'instruction sémantique des procédés morphologiques

Tester des hypothèses

- La sémantique distributionnelle permet
 - de connaître les plus proches voisins sémantiques d'un mot donné.
 - de comparer la similarité sémantique de couples de mots.
- Elle permet de tester certaines hypothèses sur les instructions sémantiques.

Désambiguïsation dérivationnelle

- Identifier l'histoire dérivationnelle de *dénationalisation* :
 - ✓ *dénationalisation* → *dénationaliser* → *nationaliser*
 - ✗ *dénationalisation* → *nationalisation* → *nationaliser*
- Est-ce que *dénationalisation* est plus proche de *dénationaliser* ou de *nationalisation* ?
- Est-ce que cette proximité dépend du corpus ?

Tester des hypothèses (2)

Complexité dérivationnelle

- L'hypothèse peut être testée pour tous les triplets (X_0, X_1, X_2) tels que morphologiquement $X_0 \rightarrow X_1 \rightarrow X_2$.
 - Est-ce qu'on a toujours
$$d(X_0, X_1) < d(X_0, X_2) ?$$
 - Pour répondre à cette question, on a besoin d'un corpus et d'un dictionnaire ou d'un lexique morphologique dérivationnel.

Proximité sémantique Verbe → Nom d'action

- L'idée que les noms d'action déverbaux (*nominals*) seraient des réalisations nominales de leur verbes de base remonte au moins à Chomsky (1970) (*Remarks on Nominalization*).
 - The enemy *destroyed* the city
 - The enemy's *destruction* of the city
- Dans un cadre différent, la même idée a été proposée et étendue par Roché (2009) aux relations entre les noms et leurs adjectifs de relation et entre le adjectifs et leurs noms de qualité.
 - protéger* → *protection*
 - nation* → *national*
 - profond* → *profondeur*
- Les lexèmes de ces couples sont maximalelement similaires sur le plan sémantique.
- Ils ne diffèrent que par leurs catégories grammaticales et leurs formes.

M1 de Marine Wauquier

Dispositif expérimental

- Étude sur le français
- Deux corpus : Wikipédia 2013 et Le Monde 1991-2000.
- Un lexique dérivationnel : Lexeur (Fabre et al., 2004).
- Un outil de sémantique distributionnelle : *word2vec* (Mikolov et al., 2013).

Lexeur

Masc	Fém	Base	Cat	Action
balanceur	balanceuse	balance	N	
balanceur	balanceuse	balancer	V	balancement balançage balance
balayeur	balayeuse	balayer	V	balayage balaieement
balbutieur	balbutieuse	balbutier	V	balbutiement
baliseur	baliseuse	baliser	V	balisage balisement

M1 de Marine Wauquier (2)

Lexeur

- Lexique morphologique dérivationnel de noms en *-eur*
 - Les entrées du lexique sont des noms en *-eur* (*calibreur*)
- Pour chaque entrée, Lexeur fournit :
 - Le nom ou les noms féminins correspondants (*calibreuse*)
 - La base qui peut être nominale (*camionneur* ← *camion*) ou verbale (*calibrer*)
 - Le ou les noms d'action associés (*calibrage*, *calibrement*)
- La base et les noms d'action ont été validés manuellement.
- Les équivalents féminins ont été générés par programme.

M1 de Marine Wauquier (3)

Expérience

- Extraction de tous les triplets $(N_{agent}, V_{base}, N_{action})$ présents dans les entrées de Lexeur. N_{agent} peut être masculin ou féminin.
- Calcul d'un modèle *word2vec* pour chaque corpus (configuration par défaut).
- Pour chaque triplet dont les 3 mots sont représentés dans le modèle, calcul des distances
 $d(N_{agent}, V_{base})$; $d(N_{action}, V_{base})$; $d(N_{agent}, N_{action})$
- Calcul des rapports $\frac{d(N_{agent}, V_{base})}{d(N_{action}, V_{base})}$ et $\frac{d(N_{agent}, N_{action})}{d(N_{action}, V_{base})}$.

M1 de Marine Wauquier (4)

Résultat attendu:

- $d(N_{action}, V_{base}) < d(N_{agent}, V_{base})$
- $d(N_{action}, V_{base}) < d(N_{agent}, N_{action})$

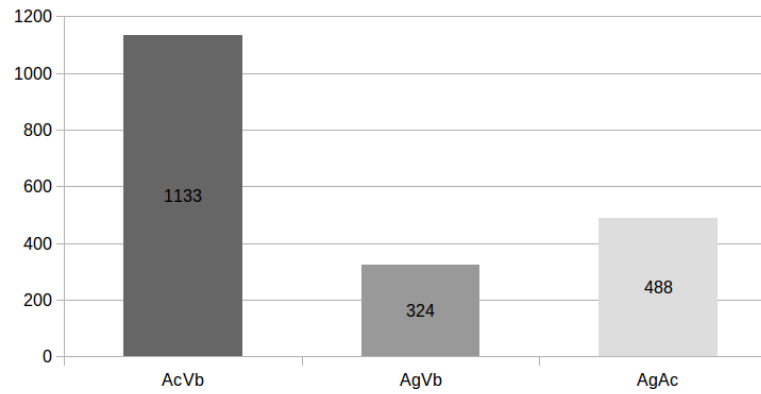
$$\frac{d(N_{agent}, V_{base})}{d(N_{action}, V_{base})} > 1$$

$$\frac{d(N_{agent}, N_{action})}{d(N_{action}, V_{base})} > 1$$

Pour tous les triplets ?

M1 de Marine Wauquier (5)

Tendances (Wikipédia 2013)

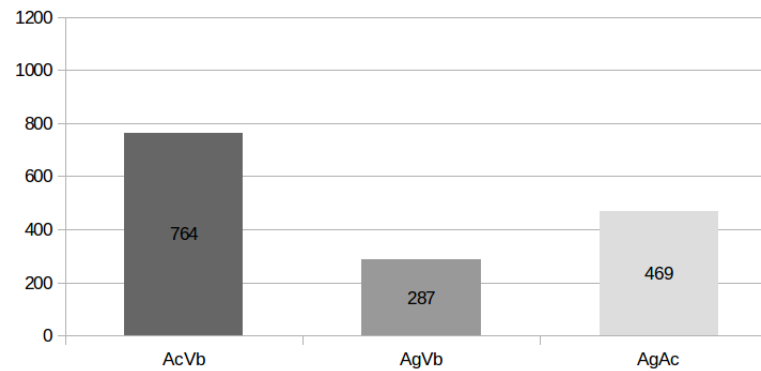


Nombre de triplets dont la plus courte distance est AcVb, AgVb, AgAc.

L'hypothèse est partiellement confirmée

M1 de Marine Wauquier (6)

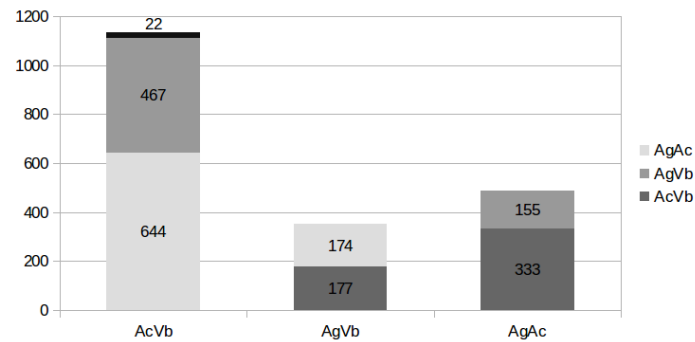
Tendances (Le Monde 1991-2000)



Nombre de triplets dont la plus courte distance est AcVb, AgVb, AgAc.

M1 de Marine Wauquier (7)

Tendances (Wikipédia 2013)



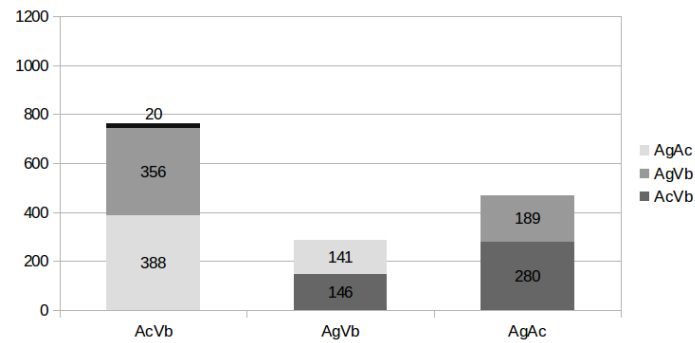
Nombre de triplets dont la 1^{re} et la 2^e plus courte distance sont :

1 = AcVb ; 2 = AgVb	1 = AgVb ; 2 = AgAc	1 = AgAc ; 2 = AgVb
1 = AcVb ; 2 = AgAc	1 = AgVb ; 2 = AcVb	1 = AgAc ; 2 = AcVb

L'hypothèse est partiellement confirmée

M1 de Marine Wauquier (8)

Tendances (Le Monde 1991-2000)



Nombre de triplets dont la 1^{re} et la 2^e plus courte distance sont :

1 = AcVb ; 2 = AgVb	1 = AgVb ; 2 = AgAc	1 = AgAc ; 2 = AgVb
1 = AcVb ; 2 = AgAc	1 = AgVb ; 2 = AcVb	1 = AgAc ; 2 = AcVb

M1 de Marine Wauquier (9)

Valeurs extrêmes du ratios

- Les triplets qui présentent des valeurs extrêmes pour les ratios $\frac{d(N_{agent}, V_{base})}{d(N_{action}, V_{base})}$ ou $\frac{d(N_{agent}, N_{action})}{d(N_{action}, V_{base})}$ ont tendance à être problématiques.
- Les distances *word2vec* peuvent être utilisées pour améliorer le lexique Lexeur.

AgVb/AcVb faible			AgVb/AcVb élevé		
tourneur	tourner	tour usinage (archaïsme)	ouvreur (rugby)	ouvrir	ouvrage (bâtiment)
menteur	mentir	menterie (archaïsme)	trieuse	trier	triage (train)
commandeur (ordres)	commander	commanderie (lieu)	échangeur (routier)	échanger	échange

Vue d'ensemble

- 1 La sémantique distributionnelle comme instrument au service de la morphologie
- 2 La sémantique distributionnelle pour la création de ressources morphologiques**
- 3 Représentation vectorielle de l'instruction sémantique des procédés morphologiques

Validation de DErivBase (Zeller et al., 2014)

Objectif

- Développer des stratégies permettant une validation sémantique de couples de mots morphologiquement apparentés.

DErivBase

- base morphologique de l'allemand
- créée au moyen de règles dérivationnelles opérant sur la forme graphémique des mots
- 280 000 lexèmes regroupés en familles dérivationnelles.

Méthode

- Combinaison des informations sémantiques fournies par un modèle distributionnel et une estimation de la fiabilité des règles dérivationnelles.

Validation de DErivBase (Zeller et al., 2014) (2)

Dispositif expérimental

- Morphologie dérivationnelle de l'allemand.
 - Utilisation d'un échantillon de couples de mots figurant à la même famille dérivationnelle dans DErivBase.
 - Les couples sont annotés manuellement :
 - R** Relation morphologie et sémantique.
 - M** Relation morphologie mais pas de relation sémantique.
 - N** Aucune relation, ni morphologique, ni sémantique
 - L** Erreur de lemmatisation
 - C** Composé
-
- Environ 25% des couples de lexèmes annotés ne sont pas sémantiquement motivés.

Validation de DErivBase (Zeller et al., 2014) (3)

Hypothèses

1. La similarité distributionnelle entre mots morphologiquement connectés indique qu'il existe entre eux une relation sémantique.
- 1'. Un rang distributionnel élevée indique qu'il existe une relation sémantique entre des mots dérivationnellement apparentés.
2. Certaines règles dérivationnelles préservent mieux le sens que d'autres.

Validation de DErivBase (Zeller et al., 2014) (4)

Dispositif expérimental

- Corpus = SdeWaC : 44 millions de phrases (~ 850 millions de mots) analysables syntaxiquement provenant de deWaC.
- Annotation morphosyntaxique et lemmatisation par le TreeTagger (et MATE tools comme solution de repli).
- Modèle distributionnel classique (contextes limités aux 10K mots les plus fréquents)

Observations

- Les relations dérivationnelles qui s'établissent entre des mots de catégories différentes ont moins de contextes partagés.
- Divergence sémantique :  *chanteur*: *chanteuse*
 *entraîneur*: *entraîneuse*

Validation de DErivBase (Zeller et al., 2014) (5)

Résultats

- Le rang dans le voisinage distributionnel est un bon indice de l'existence d'une relation dérivationnelle.
- Entraînement d'un classifieur LibSVM qui combine des indices distributionnels et des indices « structurels »
(ex. proportion de **R** dans l'ensemble des couples quiinstancient la relation dérivationnelle).
- Le classifieur améliorent de 10% l'exactitude de l'identification des relations sémantiques entre les mots morphologiquement apparentés.

Termes morphologiquement complexes (Bernier-Colborne and Drouin, 2016)

Construction de ressources terminologiques

- Comparaison de 3 modèles distributionnels
 - AD classique
 - *word2vec* CBOW
 - *word2vec* skip-gram
- Termes de 3 parties du discours (nom ; verbe ; adjectif)
- 4 relations lexicales
 - quasi-synonymie
 - antonymie
 - hyperonymie
 - dérivation morphologique

Termes morphologiquement complexes (Bernier-Colborne and Drouin, 2016) (2)

Dispositif expérimental

- Corpus = PANACEA = corpus monolingue français du domaine de l'environnement. 50 millions de mots. Lemmatisé.
- Jeu de test : 1314 couples de mots extraits du dictionnaire spécialisé DiCoEnviro
- Les relations dérivationnelles connectent des mots de POS différentes.
- Mesure pour l'évaluation = *mean average precision* ; MAP (moyenne des précisions moyennes)

Termes morphologiquement complexes (Bernier-Colborne and Drouin, 2016) (3)

Résultats

- Les modèles de l'AD classique ont globalement une meilleure MAP que les modèles *word2vec*.
 - Les résultats des modèles sont meilleurs pour la synonymie et les relations dérivationnelles que pour l'antonymie et l'hyponymie.
 - Les modèles *word2vec* sont nettement supérieurs aux modèles AD pour les relations dérivationnelles.
-
- Les modèles captent la dérivation aussi bien que la synonymie, mais les configurations optimales pour ces relations sont différentes.

Termes morphologiquement complexes (Bernier-Colborne and Drouin, 2016) (4)

Facteurs importants (AD)

facteurs	rel. dériv.	autres relations
taille	> 10	2 - 3
géométrie	rectangulaire	triangulaire
direction	indifférente	représentation séparée
pondération	MI / $\log_2$	Simple-LL / log

Facteurs importants (word2vec)

facteurs	rel. dériv.	synonymie
taille	> 10	2 - 3
architecture	skip-gram	CBOW
sous-échantillonnage	$t = 10^{-3}$	$t = 10^{-5}$

Vue d'ensemble

- 1 La sémantique distributionnelle comme instrument au service de la morphologie
- 2 La sémantique distributionnelle pour la création de ressources morphologiques
- 3 Représentation vectorielle de l'instruction sémantique des procédés morphologiques**

Modèles compositionnels

Objectif

- Disposer de représentations vectorielles pour les mots dérivés rares.

Baseline

- Les dérivés sont sémantiquement proches de leurs bases.
 - Les bases sont plus fréquentes que leurs dérivés.
- ⇒ Une méthode simple consiste à associer au dérivé la représentation de sa base.

Modèles compositionnels (2)

Compositional Distributional Semantic Models ; CDSM

- Calculer le sens d'un Dérivé en combinant le sens de sa Base et une représentation de l'instruction sémantique de l'affixation.
- Les CDSM ont été initialement conçus pour calculer le sens des syntagmes et des phrases à partir des représentations de leurs mots.
- Les CDMS de « 1^{re} génération » représentent tous les mots par des vecteurs. Le sens des syntagmes est obtenu en composant ces vecteurs (ex. addition).
- Les CDMS de « 2^e génération » représentent les sens des prédicats par des matrices ou des tenseurs.
- La méthode suppose que l'on dispose d'une analyse morphologique des dérivés (ex. *reconstruire* est formé par préfixation en *re-* à partir de *construire*)

Modèles compositionnels (3)

ADD (CDMS additifs)

- Pour une relation dérivationnelle (B, D) donnée, l'instruction sémantique du procédé dérivationnel P est considérée comme un vecteur p qui décrit le déplacement permettant d'atteindre la position de D (vecteur d) à partir de celle de B (vecteur b) :

$$d \approx b + p$$

- La moyenne $\hat{p}$ des vecteurs p est utilisée comme représentation du procédé :

$$\hat{p} = \frac{1}{N} \sum_i (d_i - b_i)$$

Modèles compositionnels (4)

Variantes WADD

$$d = \alpha b + \beta p$$

où α et β sont des constantes réelles

Variantes FULLADD

$$d = Ab + Bp$$

où A et B sont des matrices de pondération

Modèles compositionnels (5)

LEXFUN (CDMS multiplicatifs)

- L'instruction sémantique du procédé est représentée par une matrice P qui décrit une transformation linéaire qui appliquée au vecteur b donne le vecteur d :

$$d \approx Pb$$

- L'instruction sémantique du procédé est représentée par la matrice $\hat{P}$ calculée par la méthode des moindres carrés entre les vecteurs prédits $\hat{d}_i$ et les vecteurs effectifs d_i .

Lazaridou et al. (2013)

Dispositif expérimental

- Expérience réalisée sur l'anglais
- Corpus = ukWaC + Wikipédia (dump de 2009) + BNC
- 15 suffixes (*-able*, *-al*, *-er*, *-ful*, etc.) & 3 préfixes (*in-*, *re-*, *un-*) issus de la base CELEX.
- Jeu de donnée = couples (Base, Dérivé) formés par les 18 affixes (extraits de CELEX ; la fréquence de Base et Dérivé ≥ 20).
 - Jeu de test = 50×18 couples
 - Jeu d'entraînement = tout le reste
- Modèle AD classique construit pour les 20K mots lexicaux les plus fréquents + les mots du jeu d'entraînement.
- Mesure d'évaluation = distance moyenne entre les vecteurs des Dérivés et les vecteurs prédits par composition.

Lazaridou et al. (2013) (2)

Dispositif expérimental (suite)

- Estimation de la **qualité** des dérivés et de **similarité** des dérivés avec leur base réalisée par des « contributeurs » CrowdFlower.
 - Les estimations de similarité confirment l'hypothèse de Chomsky (1970) et de Roché (2009).

Résultats

- FULLADD est la méthode la plus performante pour les Dérivés qui sont fortement similaires à leurs Bases
- LEXFUN est la méthode la plus performante pour les Dérivés qui sont moins similaires à leurs Bases.

Kisselew et al. (2015)

Modèles distributionnels

- Morphologie dérivationnelle de l'allemand.
- Corpus = SDeWaC
- Entraînement d'un modèle prédictif *word2vec* (CBOW).

6 affixations

- *un*_{-A} (négation); *anti*_{-A} (opposition); *-in*_N (féminin); *-chen*_N (diminutif); *an*_{-V}; *durch*_{-V}
- Les bases et les dérivés sont de même catégorie.
- Les instructions sémantiques ne modifient pas la structure argumentale.

Kisselew et al. (2015) (2)

Données de test

- Pour chaque affixe, sélection aléatoire de 80 couples (Base, Dérivé) dans **DErivBase**.
- La fréquence du Dérivé dans SdeWaC ≥ 20 .

Modèles compositionnels

- Construction d'un modèle **ADD** et d'un modèle **LEXFUN** au moyen de **DISSECT Toolkit** (développé par l'équipe de M. Baroni).
- Baseline : la représentation du dérivé est identique à celle de la base (même méthode que Lazaridou et al. (2013)).
- Métrique : *Recall out of five* R_{oof} = nombre de fois où le dérivé apparaît parmi les 5 premiers voisins (de même catégorie) du vecteur prédit.

Kisselew et al. (2015) (3)

Résultats

- ADD produit les meilleurs résultats
- Les performances de LEXFUN sont inférieures à celles de la Baseline.
- R_{oof} est un meilleur indicateur de la performance des modèles que la similarité moyenne.
- ADD est performant pour le suffixe *-in* qui forme l'équivalent féminin des noms masculins. Forte cohésion sémantique.
- La prédiction du sens des dérivations verbales est difficile pour les 2 modèles. Forte variabilité de l'instruction sémantique.

Kisselew et al. (2015) (4)

Des modèles complémentaires

- Les modèles LEXFUN tendent à rapprocher les membres de la série dérivationnelle des dérivés (les mots construits au moyen du même affixe).
- Les modèles ADD tendent à rapprocher les mots dont le sens est proche de celui des bases (familles dérivationnelles).

Kisselew et al. (2015) (5)

Effet de la cohésion sémantique sur le modèle ADD

- 4 mesures de la cohérence pour un couple (Base, Dérivé) :
 - BCoh** similarité (cosinus) du vecteur de Base avec le centroïde de l'ensemble des bases des mots construits par l'affixation.
 - DCoh** similarité (cosinus) du vecteur de Dérivé avec le centroïde de l'ensemble des Dérivés construits par l'affixation.
 - BDSim** similarité (cosinus) entre Base et Dérivé.
 - DiffCoh** similarité (cosinus) du vecteur $d - b$ avec le centroïde des différences $d_i - b_i$ pour l'ensemble des couples (Base_i, Dérivé_i) connectés par l'affixation.
- La cohérence des bases dégrade les résultats (R_{oof}). La cohérence des dérivés n'a pas d'effet positif significatif.
- BDSim et DiffCoh ont un effet positif fort sur R_{oof} .
 - Pour DiffCoh, c'est attendu : ADD utilise le centroïde des différences.

Références

- Bernier-Colborne, G. and P. Drouin (2016). Évaluation des modèles sémantiques distributionnels: le cas de la dérivation syntaxique. In *Actes de la 23^e conférence sur le traitement automatique des langues naturelles (TALN-2009)*, Paris, pp. 125–138.
- Chomsky, N. (1970). Remarks on nominalizations. In R. Jakobs and P. Rosenbaum (Eds.), *Readings in English Transformational Grammar*. Waltham, MA: Ginn.
- Fabre, C., F. Floricic, and N. Hathout (2004). Collecte outillée pour l'analyse des emplois discordants des déverbaux en *-eur*. Communication aux journées d'étude sur *La place des méthodes quantitatives dans le travail du linguiste*. ERSS, Université de Toulouse II-Le Mirail.
- Kisselew, M., S. Padó, A. Palmer, and J. Šnajder (2015). Obtaining a better understanding of distributional models of german derivational morphology. In *11th International Conference on Computational Semantics (IWCS 2015)*, London.
- Lasserre, M. (2016). *De l'intrusion d'un lexique allogène. L'exemple des éléments néoclassiques*. Thèse de doctorat, Université de Toulouse.

Références (2)

- Lazaridou, A., M. Marelli, R. Zamparelli, and M. Baroni (2013). Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria. ACL.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 1–12.
- Roché, M. (2009). Pour une morphologie *lexicale*. In *La morphologie lexicale est-elle possible ?*, Volume 17 of *Mémoires de la Société de Linguistique, Nouvelle Série*. Leuven: Éditions Peeters.
- Zeller, B. D., S. Padó, and J. Snajder (2014). Towards semantic validation of a derivational lexicon. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (COLING 2014)*, Dublin, pp. 1728–1739.