

Sens, synonymes et définitions

Ingrid Falk¹, Claire Gardent², Evelyne Jacquy³, Fabienne Venant⁴

¹INRIA / Université Nancy 2

²CNRS / INRIA Nancy Grand-Est, Nancy

³CNRS / ATILF, Nancy

⁴Université Nancy 2 / INRIA, Nancy Grand-Est, Nancy

16ème conférence sur le Traitement Automatique des Langues
Naturelles

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives

Résumé

Objectif

Créer un dictionnaire de synonymes qui associe à chaque sens d'un mot l'ensemble des synonymes correspondants.

Présentation d'une méthode pour fusionner dictionnaires de synonymes et dictionnaires de langue.

- ▶ 5 dictionnaires de synonymes contenus dans la base de l'ATILF et
- ▶ le TLFi (Trésor de la Langue Française informatisé).

Résultat :

- ▶ Un dictionnaire de synonymes couvrant et avec définitions.

Résultats souhaités

Exemple : **achever**

Dictionnaires de synonymes :

abattre, aboutir, accomplir, aiguïser, améliorer, anéantir, assommer, boucler, cesser, clore, clôturer, compléter, conclure, conduire, consommer, continuer, couronner, estoquer, expédier, exécuter, finir, parachever, parfaire, perfectionner, raser, ruiner, réaliser, réussir, se taire, terminer, tuer.

Définitions du TLFi :

1. *Mettre la dernière main pour perfectionner*
2. *Porter un coup mortel à un animal déjà atteint physiquement; donner le coup de grâce.*
3. *Mener à sa fin, compléter l'action de.*

...

Résultats souhaités

Groupement et rattachement des synonymes aux sens donnés par les définitions du TLFi:

Dictionnaires de synonymes :

abattre, aboutir, accomplir, accomplir, aiguïser, améliorer, anéantir, assommer, boucler, cesser, clore, clôturer, compléter, compléter, conclure, conduire, consommer, continuer, couronner, estoquer, expédier, exécuter, finir, parachever, parfaire, perfectionner, raser, ruiner, réaliser, réussir, se taire, terminer, tuer.

Définitions du TLFi :

1. Mettre la dernière main pour perfectionner
2. Porter un coup mortel à un animal déjà atteint physiquement; donner le coup de grâce.
3. Mener à sa fin, compléter l'action de.

...

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche**
- 3 Ressources
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives

Approche

Entrées

Dictionnaires de synonymes : Base de l'ATILF

Dictionnaire de langue général : TLFi

TLFi \rightsquigarrow définitions \approx sens

Sortie

Un dictionnaire de synonymes qui associe à chaque sens (définition) d'un mot l'ensemble des synonymes correspondants.

Travaux similaires

- ▶ DicoSyn [Manguin et Al., 2004]
- ▶ WOLF [Sagot and Fišer, 2008]

Différences

DicoSyn : Association groupe de synonymes ↔ définition représentant le sens.

WOLF : Ensembles de synonymes obtenus à partir de données francophones.
pas de rattachement aux synset de WordNet.

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources**
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives

Ressources utilisées

5 des 7 dictionnaires de synonymes de la base des synonymes de l'ATILF.

Syn. Dic.	Verbes	Syn/verbes
Bailly	2600	1.
Benac	2656	1.5
Du Chazaud	3808	5.25
Larousse	3835	4.7
Le Petit Robert	5027	6.
total	5736	11.

mais: pas d'information catégorielle, pas de définitions.

TLFi

- ▶ 54 280 entrées, 92 997 lemmes, 271 166 définitions.
- ▶ balisé en format XML.
- ▶ textes des définitions lemmatisés et catégorisés.

mais: peu de synonymes, information non-systématique.

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources
- 4 Méthode**
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives

Méthode de fusion

Étant donné :

- ▶ Un verbe V ,
- ▶ Un ensemble de définitions $D_V^1 \dots D_V^n$ associés à V par le TLFi,
- ▶ L'ensemble de synonymes $Syn_V^1 \dots Syn_V^m$ associés à V par la base de synonymes,

Pour chaque synonyme Syn_V^k :

Quelles sont les définitions D_V^i pour lesquelles Syn_V^k est synonyme de V ?

Méthode de fusion : étapes

1. Extraction et transformation définitions TLFi \rightsquigarrow index. Index = ensemble de mots pleins lemmatisés.
2. A chaque définition et chaque synonyme est associé un index.
3. Index d'un synonyme \rightsquigarrow index de l'ensemble de ses définitions.
4. Mesure de la similarité entre index.
Quelle est la définition D_V^i de V la plus similaire de la définition de Syn_V^k ?
5. Association synonymes/définitions. Chaque synonyme est associé avec la (les) définition(s) D_V^i de V les plus similaires à son index.

Création des index.

Définitions TLFi \rightsquigarrow index

1. Les balises XML sont utilisées pour extraire les *définitions* du TLFi.
2. Les définitions ne contenant ni texte, ni synonymes, ni indications de domaine sont filtrées.
3. Index = liste des mots plein lemmatisés apparaissant dans le texte, les synonymes et les indications de domaine.

Exemples

Entrée TLFi de **projeter**

→ **PROJETER**, verbe trans.

A. — [Corresp. à *projection*]

1. Jeter loin en avant, avec force. *Il s'agit de projeter l'adversaire sur le tapis (de terre, dans un heurt un peu violent, il [le cheval] part gaiement (Jeux et sports, 19*

- 1. ... elle s'en alla, après m'avoir dévisagé, jugé, pesé, analysé de ce r
eût projeté sur les gens un de ces liquides épais dont se servent les pieu
 MAUPASS., *Contes et nouv.*, t.2, Inconnue, 1885, p.1000.

— *Au fig.* Pousser, inciter quelqu'un. V. *jaculation* ex. de Huysmans.

— *Empl. pronom.* Se lancer. *Il s'est projeté dans mes bras!* (CÉLINE, *Mort à crédit*,

2. Envoyer des rayons lumineux; représenter une image sur une surface. *Je vis m'élançai* (PROUST, *Swann*, 1913, p.35). *Jean Farou (...)* projetait sur les murs une

— **CIN., AUDIOVISUEL.** Passer dans un projecteur (v. ce mot B). *Proje*
cinématographiques de cellules agrandies de telle sorte que leur taille soit
établissements furent obligés d'échanger leurs programmes et, après quelque tem

♦ *Projeter quelque lumière* (au fig.). **Éclaircir.** *Synon. jeter quelque lumière* (v.

Exemple : le verbe *projeter*

Définitions extraites et leurs index :

Définition : Jeter loin en avant avec force.

Index : ⟨ jeter, loin, avant, force ⟩

Définition : *CIN. AUDIOVISUEL*. Passer dans un projecteur.

Index : ⟨ cinéma, audiovisuel, passer, projecteur ⟩

Définition : Éclaircir. Synon. jeter quelque lumière.

Index : ⟨ éclaircir, jeter, lumière ⟩

Mesurer la similarité entre index

Expérimentation avec des mesures numériques de similarité de 2 types :

1. Nombre de lemmes (et de séquences de lemmes) communs aux deux index.
2. Mesures vectorielles de premier et de second ordre, avec ou sans TF.IDF cutoff.

Total 6 mesures de similarité.

Identifier la meilleure mesure de similarité

- ▶ Construction d'une ressource de référence : 27 verbes, leurs définitions et pour chaque définition, les synonymes associés.
- ▶ Extraction de triplets $\langle \text{Verbe}, \text{Définition}, \text{Synonyme} \rangle$.
- ▶ À un triplet $\langle V, D_V, Syn_V \rangle$ est associée
 - la valeur 1 si Syn_V est considéré synonyme de V dans le sens donné par D_V ,
 - la valeur 0 sinon.

Exemple :

$\langle \text{achever}, \text{Mettre la dernière main pour perfectionner}, \text{accomplir} \rangle \rightsquigarrow 1$

- ▶ Associations triplets \leftrightarrow valeurs effectuées par **système** et **annotateurs**
- ▶ Comparaison des résultats système et de la référence sur la base des métriques classiques de précision, rappel et F-mesure.

Construction de la référence

3 dimensions : généricité, fréquence, polysémie

généricité position dans la hiérarchie de EuroWordNet.

fréquence extraite du lexique de Syntex (D. Bourigault).

polysémie nombre de définitions dans le TLFi.

3 valeurs :

basse, moyenne, élevée.

⇒ 27 verbes.

4 annotateurs :

- ▶ Fabienne Venant (U. Nancy 2), Mick Grzesitchak (CNRS/ATILF), Christiane Jadelot (CNRS/ATILF), Aurélie Merlot (U. Nancy 2/ATILF).
- ▶ Encadrés par Evelyne Jacquy (CNRS/ATILF).

L'accord inter-annotateurs

Syn. Dic.	Triples	Annotators	Agr.
Robert	2422	P1	80.39
		P2	85.59
		P3	75.55
		Q	59.537
Larousse	2573	P1	79.90
		P2	87.213
		P3	80.10
		Q	63.894
Du Chazaud	4893	P1	81.484
		P2	87.96
		P3	73.88
		Q	57.319
All dic.	7047	P1	81.481
		P2	87.072
		P3	76.5
		Q	63.374

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats**
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives

Les métriques

- Rappel** Nombre de triplets référence trouvé par le système / Nombre de triplets référence.
- Précision** Nombre de triplets référence trouvé par le système / Nombre de triplets trouvés par le système.
- F-mesure** Moyenne harmonique de rappel et précision.

Rappel, Précision et F-mesure

Mesure sim.	R	P	F
baseline	0.497	0.315	0.385
Recouvrement	0.725	0.508	0.598
Rec. étendu	0.723	0.508	0.597
Rec. étendu norm.	0.729	0.513	0.602
Vecteurs 1 ^{er} ordre	0.727	0.51	0.56
Vecteurs 2 nd ordre, sans tf.idf	0.715	0.503	0.59
Vecteurs 2 nd ordre, avec tf.idf	0.717	0.505	0.592

Table: Précision, rappel et F-mesure pour les différentes mesures de similarité.
Ligne de base (baseline) : association arbitraire entre synonymes et définitions.

Les usages réfléchis

- ▶ Une entrée du TLFi peut regrouper des usages distincts et en particulier, usage réfléchi (**s'abandonner**) et usage non réfléchi (**abandonner**).
- ▶ Ces usages distincts ont souvent des synonymes distincts.

Exemple

abandonner abdiquer, abjurer, abolir, accorder, aliéner, capituler, cesser, concéder, confier, céder, disparaître, donner, délaisser, ...

s'abandonner céder, faillir, fléchir, mollir, parler, s'adonner, s'effondrer, s'enfoncer, s'ouvrir, s'écouter, satisfaire, se donner, ...

- ▶ ↪ distinguer dans les entrées du TLFi les usages réfléchis des non-réfléchis
- ↪ comparer synonymes des usages (non-)réfléchis uniquement avec définitions aux usages (non-)réfléchis.

Résultats avec séparation des usages réfléchis/non-réfléchis

Mesure	Sans diff. réfl. vs. non-réfl.			Avec diff. réfl vs. non-réfl		
	R	P	F	R	P	F
baseline	0.497	0.315	0.385	0.440	0.433	0.437
Rec. 1	0.725	0.508	0.598	0.697	0.685	0.691
Rec. 2	0.723	0.508	0.597	0.697	0.685	0.691
Rec. 3	0.729	0.513	0.602	0.711	0.670	0.706
Vect. 1	0.727	0.510	0.560	0.704	0.693	0.698
Vect. 2	0.715	0.503	0.590	0.698	0.686	0.692
Vect. 3	0.717	0.505	0.592	0.701	0.689	0.695

Table: Précision, rappel et F-mesure pour les différentes mesures de similarité, avec et sans la distinction réfléchi / non-réfléchi.

Un pré-traitement linguistique plus fin améliore les résultats

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion**
- 7 Perspectives

Conclusions

- ▶ La méthode proposée permet d'associer synonymes et définitions sans intervention humaine avec une précision relativement bonne (0.706 avec 0.87 comme borne supérieure pour l'accord inter-annotateur entre humains).
- ▶ Le pré-traitement linguistique a un impact déterminant sur la qualité des résultats.
- ▶ Parce qu'elle est indépendante du dictionnaire de synonymes utilisé, la méthode permet de fusionner plusieurs dictionnaires de synonymes par le biais des regroupements sur les définitions TLFi.

Sommaire

- 1 Introduction et motivation
 - Exemple
- 2 L'Approche
- 3 Ressources
- 4 Méthode
 - Création des index
 - Similarité des index
 - La référence
- 5 Résultats
 - Métriques
 - Les usages réfléchis
- 6 Conclusion
- 7 Perspectives**

Perspectives

- ▶ Pré-traitement des collocations verbales.
- ▶ Application de la méthode à l'ensemble des dictionnaires de synonymes.
- ▶ Extension de la couverture par l'intégration de ressources externes (Wiktionary, Wikipedia, EuroWordNet).
- ▶ Validation manuelle par service web.
- ▶ Alignement avec
 - ▶ Wolf,
 - ▶ l'EuroWordNet français,
 - ▶ les sens microscopiques ou cliques des espaces sémantiques de Dicosyn.

Merci !

Bibliography



B. Sagot and D. Fišer.

Building a Free French WordNet from Multilingual Resources.

Proc. of Ontolex, 2008.



Jean-Luc Manguin, Jacques François, Rembert Eufe, Ludwig Fesenmeier, Corinne Ozouf and Morgane Sénéchal.

Le dictionnaire électronique des synonymes du CRISCO : un mode d'emploi à trois niveaux.

Les Cahiers du CRISCO, 2004.