

Densifier les réseaux de synonymie : étude de Wiktionary

M2 TAL - 1^{er} Déc. 2009

Emmanuel Navarro

Franck Sajous

Bruno Gaume

Laurent Prévot

ShuKai Hsieh

Tzu-Yi Kuo

Pierre Magistry

Chu-Ren Huang

IRIT, CNRS & Univ. de Toulouse

CLLE-ERSS, CNRS & Univ. de Toulouse

CLLE-ERSS & IRIT, CNRS & Univ. de Toulouse

LPL, CNRS & Univ. de Provence

English Department, NTNU, Taiwan

Graduate Institute of Linguistics, NTU, Taiwan

TIGP, CLCLP, Academia Sinica, GIL, NTU, Taiwan

Dept. of Chinese and Bilingual Studies,
Hong Kong Poly U., Hong Kong.



Sommaire

- 1 Réseaux sémantiques : tour d'horizon
 - Princeton WordNet
 - EuroWordNet
 - Améliorer l'existant
 - Recréer l'existant
- 2 Wiktionary et al : réseaux de synonymie
 - Présentation de Wiktionary
 - Réseau de synonymie
 - Ressources étalons
 - Comparaison
- 3 Densification du réseau
 - En exploitant sa structure de SW
 - En utilisant les traductions
- 4 Conclusions et travaux futurs

Pourquoi faire ?

Les RS sont utiles (nécessaires) aux applis TAL

- accès aux textes « *par le sens* » ;
- désambiguïsation en contexte ;
- RI, QR : représentation des documents et expansion de requêtes ;
- détection d'« impropriétés » (par rupture de cohésion dans les chaînes lexicales) [Hirst and St Onge, 1998] ;
- calcul de distances sémantiques ;
- TAO, etc.

Pour l'anglais : Princeton WordNet [Fellbaum, 1998b]

- Visée psycholinguistique ;
- Début du projet en 1985, amorcé par le corpus Brown ;
- Méthodes semi-automatiques et validation manuelle ;
- Version 1.0 en 1991. Aujourd'hui, version 3.0 disponible gratuitement (outil d'interrogation et fichiers de données)
<http://wordnet.princeton.edu/wordnet/download/>
- Taille :

POS	Littéraires	Synsets
N	117798	82115
V	11529	13767
Adj	21479	18156
Adv	4481	3621

Princeton WordNet (2)

- PWN ignoré par les psycholinguistes, omniprésent en TAL ;
- Critiques :
 - étanchéité des sous-réseaux (POS) ;
 - pas d'indication de domaine (plus vrai aujourd'hui) ;
 - densité variable ;
 - granularité trop fine ? 63 sens pour *break* ;
 - mais {*car, auto, automobile, machine, motorcar*} forment un synset car ils peuvent commuter dans un certain contexte.
- Succès : qualité de la ressource ou seule ressource existante (et disponible) ?
- PWN : *The Mother of all WordNets* [Fellbaum, 1998a].

Autre langues

EuroWordNet [Vossen, 1998]

- Dév. d'un équivalent de PWN pour les langues européennes ;
- 1^{ère} vague : néerlandais, italien, espagnol et anglais (objectif initial de 30000 synsets) ;
- français, allemand, estonien et tchèque : 2^{ème} phase, (objectif restreint à 7500 synset) ;
- *language specific WordNets* développés séparément ;
- Ressources locales liées par un index interlangues (ILI) ;
- Déf. d'un noyau de concepts de base (*top-ontology*)
- Deux démarches inverses (ascendantes et descendantes) ;

Autre langues

EuroWordNet (2)

- Le développement de chaque ressource locale dépend fortement des ressources lexicales et des outils préexistants ;
- Seulement Noms et Verbes ;
- WN français : dév. par Univ. d'Avignon, Memodata et Bertin. Arrivé sur le tard, manque de temps → traduction de synsets de PWN 1.5 ;
- Qualité critiquée. . . le caractère payant aussi ;
- Ressource figée depuis la fin du projet.

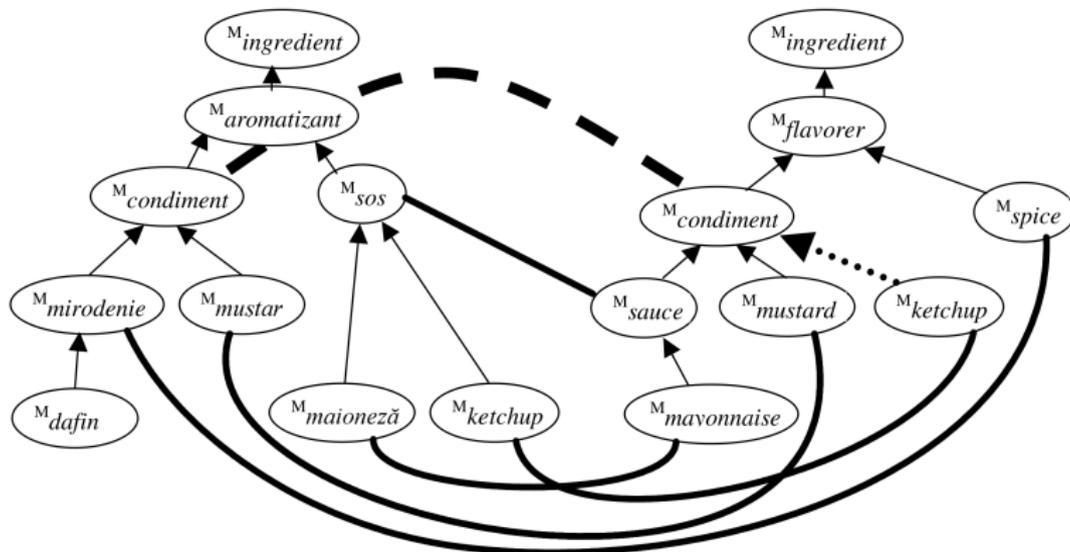
Un peu plus à l'Est. . .



Balkanet [Tufiş and Cristea, 2002]

- Comme EWN, projet financé par la CEE ;
- grec, turc, roumain, tchèque, bulgare et serbe ;
- même structure : réseaux monolingues + ILI et top-ontology ;
- une base pour les trads : corpus alignés (*1984* et *La République*) ;
- noyau de concepts extrait d'EWN ;
- construction d'un lexique bilingue anglais-roumain.

Principe de préservation de la hiérarchie



Translational links and consistency checks [Tufiş and Cristea, 2002]

Améliorer EWN-FR [Jacquin et al., 2007]

Constat :

- relations manquantes
eg. *citronnier* orphelin alors que le concept *arbre fruitier* existe ;
- hiérarchie de concepts plate ;
- absence de définitions ;
- erreurs de transcription ;
- incohérences dans les formats de fichiers.

Améliorer EWN-FR (2)

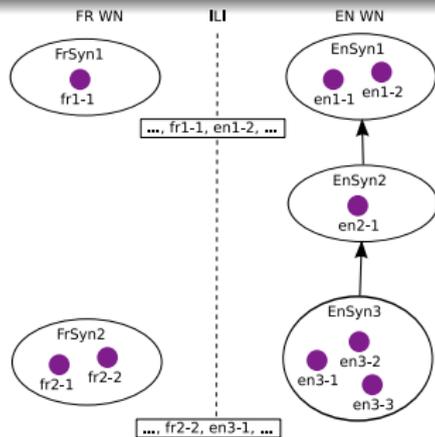
Méthodes

- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;

Améliorer EWN-FR (2)

Méthodes

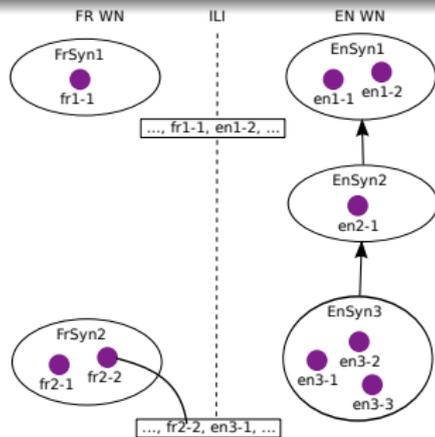
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

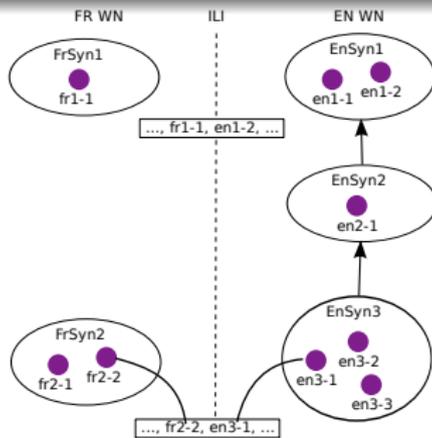
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

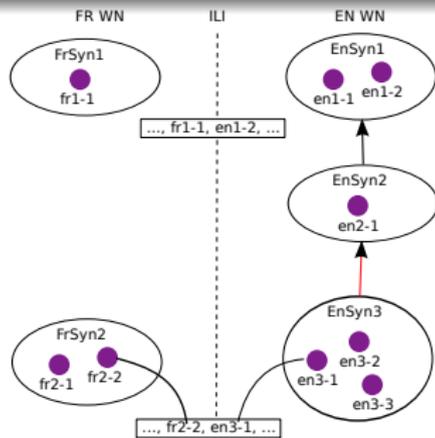
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

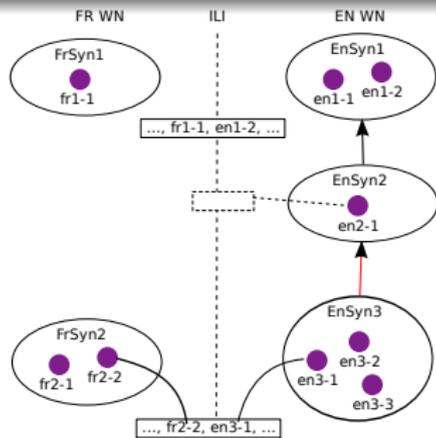
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

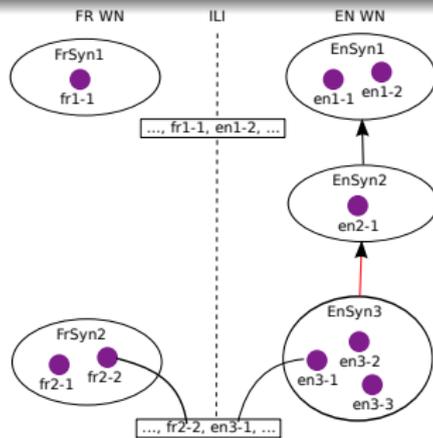
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

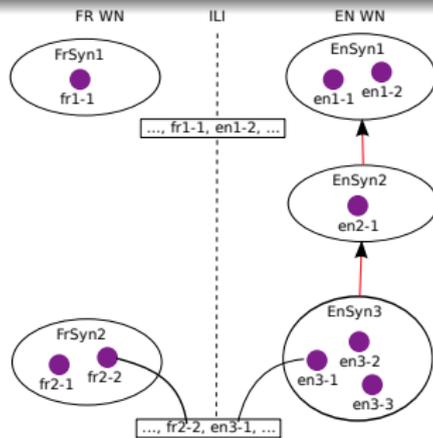
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

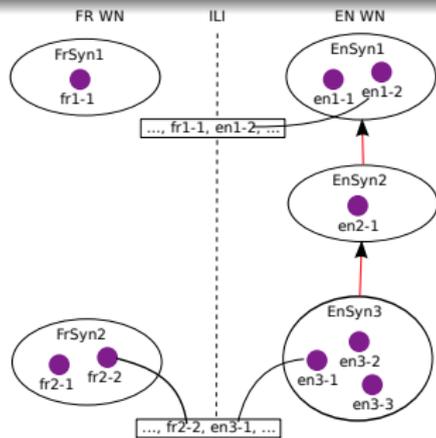
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

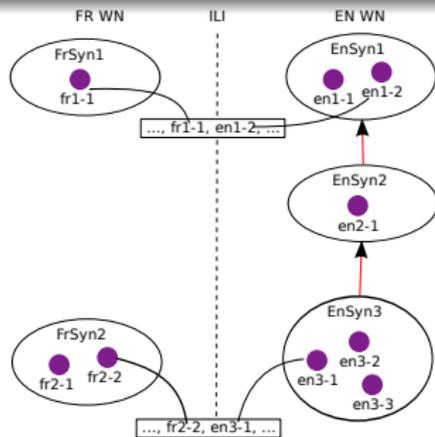
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

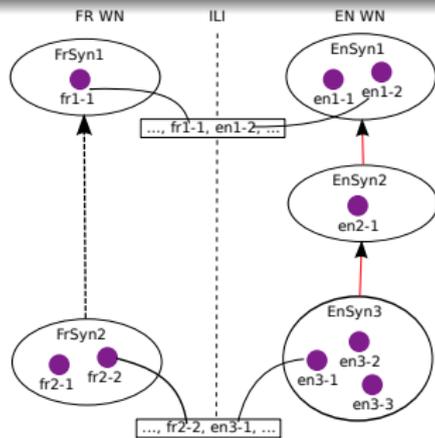
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

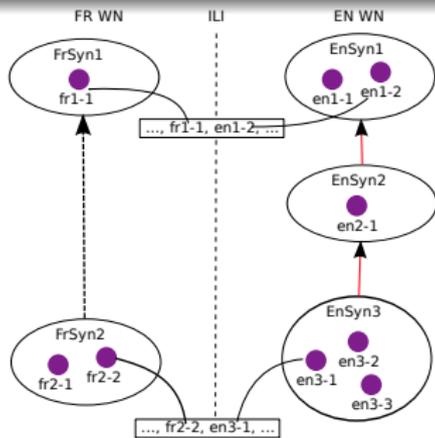
- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Améliorer EWN-FR (2)

Méthodes

- définitions : Wikipédia (traitement particulier pour pages d'homonymie) ;
- relations manquantes : utiliser liens de trad + relations sémantiques (éventuellement transitives) dans la langue cible ;
- [Tufiş and Cristea, 2002] ont « montré » que ça marche ;



Relations manquantes

- validation manuelle ;
- 0.3% de relations d'hyponymie ajoutées ;
- nb de relations candidates examinées non mentionné.

Créer une ressource libre : WOLF [Sagot and Fišer, 2008]

Méthode 1 : traduction de PWN

- extraction d'un lexique FR-EN grâce aux liens de trad de Wikipédia (→ 300000 liens)
- + lexique FR-EN dans Wiktionary (24000 pour Wikt.EN et 24000 pour Wikt.FR)
- + Wikispecies + Eurovoc (thésaurus multilingue UE)
- projection de ces lexiques sur les littéraux monosémiques de PWN
- → 18000 synsets pour Wikipédia, 10500 pour Wikt., 7000 pour Wikispecies, 1000 pour Eurovoc

Créer une ressource libre : WOLF (2)

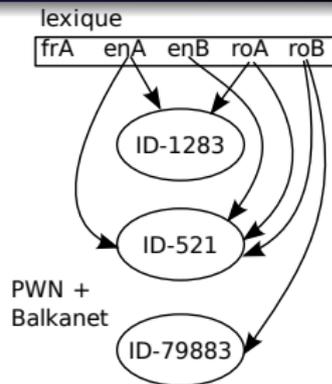
Méthode 2 : alignement + Balkanet

- corpus aligné en phrases : JRC (français, anglais, roumain, tchèque et bulgare) ;
- lemmatisation par TT ;
- Uplug(GIZA++) → alignement en mots, construction d'un lexique multilingue ;
- construction des synsets français

Créer une ressource libre : WOLF (2)

Méthode 2 : alignement + Balkanet

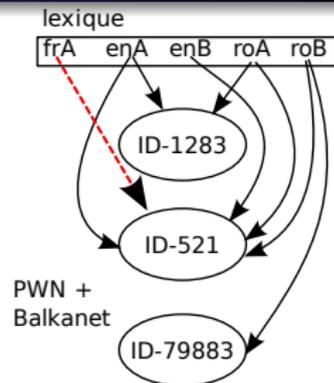
- corpus aligné en phrases : JRC (français, anglais, roumain, tchèque et bulgare) ;
- lemmatisation par TT ;
- Uplug(GIZA++) → alignement en mots, construction d'un lexique multilingue ;
- construction des synsets français



Créer une ressource libre : WOLF (2)

Méthode 2 : alignement + Balkanet

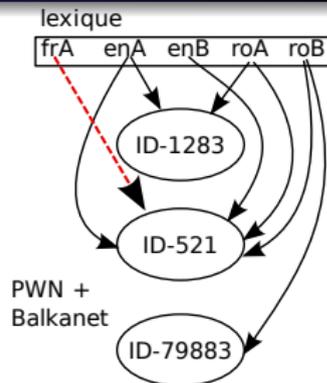
- corpus aligné en phrases : JRC (français, anglais, roumain, tchèque et bulgare) ;
- lemmatisation par TT ;
- Uplug(GIZA++) → alignement en mots, construction d'un lexique multilingue ;
- construction des synsets français



Créer une ressource libre : WOLF (2)

Méthode 2 : alignement + Balkanet

- corpus aligné en phrases : JRC (français, anglais, roumain, tchèque et bulgare) ;
- lemmatisation par TT ;
- Uplug(GIZA++) → alignement en mots, construction d'un lexique multilingue ;
- construction des synsets français



Fusion

- 32000 synsets/38000 littéraux (N, V, Adj, Adv) ;
- import des synsets de PWN pour boucher les « trous de la hiérarchie » et projection des relations de PWN
- évaluation auto WOLF vs EWN. Noms : P=80%, R=74% ;
- évaluation manuelle sur 100 littéraux → absences dans EWN.

WOLF (3) - exemple

```
<SYNSET>
<ID> ENG20-00661907-v</ID>
<POS>v</POS>
<SYNONYM>
  <LITERAL>comprendre<SENSE>122/5:fr.csbgen,fr.csen,fr.rocsbgen,fr.rocsen,fr.roen</SENSE></LITERAL>
  <LITERAL>inclure<SENSE>180/5:fr.csbgen,fr.csen,fr.rocsbgen,fr.rocsen,fr.roen</SENSE></LITERAL>
  <LITERAL>inscrire<SENSE>12/3:fr.csbgen,fr.csen,fr.rocsbgen,fr.rocsen,fr.roen</SENSE></LITERAL>
</SYNONYM>
<ILR><TYPE>hypernym</TYPE>ENG20-00667530-v</ILR>
<ILR><TYPE>near_antonym</TYPE>ENG20-00595585-v</ILR>
<ILR><TYPE>eng_derivative</TYPE>ENG20-00356412-n</ILR>
<DEF>consider as part of something</DEF>
<USAGE>I include you in the list of culprits</USAGE>
<BCS>2</BCS>
<DOMAIN>factotum</DOMAIN>
<SUMO>Classifying<TYPE>=</TYPE></SUMO>
</SYNSET>
<SYNSET>
<ID> ENG20-00570184-v</ID>
<POS>v</POS>
<SYNONYM>
  <LITERAL>entendre<SENSE>6/1:fr.roen</SENSE></LITERAL>
  <LITERAL>comprendre<SENSE>17/5:fr.csbgen,fr.csen,fr.rocsbgen,fr.rocsen,fr.roen</SENSE></LITERAL>
</SYNONYM>
<ILR><TYPE>eng_derivative</TYPE>ENG20-05469809-n</ILR>
<DEF>know and comprehend the nature or meaning of</DEF>
<USAGE>She did not understand her husband</USAGE>
<USAGE>I understand what she means</USAGE>
<BCS>3</BCS>
<DOMAIN>psychology</DOMAIN>
<SUMO>Interpreting<TYPE>=</TYPE></SUMO>
</SYNSET>
```

Diagram illustrating relationships between the two synsets:

- Red arrow from `<ILR><TYPE>near_antonym</TYPE>ENG20-00595585-v</ILR>` to `exclude`
- Red arrow from `<DEF>consider as part of something</DEF>` to `inscription, inclusion`
- Red arrow from `<ILR><TYPE>hypernym</TYPE>ENG20-00667530-v</ILR>` to `concerner, considérer, estimer, juger`

WOLF (4) - exemple

```
<SYNSEMANTIC>
<ID>ENG20-01382171-v</ID>
<POS>v</POS>
<SYNONYM>
<LITERAL>raccorder<SENSE>0/1:frwiktionary</SENSE></LITERAL>
<LITERAL>boucher<SENSE>0/1:frwiktionary</SENSE></LITERAL>
</SYNONYM>
<ILR><TYPE>hypernym</TYPE>ENG20-01382570-v</ILR>
<ILR><TYPE>subevent</TYPE>ENG20-01380907-v</ILR>
<ILR><TYPE>eng_derivative</TYPE>ENG20-03817703-n</ILR>
<DEF>fill or close tightly with or as if with a plug</DEF>
<USAGE>plug the hole</USAGE>
<USAGE>stop up the leak</USAGE>
<DOMAIN>factotum</DOMAIN>
<SUMO>Putting<TYPE>+</TYPE></SUMO>
</SYNSEMANTIC>
<SYNSEMANTIC>
<ID>ENG20-01436675-v</ID>
<POS>v</POS>
<SYNONYM>
<LITERAL>encombrer<SENSE>0/1:frwiktionary</SENSE></LITERAL>
<LITERAL>raccorder<SENSE>0/1:frwiktionary</SENSE></LITERAL>
<LITERAL>boucher<SENSE>0/1:frwiktionary</SENSE></LITERAL>
</SYNONYM>
<ILR><TYPE>hypernym</TYPE>ENG20-01434634-v</ILR>
<ILR><TYPE>near_antonym</TYPE>ENG20-01437192-v</ILR>
<ILR><TYPE>eng_derivative</TYPE>ENG20-02938166-n</ILR>
<ILR><TYPE>eng_derivative</TYPE>ENG20-12676976-n</ILR>
<ILR><TYPE>eng_derivative</TYPE>ENG20-13153074-n</ILR>
<DEF>become or cause to become obstructed</DEF>
<USAGE>The leaves clog our drains in the Fall</USAGE>
<USAGE>The water pipe is backed up</USAGE>
<DOMAIN>factotum</DOMAIN>
<SUMO>Process<TYPE>+</TYPE></SUMO>
</SYNSEMANTIC>
```

WOLF (5) - exemple

```
<SYNSET >
  <ID>ENG20-07193347-n</ID>
  <POS>n</POS>
  <SYNONYM >
    <LITERAL>bœuf haché</SENSE> 0/1:enwikipedia</SENSE> </LITERAL>
  </SYNONYM >
  <ILR> <TYPE>hypernym</TYPE>ENG20-07192932-n</ILR>
  <ILR> <TYPE>holo_part</TYPE>ENG20-07225964-n</ILR>
  <DEF>fait à partir de morceaux de bœuf, haché soit par le boucher soit industriellement</DEF>
  <DOMAIN>gastronomy</DOMAIN>
  <SUMO>Meat<TYPE>+ </TYPE></SUMO>
</SYNSET >
<SYNSET >
  <ID>ENG20-09446276-n</ID>
  <POS>n</POS>
  <SYNONYM >
    <LITERAL>poissonnerie</SENSE> 0/1:enwikipedia</SENSE> </LITERAL>
  </SYNONYM >
  <ILR> <TYPE>hypernym</TYPE>ENG20-10021153-n</ILR>
  <DEF>marché de poissons et de fruits de mer, alors que la boucherie en est un de
    viandes d'oiseaux et de mammifères</DEF>
  <DOMAIN>commerce</DOMAIN>
  <SUMO>Position<TYPE>+ </TYPE></SUMO>
</SYNSET >
```

WOLF (5) - exemple

```
<SYNSET>
  <ID>ENG20-07193347-n</ID>
  <POS>n</POS>
  <SYNONYM>
    <LITERAL>bœuf haché</LITERAL><SENSE>0/1:enwikipedia</SENSE></LITERAL>
  </SYNONYM>
  <ILR><TYPE>hypernym</TYPE>ENG20-07192932-n</ILR>
  <ILR><TYPE>holo_part</TYPE>ENG20-07225964-n</ILR>
  <DEF>fait à partir de morceaux de bœuf, haché soit par le boucher soit industriellement</DEF>
  <DOMAIN>gastronomy</DOMAIN>
  <SUMO>Meat<TYPE>+</TYPE></SUMO>
  </SYNSET>
  <SYNSET>
    <ID>ENG20-07225964-n</ID>
    <POS>n</POS>
    <SYNONYM></SYNONYM>
    <ILR><TYPE>hypernym</TYPE>ENG20-07224829-n</ILR>
    <DEF>a fried cake of minced beef served on a bun</DEF>
    <DOMAIN>gastronomy</DOMAIN>
    <SUMO>Food<TYPE>+</TYPE></SUMO>
  </SYNSET>
  <DEF>marché de poissons et de fruits de mer, alors que la boucherie en est un de
    viandes d'oiseaux et de mammifères</DEF>
  <DOMAIN>commerce</DOMAIN>
  <SUMO>Position<TYPE>+</TYPE></SUMO>
</SYNSET>
```

- 1 Réseaux sémantiques : tour d'horizon
 - Princeton WordNet
 - EuroWordNet
 - Améliorer l'existant
 - Recréer l'existant
- 2 Wiktionary et al : réseaux de synonymie
 - Présentation de Wiktionary
 - Réseau de synonymie
 - Ressources étalons
 - Comparaison
- 3 Densification du réseau
 - En exploitant sa structure de SW
 - En utilisant les traductions
- 4 Conclusions et travaux futurs

Wiktionary

Caractéristiques

- multilingue ;
- étymologie, POS, défs, exemples, relations sémantiques, traductions ;
- accessible librement (version téléchargeable, licence CC By-SA).

Un bon candidat ? Oui, mais...

- pas de validation par des experts ;
- des incohérences ;
- ressource non triviale à extraire.

Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »

Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »
- <http://fr.wiktionary.org/wiki/boucher>
(article "boucher" en français) pointe vers :



Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »
- <http://fr.wiktionary.org/wiki/boucher>
(article "boucher" en français) pointe vers :
 - <http://en.wiktionary.org/wiki/boucher>
article **en anglais** sur **boucher** via lien interlangue



Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »
- <http://fr.wiktionary.org/wiki/boucher>
(article "boucher" en français) pointe vers :
 - <http://en.wiktionary.org/wiki/boucher>
article **en anglais** sur **boucher** *via* lien interlangue
 - et : <http://fr.wiktionary.org/wiki/butcher#Anglais>
article **en français** sur **butcher** *via* section *traductions*
(*butcher* entrée du wiktionnaire français !)



Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »
- <http://fr.wiktionary.org/wiki/boucher>
(article "boucher" en français) pointe vers :
 - <http://en.wiktionary.org/wiki/boucher>
article **en anglais** sur **boucher** *via* lien interlangue
 - et : <http://fr.wiktionary.org/wiki/butcher#Anglais>
article **en français** sur **butcher** *via* section *traductions*
(*butcher* entrée du wiktionnaire français !)
 - mais pas vers : <http://en.wiktionary.org/wiki/butcher>
(article **en anglais** sur **butcher**)



Multilinguisme, une taille... gonflée !

- 2009 : « *Actuellement 1 542 003 articles décrivent en français les mots de plus de 700 langues* »
- <http://fr.wiktionary.org/wiki/boucher>
(article "boucher" en français) pointe vers :
 - <http://en.wiktionary.org/wiki/boucher>
article **en anglais** sur **boucher** *via* lien interlangue
 - et : <http://fr.wiktionary.org/wiki/butcher#Anglais>
article **en français** sur **butcher** *via* section *traductions*
(*butcher* entrée du wiktionnaire français !)
 - mais pas vers : <http://en.wiktionary.org/wiki/butcher>
(article **en anglais** sur **butcher**)



	Total	Meta	Lang. et autres	Pertinents	
fr	728,266	25,244	369,948	337,074	46%
en	905,963	46,202	667,430	192,331	21%
de	88,912	7,235	49,672	32,005	36%
pl	110,369	4,975	95,241	10,153	9%
zh	131,752	8,195	112,520	1,037	0.7%

(2008)

Contenu et structure

Contenu des articles

boot

Etymology 1

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe

- * French: botte
- * Spanish : bota

kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie

boot

Etymology 1

etymology

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe

- * French: botte
- * Spanish : bota

kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours

boot

Etymology 1

etymology

Middle English, from Old French *bote*

Noun

part of speech (noun)

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe

- * French: botte
- * Spanish : bota

kick - see kick

Verb

part of speech (verb)

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples

boot

Etymology 1 *etymology*
Middle English, from Old French *bote*

Noun *part of speech (noun)*

1. A heavy shoe *wordsenses*
2. A blow with the foot; a kick.

Synonyms
* (*shoe*): buskin, mukluk
* (*blow with foot*): kick

Translations
shoe
* French: botte
* Spanish: bota
kick - see kick

Verb *part of speech (verb)*

1. To kick *wordsense #1*
*I **booted** the ball*

2. To disconnect *wordsense #2*
*I got **booted** from the chatroom*

Synonyms
* (*kick*): kick
* (*disconnect*): kick

Translations [...]

Etymology 2
Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques

boot

Etymology 1 *etymology*
Middle English, from Old French *bote*

Noun *part of speech (noun)*

1. A heavy shoe *wordsenses*
2. A blow with the foot; a kick.

Synonyms *synonyms*

- * (shoe): buskin, mukluk
- * (blow with foot): kick

Translations

shoe
* French: botte
* Spanish: bota
kick - see kick

Verb *part of speech (verb)*

1. To kick *wordsense #1*
*I **booted** the ball*

2. To disconnect *wordsense #2*
*I got **booted** from the chatroom*

Synonyms

- * (kick): kick
- * (disconnect): kick

Translations [...]

Etymology 2
Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques
- traductions

boot	
Etymology 1	<i>etymology</i> Middle English, from Old French <i>bote</i>
Noun	<i>part of speech (noun)</i>
1. A heavy shoe	<i>wordsenses</i>
2. A blow with the foot; a kick.	
Synonyms	<i>synonyms</i>
* (shoe): buskin, mukluk	
* (blow with foot): kick	
Translations	<i>translations</i>
shoe	
* French: botte	
* Spanish: bota	
kick - see kick	
Verb	<i>part of speech (verb)</i>
1. To kick	<i>wordsense #1</i>
<i>I booted the ball.</i>	
2. To disconnect	<i>wordsense #2</i>
<i>I got booted from the chatroom.</i>	
Synonyms	
* (kick): kick	
* (disconnect): kick	
Translations [...]	
Etymology 2	
Akin to Old Norse <i>bót</i>	

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques
- traductions

Le cas "régulier", mais...

- contenu & structure hétérogènes d'un langage à l'autre et même au sein d'un même langage (déviations par rapport à un standard qui n'existe pas).

boot

Etymology 1

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe

- * French: botte
- * Spanish : bota

kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques
- traductions

Le cas "régulier", mais...

- contenu & structure hétérogènes d'un langage à l'autre et même au sein d'un même langage (déviations par rapport à un standard qui n'existe pas).
- $X \in \text{syns}(Y) \not\Rightarrow Y \in \text{syns}(X)$,
 $X \in \text{trads}(Y) \not\Rightarrow Y \in \text{trads}(X)$, etc.

boot

Etymology 1

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

- shoe
- * French: botte
 - * Spanish : bota
- kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques
- traductions

Le cas "régulier", mais...

- contenu & structure hétérogènes d'un langage à l'autre et même au sein d'un même langage (déviations par rapport à un standard qui n'existe pas).
- $X \in \text{syns}(Y) \not\Rightarrow Y \in \text{syns}(X)$,
 $X \in \text{trads}(Y) \not\Rightarrow Y \in \text{trads}(X)$, etc.
- parfois (eg. en italien), une partie relation sém (eg. syn) correspond à tous les POS.

boot

Etymology 1

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe
* French: botte
* Spanish : bota
kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Contenu et structure

Contenu des articles

- étymologie
- parties du discours
- définitions, exemples
- relations sémantiques
- traductions

Le cas "régulier", mais...

- contenu & structure hétérogènes d'un langage à l'autre et même au sein d'un même langage (déviations par rapport à un standard qui n'existe pas).
- $X \in \text{syns}(Y) \not\Rightarrow Y \in \text{syns}(X)$,
 $X \in \text{trads}(Y) \not\Rightarrow Y \in \text{trads}(X)$, etc.
- parfois (eg. en italien), une partie relation sém (eg. syn) correspond à tous les POS.
- gestion des sous-sens... fantaisiste.

boot

Etymology 1

Middle English, from Old French *bote*

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Translations

shoe

- * French: botte
- * Spanish : bota

kick - see kick

Verb

1. To kick
*I **booted** the ball*
2. To disconnect
*I got **booted** from the chatroom*

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Translations [...]

Etymology 2

Akin to Old Norse *bót*

Extraction du réseau de synonymie de Wiktionary

Modélisation

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Verb

1. To kick
2. To disconnect

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Verb

1. To kick
2. To disconnect

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

en.N.boot



en.V.boot



Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Verb

1. To kick
2. To disconnect

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick

en.N.boot



en.N.buskin



en.N.mukluk

en.V.boot



en.N.kick



en.V.kick

Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

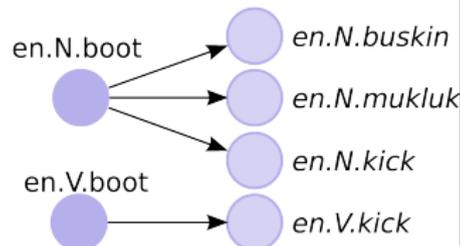
- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Verb

1. To kick
2. To disconnect

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick



Extraction du réseau de synonymie de Wiktionary

Modélisation

- chaque POS d'une entrée
→ un sommet
- liens de synonymie au sein d'une
POS donnée
- sous-sens aplatis
- liens symétrisés

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

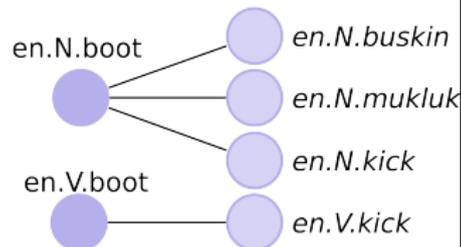
- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Verb

1. To kick
2. To disconnect

Synonyms

- * (*kick*): kick
- * (*disconnect*): kick



Sous-sens aplatis. . .

Pourquoi ?

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

- * (*shoe*): buskin, mukluk
- * (*blow with foot*): kick

?

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

boot

English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick

en.N.buskin

en.N.boot

en.N.mukluk

en.N.kick

Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

boot

English

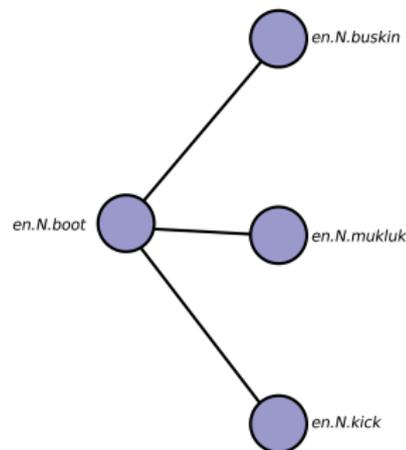
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

boot

English

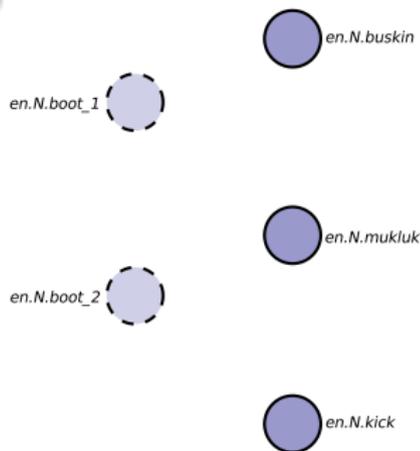
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

boot

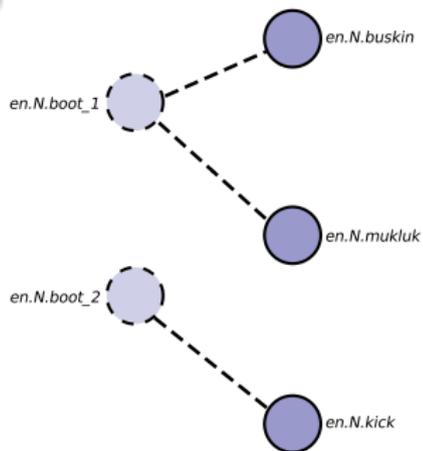
English

Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

? * buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

buskin (N)

- 1 A half-boot
- 2 A type of boot worn by the ancient Athenian tragic actors

mukluk (N)

- 1 A soft boot made of reindeer skin or sealskin and worn by Inuit.

kick (N)

- 1 A hit or strike with the leg or foot
- 2 The action of swinging a foot or leg
- 3 Sth that tickles the fancy
- 4 (Internet) The removal of a person from an online activity
- 5 (figuratively) Any bucking motion of an object that lacks legs or feet

boot

English

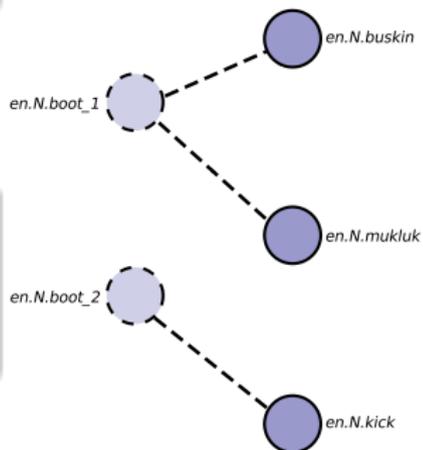
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

buskin (N)

- 1 A half-boot
- 2 A type of boot worn by the ancient Athenian tragic actors

mukluk (N)

- 1 A soft boot made of reindeer skin or sealskin and worn by Inuit.

kick (N)

- 1 A hit or strike with the leg or foot
- 2 The action of swinging a foot or leg
- 3 Sth that tickles the fancy
- 4 (Internet) The removal of a person from an online activity
- 5 (figuratively) Any bucking motion of an object that lacks legs or feet

boot

English

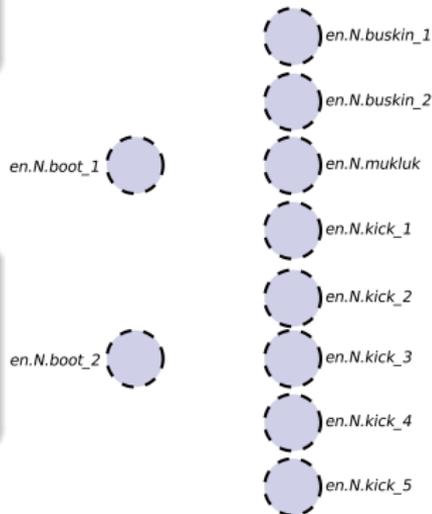
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

buskin (N)

- 1 A half-boot
- 2 A type of boot worn by the ancient Athenian tragic actors

mukluk (N)

- 1 A soft boot made of reindeer skin or sealskin and worn by Inuit.

kick (N)

- 1 A hit or strike with the leg or foot
- 2 The action of swinging a foot or leg
- 3 Sth that tickles the fancy
- 4 (Internet) The removal of a person from an online activity
- 5 (figuratively) Any bucking motion of an object that lacks legs or feet

boot

English

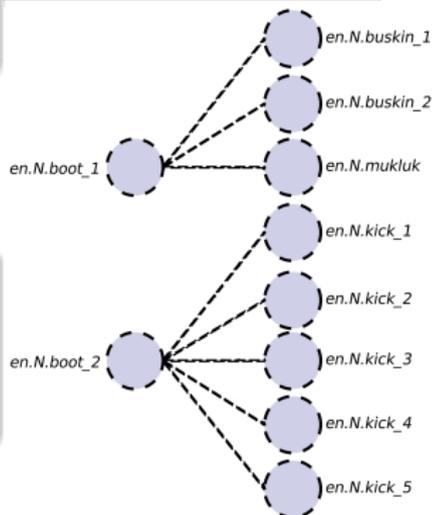
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?
?
?

buskin, mukluk
kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

buskin (N)

- 1 A half-boot
- 2 A type of boot worn by the ancient Athenian tragic actors

mukluk (N)

- 1 A soft boot made of reindeer skin or sealskin and worn by Inuit.

kick (N)

- 1 A hit or strike with the leg or foot
- 2 The action of swinging a foot or leg
- 3 Sth that tickles the fancy
- 4 (Internet) The removal of a person from an online activity
- 5 (figuratively) Any bucking motion of an object that lacks legs or feet

boot

English

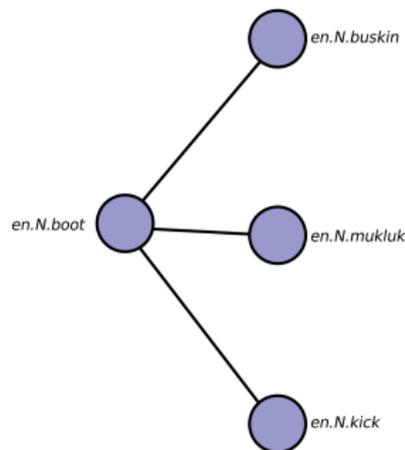
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

* buskin, mukluk
* kick



Sous-sens aplatis. . .

Pourquoi ?

- pas un choix théorique sous-jacent de notre part, mais :
- contraintes imposées par la ressource
- les sous-sens peuvent apparaître dans les définitions et pas dans les relations
- sous-sens du sommet cible ?

buskin (N)

- 1 A half-boot
- 2 A type of boot worn by the ancient Athenian tragic actors

mukluk (N)

- 1 A soft boot made of reindeer skin or sealskin and worn by Inuit.

kick (N)

- 1 A hit or strike with the leg or foot
- 2 The action of swinging a foot or leg
- 3 Sth that tickles the fancy
- 4 (Internet) The removal of a person from an online activity
- 5 (figuratively) Any bucking motion of an object that lacks legs or feet

Autre raison :

Un de nos étalons (Dicosyn) a ses sous-sens aplatis

boot

English

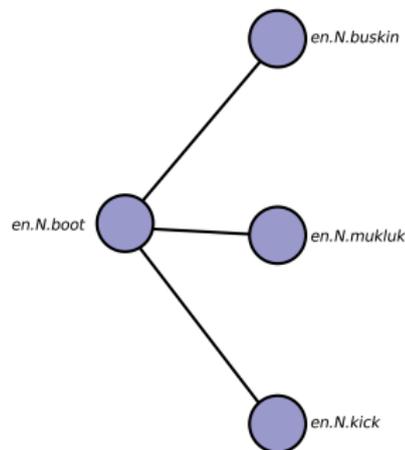
Noun

1. A heavy shoe
2. A blow with the foot; a kick.

Synonyms

?

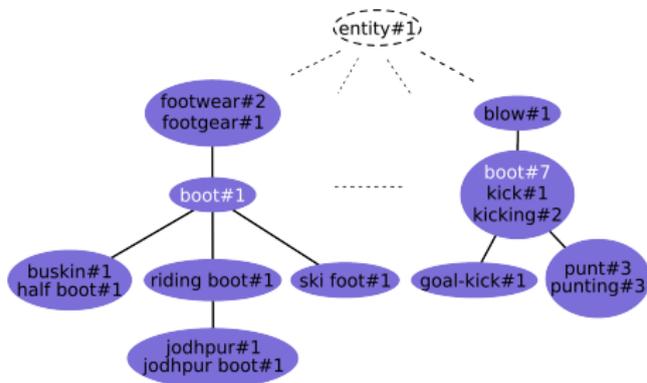
* buskin, mukluk
* kick



Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné



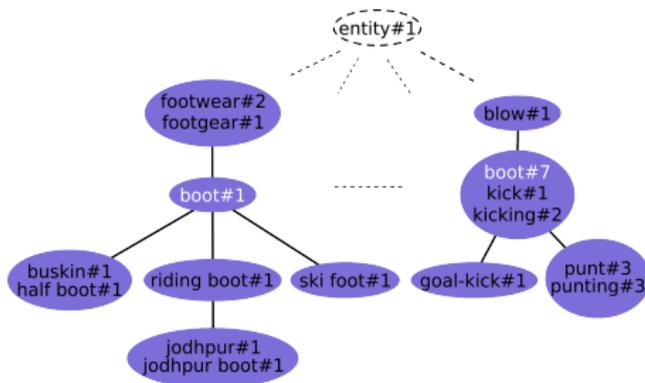
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets \leftrightarrow lexèmes
- lien entre tous les mots d'un synset donné



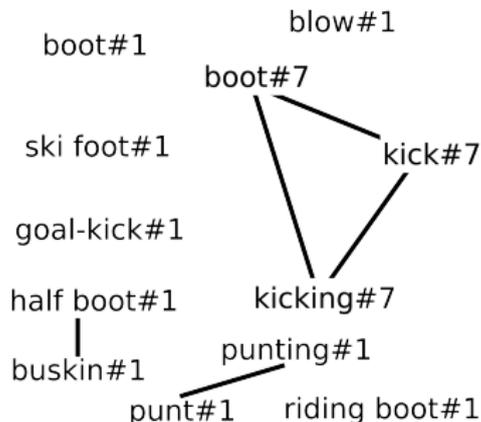
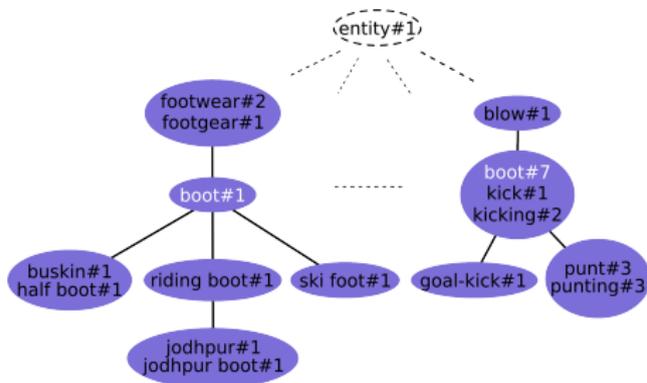
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets ↔ lexèmes
- lien entre tous les mots d'un synset donné



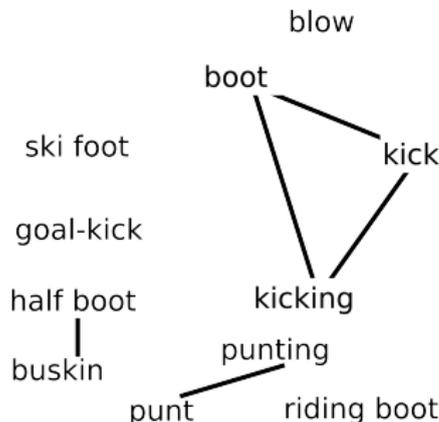
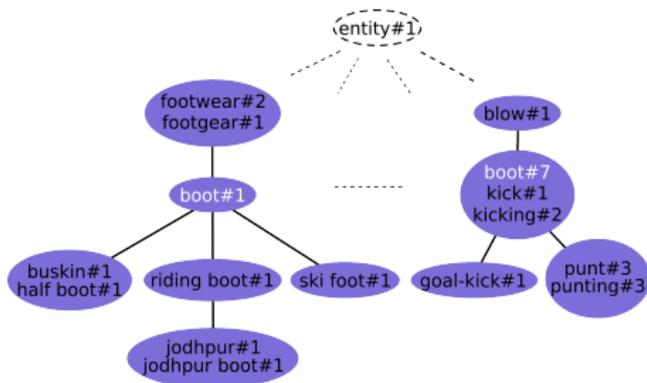
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets \leftrightarrow lexèmes
- lien entre tous les mots d'un synset donné



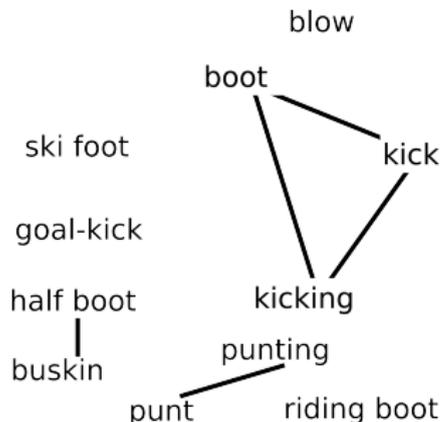
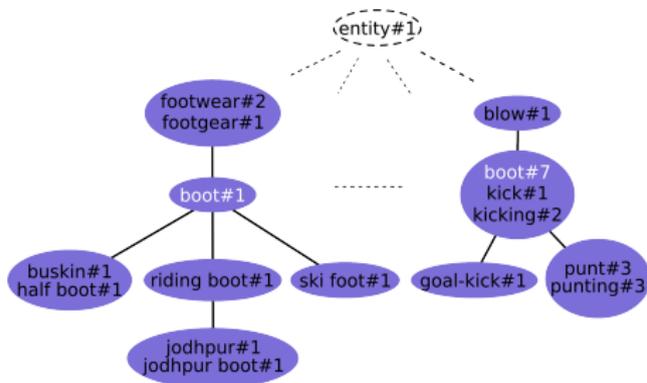
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets ↔ lexèmes
- lien entre tous les mots d'un synset donné
- + recours à l'hyponymie pour les synsets "feuilles" contenant un seul mot



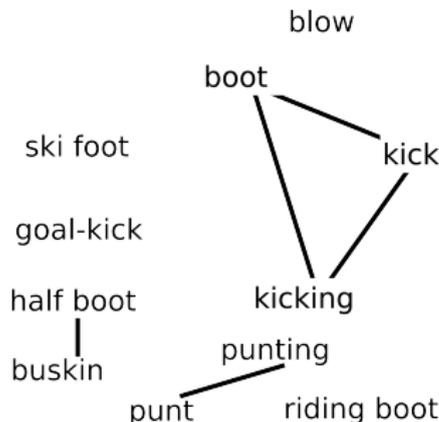
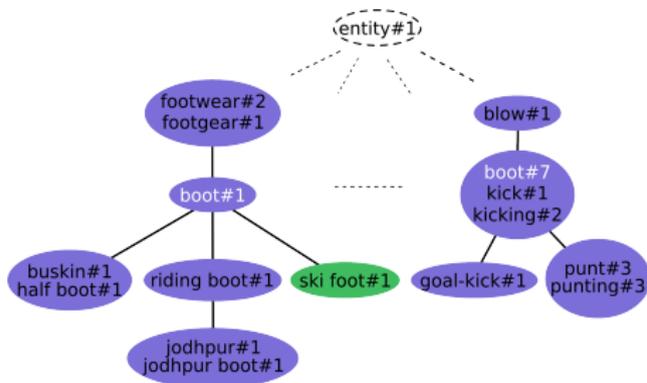
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets ↔ lexèmes
- lien entre tous les mots d'un synset donné
- + recours à l'hyponymie pour les synsets "feuilles" contenant un seul mot



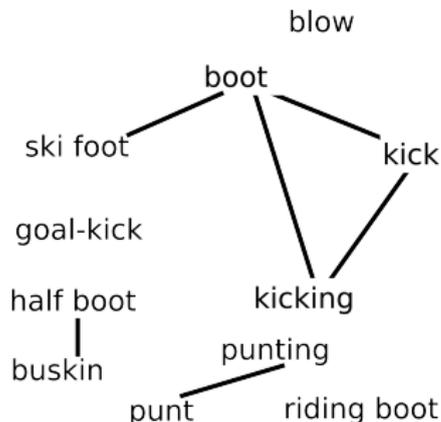
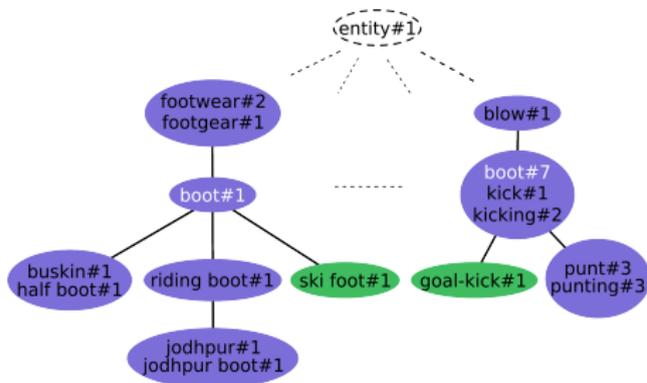
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets ↔ lexèmes
- lien entre tous les mots d'un synset donné
- + recours à l'hyponymie pour les synsets "feuilles" contenant un seul mot



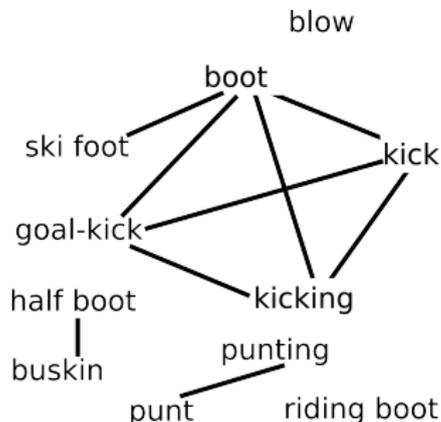
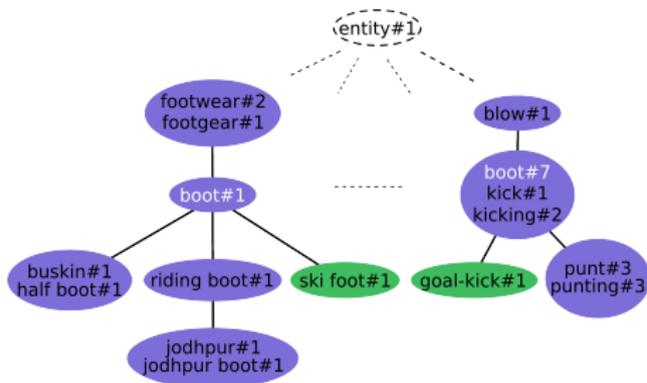
Extraction du réseau de synonymie de WordNet

WordNet

- synonymie entre sous-sens
- relations déjà symétriques
- même POS dans un synset donné

Modélisation

- sommets \leftrightarrow lexèmes
- lien entre tous les mots d'un synset donné
- + recours à l'hyponymie pour les synsets "feuilles" contenant un seul mot



Extraction du réseau de synonymie de Dicosyn

Dicosyn

- compilation des relations de synonymie extraites de 7 dictionnaires (*Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert*);
- créé par l'ATILF, corrigé au CRISCO :
<http://elsap1.unicaen.fr/dicosyn.html>
- sous-sens écrasés ;
- parties du discours aussi ! ;
- → recatégoriser et symétriser.

Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;

Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;

Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

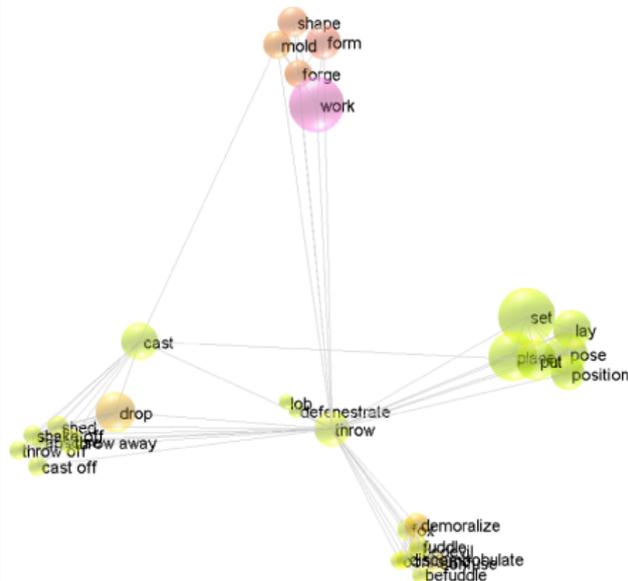
- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.

Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.

Eg. : to throw (WordNet)

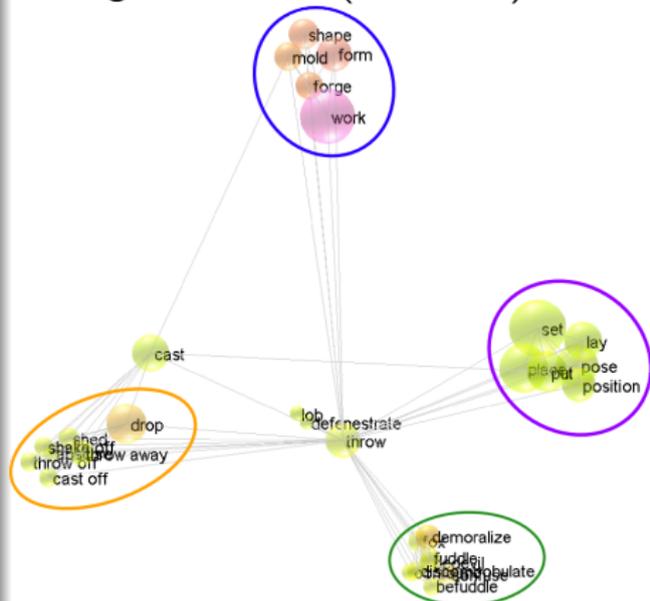


Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.

Eg. : to throw (WordNet)

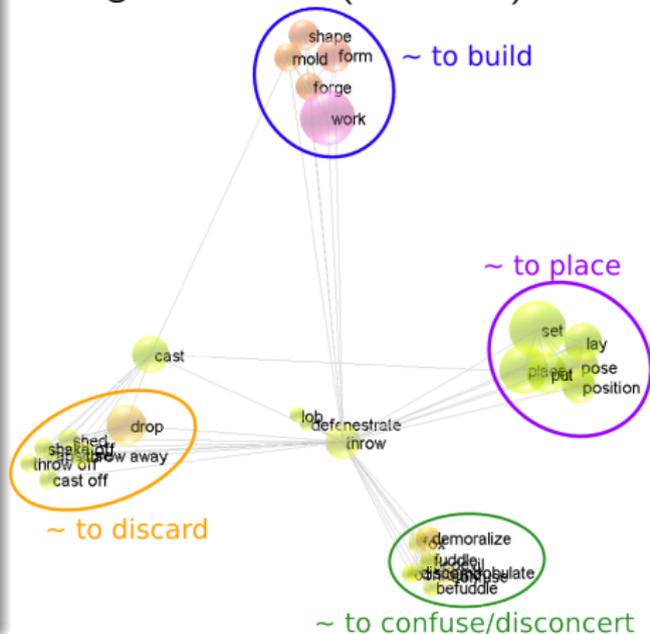


Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.

Eg. : to throw (WordNet)

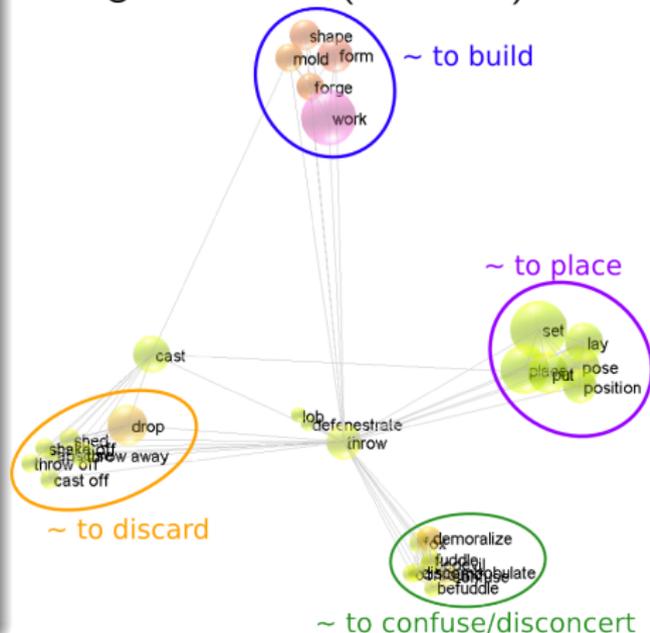


Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.
- Wiktionary, WordNet et Dicosyn sont des SW

Eg. : to throw (WordNet)

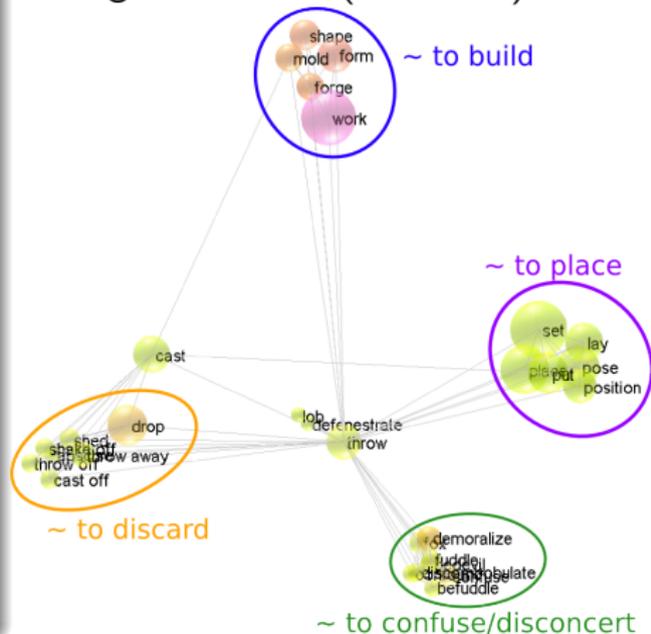


Small Worlds (SW)

Les réseaux sémantiques sont
(souvent) des SW

- globalement creux, localement denses (fort clustering) ;
- longueur moyenne des chemins entre 2 sommets courte ;
- distribution des sommets d'incidence : loi de puissance.
- Wiktionary, WordNet et Dicosyn sont des SW
- → on va tirer parti de leur structure clusterisée

Eg. : to throw (WordNet)

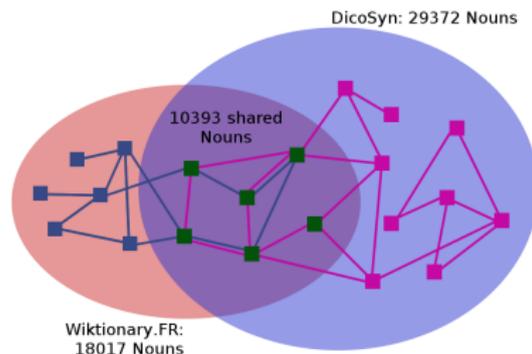


Wiktionary FR/Dicosyn

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	DicoSyn	Inter.	P	R
N	18017	29372	10393	58%	35%
A	5411	9452	3076	57%	33%
V	3897	9147	2966	76%	32%

Exemple : le réseau des noms



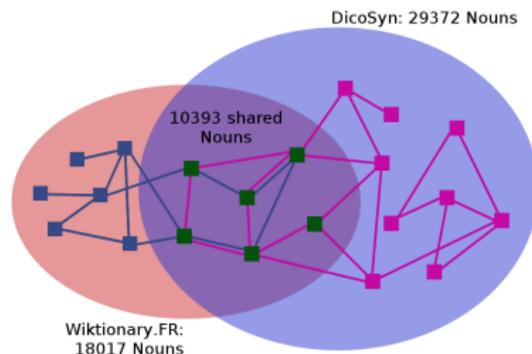
Wiktionary FR/Dicosyn

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	DicoSyn	Inter.	P	R
N	18017	29372	10393	58%	35%
A	5411	9452	3076	57%	33%
V	3897	9147	2966	76%	32%

Relations

Exemple : le réseau des noms



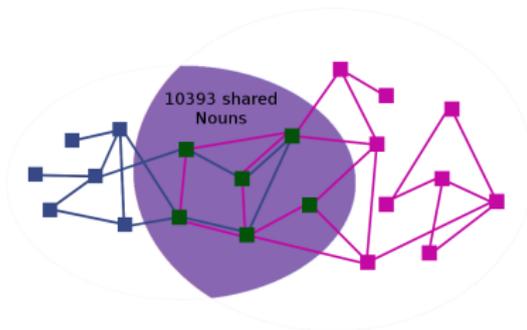
Wiktionary FR/Dicosyn

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	DicoSyn	Inter.	P	R
N	18017	29372	10393	58%	35%
A	5411	9452	3076	57%	33%
V	3897	9147	2966	76%	32%

Relations

Exemple : le réseau des noms



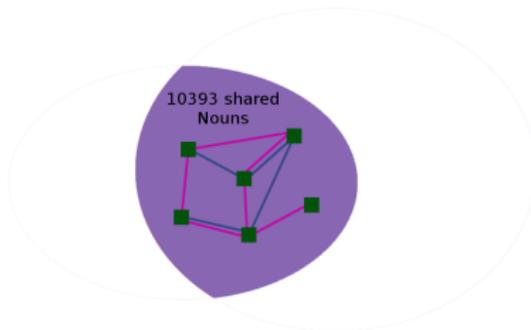
Wiktionary FR/Dicosyn

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	DicoSyn	Inter.	P	R
N	18017	29372	10393	58%	35%
A	5411	9452	3076	57%	33%
V	3897	9147	2966	76%	32%

Relations

Exemple : le réseau des noms



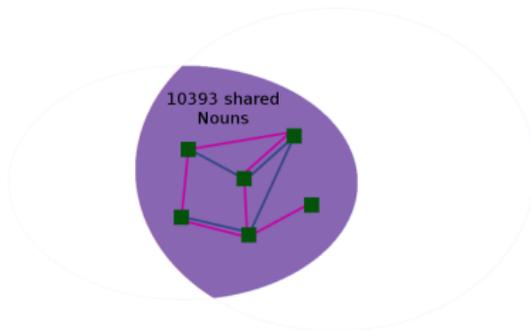
Wiktionary FR/Dicosyn

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	DicoSyn	Inter.	P	R
N	18017	29372	10393	58%	35%
A	5411	9452	3076	57%	33%
V	3897	9147	2966	76%	32%

	Relations				
	Wikt.	DicoSyn	Inter.	P	R
N	3510	44501	3510	69%	8%
A	1300	17404	1677	78%	7%
V	899	23968	1267	71%	4%

Exemple : le réseau des noms



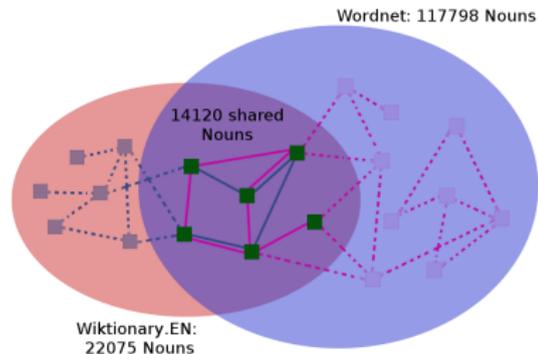
Wiktionary EN/WordNet

Couverture lexicale/Réseau de synonymie

	Mots				
	Wikt.	WordNet	Inter	P	R
N	22075	117798	14120	64%	12%
A	8437	21479	5874	70%	27%
V	6368	11529	5157	81%	45%

	Relations				
	Wikt.	Wordnet	Inter.	P	R
N	6453	18440	2763	43%	15%
A	3139	12792	1314	42%	10%
V	2667	18725	993	37%	5%

Exemple : le réseau des noms



Commentaire préliminaire

Étalons, precision&rappel

- une comparaison un peu grossière, mais :

Commentaire préliminaire

Étalons, precision&rappel

- une comparaison un peu grossière, mais :
- à quel point ces mesures sont-elle significatives (dans ce cas) ?

Commentaire préliminaire

Étalons, precision&rappel

- une comparaison un peu grossière, mais :
- à quel point ces mesures sont-elle significatives (dans ce cas) ?
- ↔ à quel point les étalons sont-ils parfaits ?

Commentaire préliminaire

Étalons, precision&rappel

- une comparaison un peu grossière, mais :
- à quel point ces mesures sont-elle significatives (dans ce cas) ?
- \leftrightarrow à quel point les étalons sont-ils parfaits ?
- exemples :
 - '*to horn*' \in WN et \notin Wikt.en.N₂₀₀₈, apparaît en 2009
 - '*to poo*', '*to prefetch*', '*to google*' \in Wikt et \notin WN \rightarrow bruit ?
 - relations réellement manquantes : $\{to\ act \leftrightarrow to\ play\} \in$ WN et \notin Wikt
 - mais : $\{to\ reduce \leftrightarrow to\ decrease\}$ et $\{to\ cook \leftrightarrow to\ microwave\} \in$ Wikt et \notin WN \rightarrow bruit ?

Commentaire préliminaire

Étalons, precision&rapport

- une comparaison un peu grossière, mais :
- à quel point ces mesures sont-elle significatives (dans ce cas) ?
- \leftrightarrow à quel point les étalons sont-ils parfaits ?
- exemples :
 - '*to horn*' \in WN et \notin Wikt.en.N₂₀₀₈, apparaît en 2009
 - '*to poo*', '*to prefetch*', '*to google*' \in Wikt et \notin WN \rightarrow bruit ?
 - relations réellement manquantes : $\{to\ act \leftrightarrow to\ play\} \in$ WN et \notin Wikt
 - mais : $\{to\ reduce \leftrightarrow to\ decrease\}$ et $\{to\ cook \leftrightarrow to\ microwave\} \in$ Wikt et \notin WN \rightarrow bruit ?
- \rightarrow (*on suppose qu'*) avec le temps, le rappel va augmenter

Commentaire préliminaire

Étalons, precision&rappel

- une comparaison un peu grossière, mais :
- à quel point ces mesures sont-elle significatives (dans ce cas) ?
- \leftrightarrow à quel point les étalons sont-ils parfaits ?
- exemples :
 - '*to horn*' \in WN et \notin Wikt.en.N₂₀₀₈, apparaît en 2009
 - '*to poo*', '*to prefetch*', '*to google*' \in Wikt et \notin WN \rightarrow bruit ?
 - relations réellement manquantes : $\{to\ act \leftrightarrow to\ play\} \in$ WN et \notin Wikt
 - mais : $\{to\ reduce \leftrightarrow to\ decrease\}$ et $\{to\ cook \leftrightarrow to\ microwave\} \in$ Wikt et \notin WN \rightarrow bruit ?
- \rightarrow (*on suppose qu'*) avec le temps, le rappel va augmenter
- \rightarrow est-il possible d'évaluer (automatiquement) la précision ?

- 1 Réseaux sémantiques : tour d'horizon
 - Princeton WordNet
 - EuroWordNet
 - Améliorer l'existant
 - Recréer l'existant
- 2 Wiktionary et al : réseaux de synonymie
 - Présentation de Wiktionary
 - Réseau de synonymie
 - Ressources étalons
 - Comparaison
- 3 **Densification du réseau**
 - En exploitant sa structure de SW
 - En utilisant les traductions
- 4 Conclusions et travaux futurs

Méthode des voisins

Intuition : la synonymie, une relation transitive ?

- *“Les voisins de mes voisins devraient être dans mon voisinage”*

Méthode des voisins

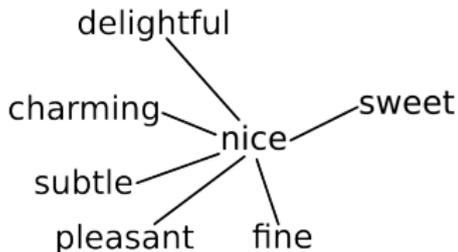
Intuition : la synonymie, une relation transitive ?

- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.

Méthode des voisins

Intuition : la synonymie, une relation transitive ?

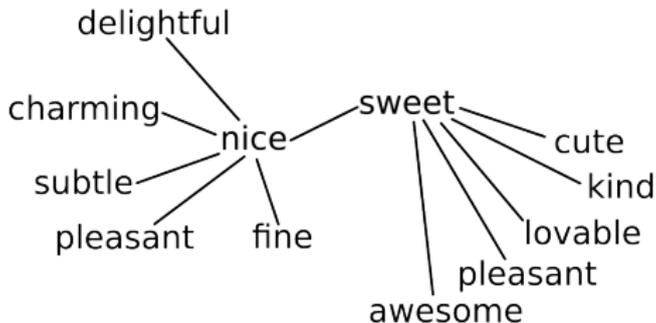
- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.



Méthode des voisins

Intuition : la synonymie, une relation transitive ?

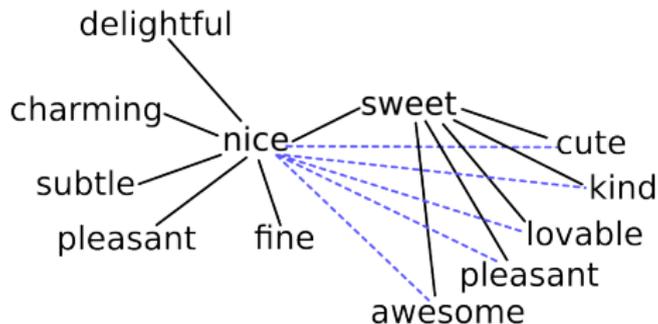
- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.



Méthode des voisins

Intuition : la synonymie, une relation transitive ?

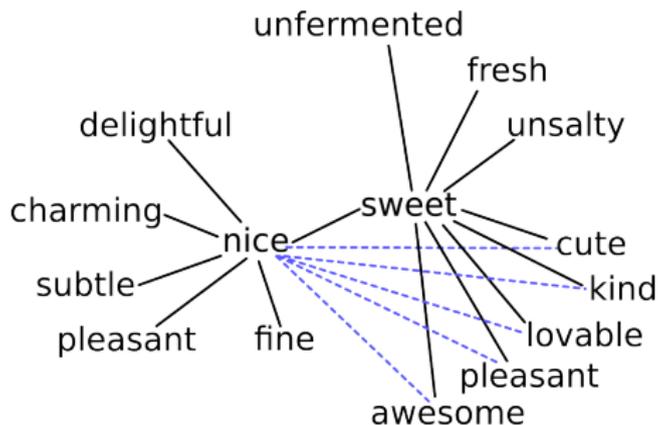
- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.



Méthode des voisins

Intuition : la synonymie, une relation transitive ?

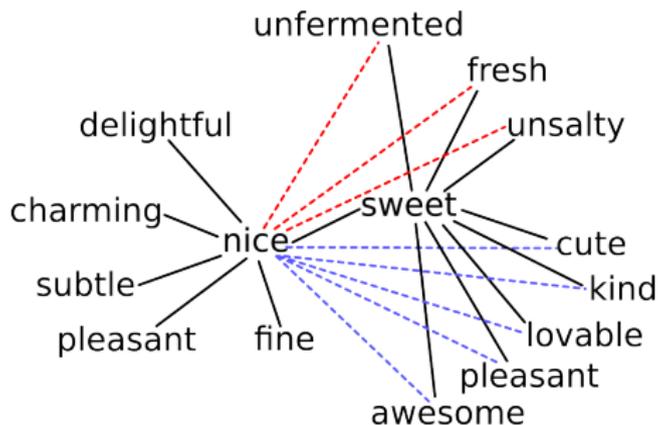
- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.



Méthode des voisins

Intuition : la synonymie, une relation transitive ?

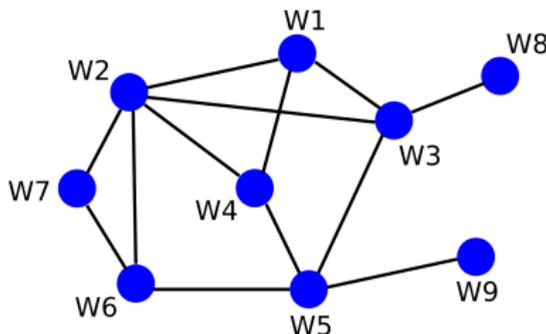
- *“Les voisins de mes voisins devraient être dans mon voisinage”*
- → un voisin de mes voisins qui n'est pas dans mon voisinage devrait être un bon voisin candidat.



Méthode Prox

Prox [Gaume, 2004] : `prox.irit.fr`

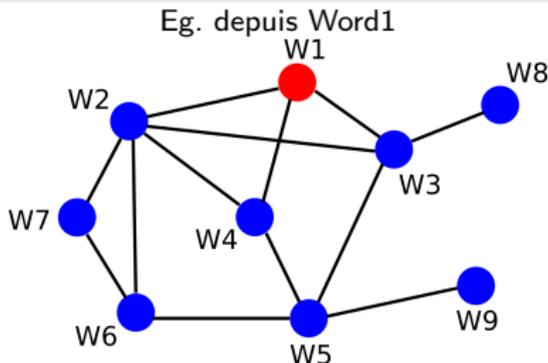
- Une méthode stochastique pour étudier les PM hiérarchiques
- Métrique : pour une paire de sommets (u, v) , on calcule *“la probabilité qu’une particule se déplaçant aléatoirement depuis u atteigne v après k étapes.”*



Méthode Prox

Prox [Gaume, 2004] : `prox.irit.fr`

- Une méthode stochastique pour étudier les PM hiérarchiques
- Métrique : pour une paire de sommets (u, v) , on calcule *“la probabilité qu’une particule se déplaçant aléatoirement depuis u atteigne v après k étapes.”*

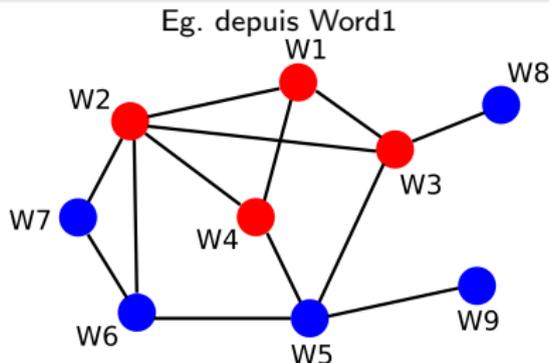


	1	2	3	4	5	6	7	8	9
Initial	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Méthode Prox

Prox [Gaume, 2004] : `prox.irit.fr`

- Une méthode stochastique pour étudier les PM hiérarchiques
- Métrique : pour une paire de sommets (u, v) , on calcule *“la probabilité qu’une particule se déplaçant aléatoirement depuis u atteigne v après k étapes.”*

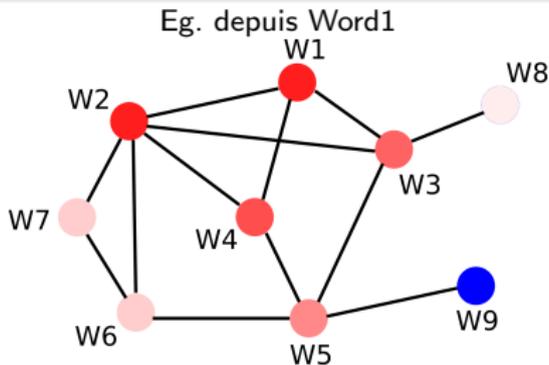


	1	2	3	4	5	6	7	8	9
Step 1	0.25	0.25	0.25	0.25	0.00	0.00	0.00	0.00	0.00

Méthode Prox

Prox [Gaume, 2004] : `prox.irit.fr`

- Une méthode stochastique pour étudier les PM hiérarchiques
- Métrique : pour une paire de sommets (u, v) , on calcule *“la probabilité qu’une particule se déplaçant aléatoirement depuis u atteigne v après k étapes.”*

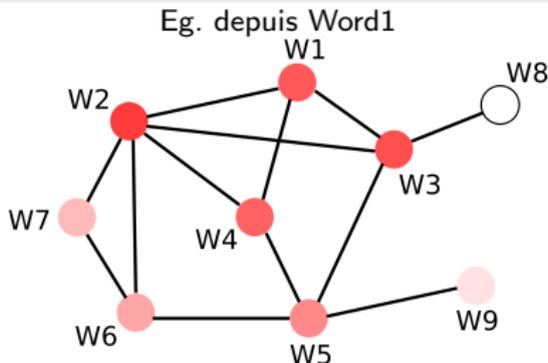


	1	2	3	4	5	6	7	8	9
Step 2	0.22	0.22	0.15	0.17	0.11	0.04	0.04	0.01	0.00

Méthode Prox

Prox [Gaume, 2004] : prox.irit.fr

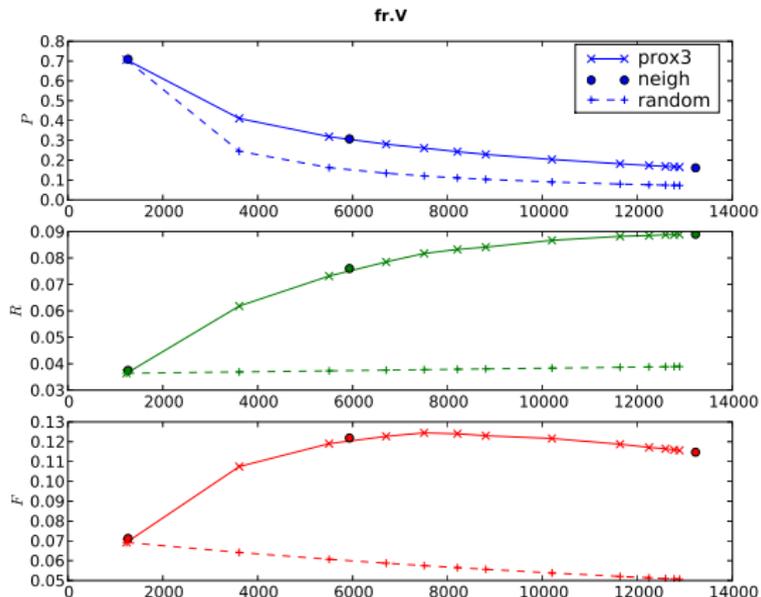
- Une méthode stochastique pour étudier les PM hiérarchiques
- Métrique : pour une paire de sommets (u, v) , on calcule *“la probabilité qu’une particule se déplaçant aléatoirement depuis u atteigne v après k étapes.”*



	1	2	3	4	5	6	7	8	9
Step 3	0.16	0.19	0.17	0.15	0.11	0.08	0.06	0.00	0.02

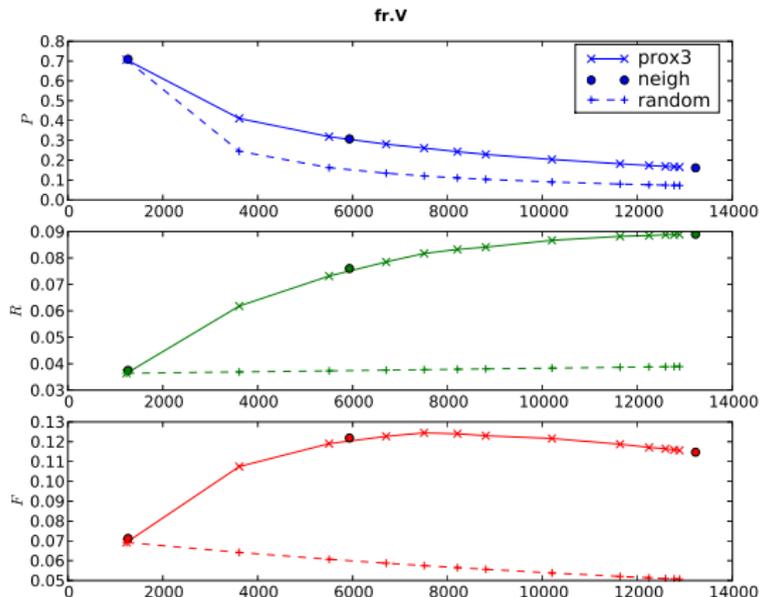
Résultats

Eg. évaluation sur les verbes FR /rap Dicosyn



Résultats

Eg. évaluation sur les verbes FR /rap Dicosyn



Commentaires

- Prox produit des liens pertinents (et ordonnés)
eg. *'to absolve'* ↔ *'to forgive'*, absent from WN
- faux positifs potentiellement intéressants à considérer :
 - *'to uncover'* ↔ *'to peel'* (hyponymie)
 - *'to skin'* ↔ *'to peel'* ('synonymie inter-domaines')

Méthode des liens de traduction

Intuition

*2 mots partageant
beaucoup de
traductions dans
différents langages
sont probablement
synonymes*

Méthode des liens de traduction

Intuition

2 mots partageant beaucoup de traductions dans différents langages sont probablement synonymes

Méthode

- soit T_w l'ensemble des traductions du mot w
- pour chaque paire de mots (w, w') :

$$Jaccard(w, w') = \frac{|T_w \cap T_{w'}|}{|T_w \cup T_{w'}|}$$

- ajout incrémental de relations suivant leur Jaccard, jusqu'à un certain seuil

Méthode des liens de traduction

Intuition

2 mots partageant beaucoup de traductions dans différents langages sont probablement synonymes

Méthode

- soit T_w l'ensemble des traductions du mot w
- pour chaque paire de mots (w, w') :

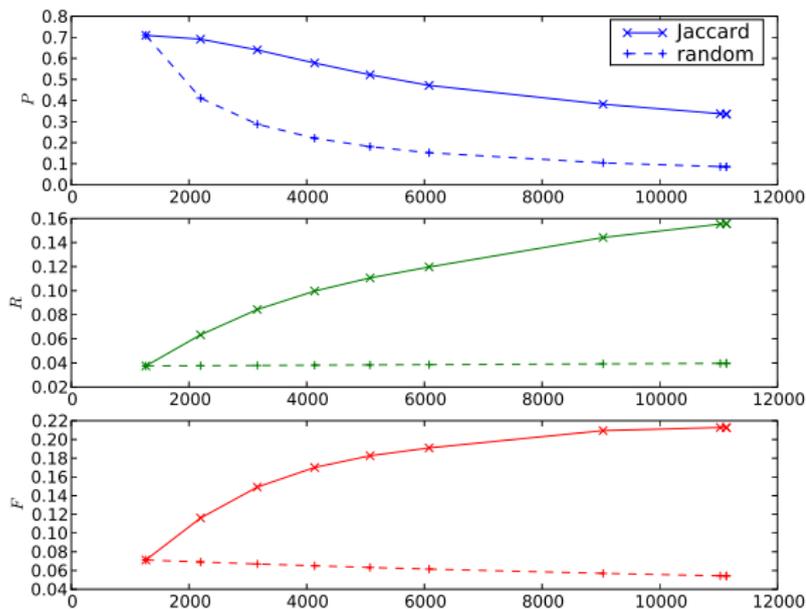
$$Jaccard(w, w') = \frac{|T_w \cap T_{w'}|}{|T_w \cup T_{w'}|}$$

- ajout incrémental de relations suivant leur Jaccard, jusqu'à un certain seuil



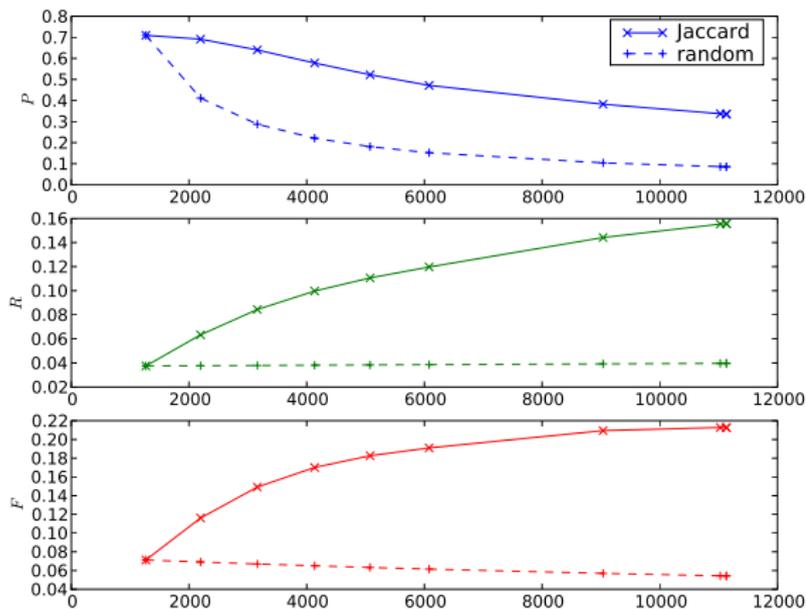
Résultats

figure 2 (French Verb)



Résultats

figure 2 (French Verb)

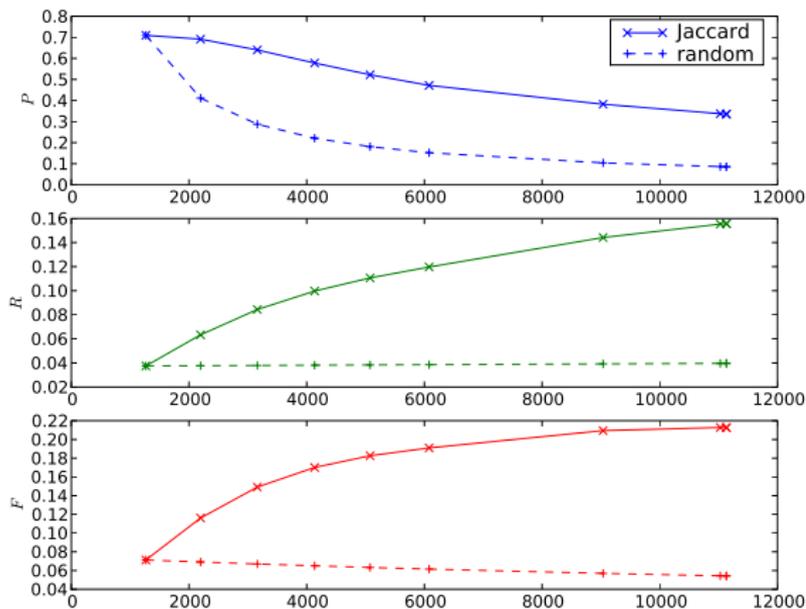


Commentaires

- ajout des 1000 premiers liens (+55%) → perte de 2% de précision
- liens ajoutés avec trads \neq liens ajoutés avec Prox

Résultats

figure 2 (French Verb)



Commentaires

- ajout des 1000 premiers liens (+55%) → perte de 2% de précision
- liens ajoutés avec trads \neq liens ajoutés avec Prox

Idée : trads + Prox

- utiliser la méthode des traductions pour densifier le graphe
- utiliser ensuite la structure des clusters avec Prox

Conclusions

Hypothèses confirmées

- beaucoup de liens manquants devraient être ajoutés entre membres d'un même cluster
- des mots partageant beaucoup de traductions ont de bonnes chances d'être synonymes

Conclusions

Hypothèses confirmées

- beaucoup de liens manquants devraient être ajoutés entre membres d'un même cluster
- des mots partageant beaucoup de traductions ont de bonnes chances d'être synonymes

Nos méthodes marchent. . .

- mais juste un début : grosse marge d'amélioration.
- → dans l'immédiat : combiner les deux méthodes.

Conclusions

Hypothèses confirmées

- beaucoup de liens manquants devraient être ajoutés entre membres d'un même cluster
- des mots partageant beaucoup de traductions ont de bonnes chances d'être synonymes

Nos méthodes marchent. . .

- mais juste un début : grosse marge d'amélioration.
- → dans l'immédiat : combiner les deux méthodes.

Application directe

- support pour l'édition collaborative :
une liste de synonymes ordonnés par pertinence peut être proposée au contributeur
- → module à inclure dans le framework Wiktionary ?

Travaux futurs

Étude en diachronie

- étudier comment évolue Wiktionary
- → prévoir les besoins des contributeurs en terme de TAL

Travaux futurs

Étude en diachronie

- étudier comment évolue Wiktionary
- → prévoir les besoins des contributeurs en terme de TAL

Invariants and variabilité

- étude de (la|l'in)variabilité des paires sémantiques
(Wiktionary comme un réseau de synonymie multilingue)
eg. *house/family, child/fruit, feel/know*

Travaux futurs (2)

Le problème récurrent de l'évaluation. . .

- évaluer les ressources entre elles :
 - toujours la même conclusion : les golds ne sont pas parfaits.
 - cela a-t-il un sens quand conçues dans optiques différentes ?
(→ granularité différente).

Travaux futurs (2)

Le problème récurrent de l'évaluation. . .

- évaluer les ressources entre elles :
 - toujours la même conclusion : les golds ne sont pas parfaits.
 - cela a-t-il un sens quand conçues dans optiques différentes ?
(→ granularité différente).
- évaluation orientée tâche : oui, mais si on veut construire des ressources non-orientées tâches ?

Travaux futurs (2)

Le problème récurrent de l'évaluation. . .

- évaluer les ressources entre elles :
 - toujours la même conclusion : les golds ne sont pas parfaits.
 - cela a-t-il un sens quand conçues dans optiques différentes ?
(→ granularité différente).
- évaluation orientée tâche : oui, mais si on veut construire des ressources non-orientées tâches ?
- évaluation manuelle sur échantillon
 - question « synonymes O/N ? » → accord inter-juges faible [Muller et al., 2006] ;
 - revient à se demander ce qu'est la synonymie
"natural languages abhor absolute synonymy just as nature abhors a vacuum"
(Cruse 1986) cité par [Edmonds and Hirst, 2002]
 - tenter une classification ?

Travaux futurs (3)

Examples of near-synonymic variation [Edmonds and Hirst, 2002]

Type of variation	Example
Abstract dimension	seep : drip
Emphasis	enemy : foe
Denotational, indirect	error : mistake
Denotational, fuzzy	woods : forest
Stylistic, formality	pissed : drunk : inebriated
Stylistic, force	ruin : annihilate
Expressed attitude	skinny : thin : slim, slender
Emotive	daddy : dad : father
Collocational	task : job
Selectional	pass away : die
Subcategorization	give : donate

Travaux futurs (4)

Évaluation (2)

- E&H : classif. pour distinguer les quasi-synonymes ;
- notre pb à nous : classifier entre « (quasi-)synonymes », « autre relation potentiellement pertinente » ou « pas de relation du tout » ;
- tenter une évaluation manuelle avec (à qqch près) :
synonymie / antonymie / hyperonymie / métonymie /
métaphore / troponymie / rien du tout / ne sait pas

Travaux futurs (4)

Évaluation (2)

- E&H : classif. pour distinguer les quasi-synonymes ;
- notre pb à nous : classifier entre « (quasi-)synonymes », « autre relation potentiellement pertinente » ou « pas de relation du tout » ;
- tenter une évaluation manuelle avec (à qqch près) :
synonymie / antonymie / hyperonymie / métonymie /
métaphore / troponymie / rien du tout / ne sait pas

Autre approche possible : Prox + corpus

- projeter des patrons lexico-syntaxiques sur corpus pour extraire des couples candidats
croiser avec les candidats fournis par Prox ;
- patrons pour l'hyperonymie : [Hearst, 1992] et [Morin, 1998]
- pour la synonymie : ?

Bibliographie : autour de WordNet I

- [Fellbaum, 1998a] Fellbaum, C. (1998a). A semantic network of english : the mother of all wordnets. pages 137–148.
- [Fellbaum, 1998b] Fellbaum, C., editor (1998b). *WordNet : An Electronic Lexical Database*. MIT Press.
- [Habert, 1998] Habert, B. (1998). Compte rendu de Christiane Fellbaum (ed.) *WordNet, an electronic lexical database*, The MIT Press, 1998. *TAL*, 39(2) :152–155.
<http://www.limsi.fr/Individu/habert/Publications/Fichiers/habert98a.html>.

Bibliographie : autour de WordNet II

[Habert, 1999] Habert, B. (1999). Compte rendu de Piek Vossen (ed.) *EuroWordNet : a multilingual database with lexical semantic networks*, Dordrecht, Kluwer Academic Publishers, 1998. *TAL*, 40(1) :170–172.

<http://www.limsi.fr/Individu/habert/Publications/Fichiers/habert99b/habert99b.html>.

[Hirst and St Onge, 1998] Hirst, G. and St Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum, C., editor, *WordNet : An Electronic Lexical Database*, pages 305–332. MIT Press.

Bibliographie : autour de WordNet III

- [Jacquin et al., 2007] Jacquin, C., Desmontils, E., and Monceaux, L. (2007). French EuroWordNet Lexical Database Improvements. In *CICLing '07 : Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing*, pages 12–22. Springer-Verlag.
- [Rodríguez et al., 1998] Rodríguez, H., Climent, S., Vossen, P., Bloksma, L., Peters, W., Alonge, A., Bertagna, F., and Roventini, A. (1998). The Top-Down Strategy for Building EuroWordNet : Vocabulary Coverage, Base Concepts and Top Ontology. *Computers and the Humanities*, 32 :117–152.
- [Sagot and Fišer, 2008] Sagot, B. and Fišer, D. (2008). Building a free french wordnet from multilingual resources. In *Proceedings of OntoLex 2008*, Marrackech.
<http://alpage.inria.fr/~sagot/pub/OntoLex08.pdf>.

Bibliographie : autour de WordNet IV

- [Tufiş and Cristea, 2002] Tufiş, D. and Cristea, D. (2002). Methodological issues in building the Romanian Wordnet and consistency checks in BalkaNet. In Christodoulakis, D. N., Kunze, C., and Lemnitzer, L., editors, *Proceedings of LREC 2002 Workshop on Wordnet Structures and Standardisation*, pages 35–41, Las Palmas, Spain.
- [Vossen, 1998] Vossen, P., editor (1998). *EuroWordNet : a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, Norwell, MA, USA.

Bibliographie : exploitation de Wiktionary I

- [Navarro et al., 2009] Navarro, E., Sajous, F., Gaume, B., Prévot, L., Hsieh, S., Kuo, I., Magistry, P., and Huang, C.-R. (2009). Wiktionary and NLP : Improving synonymy networks. In *Proceedings of the 2009 ACL-IJCNLP Workshop on The People's Web Meets NLP : Collaboratively Constructed Semantic Resources*, pages 19–27, Suntec, Singapore.
<http://w3.erss.univ-tlse2.fr/textes/pagespersos/fsajous/papers/ACL-IJCNLP2009/navarroEtAl-ACL-IJCNLP2009.pdf>.

Bibliographie : exploitation de Wiktionary II

- [Zesch et al., 2008a] Zesch, T., Müller, C., and Gurevych, I. (2008a). Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, Marrakech. http://www.lrec-conf.org/proceedings/lrec2008/pdf/420_paper.pdf.
- [Zesch et al., 2008b] Zesch, T., Muller, C., and Gurevych, I. (2008b). Using wiktionary for computing semantic relatedness. In *Proceedings of 23rd AAAI Conference on Artificial Intelligence*. <http://www.dit.unitn.it/~p2p/RelatedWork/Matching/Zesch.pdf>.

Bibliographie : modélisation de la synonymie I

- [Edmonds and Hirst, 2002] Edmonds, P. and Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, 28(2) :105–144. <http://www.philipedmonds.com/papers/Edmonds+Hirst-2002-Near-synonymy.pdf>.
- [Kilgarriff, 1997] Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the humanities*, 31(2) :91–113. <http://www.ims.uni-stuttgart.de/~schulte/Teaching/ESSLLI-07/Referenzen/Senses/kilgarriff-1997.pdf>.

Bibliographie : misc I

- [Gaume, 2004] Gaume, B. (2004). Balades aléatoires dans les petits mondes lexicaux. *I3 : Information Interaction Intelligence*, 4(2). <http://w3.erss.univ-tlse2.fr/textes/pagespersos/gaume/resources/I3.impression.5.pdf>.
- [Hearst, 1992] Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING)*, pages 539–545, Nantes. <http://people.ischool.berkeley.edu/~hearst/papers/coling92.pdf>.

Bibliographie : misc II

[Morin, 1998] Morin, E. (1998). Prométhée : un outil d'aide à l'acquisition de relations sémantiques entre termes. In *5ème Conférence annuelle sur le Traitement Automatique des Langues Naturelles*, page 172 –181, Paris.

<http://www.sciences.univ-nantes.fr/info/perso/permanents/morin/article/morin-taln98.pdf>.

[Muller et al., 2006] Muller, P., Hathout, N., and Gaume, B. (2006). Synonym extraction using a semantic distance on a dictionary. In *Proceedings of the HLT/NAACL workshop Textgraphs*, New York. <http://www.irit.fr/~Philippe.Muller/Publis/textgraphs06.pdf>.