

# Random Walk for Knowledge-Based Word Sense Disambiguation

## Random Walks for Knowledge-Based Word Sense Disambiguation

Eneko Agirre\*

IXA NLP group

University of the Basque Country

Oier López de Lacalle\*\*

University of Edinburgh

IKERBASQUE

Basque Foundation for Science

Aitor Soroa†

IXA NLP group

University of the Basque Country

*Word Sense Disambiguation (WSD) systems automatically choose the intended meaning of a word in context. In this article we present a WSD algorithm based on random walks over large Lexical Knowledge Bases (LKB). We show that our algorithm performs better than other graph-based methods when run on a graph built from WordNet and eXtended WordNet. Our algorithm and LKB combination compares favorably to other knowledge-based approaches in the literature that use similar knowledge on a variety of English data sets and a data set on Spanish. We include a detailed analysis of the factors that affect the algorithm. The algorithm and the LKBs used are publicly available, and the results easily reproducible.*

*Computational Linguistics, Volume 40, Number 1 (2014)*

- Introduction à WordNet
- PageRank et Personalized PageRank (PPR)
- Application de PPR sur le graphe WordNet

# Word Sense Disambiguation

*Word Sense Disambiguation* (WSD) est une tâche consistant à choisir le sens attendu d'un mot en contexte

(1) *Nadal is sharing a house with his uncle and **coach**, Toni, and his physical trainer, Rafael Maymo.*  
*= sports coach*

(2) *Our fleet comprises **coaches** from 35 to 58 seats.*  
*= transport service*

# Wordnet

*Wordnet est une base de données lexicale de l'anglais qui regroupe des noms, des verbes, des adjectifs et des adverbes en ensembles de synonymes qui expriment chacun un concept (ou sens) distincts. Ces concepts sont appelés synsets*

<http://wordnetweb.princeton.edu/perl/webwn>

## WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

### Noun

- [S: \(n\) coach](#), [manager](#), [handler](#) ((sports) someone in charge of training an athlete or a team)
- [S: \(n\) coach](#), [private instructor](#), [tutor](#) (a person who gives private instruction (as in singing, acting, etc.))
- [S: \(n\) passenger car](#), [coach](#), [carriage](#) (a railcar where passengers ride)
- [S: \(n\) coach](#), [four-in-hand](#), [coach-and-four](#) (a carriage pulled by four horses with one driver)
- [S: \(n\) bus](#), [autobus](#), [coach](#), [charabanc](#), [double-decker](#), [jitney](#), [motorbus](#), [motorcoach](#), [omnibus](#), [passenger vehicle](#) (a vehicle carrying many passengers; used for public transport) "he always rode the bus to work"

### Verb

- [S: \(v\) coach](#), [train](#) (teach and supervise (someone); act as a trainer or coach (to), as in sports) "He is training our Olympic team"; "She is coaching the crew"
- [S: \(v\) coach](#) (drive a coach)

# Wordnet

## WordNet Search - 3.1

- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

### Noun

*synset*

*paraphrase descriptive*

- [S: \(n\)](#) **coach**, [manager](#), [handle](#) ((sports) someone in charge of training an athlete or a team)
- [S: \(n\)](#) **coach**, [private instructor](#), [tutor](#) (a person who gives private instruction (as in singing, acting, etc.))
- [S: \(n\)](#) [passenger car](#), **coach**, [carriage](#) (a railcar where passengers ride)
- [S: \(n\)](#) **coach**, [four-in-hand](#), [coach-and-four](#) (a carriage pulled by four horses with one driver)
- [S: \(n\)](#) [bus](#), [autobus](#), **coach**, [charabanc](#), [double-decker](#), [jitney](#), [motorbus](#), [motorcoach](#), [omnibus](#), [passenger vehicle](#) (a vehicle carrying many passengers; used for public transport) "*he always rode the bus to work*"

### Verb

- [S: \(v\)](#) **coach**, [train](#) (teach and supervise (someone); act as a trainer or coach (to), as in sports) "*He is training our Olympic team*"; "*She is coaching the crew*"
- [S: \(v\)](#) **coach** (drive a coach)

# Wordnet

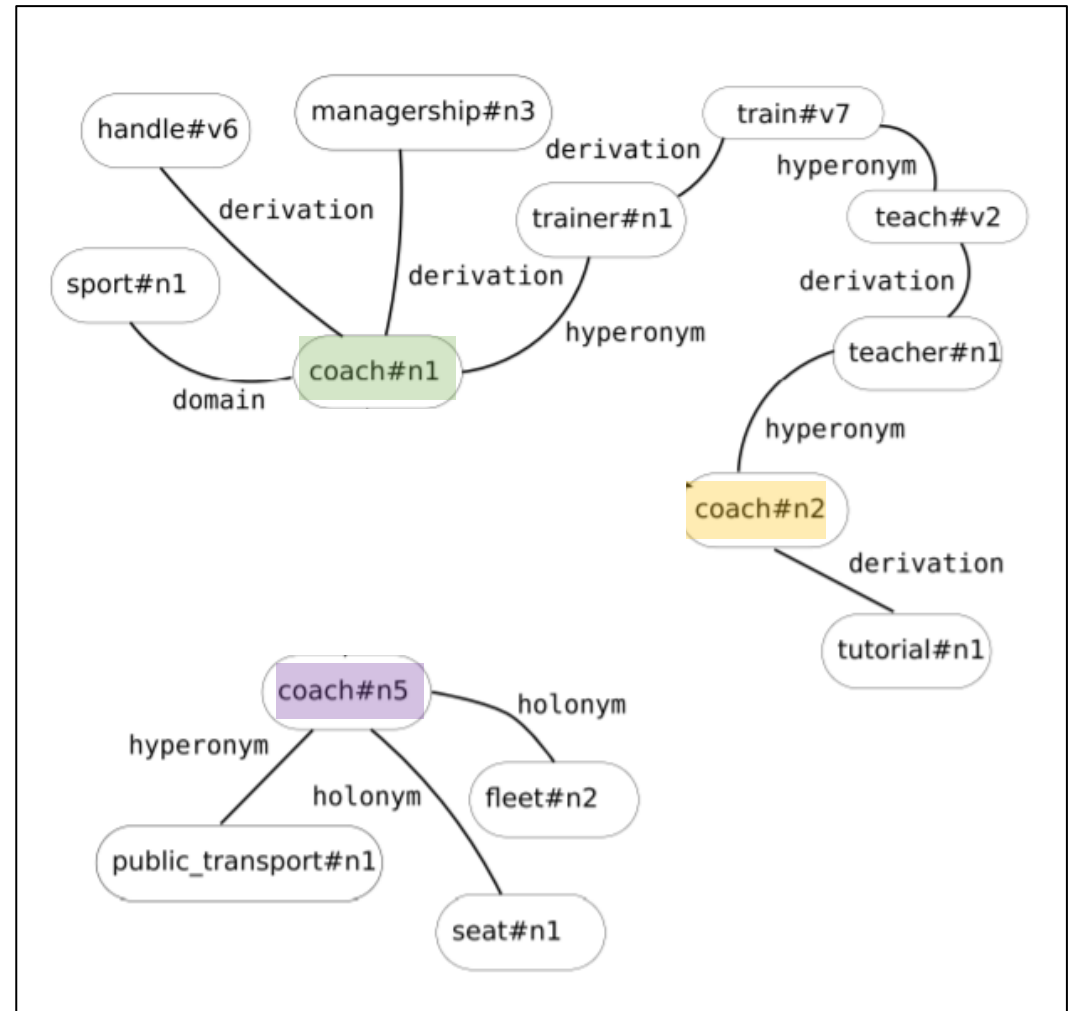
*Wordnet définit des relations entre les synsets*

- **S: (n) coach, manager, handler** ((sports) someone in charge of training an athlete or a team)
  - direct hyponym / full hyponym
    - **S: (n) baseball coach, baseball manager** (a coach of baseball players)
    - **S: (n) basketball coach** (a coach of basketball players)
    - **S: (n) conditioner** (a trainer of athletes)
    - **S: (n) football coach** (a coach of football players)
    - **S: (n) hockey coach** (a coach of hockey players)
    - **S: (n) tennis coach** (a coach of tennis players)
    - **S: (n) McGraw, John McGraw, John Joseph McGraw** (United States baseball player and manager (1873-1934))
  - domain category
    - **S: (n) sport, athletics** (an active diversion requiring physical exertion and competition)
  - has instance
    - **S: (n) McGraw, John McGraw, John Joseph McGraw** (United States baseball player and manager (1873-1934))
  - direct hypernym / inherited hypernym / sister term
    - **S: (n) trainer** (one who trains other persons or animals)
  - derivationally related form
    - **W: (v) coach** [Related to: **coach**] (teach and supervise (someone); act as a trainer or coach (to), as in sports) *"He is training our Olympic team"; "She is coaching the crew"*
    - **W: (n) managership** [Related to: **manager**] (the position of manager)
    - **W: (v) manage** [Related to: **manager**] (watch and direct) *"Who is overseeing this project?"*
    - **W: (v) handle** [Related to: **handler**] (show and train) *"The prize-winning poodle was handled by Mrs. Priscilla Prescott"*

# Créer un graphe à partir de Wordnet

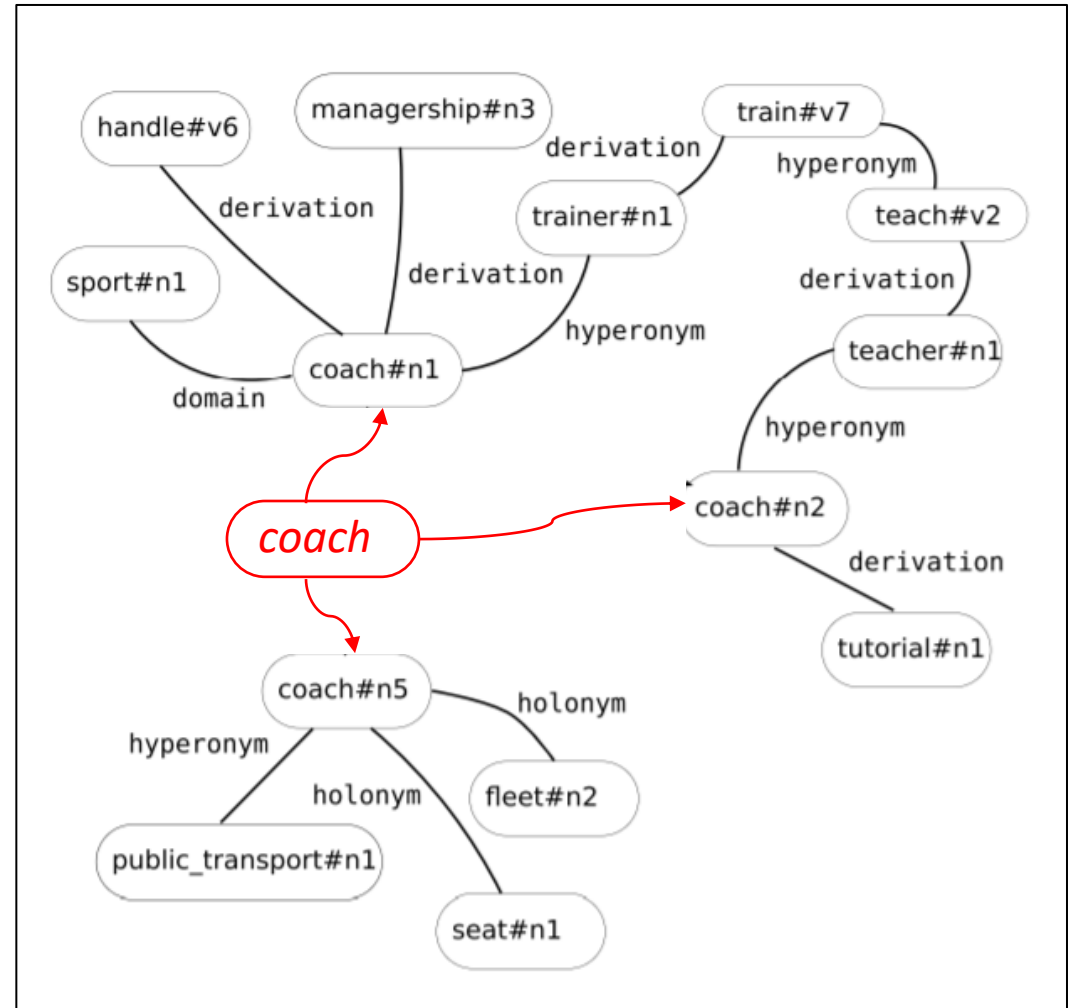
1) Ajouter des nœuds correspondant à **chacun des synsets** des mots et des liens (non dirigés) pour chacune des relations entre synsets (domaine, hyponymie, hyperonymie, holonymie, relations liées aux paraphrases, etc.)

- <coach#n1, manager#n2, handler#n3>
- <coach#n2, private instructor#n1, tutor#n1>
- <coach#n3, passenger car#n1, carriage#n1>
- <coach#n4, four-in-hand#n2, coach-and-four#n1>
- <coach#n5, bus#n1, autobus#n1, charabanc#n1, double-decker#n1, jitney#n1 ...>
- <coach#v1, train#v7>
- <coach#v2>



# Créer un graphe à partir de Wordnet

2) Ajouter des nœuds correspondant aux mots du dictionnaire et les lier à leurs synsets avec un lien dirigé





# PageRank

Soit  $G = (V, E)$  un graphe dirigé composé de  $N$  nœuds ( $V = \{v_1, \dots, v_N\}$ )

On note  $d_i$  le degré sortant d'un nœud ;

Soit  $M$  une matrice  $N \times N$  de probabilité de transition, où  $M_{ji} = \frac{1}{d_i}$  s'il existe un lien de  $i$  vers  $j$ , 0 sinon.

Considérons une marche aléatoire sur le graphe  $G$  avec une probabilité de passer de  $i$  à  $j$  égale à  $\frac{1}{d_i}$ .

Si on note  $X_0, X_1, \dots, X_T$  les nœuds visités par la marche aléatoire, on a :

$$\forall t, \forall i = 1 \dots N, \quad P(X_t = v_i) = \sum_{j \rightarrow i} P(X_{t-1} = v_j) * \frac{1}{d_j}$$

Et donc :

$$\forall t, \quad P_t = MP_{t-1}$$

Si le graphe  $G$  est fortement connecté (d'un nœud donné, on peut atteindre n'importe quel autre nœud du graphe), Alors la suite  $(P_t)$  converge vers une distribution stationnaire  $P$  telle que :

$$P = MP$$

# Personalized PageRank (PPR)

Dans un cas plus général, on peut introduire la possibilité, au cours de la marche aléatoire, de se déplacer vers n'importe quel nœud  $i$  avec une probabilité  $r_i$ . (on ajoute un facteur  $\alpha$  – *damping factor* – qui définit la manière dont les probabilités sont combinées à chaque étape). Usuellement,  $\alpha \in [0.85, 0.95]$ ).

Si on définit  $r = (r_1, r_2, \dots, r_n)$ , alors:

$$\forall i = 1 \dots n, \quad P(X_t = v_i) = \alpha \left( \sum_{j \rightarrow i} P(X_{t-1} = v_j) * \frac{1}{d_j} \right) + (1 - \alpha)r_i$$

Et donc :

$$\forall t, \quad P_t = \alpha M P_{t-1} + (1 - \alpha)r$$

Et on a donc :

$$P = \alpha M P + (1 - \alpha)r$$

# Personalized PageRank (PPR)

## Static PageRank :

Habituellement, on fixe  $r_i = \frac{1}{N}, \forall i \in \llbracket 1, N \rrbracket$

On assigne donc la même probabilité pour chaque nœud en cas de saut aléatoire.

$$P = \alpha MP + (1 - \alpha) \frac{\mathbb{1}_N}{N}$$

## Personalized PageRank:

On va appeler Personalized PageRank tout calcul de  $P$  pour lequel  $V$  n'est pas la distribution uniforme.

Supposons que  $r$  définisse une distribution où toute la masse probabiliste soit concentrée sur un nœud  $i$  :

$$\begin{cases} r_i = 1 \\ \forall j \neq i, r_j = 0 \end{cases}$$

Tous les sauts aléatoires retourneront donc au nœud  $i$ , son rang sera donc plus haut et, par propagation, celui des nœuds de son voisinage augmentera aussi.

Ainsi, dans ce cas le vecteur  $P$  peut être vu comme étant la pertinence des nœuds du graphes *du point de vue du nœud  $i$* .

# Random Walk for WSD

## (application de Static PageRank)

*Soit un texte dont on extrait  $m$  mots  $W_i, i = 1, \dots, m$  qui appartiennent au dictionnaire de WordNet et peuvent donc être liés à des concepts dans la base de connaissance lexicale.*

*Objectif : appliquer PPR au graphe WordNet puis sélectionner, pour chaque  $W_i$ , le concept de plus haut score.*

Si on applique *Static PageRank* au graphe de WordNet, on obtient un classement des sens des mots indépendant de tout contexte.

Si on cherche à lever l'ambiguïté concernant le sens d'un mot  $w$ , il suffit de chercher le synset associé à ce mot de plus haut rang.

Dans la mesure où la valeur calculée *via* PageRank est fortement lié au degré des nœuds, la version statique de PageRank tend à renvoyer le sens qui a le plus de relations (qui est généralement le « sens le plus fréquent » dans un corpus général).

# Random Walk for WSD (application de PPR)

*Soit un texte dont on extrait  $m$  mots  $W_i, i = 1, \dots, m$  qui appartiennent au dictionnaire de WordNet et peuvent donc être liés à des concepts dans la base de connaissance lexicale.*

*Objectif : appliquer PPR au graphe WordNet puis sélectionner, pour chaque  $W_i$ , le concept de plus haut score.*

## Application (1) de PPR :

- 1) Ajout des mots de contexte  $W_i, i = 1, \dots, m$  dans le graphe et de liens dirigés qui les lient à leurs concepts.
- 2) Définition de la distribution  $r$  de sorte à ce que la masse probabiliste initiale soit **répartie uniformément sur les nœuds correspondant aux mots du contexte** et qui, du fait des liens dirigés (sortants) agissent comme source de masse et donc accroissent la pertinence des concepts liés.
- 3) Calcul de PPR. Le vecteur  $P$  calculé peut être vu comme une mesure de la pertinence des concepts en présence du contexte (les  $W_i$ )

=> **Problème si des concepts d'un même mot sont directement liés et se renforcent mutuellement.**

## Application (2) de PPR :

**Pour chaque mot cible** dont on doit lever l'ambiguïté, on réalise le calcul précédent en ajoutant dans le graphe tous les mots de contexte **excepté le mot cible**.

=> **Fonctionne mieux** mais **computationnellement moins efficace** (calcul de PPR pour chacun des mots cibles)