



Wordnet

C. Fellbaum, G. Miller

Claire Gardent

Wordnet

- Dictionnaire informatisé dont l'unité de base est le concept (non le mot)
- Version 1.2 (1992) : 102 000 mots et 52 000 synsets; 100 utilisateurs
- Version 1.5 (1997) : 168 000 mots et 91 600 synsets; 1000 utilisateurs; site web
- Version 1.6 : 94.000 formes nominales, 10.000 formes verbales, 20.000 formes adjectivales et 4.500 formes adverbiales
- 5 catégories : noms, verbes, adjectifs, adverbes, mots grammaticaux (les mots gr. ne sont pas inclus dans le système)
- Répertoire mots, collocations et idiomes

Sens lexical

Wordnet utilise deux moyens pour définir le sens d'un mot :

- Les **synsets**
- Les **relations lexicales**

Synset

Le sens d'un mot est représenté par (i) l'ensemble des mots utilisés pour exprimer ce sens i.e., par un ensemble de synonymes (les **synsets**) et (ii) une définition.

Exemple : board

Synset : { board, plank }

Définition : morceau de bois

Relations lexicales

Le sens d'un mot est aussi déterminé par ses relations sémantiques avec d'autres sens (ex: chien/caniche, doberman etc.)

Wordnet est structuré par des relations entre synsets et entre mots; les relations suivantes sont utilisées :

Synonymie : Deux mots sont synonymes s'ils sont interchangeables dans **certains** contextes linguistiques

Antonymie : relation lexicale entre mots (**rise/fall** mais pas **rise/descend**); très importantes pour les adjectifs et les adverbes;
Difficile à définir

Hyponymie : relation lexicale entre sens; X est un hyponyme de Y si *Un X est une sorte de Y* est vrai; relation transitive et asymétrique qui génère une hiérarchie d'héritage; centrale pour

Relations sémantiques

Méronymie : relation lexicale entre sens; X est un méronyme de Y si
Un Y a une X (comme partie) ou Un X est une partie de Y est vrai;
relation transitive et asymétrique qui génère une hiérarchie
d'héritage;

Morphologie : La morphologie est nécessaire pour permettre un
accès simple aux lèmes (arbres → arbre); le traitement de la
morphologie inflectionnel fait partie de l'interface entre WN et
utilisateur

Connaitre la place d'un mot dans ce réseau de relations, c'est connaître
son sens

Les noms dans WN

- Version 1.5 : 80 000 noms (dont des collocations), 60 000 concepts
- Synset = concept lexicalisé
- Hyponymie = relation entre concepts lexicalisés
 $\{ robin, redbreast \} \rightarrow \{ bird \}$

Les noms dans WN

La relation d'hyponymie structure les noms en une hiérarchie d'héritage ce qui permet:

- d'éviter la circularité (structure arborescence)
- d'éviter les répétitions: Tout concept dans la hiérarchie hérite des propriétés de ses superconcepts
- de favoriser l'exhaustivité (la hiérarchie peut être utilisée pour obtenir aussi bien les superconcepts que les sousconcepts d'un concept)

Hiérarchies nominales

- WN divise les noms en plusieurs hiérarchies (12) : organism, entity, abstraction, psychol. feature, nat. phenomenon, activity, event, group, location, possession, shape, state
- Chaque hiérarchie correspond à un champ sémantique distinct
- Le nombre maximum de niveaux à l'intérieur de ces hiérarchies est 12

Exemple : shetland pony/horse/equid/odd-toe ungulate/placental mammal/mammal/vertebrate/chordate/animal/organism/entity

Les limites de WN

- The tennis problem : il est impossible d'identifier dans WN un vocabulaire spécifique e.g., le vocabulaire du tennis.
- WN ne contient aucun savoir syntagmatique i.e., aucune information sur les contextes dans lesquels les mots apparaissent
- WN ne dit rien sur les exceptions
Exemple : les pingouins sont des oiseaux mais ne volent pas

Les limites de WN

- La relation d'hyponymie implémentée dans WN correspond en fait à plusieurs relations sémantiques distinctes e.g., IS A KIND OF (reln formelle/taxonomique) et IS USED AS A KIND OF (relation télélique/fonctionnelle)

Exemples : a thrush IS A KIND OF bird; an adornment IS USED AS A KIND OF decoration

- WN n'utilise pas un nombre suffisamment grand de relations sémantiques pour réellement définir le sens des mots; chaque synset est donc pourvu d'une glose décrivant son sens

Méronymie

- WN encode également les relations de méronymie (partie/tout) entre concepts nominaux
- Ces relations se trouvent principalement dans les hiérarchies pour **body, artéfact, quantité**
- Les méronymes sont des attributs hérités par le biais de l'hyponymie; cela peut mener à l'introduction de concepts intermédiaires

Exemple : si les véhicules ont des roues, les luges ne sont pas des véhicules; on introduit donc le concept de **véhicules à roues**

Méronymie

- WN inclut trois types de méronymies :
 - (1) X **est une composante** de Y
 - (2) X **est un élément** de Y
 - (3) X **est le matériau** dont Y est constitué

Attributs et modificateurs

- WN inclut les associations entre attributs et valeurs possible pour un concept donné

Exemple : attribut = taille → petit, grand

- Les associations fréquentes entre noms et adjectifs ne sont pas traitées (e.g., gentil chien vs. chien étroit)
- WN inclut les collocations Adj+Nom (**easy chair, electric chair, high chair** etc.) et Nom+Nom (**barber chair, beach chair** etc.) en particulier parce que celles si sont nécessaires pour définir la taxonomie basée sur l'hyponymie

Grouper les sens des mots ambigus

- Dans un dictionnaire, les sens d'un mots sont hiérarchisés; au top niveau, on trouve les différences de sens les plus importantes, en dessous plutôt des nuances.
- WN ordonne ces sens d'après leur fréquences relatives en corpus
- Autres mécanismes de groupement de sens proches: les cousins, les jumeaux et les soeurs

Cousins, jumeaux et soeurs

- Cousins: deux sens similaires du même mot dont le superconcept direct a des sous-concepts reliés par la même notion de similarité (ex: sole-poisson et sole-nourriture mais pas sole-poisson et sole-partie de pied)
- Jumeaux: synsets ayant plus de 2 mots en commun (ex: duo-groupe musical et duo-composition, { duo, duet, duette })
- Soeurs: deux mots ayant un superconcept direct en commun (ex: flounder-1 et flounder-2 qui expriment deux types de “flatfish”)

Les modificateurs dans WN

- Les adjectifs : modifient le sens des noms
- Les adverbes : modifient le sens des autres parties du discours
(verbes, adjectifs, adverbes, phrases)

Adjectifs

- WN 1.5 contient 16 428 synsets d'adjectifs (où des noms, des participes et des GPs sont inclus qui fonctionnent comme modificateurs e.g., **home** dans **home cooking**)
- WN divise les adjectifs en deux grandes catégories : les adjectifs descriptifs et les adjectifs relationnels
- Les **adjectifs descriptifs** forment de loin la plus grande catégorie
- Les **adjectifs relationnels** (i.e., reliés par dérivation à un nom e.g., **electrical**) forment une classe bcp plus petite

Adjectifs descriptifs

- Un adjectif descriptif assigne à un attribut d'un nom une valeur
Exemple : le nom **paquet** a pour attribut **poids** dont la valeur peut être spécifiée par l'adjectif **lourd**
- WN associe aux adjectifs descriptifs des pointeurs vers les noms dont ils spécifient la valeur des attributs
- La relation sémantique qui structure la classe des adjectifs descriptifs est la relation d' **antonymie**

Antonymie

- L'antonymie est une relation entre mots (pas entre synsets i.e., sens) e.g., **heavy** est l'antonyme de **light** pas mais de **weightless**
- Ainsi de nombreux antonymes sont formés par dérivation morphologique i.e., par des règles qui s'appliquent aux mots et non aux sens (e.g., **tidy/untidy**)
- Des adjectifs ayant des sens similaires peuvent avoir des antonymes différents
light/heavy, weight/weightless
- De nombreux adjectif descriptifs n'ont pas d'antonymes e.g., **ponderous**.
- L'antonymie est à distinguer de l'**opposition conceptuelle** qui est
une relation entre sens

Structure dans la classe des adjectifs

- Deux relations sont utilisées pour structurer la classe des adjectifs : la **similarité** et l'**antonymie**
- La similarité reflète une relation de spécialisation : deux adjectifs A1 et A2 sont **similaires** ssi la classe des noms modifiée par A1 est un sous ensemble de la classe des noms modifiée par A2
e.g., ponderous/heavy
- La similarité sert à créer des “(demi)-clusters”
- Les demi-clusters vont par paires et sont reliés par une relation d'antonymie
- Une relation d'opposition conceptuelle (ou antonymie indirecte) relie les éléments de demi-clusters qui ne sont pas en relation d'antonymie

Cas problématiques

- Certains adjectifs n'ont pas d'antonymes e.g., **angry, enrage, rate, incensed, furious**
- Dans ce cas, WN inclut la négation de l'adjectif i.e., **not angry**
- Cette stratégie permet également de rendre compte de paires telles que **addictive/nonaddictive, combustible/noncombustible**
- WN recense 1732 clusters bipolaires soit 3464 demi-clusters d'adjectifs au sens similaire

Quantificateur

- **all, some, many, less, few**
- Un quantificateur quantifie un terme
- Les quantificateurs ressemblent aux adjectifs
 - Ils ont des antonymes : **much/little** etc.
 - Ils sont graduables : **very much, very many, little/less/least**
- Dans WN, les quantificateurs sont traités comme des adjectifs

Adjectifs participes

- formes terminées par **-ed, ing**
- position prénominale
- sens statique “étant dans un état X”
- n’ont généralement pas d’antonyme
- Dans WN, ces adjectifs sont inclus dans la catégorie des adjectifs descriptifs mais ne sont pas organisés en clusters et sont reliés au verbe dont ils sont dérivés (88 synsets)
- Certains participes ont des antonymes (**married/unmarried**); ceux la forment des clusters et ne sont pas reliés aux verbes dont ils sont dérivés
- Enfin, certains participes sont dérivés d’un nom plutôt que d’un verbe (**skilled**)

Adjectifs relationnels

- reliés sémantiquement et morphologiquement à des noms
- fonction = classificateur e.g., **musical instrument** identifie un type d'instrument
- ne réfèrent pas à un attribut du nom qu'ils modifient
- non graduable : * **very dental**
- n'ont souvent pas d'antonymes
- Un adjectif peut être à la fois descriptif et relationnel e.g., **criminal** : criminal law, criminal behaviour
- WN regroupe les adjectifs relationnels (2 832 synsets) et les relie par des pointeurs aux noms dont ils sont dérivés

Adverbes

- souvent dérivés d'adjectifs par suffixation (quickly, northward etc.)
- héritent souvent de ces adjectifs l'antonymie et la graduabilité (**specific(ally)/general(ly)**)
- WN contient 3 242 synsets d'adverbes
- Ces synsets sont structurés par la relation entre adverbe et adjectif, la synonymie et l'antonymie
- Chaque synset d'adverbe contient un adverbe, l'adjectif dont il est dérivé, synonymes et antonymes et une glose incluant des exemples d'usage

Les verbes dans WN

- Les verbes sont classés par champ sémantique
- Trois grandes classes : **actions, évènements, états**
- Actions et évènements sont divisés en 14 sous classes : verbes de motion, de perception, de contact, de communication, de compétition, de changement, de consommation, de création, d'émotion, de possession, de comportement social etc.
- Dans un même sémantique champ, on peut avoir plusieurs hiérarchies indépendantes (pas d'origine unique)
Exemple : les verbes de possession sont regroupés au sein de 3 hiérarchies distinctes dont les origines sont {**give, transfer**}, {**take, receive**} et { **have, hold**}

Les synsets verbaux

- Peu de vrais synonymes du fait des différences de registres (**end/terminate**) ou de restrictions de sélection (**rise,fall v. descend, ascend**)
- Les synsets ignorent les différences de registre mais tiennent compte des différence sélectionnelles
- Idioms et sens métaphoriques sont inclus dans les synset
Exemple : { **die, break, kick the bucket** }

Relations sémantiques entre verbes

- Issues des définitions dictionnairiques
- “shuffle : to move or walk in a ... way” (**hyponymie**, specif. de manière)
- “to loose : to fail to keep” (**opposition sémantique**)
- “to x while y-ing” (**implication**)
- “to become or make Y” (**relation causale**)

Implication

- Implication : V1-er implique V2-er (ronfler/dormir)
- Synonymie : V1-er implique V2-er et V2-er implique V1-er (shut/close)
- La relation d'implication permet de structurer les verbes en une hiérarchie
- Un verbe est temporellement inclus dans son super-verbe

Troponymie

- La relation entre verbe et hyper-verbe est une relation de modification par la manière (**move/amble**), la cause (**slide/pull**), la vitesse (**walk/run**), l'intensité de l'action (**drowse/sleep**) etc.
- Cette relation est appelée troponymie : V1 est un **troponyme** de V2 ssi V2-er c'est V1-er d'une certaine façon
- WN laisse la composante sémantique reliant un verbe à son troponyme sous-spécifiée
- Cette composante sémantique est souvent spécifique à un champ sémantique donné
fight/battle,war,tourney,duel,feud (fight on a particular occasion)
communicate/fax,email,phone,telex (communiquer par un certain medium)

Opposition sémantique

- Plusieurs types d'opposition
- Contraires (**donner/prendre, vendre/acheter**) : même action, différente réalisation des actants sémantiques; même champ sémantique
- Comme les adjectifs, les verbes de changement (**allonger/raccourcir**) et d'état (**inclure/exclure**) ne sont reliés que par des relations de synonymie et d'antonymie
- Les verbes opposés sont parfois des co-troponymes i.e. des troponymes avec le même super-verbe e.g., **monter/descendre , courrir/marcher**.
- Certains verbes opposés partagent une implication e.g., **échouer/réussir** implique **essayer**

La relation causale

- La relation de cause relie un verbe causatif (e.g., **donner**) à un verbe résultatif (e.g., **avoir**)
- WN inclut aussi une reln de cause entre verbe causatif transitif et correspondant intransif e.g.,
(1) a. John broke the door
b. The door broke.

La relation causale

- WN inclut une reln de cause entre verbe de motion causatif et correspondant non-causatif e.g.,
(2) a. John blew the soap in her face
b. The soap bubble blew in her face.
- La relation causale est une forme spécifique d'implication (**donner** implique **avoir**) où l'action dénotée par le causatif précède l'état dénoté par le résultatif

Comparaison avec d'autres lexiques

- **Champs sémantique** : le sens d'un mot est déterminé par ses relations (paradigmatiques et syntagmatiques) de contraste et de similarité avec d'autres mots appartenant au même champ sémantique
- Dans WN les verbes sont bien regroupés en champs sémantiques mais les relns utilisées sont toute paradigmatiques; les préférences sélectionnelles sont explicitées dans la glose des synsets.
- **Schemata** : spécifie la valence syntaxique et sémantique d'un verbe.
- Dans WN, la valence syntaxique des verbes est spécifiée mais pas le type sémantique et le rôle thématique des actants sémantiques

Comparaison avec d'autres lexiques

- **Analyse compositionnelle** : décompose le sens des mots en utilisant des primitifs sémantiques (cf. Jackendoff, Melcuk)
- Dans WN, les verbes apparaissant au sommet de la hiérarchie sont des composantes sémantiques (e.g., **change, go, stay, be, do**) ainsi que la relation causale (**donner** = causer **avoir**); certaines composante sémantiques sont présentes implicitement via la reln de troponymie (**path, speed** comme sous-reln de la troponymie)

Syntaxe et Sémantique

- Levin a montré que valence syntaxique et sens sont liés : les verbes partageant la même valence, ont aussi un commun une composante sémantique
- De même dans WN, les arbre de verbes créés sur une base sémantique coorespondent à un certain comportement syntaxique

Syntaxe et Sémantique

- Des différences de comportement syntaxique peuvent motiver la division d'un arbre en plusieurs sous-arbres

(3) a. John wove the wool into a rug. BUILD

b. John wove a rug from the wool.

c. She composed a quartet out of the song. CREATE

d. ?? She composed the song into a quartet.

- Les classes sémantiques peuvent expliquer certains comportements syntaxiques

(4) a. * This vegetable cooks / This vegetable cooks well.

b. This vegetable microwaves.

microwaves est un troponyme de **cook** exprimant la manière de cuire; l'adverbe n'est donc pas nécessaire