

WordNet

WordNet, an electronic lexical database, is considered to be the most important resource available to researchers in computational linguistics, text analysis, and many related areas. Its design is inspired by current psycholinguistic and computational theories of human lexical memory. English nouns, verbs, adjectives, and adverbs are organized into synonym sets, each representing one underlying lexicalized concept. Different relations link the synonym sets. Applications of WordNet include word sense identification, information retrieval, selectional preferences of verbs, and lexical chains.

WordNet¹ is een lexicale database waarin de woorden op een geheel andere manier zijn georganiseerd dan in een 'gewoon' woordenboek. WordNet bevat informatie over Engelse zelfstandige naamwoorden, werkwoorden, bijvoeglijke naamwoorden en bijwoorden en is georganiseerd op basis van zgn. *synsets*. Een synset is een groep woorden (met dezelfde woordklasse) die in een bepaalde context uitwisselbaar zijn, en dus in hoge mate synoniem zijn. Zo vormen de woorden {*car*, *auto*, *automobile*, *machine* en *motorcar*} een synset omdat ze gebruikt kunnen worden om naar hetzelfde concept te verwijzen. Aan een synset is meestal een verklarend commentaar toegevoegd (een soort eenvoudige betekenisomschrijving), zoals "4-wheeled; usually propelled by an internal combustion engine". Een heel belangrijk aspect van WordNet, tenslotte, is dat synsets met elkaar verbonden zijn via diverse semantische relaties. Het type van deze relaties verschilt per woordsoort. De bovengenoemde synset, bijvoorbeeld, is gerelateerd aan:

- meer algemene concepten (ook wel hyperoniemen genoemd): {*motor vehicle*; *automotive vehicle*};
- meer specifieke concepten (ook wel hyponiemen genoemd): b.v. {*cruiser*; *squad car*; *patrol car*; *police car*; *prowl car*} en {*cab*; *taxi*; *hack*; *taxicab*};
- onderdelen waaruit het bestaat (ook wel meroniemen genoemd): b.v. {*bumper*}; {*car door*}; {*car mirror*} en {*car windows*}.

Elk van deze synsets is zelf ook weer gerelateerd aan andere synsets, zoals in het onderstaande figuur is te zien.

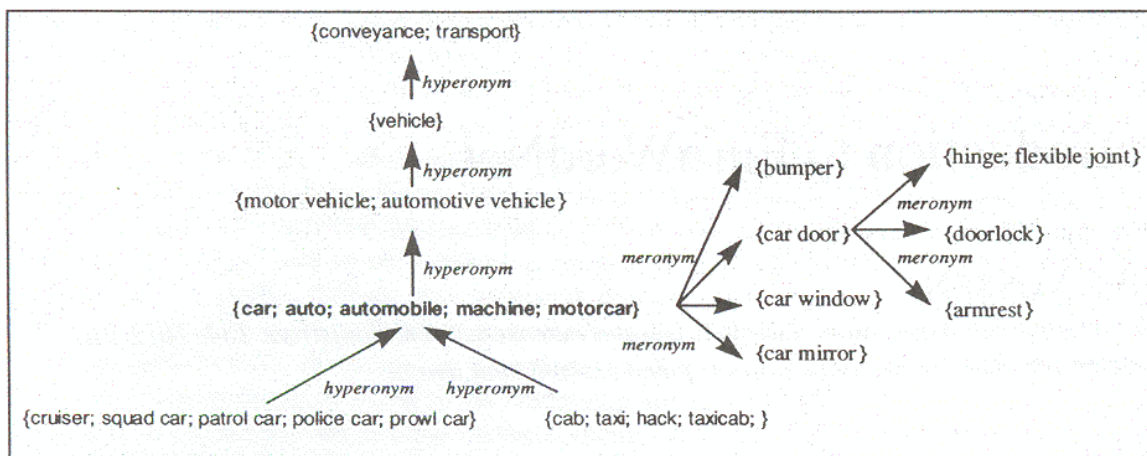


Figure 1. Synsets related to *car* in its first sense in WordNet1.5.

¹ Deze korte beschrijving van WordNet is grotendeels gebaseerd op de eerste paragraaf van het artikel 'Introduction to EuroWordNet', door P. Vossen, in *Computers and the Humanities* 32 (1998): 73-89. Daaruit is ook het figuur afkomstig.

Door middel van deze relaties zijn alle betekenissen onderling verbonden en vormen ze samen een semantisch netwerk, oftewel *wordnet*. Zo'n wordnet kan gebruikt worden om semantische gevolgtrekkingen te maken over de betekenissen van woorden (b.v. welke betekenissen kunnen worden geïnterpreteerd als *vehicles*), om alternatieve uitdrukkingen of bewoordingen voor bepaalde woorden te vinden (vgl. een thesaurus), of in programma's voor *information retrieval* om woorden uit te breiden tot sets van woorden die semantisch aan elkaar gerelateerd zijn. Op het gebied van de theoretische lexicologie zijn semantische netwerken interessant omdat ze informatie kunnen bieden over de conceptuele dichtheid van bepaalde gebieden van de woordenschat van een taal en over de distributie van semantische verschillen en relaties op verschillende terreinen van die woordenschat.