Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use

This article was originally published in the *Encyclopedia of Language & Linguistics, Second Edition*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

Fellbaum C (2006), WordNet(s). In: Keith Brown, (Editor-in-Chief) *Encyclopedia of Language & Linguistics, Second Edition*, volume 13, pp. 665-670. Oxford: Elsevier.

McCawley J (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Nespor M & Vogel I (1986). *Prosodic phonology*. Dordrecht: Foris Publications.
Odden D (1988). 'Predictable tone systems in Bantu.' In van der Hulst H & Smith N (eds.) *Autosegment studies on pitch accent*. Dordrecht: Foris publications. 225–252.
Peperkamp S (1997). *Prosodic words*. LOT/HIL dissertation # 34. The Hague: Holland Academic Graphics.
Prince A (1983). 'Relating to the grid.' *Linguistic Inquiry 14*, 19–100.
Prince A S (1985). 'Improving tree theory.' *BLS 11*, 471–490.
Prince A & McCarthy J (1990). 'Foot and word in prosodic morphology: the Arabic broken plural.' *Natural Language and Linguistic Theory 8*, 209–283.
Prince A & Smolensky P (1993). 'Optimality Theory: constraint interaction in Generative Grammar' Technical Report #2 of the Rutgers Center for Cognitive Science. Piscataway, NJ: Rutgers University.
Revithiadou A (1998). *The prosody–morphology interface*. LOT/HIL dissertation #15. The Hague: Holland Academic Graphics.
Rice C (1992). 'Binarity and ternarity in metrical theory: parametric extensions.' Ph.D. diss., University of Texas.
Rischel J (1972). 'Compound stress in Danish without a cycle.' *Annual report of the Institute of Phonetics of the University of Copenhagen 6*, 211–228.
Roca I (1986). 'Secondary stress and metrical rhythm.' *Phonology Yearbook 3*, 341–370.
Roca I (1999). 'Stress in the Romance languages.' In van der Hulst H (ed.) *Word prosodic systems in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 659–812.
Selkirk E (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge: MIT Press.
Visch E (1999). 'The rhythmic organization of compounds and phrases.' In van der Hulst H (ed.) *Word prosodic systems in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 161–232.
Vijver R van de (1997). *The iambic issue*. LOT/HIL dissertation #37. The Hague: Holland Academic Graphics.
Zonneveld W & Trommelen M (1999). 'Word Stress in west-Germanic languages.' In van der Hulst H G (ed.) *Word-prosodic systems of European languages*. Berlin & New York: Mouton de Gruyter. 478–514.

WordNet(s)

C Fellbaum, Princeton University, Princeton, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

The Princeton WordNet

Background and Motivation

WordNet, a manually constructed electronic lexical database for English, was conceived in 1986 at Princeton University, where it continues to be developed. Experiments by researchers in Artificial Intelligence (Collins and Quillian, 1968, inter alia) probing human semantic memory inspired the psycholinguist George A. Miller to test the underlying theories on a large scale.

Design and Contents

WordNet is a large semantic network interlinking words and groups of words by means of lexical and conceptual relations represented by labeled arcs. WordNet's building blocks are synonym sets (synsets), unordered sets of cognitively synonymous words and phrases (Cruse, 1986). Each member of a given synset expresses the same concept, though not all synset members are interchangeable in all contexts. Examples are {*car, automobile*}, {*hit, strike*}, and {*big, large*}. All synsets further contain a brief definition, and most include one or more sentences illustrating the synonyms' usage. A domain label (*sports, medicine, biology*) marks many synsets.

Joint membership of words in a given synset illustrates the phenomenon of synonymy. Membership of a word in multiple synonyms reflects that word's polysemy, or multiplicity of meaning. Thus, *trunk* appears in WordNet in several different synsets, including {*trunk, tree trunk*}, {*trunk, torso*}, and {*trunk, proboscis*}.

Coverage

WordNet consists of four separate components, each containing synsets with words from the major, open-class, syntactic categories: nouns, verbs, adjectives, and adverbs. WordNet 2.1 contains almost 118 000 synsets, comprising more than 81,000 noun synsets, 13 600 verb synsets, 19 000 adjective synsets, and 3 600 adverb synsets.

Relations

Synonymy is the major lexical relation among individual word forms; another is antonymy, as between the pairs {*wet*} and {*dry*} and {*rise*} and {*fall*}. Morpho-semantic relations link words from all four parts of speech that are both morphologically and semantically related (Fellbaum and Miller, 2003). For example, the semantically related senses of *interrogation,*

Author's Personal Copy

interrogator, interrogate, and *interrogative* are interlinked. Conceptual-semantic relations link not just single word forms but entire synsets.

Nouns in WordNet

Hyponymy

Concepts expressed by nouns are densely interconnected by the hyponymy relation (or hyperonymy, or subsumption, or the *ISA* relation), which links specific concepts to more general ones. For example, the synset {*mailbox, letterbox*} is a hyponym, or subordinate, of {*box*}, which in turn is a hyponym of {*container*}. {*Mailbox, letter box*} is a hypernym, or superordinate, of {*pillar box*}, which denotes a specific type of mailbox. Hyponymy builds hierarchical 'trees' with increasingly specific 'leaf' concepts growing from an abstract 'root.' All noun synsets ultimately descend from {*entity*}. (See **Figure 1.**)

Types vs. instances Among the concepts represented by nouns, WordNet distinguishes types and instances. Common nouns are types: *tree* is a type of *plant*, *china* is a type of *crockery*. Proper names are instances: *China* is an instance, rather than a type of a *country* (Miller and Hristea, 2004).

Meronymy

Another major relation among noun synsets is meronymy, which links synsets denoting parts, components, or members to synsets denoting the whole. Thus, {*finger*} is a meronym of {*hand*}, which in turn is a meronym of {*arm*}, and so forth. Meronymy in WordNet actually encompasses three distinct part-whole relations. One holds among proper parts or components, such as {*leg*} and {*table*}. Another links substances that are constituents of other substances: {*oxygen*} is a part of {*water*} and {*air*}. Members such as {*tree*} and {*parent*} are parts of groups such as {*forest*} and {*family*}.

Personal Copy

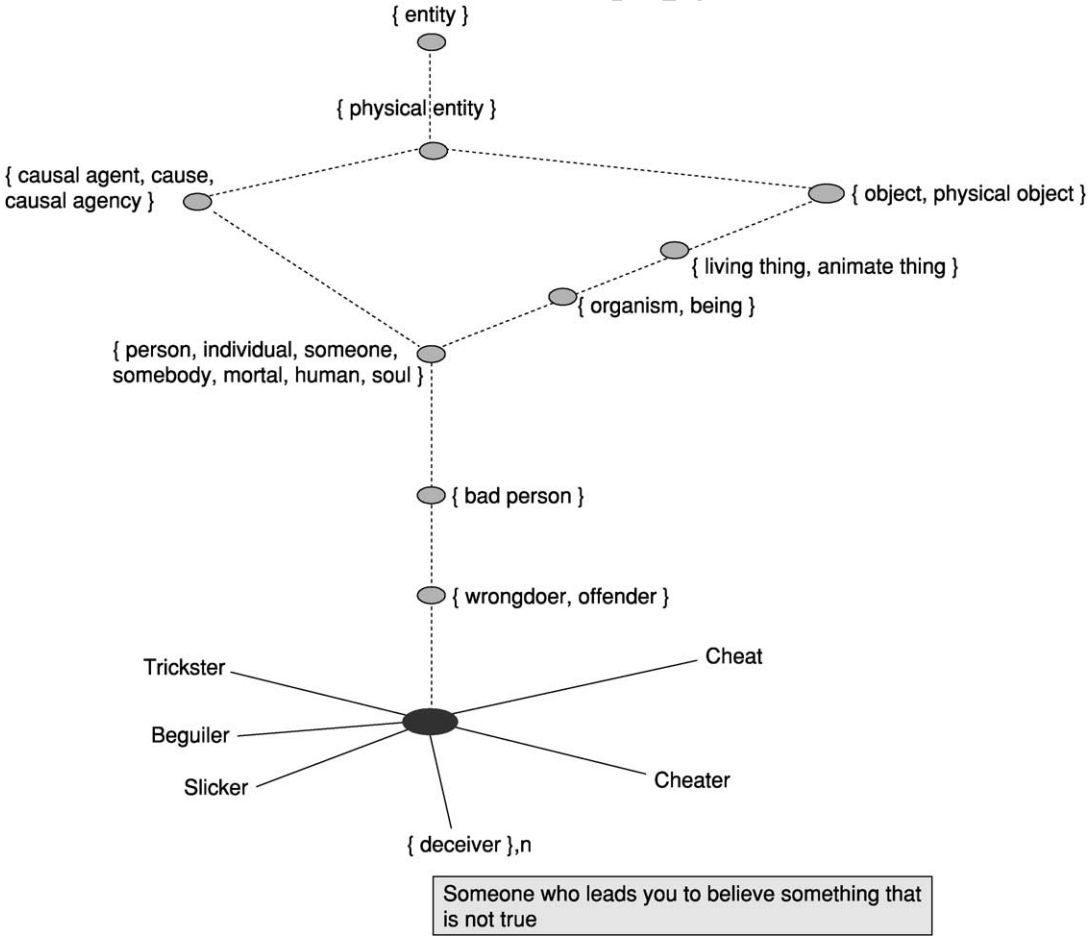

Figure 1 A WordNet noun tree.

Verbs

Verbs are organized by a several entailment relations (Fellbaum and Miller, 1990; Fellbaum, 1998b). The most prevalent is troponymy, which relates synset pairs such that one expresses a particular manner of the other (e.g., {*whisper*}-{*talk*} and {*punch*}-{*strike*}) (Fellbaum, 2002). Like hyponymy, troponymy builds hierarchies of several levels of specificity. Other relations are backward entailment (*divorce-marry*), presupposition (*buy-pay*), and cause (*show-see*). (See Figure 2.)

Adjectives

WordNet distinguishes descriptive and relational adjectives. Descriptive adjectives are organized into direct antonym pairs, such as *wet-dry* and *long-short*. Each member of a direct antonym pair is associated with a number of 'semantically similar' adjectives. *Damp* and *drenched* are semantically similar to *wet*, and *arid* to *dry*. These concepts are said to be indirect antonyms of the direct antonym of their central members, i.e., *drenched* is an indirect antonym of *dry*, and *arid* is an indirect antonym of *wet* (Miller, 1998; Gross *et al*., 1989). (See Figure 3.)

Relational adjectives (*atomic, nuclear*) are linked to the corresponding morphologically and semantically related nouns (*atom, nucleus*). Most adverbs point to their base adjectives (*rapid-rapidly, slow-slowly*).

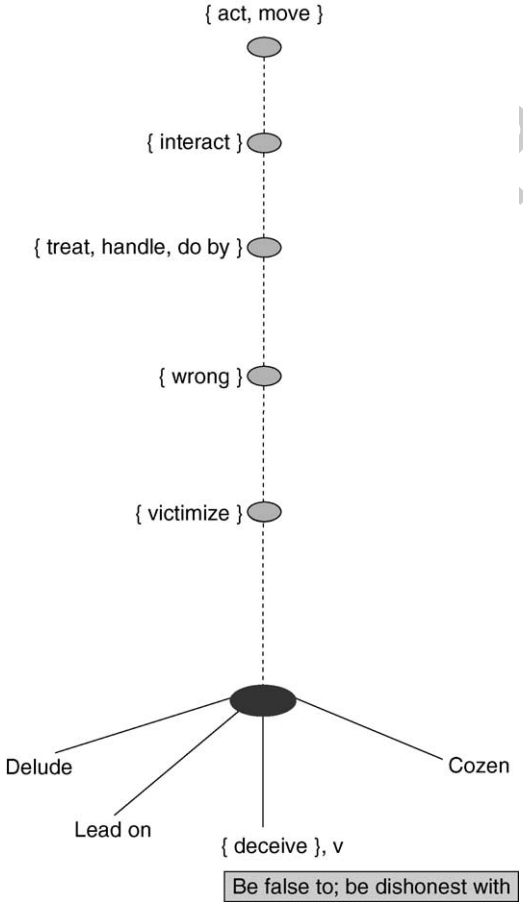

Figure 2 A WordNet verb tree.

Inheritance and Reversibility

Two important components of WordNet's design are inheritance and reversibility. Inheritance applies to hierarchy-building relations. If {*mailbox, letter box*} is encoded as a hyponym of {*box*}, and {*box*} as a hyponym of {*container*}, then {*mailbox, letter-box*} is automatically recorded as a hyponym of {*container*}, via the principle of inheritance. Similarly, if {*finger*} is a part of {*arm*}, and {*hand*} is part of {*arm*}, then {*finger*} is necessarily a part of {*arm*}, too. Many concepts are assigned to both types of hierarchy.

Relations are encoded in WordNet only once between a given pair of synsets or words. The pointer gets automatically reversed, so if {*tree*} is manually encoded as a meronym of {*forest*}, then {*forest*} will automatically become a holonym of {*tree*}. And if {*mailbox*} is manually encoded as hyponym of {*box*}, then {*box*} will automatically become a hypernym of {*mailbox*}. The lexical (word-word) relations are bidirectional, too.

WordNet as a Thesaurus

While paper dictionaries are necessarily organized orthographically, WordNet's structure centers upon a word's semantics. The digital format enables targeted look-up for meaning-related words and concepts from multiple access points. A browser with a pull-down menu lets the user search for a keyword's hyponyms, hypernyms, meronyms, antonyms, morphologically derived words, etc. Unlike a traditional thesaurus such as Roget's, the arcs among WordNet's words and synsets express a finite number of well-defined relations.

WordNet as a Tool for Disambiguation

WordNet's design has proved useful for a range of Natural Language Processing tasks that involve the challenge of word sense identification. While lexical polysemy can be resolved in part by statistical approaches, WordNet facilitates alternative or complementary symbolic approaches that exploit its encoding of semantic similarity. Given a polysemous word such as *trunk*, the information that its synonym

Personal Copy

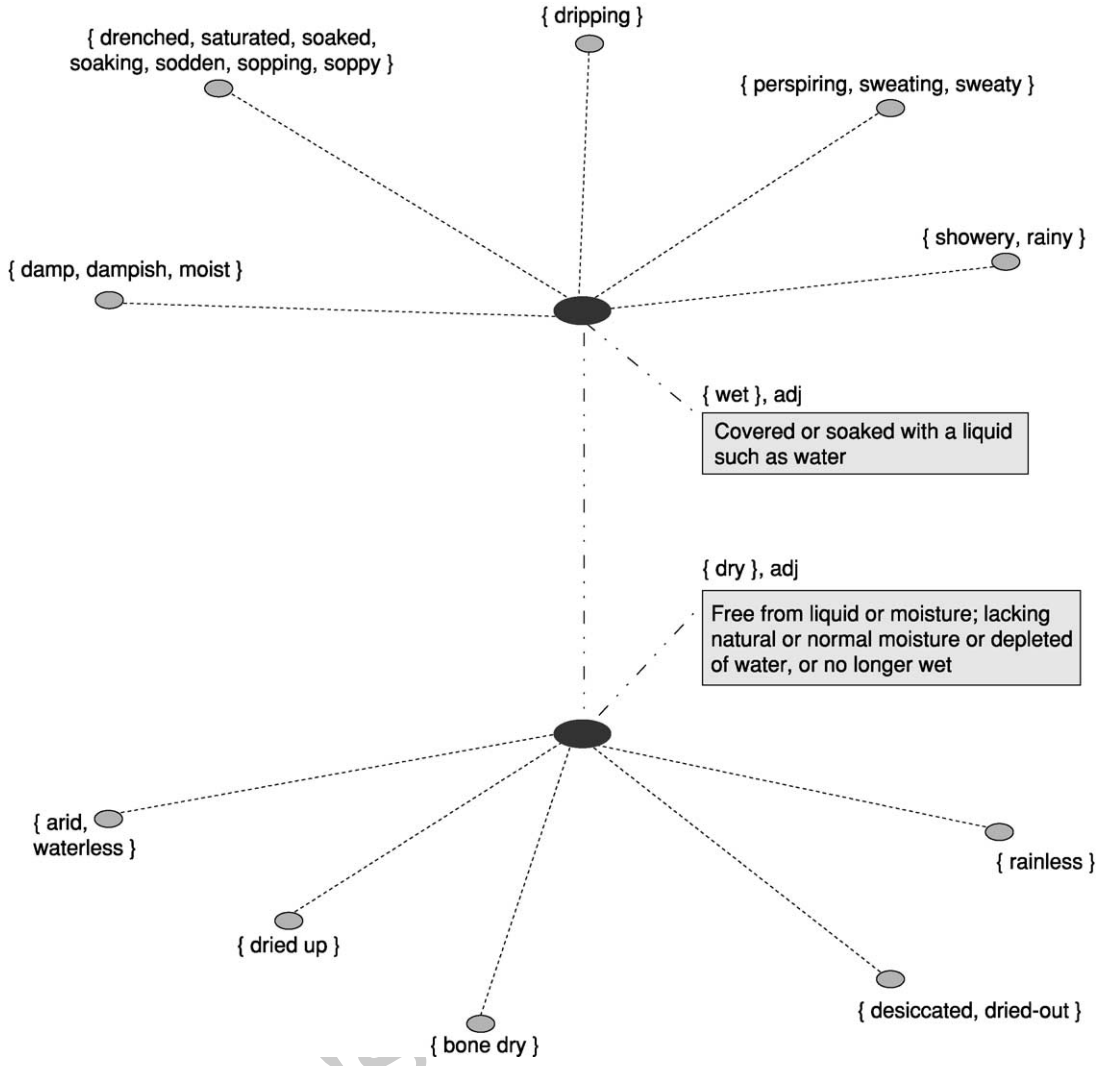

Figure 3 An adjective cluster in WordNet.

is either *tree trunk* or *torso* or *proboscis*, or that either *stem* or *body part* or *snout* are its superordinate, limits the possible readings to one.

Homonyms such as *trunk* have senses that are clearly distinguishable, but the dozens of senses of polysemous words such as *run* are less sharply differentiable, and the divergent entries in standard dictionaries indicate the impossibility of an agreed-upon sense inventory. WordNet's senses are no more fine-grained than those of a collegiate dictionary, but the requirements for disambiguation, rather than look-up of unfamiliar senses, highlight the constraints of an enumerative lexicon (Fellbaum *et al.*, 1997). Many contexts do not allow the singling out one of several overlapping senses. A solution, adopted by the Princeton WordNet and multilingual wordnets, is to group related senses together into one 'coarse' underspecified sense (Palmer *et al.*, 2005) for manual and automatic word sense identification.

Limitations of WordNet

WordNet is a lexical resource and as such does not contain any syntactic information. But there is strong evidence that, at least for verbs, semantic makeup and syntactic behavior are correlated (Levin, 1993). The extent to which WordNet's organization reflects syntactic classes is an independent test of this correlation (Kipper *et al.*, 2000). WordNet does not consider syntagmatic relations. Thematic and semantic roles of nouns functioning as arguments of specific verbs are not encoded, as in FrameNet (Fillmore *et al.*, 2003).

In 1986, digital corpora were not available, and WordNet's contents are largely derived from its creators' intuitions. More recently, the illustrative sentences have been based on web data. But WordNet, unlike some of its descendants, is not a corpus-induced dictionary.

Author's

Other Wordnets

Since the 1990s, wordnets are being built in other languages. EuroWordNet (Vossen, 1998) (EWN) encompasses eight languages, including non-Indo-European Estonian. EuroWordNet introduced some fundamental design changes that are the standard for subsequent wordnets. Other wordnets are linked to the Princeton WordNet.

The EuroWordNet Model (EWN)

Each wordnet relates to three language-neutral components: the Interlingual Lexical Index (ILI), the Domain Ontology, and the Top Concept Ontology. The Top Concept Ontology is a hierarchically organized set of about 1,000 language-independent core concepts expressed in all wordnets. The Domain Ontology consists of a set of topical concepts such as *traffic* and *illness*; unlike the domain labels in the Princeton WordNet, it is hierarchically structured. The ILI is an unstructured, flat list of lexical meanings consisting of a synset, an English gloss, and a reference to its source.

The language-specific wordnets consist of lexical items indexed to a set of synsets in that language. Each is related to a synset in the ILI, which functions as a language-independent ontology, or interlingua, that mediates among the synsets of the individual languages. To find equivalents across languages requires going through the ILI. Fine-grained ILI senses are clustered when necessary into coarser senses to allow unambiguous mapping to senses in other languages. The ILI includes a subset of Princeton WordNet's synsets plus concepts that are lexicalized in a given language where English shows a lexical gap. Thus, the ILI constitutes the superset of all concepts included in wordnets.

No relations link the ILI records, but the synsets in each language-specific wordnet are interconnected via independent semantic and lexical relations for each language. This design feature avoids the problem of crosslinguistic mismatches in the patterns of lexicalization and hierarchical structure.

In contrast to the Princeton WordNet's strict limitation to paradigmatic relations, connections in EWN are encoded among nouns and verbs that are syntagmatically associated, such as the pair *student* and *learn*. Another innovation is conjunctive and disjunctive relations. Conjunctive relations allow a synset to have multiple superordinates. Thus *knife* is both a kind of a *tool* and a kind of a *weapon*. This double parenthood captures the type vs. role (or function) distinction (Pustejovsky, 1995). Another example is *albino*, which can be a kind of *person*, *animal*, or *plant*. A disjunctive relation exists between *airplane* and its possible meronyms *propeller* and *jet*; a given type of airplane has either one, but not both, parts (Vossen, 1998).

As in the Princeton WordNet, words from different syntactic categories are linked when they are semantically and morphologically related, as are the corresponding senses of *visit* (verb), *visit* (noun), and *visitor*.

The wordnet web constitutes a powerful tool for multilingual Natural Language Processing (NLP) applications and crosslinguistic study of lexicalization patterns.

Global WordNets

Databases exist for 35 languages. Descriptions of many wordnet projects can be found in Singh (2002) and Sojka *et al*. (2004).

Wordnets for typologically adverse languages pose novel problems, especially with respect to the concept of *word*, which must be defined to determine synset membership. Challenges include the morphology of agglutinative languages such as Turkish and Estonian (Bilgin *et al*., 2004; Kahusk and Vider, 2002) and languages such as Hebrew and Arabic, where words are generated from a root that constitutes a kind of 'super-concept' (Black and ElKateb, 2004).

For critical reviews of WordNet, see Kilgarriff (2000) and Lin (1999).

Acknowledgments

This work was partially supported by a grant from the Advanced Research and Development Activity.

See also: Antonymy and Incompatibility; Hyponymy and Hyperonymy; Lexical Semantics: Overview; Synonymy.

Bibliography

Bilgin O, Cetinoglu O & Oflazer K (2004). 'Morphosemantic relations in and across WordNet.' In Sojka *et al*. (eds.). 60–66.

Black W J & ElKateb S (2004). 'A protoype English-Arabic dictionary based on WordNet.' In Sojka *et al*. (eds.). 67–74.

Collins A M & Quillian M R (1969). 'Retrieval time from semantic memory.' *Journal of Verbal Learning and Verbal Behavior 8*, 240–247.

Cruse D A (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Fellbaum C (ed.) (1998a). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fellbaum C (1998b). 'A semantic network of English verbs.' In Fellbaum (1998a). 69–104.

Fellbaum C (2002). 'On the semantics of troponymy.' In Green R, Myang S & Bean C (eds.) *Relations*. Dordrecht: Kluwer. 23–34.

Author's Personal Copy

Fellbaum C, Grabowski J & Landes S (1997). 'Analysis of a hand-tagging task.' In *Proceedings of the Workshop on Tagging Text with Lexical Semantics*. Washington, DC: Association for Computational Linguistics. 34–40.
Fellbaum C & Miller G A (1990). 'Folk psychology or semantic entailment? A reply to Rips and Conrad.' *The Psychological Review 97,* 565–570.
Fellbaum C & Miller G A (2003). 'Morphosemantic links in WordNet.' *Traitement Automatique des Langues 44,* 69–80.
Fillmore C J, Johnson C R & Petruck M (2003). 'Background to Framenet.' *International Journal of Lexicography 16(3),* 235–250.
Gross D, Fischer U & Miller G A (1989). 'The organization of adjectival meaning.' *Journal of Memory and Language 28,* 92–106.
Kahusk N & Vider K (2002). 'Estonian Wordnet benefits from word sense disambiguation.' In Singh *et al*. (2004). 26–31.
Kilgarriff A (2000). 'Review of WordNet: An electronic lexical database.' *Language 76(3),* 706–708.
Kipper K, Dang H T & Palmer M (2000). 'Class-Based Construction of a Verb Lexicon.' In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*. Austin, TX.
Levin B (1993). *English verb class alternations*. Chicago, IL: University of Chicago Press.
Lin D (1999). 'Review of WordNet: An electronic lexical database.' *Computational Linguistics 25(2),* 292–296.
Miller G A (ed.) (1990). WordNet: 'An on-line lexical database.' Special issue, *International Journal of Lexicography 3(4)* 235–312.
Miller G A (1995). 'WordNet: A lexical database for English.' *Communications of the ACM 38(11),* 39–41.
Miller G A (1998). 'Nouns in WordNet.' In Fellbaum (1998a). 23–46.
Miller G A & Fellbaum C (1991). 'Semantic networks of English.' *Cognition 41,* 197–229.
Miller G A & Hristea F (2004). 'WordNet nouns: Classes and instances.' Manuscript: Princeton University.
Miller K J (1998a). 'Modifiers in WordNet.' In Fellbaum (1998). 47–67.
Palmer M S, Dang H T & Fellbaum C (2005). 'Making fine-grained and coarse-grained sense distinctions, both manually and automatically.' *Natural Language Engineering*.
Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
Singh U N, Jayaram B D, Fellbaum C & Vossen P (eds.) (2002). *Proceedings of the first international Global WordNet conference*. Mysore, India: Central Institute of Indian Languages.
Sojka P, Pala K, Smrz P & Fellbaum C (eds.) (2004). *Proceedings of the second international Global WordNet conference*. Mazaryk University Brno, Czech Republik.
Tufis D (ed.) (2004). 'Special Issue on Balkanet.' *Romanian Journal of Information Science and Technology 7*, 1–2.
Vossen P (1998). *EuroWordNet: A multilingual database with lexical semantic networks*. Dordrecht, Holland: Kluwer.

Relevant Websites

http://www.globalwordnet.org – Global WordNet.
http://www.cogsci.princeton.edu – WordNet.

Author's Personal Copy

Wrede, Ferdinand (1863–1934)

W H Veith, Universität Mainz, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Ferdinand Wrede was born in Berlin as the son of a Music Director. He studied Germanistics and history in Berlin with professors who were very famous at that time. He received his doctor's degree from the university of Berlin in 1886. From 1887 on, he was a collaborator of the *Sprachatlas des Deutschen Reichs [Linguistic atlas of the 'Deutsches Reich']* project, with Georg Wenker as director, in Marburg. In 1890, he qualified as a university lecturer; thereafter (1911–1920), he was honorary Professor, afterwards full Professor of Germanic philology in Marburg. Parallel to his university career, there was his work at the university library of Marburg, ending up with the position of a chief librarian (until 1920). After the death of Georg Wenker (1911), he became the director of the *Linguistic atlas of the Deutsches Reich*, and in 1920, he managed to transfer this project into a permanent institution: the Zentralstelle für den Sprachatlas des Deutschen Reichs und deutsche Mundartforschung [Center for the linguistic atlas of the Deutsches Reich and for German dialect research].

Wrede's fundamental studies first were concerned with Germanic: his dissertation on the Vandals (1886) and his second doctor's degree, 1890–1891 on the Ostrogoths in Italy. However, in the context of these studies, the dialectological aspect was always present, resulting in many activities of this kind. After having become a collaborator of the *Sprachatlas des Deutschen Reichs* project, he started commenting on Wenker's dialect maps (Wrede, 1892–1902). In 1908,