Lexical Nets

Miriam Butt
December 2002

WordNet

Main Researchers: George Miller (Princeton), Christiane Fellbaum (Princeton)

Aim: to create a lexical thesaurus (not a dictionary) which models the lexical organization used by humans.

WordNet is free and runs on Unix, Windows and Macintosh

http://www.cogsci.princeton.edu/~wn/

It is designed to be compatible with a range of applications (i.e, one module among many).

WordNet

To Date: 95 600 different word forms (51 500 simple words, 44 100 collocations)

Types of Words: nouns, verbs, adjectives, adverbs (no function words)

Each of these word classes is organized by different principles.

Organization by Likeness

Fillenbaum and Jones (1965): If you give a person a/n
noun, 79% of the time you get a noun
adjective, 65% of the time you get an adjective
verb, 43% of the time you get a verb

Synsets: words are arranged in clusters of synonym sets to help identify the meaning and differentiate it from other meanings.

{ board, plank}

Organization by Likeness

Semantic Relations: The overall organizing principle of WordNet is in terms of semantic relations (rather than orthography). So, one can ask WordNet about all these.

Synonymy (dog, canine)

Antonymy (rich, poor)

Hyponomy (tree, maple) --- ISA Relation

Meronymy (tree, limb) --- HASA Relation

Entailments (snore, sleep), for verbs

Nouns

Basic Organizing Principle: Lexical Inheritance System

To Date: 57 000 word forms organized in 48 800 synsets.

Some collocations, but no proper nouns.

@→ is equivalent to ISA relation

oak @→ tree @→ plant @→ organism

~→ inverse ISA relation

Nouns

ISA Top-Down organizaton: 25 unique "beginners" (Table 1)

Additional Organization within Beginners: this was found to be useful (Figure 1)

Planned: add details which distinguish concepts. E.g., for *canary*:

(1) Attributes: *small, yellow* (adjectives)

(2) Parts: *beak, wings* (nouns)

(3) Functions: *sing, fly* (verbs)

Nouns

Parts and Meronymy: can definite part-whole relations in a number of ways and can ask WordNet about them (p is most frequent).

Component Parts: W_m#p →W_h, W_m is component of W_h
(door, car)

Member: W_m#m →W_h, W_m is member of W_h
(tree, forest)

Stuff: W_m#s →W_h, W_m is the stuff W_h is made of
(aluminum, airplane)

Nouns

The organization of nouns also includes information about antonyms (man, woman), all this makes for a relatively complex network (Figure 2).

Adjectives

Basic Organizing Principle: Related Senses (similarity)

To Date: 19 500 word forms organized into 10 000 synsets

Descriptive, relational, color adjectives.

!→ expresses binary opposition

heavy !→ light and also light ! → heavy

Adjectives

Not all adjectives have direct antonyms. But one can find indirect antonyms via the similarity pointer.

&→similarity pointer

moist &→ wet ! → dry

See Figure 1 for an example network.

One possible organizing principle could be gradation (Table 1), but this was found not to be useful (only 2% of adjectives work this way).

Adjectives

Relational Adjectives mean something like "of, relating/pertaining to, or associated with" and are generally derived from nouns. WordNet provides a pointer to the relevant noun (chemical, chemistry).

Syntactic Restrictions: some adjectives can only appear in attributive position (*the alert child*), some only in predicate position (t*he child is astir*).

See attached handout for some sample adjective codings.

Verbs

Basic Organizing Principle: Lexical Entailments

To Date: 21 000 word forms (13 000 unique) organized into 8 400 synsets (generally, languages seem to have less verbs than nouns).

Simple Verbs, but also particle verbs (*look up*)

Verbs

Verb Organization: Verbs are sorted into 15 files, based on semantic criteria (cf. also Levin (1993) on English verb classes).

1) Bodily Function and Care: 275 synsets, *sweat, shiver, faint, freeze*

2) Change: 750 synsets, *change, alter, vary, modify*

3) Communication: 710 synsets, *beg, order, lisp, neigh*

4) Competition: 200 synsets, *duel, face-off, fight, referee*

5) Consumption: 130 synsets, *drink, eat*

Verbs

6) Contact: 820 synsets, *fasten, attach, rub, paw, box*

7) Cognition: *deduce, induce, infer, guess*

8) Creation: 250 synsets, *create, invent, sew, bake*

9) Motion: 500 synsets, *move, travel, run, gallop, fly*

10) Emotion/Psych: *fear, miss, love, amuse, anger*

Verbs

11) Stative ("other" category): 200 synsets, *equal, differ, be*

12) Perception: 200 synsets): *watch, spy, gaze, ache, hurt*

13) Possession: 300 synsets): *have, give, take, rob*

14) Social Interaction: 400 synsets,*impeach, petition, quarrel*

15) Weather (smallest file): 66 synsets, *rain, thunder*

Verbs

WordNet takes a *Relational* rather than a *Decompositional* Approach

Decompositional: bake = Cause(x, Become (y,BAKED))

Relational: Relate the differing verbs in clusters using the factors identified by Decompositional Semantics as crucial: Cause, Become (=Change), Manner.

Example: *swim* is a type/manner of TRAVEL through water

Verbs

Lexical Entailments: if you snore, that entails you are sleeping. As this seems to be a psychologically useful way to model things, take advantage of it.

Further Factors:

1) temporal inclusion/co-extensiveness

2) Causal Relation/Backward Presupposition

Example: see Figure 3

Familiarity Index

It would be nice to know how "familiar" or "common" a word is.

Frequency has often been implicated in finding this out, but it is hard to calculate frequence for all of English.

Another idea is to take Polysemy as an indicator of familiarity. WordNet calculates this and you can ask it about the *Familiaity Index* of a given word.

Example: horse is common as a noun, polysemy count = 6

Morphy

WordNet has the same problem as many other applications: one needs to figure out the lemma for any given word form.

Solution: home grown stemmer called *Morphy*, which relies on a lexicon and a crude knowledge of English morphology (Table 4) .

FrameNet

Main Researchers: Charles Fillmore (Berkeley)

Aim: to create a computational lexicon which describes the "semantic frames" and valencies of verbs, nouns and adjectives.

FrameNet is not yet accessible, though partner projects ar contributing to the effort (e.g., IMS Stuttgart).

http://www.icsi.berkeley.edu/~framenet

FrameNet

A Typical Frame:

frame(CommericalTransaction)
frame-elements{BUYER, SELLER, PAYMENT, GOODS}
scenes(BUYER gets GOODS, SELLER gets PAYMENT)

Frames can be inherited:

frame(RealEstateTransaction)
inherits(CommercialTransaction)
link(BORROWER=BUYER, LOAN=PAYMENT)
frame-elements{BORROWER, LOAN, LENDER}
scenes(LOAN (from LENDER) creates PAYMENT, BUYER gets LOAN)

FrameNet

Plan: One would like to automatically tag texts with semantic frames. See Figures 1 and 2 (from Coling 1998).

First Experiment: Gildea and Jurafsky (2001)

See Tables 8 and 9