WordNet

A conceptual search enabled electronic lexical database

Nirmal Kumar S

Computer Science and Engineering
The LNM Institute of Information Technology

Electronic Lexical Database

An electronic lexical database is a database consisting of one or several dictionaries that enable searching for the words electronically.

Some Numbers

POS	Unique Synsets Strings		Total Word-Sense Pairs
Noun	117798	82115	146312
Verb	11529	13767	25047
Adjective	21479	18156	30002
Adverb	4481	3621	5580
Totals	155287	117659	206941

Some Numbers

POS	Monosemous Words and Senses	Polysemous Words	Polysemous Senses
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Totals	128391	26896	79450

Conventional Organization

Standard alphabetical procedures for organizing lexical information put together words that are spelled alike and scatter words with similar or related meanings haphazardly through the list.

What is Special about WordNet?

WordNet allows searching the database conceptually, rather than alphabetically.

Why this organization?

The design is inspired by current psycholinguistic theories of human lexical memory.

Psycholinguistics

Psycholinguistics is an interdisciplinary field of research that emerged in 20th Century that is concerned with the cognitive bases of linguistic competence.

Limitations

WordNet does not include information about the etymology or the pronunciation of words and it contains only limited information about usage.

WordNet aims to cover most of everyday English and does not include much domain-specific terminology.

Inception of WordNet

In 1985 a group of psychologists and linguists at Princeton University undertook to develop a lexical database along the lines suggested by some of the studies in their fields.

The initial idea was to provide an aid to use in searching dictionaries conceptually, rather than merely alphabetically – it was to be used in close conjunction with an on-line dictionary of the conventional type.

Inception of WordNet

As the work proceeded, it demanded a more ambitious formulation of its own principles and goals. WordNet is the result.

Organization

WordNet divides the lexicon into five categories:

- Nouns
- Verbs
- Adjectives
- Adverbs
- Function words

Organization

WordNet organizes lexical information in terms of word meanings, rather than word forms.

Lexical Matrix in WordNet

Illustrating the Concept of a Lexical Matrix:

F_1 and F_2 are synonyms; F_2 is polysemous

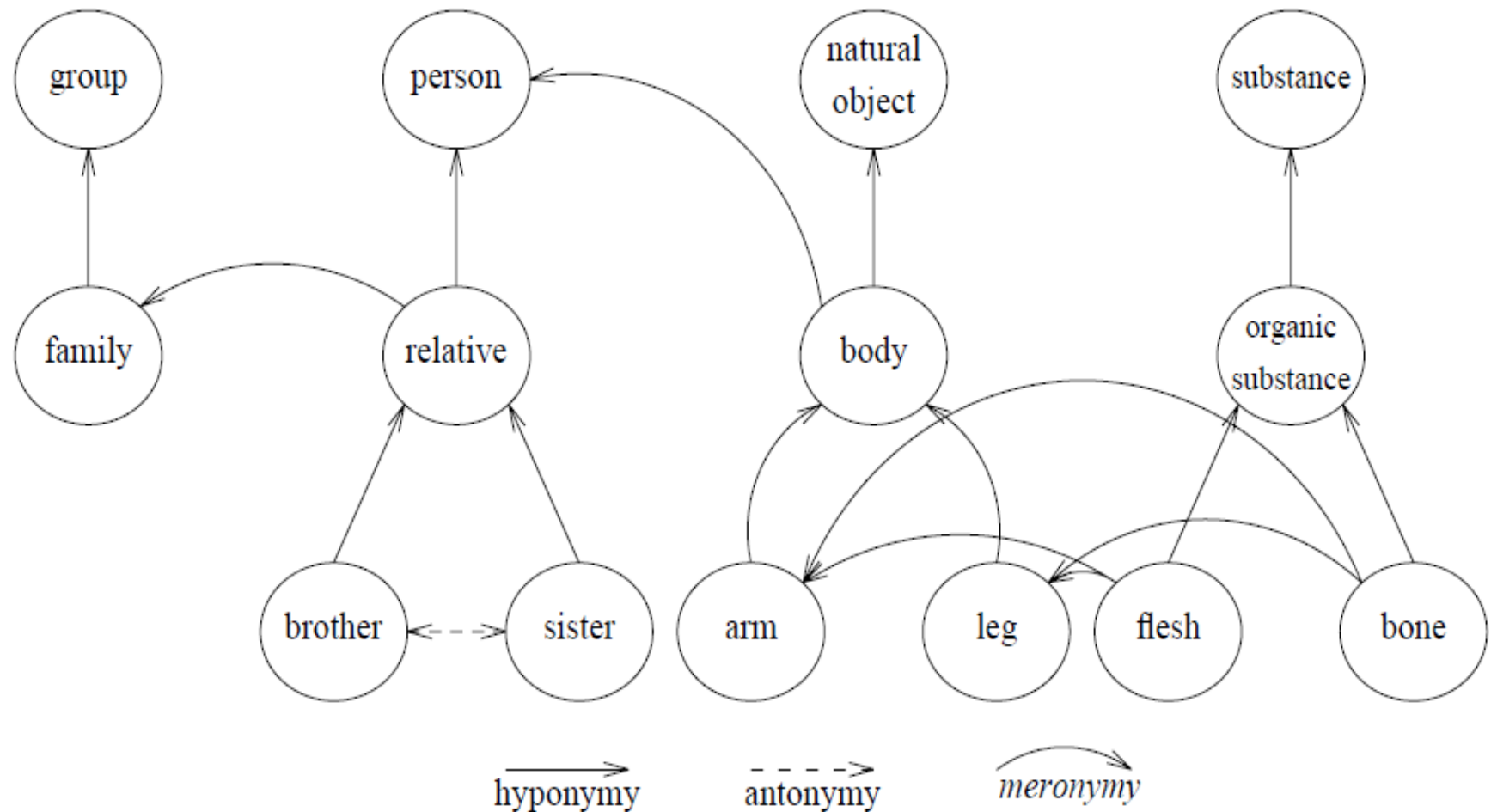| Word Meanings	Word Forms				
	F_1	F_2	F_3	. . .	F_n
M_1	$E_{1,1}$	$E_{1,2}$			
M_2		$E_{2,2}$			
M_3			$E_{3,3}$		
$\vdots$				. . .	
M_m					$E_{m,n}$

Organization of the Concepts

English nouns, verbs, and adjectives are organized into synonym sets, each representing one underlying lexical concept.

Different relations link the synonym sets.

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gin, martini ship, fleet
Troponomy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry

Note: *N = Nouns* *Aj = Adjectives* *V = Verbs* *Av = Adverbs*

LNMIIT
The LNM Institute of
Information Technology

Network representation of three semantic relations among an illustrative variety of lexical concepts

Nouns

Organization of Nouns

One way to construe the hierarchical principle is to assume that all nouns are contained in a single hierarchy.

With some vague abstraction designated, say, {*entity*}, at the top; to make {*object, thing*} and {*idea*} its immediate hyponyms, and so to continue down to more specific meanings, thus pulling all nouns together into a single hierarchical memory structure.

Organization of Nouns

The alternative is to partition the nouns with a set of semantic primes – to select a (relatively small) number of generic concepts and to treat each one as the unique beginner of a separate hierarchy.

Since the features that characterize a unique beginner are inherited by all of its hyponyms, a unique beginner can be regarded as a primitive semantic component of all words in its hierarchically structured semantic field.

Organization of Nouns

Partitioning the nouns reduces the size of the files that the lexicographers must work with, and makes it possible to assign the writing and editing of different files to different lexicographers.

List of 25 unique beginners for WordNet nouns

{act, action, activity}   {natural object}
{animal, fauna}   {natural phenomenon}
{artifact}   {person, human being}
{attribute, property}   {plant, flora}
{body, corpus}   {possession}
{cognition, knowledge}   {process}
{communication}   {quantity, amount}
{event, happening}   {relation}
{feeling, emotion}   {shape}
{food}   {state, condition}
{group, collection}   {substance}
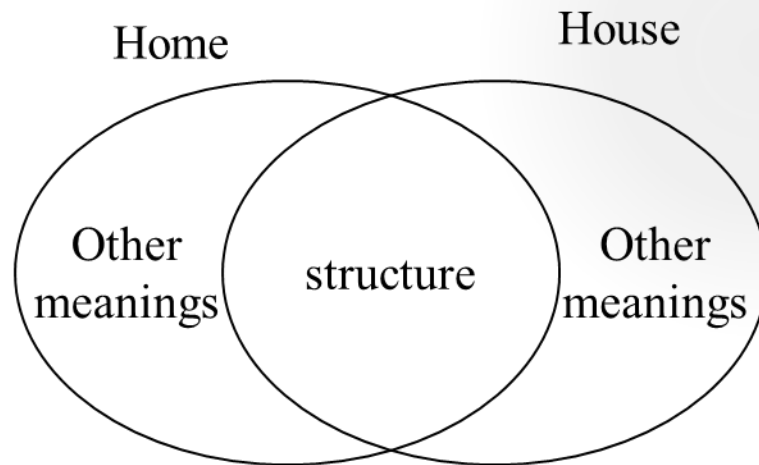{location, place}   {time}
{motive}

Synonyms

Home

Home page

House

Parliament house

Handling polysemy

- *{cabinet, case4, console, cupboard,@* (a cupboard with doors and shelves)}

- *{case5, container,@* (a small portable metal container)}

- *{shell, shell plating, case6, casing1, outside surface,@* (the outer covering or housing of something)}

- *{casing, case7, framework,@* (the enclosing frame around a door or window opening)}

Handling polysemy

- *{carton, case0, box,@* (a box made of cardboard; opens by flaps on the top)}

- *{case1, bag,@* (a portable bag for carrying small objects)}

- *{case2, pillowcase, pillowslip, slip2, bed linen,@* (a removable and washable cover for a pillow)}

- *{bag1, case3, grip, suitcase, traveling bag,@* (a portable rectangular traveling bag for carrying clothes)}

Verbs

Verbs in English

- English sentences require a verb though not necessarily a noun
  - eg: cheating is bad

- language has far fewer verbs than nouns
  - *Collins English Dictionary* lists 43,636 different nouns and 14,190 different verbs

- Verbs are more polysemous than nouns
  - the nouns in *Collins* have on the average 1.74 senses, whereas verbs have on the average 2.11 senses

Verbs in WordNet

WordNet contains over 21,000 verb word forms (of which over 13,000 are unique strings) and approximately 8,400 word meanings (synsets)

Verb Categories

Bodily care and functions

Change

Cognition

Communication

Competition

Consumption
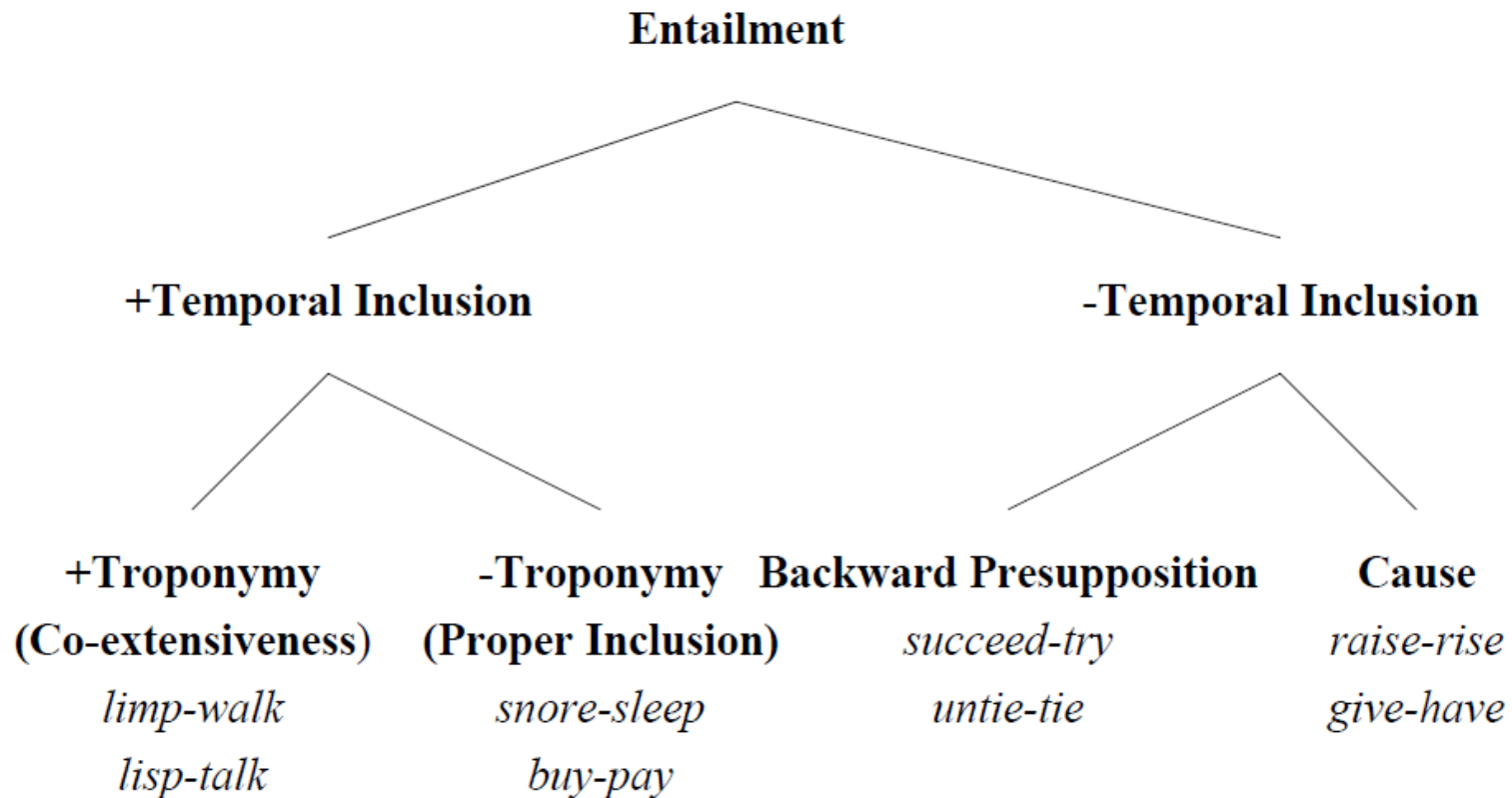
Contact

Creation

Emotion

Motion

Perception

Possession

Social interaction

Weather verbs.
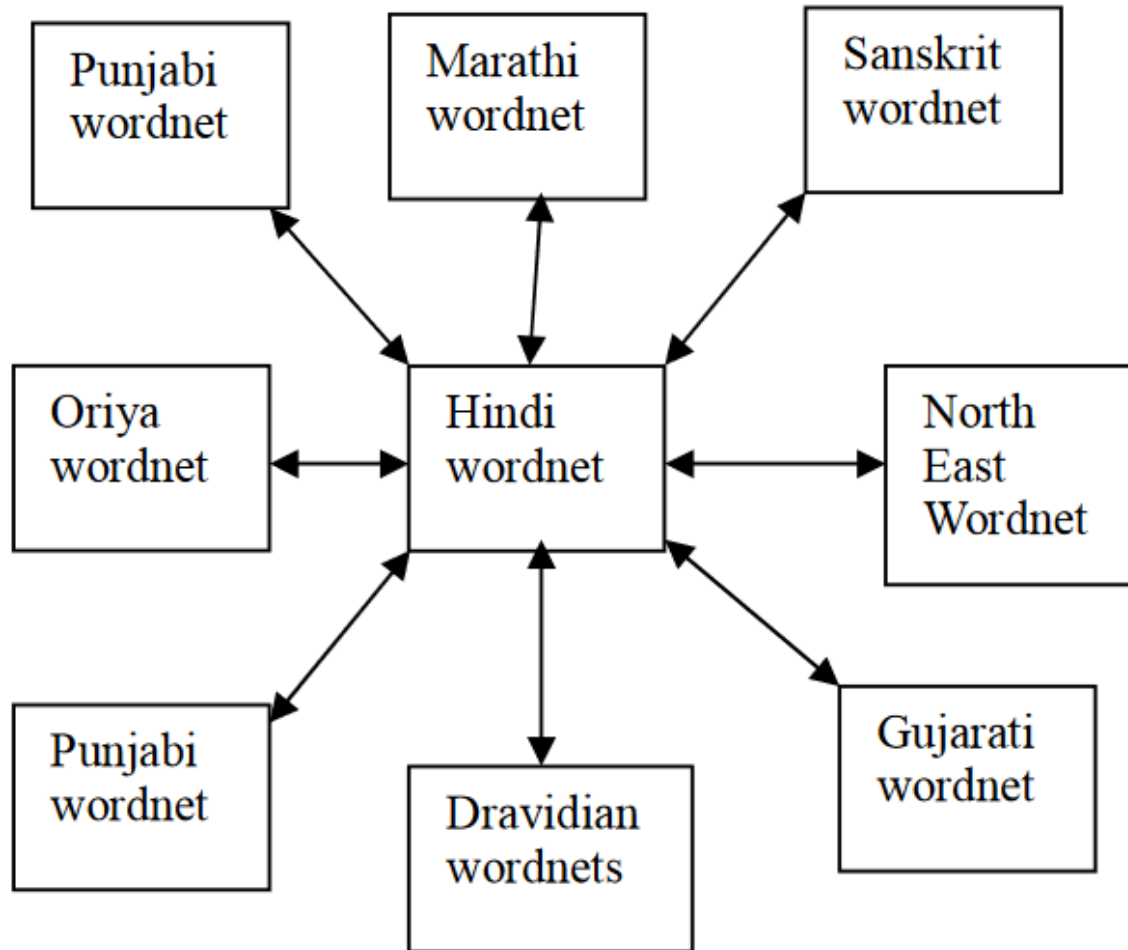
verbs referring to states

Entailment

IndoWordNet

Wordnet –Language	Institute(s)
Assamese	Guahati University, Assam
Bengali	Indian Statistical Institute Kolkata, IIT Kharagpur and Jadavpur University
Bodo	Guahati University, Assam
Gugarati	DDU, Nadiad, Gujarat
Hindi	IIT Bombay
Kannad	Amrita University, Koimbatore
Kashmiri	Kashmir University, Srinagar
Malayalam	Amrita University, Koimbatore
Manipuri	Manipur University, Imphal, Manipur

Marathi	IIT Bombay
Nepali	Assam University, Silchar, Assam
Oriya	University of Hyderabad
Punjabi	Thapar Institute and Punjabi University, Patiala
Sanskrit	IIT Bombay
Tamil	Tamil University, Thanjavur and Amrita University
Telugu	University of Hyderabad and Dravidian University, Kuppam
Urdu	University or Hyderabad and International Institute of Information Technology, Allahabad

Wordnet –Language	#synsets/unique-words
Assamese	3530/19609
Bengali	8679/ 18563
Bodo	3837/13357
Gugarati	970/2125
Hindi	33900/82000
Kannad	5920/7344
Kashmiri	6569/8674
Malayalam	6154/8622
Manipuri	2744/5231
Marathi	9739/21223
Nepali	5802/10278
Oriya	To start
Punjabi	To start
Sanskrit	3340/17820
Tamil	4750/9821
Telugu	10639/18250
Urdu	6123/9641

Other WordNets

Other WordNets

Chinese

French

German

Polish

Bulgarian

Finnish

Arabic

Euro

BalkaNet
    Bulgarian
    Czech
    Greek
    Romanian
    Turkish
    Serbian

Thank you!

The LNMIIT – Where young dreams take shape...