

# Words, Concepts, Meanings

Christiane Fellbaum

© Christiane Fellbaum

# Big question

(How) are language and ontology related?

Words and concepts?

# Some Definitions

- The **Lexicon**: the component of grammar that includes **speakers' knowledge of words**
- Very large (> 40 000 words) Do the dictionary test!
- Open-ended
- Dynamic--never static
- Acquisition is on-going, lifelong (so is loss)
- No two speakers have exactly the same lexicon
- Terminology, regional differences, slang, youth language,...

# Knowledge of words includes:

- **Sound** (pronunciation)
- **Morphology**
  - e.g., plural formation: *woman=>women*,  
*house=>houses*
  - verbal inflections: *I go/he goes/I went,...*
- **Syntax**
  - e.g., preposition selection for adjectives: proud  
of/\*in
- For many languages: **written representation**  
(spelling, mapping to graphemes)

# Knowledge of words includes:

- Selectional restrictions/preferences:

*strong vs. ?powerful tea*

*dog house vs. ?canine house/domicile*

# Knowledge of words

- Meaning/concept behind the word
  - Some languages: class membership marker
- E.g., Japanese –*hiki* for counting small animals

# Word vs. world knowledge

Lexical vs. “encyclopedic” knowledge

Fuzzy boundary. Example: “radon”

*World dictionary:*

a colourless radioactive element of the rare gas group, the most stable isotope of which, radon-222, is a decay product of radium. It is used as an alpha particle source in radiotherapy. Symbol: Rn; atomic no: 86; half-life of 222 Rn: 3.82 days; valency: 0; density: 9.73 kg/m<sup>3</sup>; melting pt: --71°C; boiling pt: --61.7°C

*Wikipedia:*

Radon (play /'reɪdɒn/ ray-don) is a chemical element with symbol Rn and atomic number 86. It is a radioactive, colorless, odorless, tasteless noble gas, occurring naturally as the decay product of uranium. It is one of the densest substances that remains a gas under normal conditions and is considered to be a health hazard due to its radioactivity. Its most stable isotope, 222Rn, has a half-life of 3.8 days. Due to its intense radioactivity, it has been less well-studied by chemists, but a few compounds are known...

# Some definitions: **Word**

Intuitively easy, but...

(1) Word = concept/unit of meaning?

Nootka verb (Sapir):

*I have been accustomed to eating twenty round objects while engaging in...*



# Word

(2) Word = sequence of letters between white empty spaces

Not all languages use letters (~phonemic representation) or segment words  
e.g., Chinese runs together all characters

# Word

(3) Word = minimal free morpheme

Status of bound morphemes (*un-*, *-ed*)?

Compounds; compound frames (*Miami-to-Montreal* train, *Newark-to-Verona* flight)?

Multi-Word Units (*put up*, *end-of-life care*)

idioms (*hit the ceiling*, *won't hear of it*, *blow one's stack*)

Non-compositional MWUs: *hot dog*, *hothouse*

# A useful definition

Words = form-meaning mappings

# Less useful definition

..but one we will adopt here for practical purposes:

Word meaning = “concept”

(Barry Smith is not in the room...)

# What gets labeled?

- Some concepts may be innate
- Others bootstrapped from innate concepts
- Hard-wired?
- Based on perception?
- Quine's rabbit

# What gets labeled?

All humans perceive same color spectrum

But languages don't "cut up," label colors in the same way

e.g., Russian has two words for blue

Interior designers have hundreds of words for *red, blue, etc.*

Berlin & Kay color study

# Things we will ignore

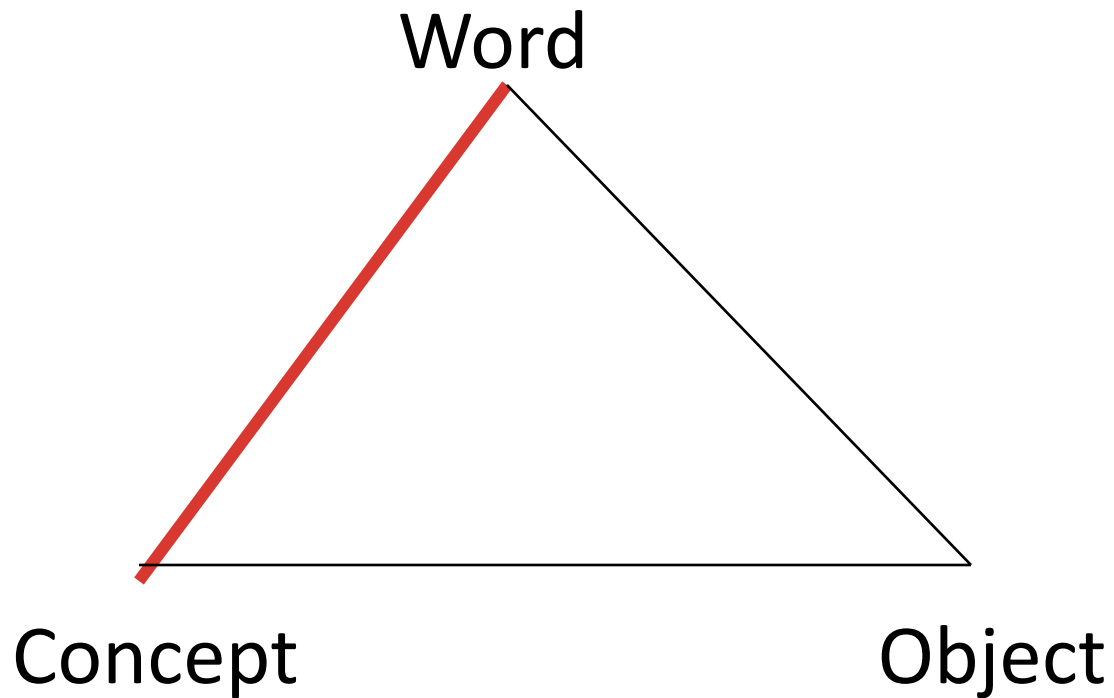
No two speakers associate precisely the same meaning(s) with a word form, but there is a core meaning that is agreed upon by convention within the speaker community (e.g., your concept of *house* may not be the same as mine but we are likely to call the same instances *houses*)

# More things we will ignore

- Denotation: core meaning, independent of particular contexts, speaker associations, etc.
- Connotation: associated meaning facets, not always stable; can vary across speakers (e.g., “mother” may connote love, protection, etc. for many but not all speakers; “socialism” generally has positive connotations in Western Europe but very negative connotations in the U.S.)



# Semiotic triangle (Peirce, Osgood)



# Saussure

Arbitrariness of the sign

Form is arbitrary, agreed upon by convention

# Words

Two important properties of words  
(universal?)

Polysemy/Homonymy

Synonymy

# The lexicon

A formal definition:

Repository of all form-meaning pairs

Ideal (?) situation: one form, one meaning:

Every word form has exactly one meaning

Every meaning is expressible by exactly one  
word form

Real Life: (mostly) many-to-many mappings

# Synonymy

One meaning/concept is expressed by several different word forms:

{beat, hit, strike}, {shut, close}

{car, motorcar, auto, automobile}

{big, large}, {difficult, hard}

# Polysemy

One word form expresses multiple meanings

{*table*, tabular\_array}

{*table*, piece\_of\_furniture}

{*table*, table\_mountain, mesa}

{*table*, postpone}

# Polysemy, homonymy

Homonymy: meanings unrelated to everyday speaker

*pitch* (acoustic property/tar)

*bar* (saloon/stick)

*bass*(fish, musical instrument)

Polysemy: related meanings:

*bass* (vocal range, singer)

Borderline can be fuzzy.

# Polysemy test

Zeugma: conjoining of different senses of a word is odd

?He *left* Rome and the bills on the table

*leave<sub>1</sub>: go away from a place*

*leave<sub>2</sub>: leave behind, cause to remain in a specific place*

*cf.:*

*He left Rome and, later, the country*

*He left the bills and a bottle of beer on the table*



# Polysemy

Regular, systematic polysemy:

*book, newspaper,...* (object, content)

*chicken, beans,...*(animal/plant, food/dish)

Metonymy (part-whole polysemy)

*The House of Saud issued a statement*

*The office isn't answering the phone*

# Polysemy: Metaphor

Conventionalized:

“Time is Money”

*Save/spend/invest/waste/lose/gain time*

(Lakoff and Johnson 1980)

Unconventional, ad hoc (though readily interpretable):

*My surgeon is a butcher*

*The restaurant is a zoo on weekends*

(we’ll ignore this)

# The lexical matrix

A “map” of the lexicon

Plotting polysemy and synonymy

Most cells are empty

pitch	tar				
pitch		frequency			
pitch			throw		toss
	tar			tarball	

# Synonymy, polysemy

perfect for poets: great power of expression  
(synonymy) and deliberate ambiguity or  
vagueness (polysemy)

Terrible for computers!

# Zipf's Law

A power law:

Frequency of a word is inversely proportional to its rank in a frequency table

Most frequent word occurs twice as often as the second most frequent word

Three times as often as the third most frequent word, etc.

# This is not quite true...

## Frequency rank lists and frequencies

Brown Corpus (1 mio words)	British National Corpus (100 mio)
the 69975	the 6055159
be 39175	be 654445 (all inflected forms)
of 36432	of 3051609
and 28872	and 2632194
a 23073	a 2168817
in 20870	in 1944328

(Zipf's Law holds for the middle of the curve; there's a very long tail with many words that are used only once in a corpus)

# To make things worse...

The most frequent word forms are the most polysemous  
Frequency/sense distribution follows a power law (roughly)



# Interim conclusions

The lexicon (mapping of concepts to word)

- is not systematic (not 1:1)
- is dynamic, but its development is uncontrolled and seemingly random
- is organic, not designed or artificially constructed
- But serves human speakers perfectly!

# Questions

- How well does the lexicon reflect speakers' conceptual inventory?
- How can we discover the concepts behind the words?
- Is the conceptual inventory as messy as the lexicon?

# Lexicon and ontology

- Ontology is logical, systematic
- Formal, ontological perspective can “make sense” of the lexicon
- Mapping of lexicon and ontology clearly reveals differences between the two
- More on this later...

# Next questions

- Why is the lexicon (of all human languages) so seemingly messy?
- What does the lexicon show about the human mind and human cognition?