

Semantik

Ola Knutsson
DH2418, Språkteknologikursen

“The word is a thing in our consciousness ... that is absolutely impossible for one person, but that becomes a reality for two” (Vygotsky, 1987)

Ordens betydelser beror av de andra ordens betydelser

You shall know a word by the company it keeps (Firth, 1957).

Is this paper any good at all? (Fillmore (1967)

Inlärare måste “veta” 95 % av ordens betydelser i en text för att förstå den, men endast 2000 lexem för att förstå 80 %.

Lexikal semantik

- Har orden någon betydelse?
- Har vårt lexikon en systematisk uppbyggnad?
- Är lexikonet statiskt?
- Om vi vet vad orden betyder vet vi då vad texten handlar om?
- Är alla lexem är definierade genom andra lexem?
röd --> blodets färg
blod --> en röd vätska som cirkulerar i kroppen

”Any word exists for the speaker in three aspects: as a neutral word of a language, belonging to nobody; as an other's word, which belongs to another person and is filled with echoes of the other's utterance, and, finally, my word, for, since I am dealing with it in a particular situation, with a particular speech plan, it is already imbued with my expression” (Bakhtin, 1986).

- Syntagmatiska relationer: ord som **samförekommer**, kombinatoriska relationer (fraser, meningar, stycken, dokument)
- Paradigmatiska relationer: ord som **inte samförekommer**, ersättningsorienterade relationer (samförekommer med samma **andra** ord).

• Läs mer: Sahlgren, Magnus (2006) *The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, Doctoral thesis, Stockholm University.

Struktur inom lexikala begrepp

Defintion av *änka*:

X är/var en änka vid tidpunkten t omm

X är en kvinna och det existerade en person Y, sådan att:

Y var gift med X vid tidpunkten u

Y dog vid tidpunkten u och

$u > t$ (u kommer före t)

Hur fungerar vårt mentala lexikon?

Formlikhet, ord som rimmar:

Häst - präst

Men vanligast är semantiskt inflytande (gäller även felsägningar)

Svart – vit

Mat – äta

Formrelationer: Homonymi

- Homonymi – olika betydelse identisk form
- Homografer hov (hästhov) – hov (kungens hov)
- Homofoner hov (kungens) – håv (fjärilshåv)
- Homografer & Homofoner ro (en båt) - ro (lugn och ro)

Polysemi

Flera olika närbesläktade betydelser baserade på samma lemma.

hänga

- hållas uppe
- anbringa
- avliva
- vara direkt beroende av
- vistas

Problem i tillämpningar?

Real word errors

Taligenkänning

Text-till-tal

IR

Betydelserelationer

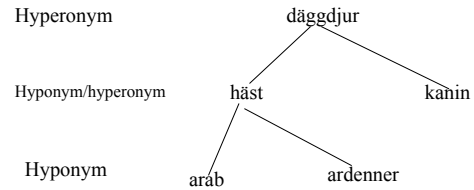
- Synonymer – två ord med samma betydelse
- Utbytbara i alla kontexter? *Båda – bägge*
Ej *häst, springare, kuse*.

Utbytbarhet: synonymer om orden kan bytas ut utan förändringar i betydelse eller acceptans. I praktiken: i någon språklig kontext.

Hyponymi

Hyponym – underordnade uttrycket

Hyperonym - överordnade uttrycket



Hyponymi-test

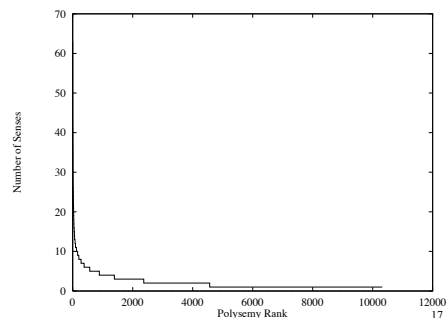
Det där är en x => Det där är en y.

Det där är en ardenner => Det där är en häst

<i>Semantiska relationer</i>	<i>Exempel</i>
ÖVER-/UNDERORDNADE BEGREPP (HYPONYMI)	fordon => bil => skåpbil; förmimma => se => skymta;
HELHET/DEL (MERONYMI)	
HELHET/KOMPONENT	gren => kvist =>blad;
GRUPP/MEDLEM	svärm=>bi; förening=>medlem; flotta=>skepp
FÖREMÅL/ÄMNE, MATERIAL	pinne => trä; klippa=>sten, mineral
MOTSATSORD (ANTONYMI)	våt - torr; köpa - sälja; pojke - flicka
SYNONYMGRUPPER	blöt, våt; pojke, grabb, kille

Princeton WordNet (version 1.6)

Kategori	Unika former	Antal betydelser
Substantiv	94474	116317
Verb	10319	22066
Adjektiv	20170	29881
Adverb	4546	5677



EuroWordNet

Synsets kan länkas till en “Top-concept”-ontologi med 63 språkoberoende koncept, t.ex. Plant, Human, Animal, Vehicle, Furniture m.fl.

Holländska, italienska, spanska, tyska och franska finns, fler är under utveckling.

Svenskt ordnät (SWordNet)

- Utvecklat i Lund och Uppsala
- 15000 synsets (2001), främst substantiv och verb
- Skillnader mellan språken, delvis p.g.a. olika kodningsprinciper: Holländska: *mens* 572 direkta hyponymer, Tyska *Mensch* 12 direkta hyponymer och svenska *människa* ett tjugotal.
- Det svenska verbet *få* (rang 5) saknar direkt ekvivalent i många andra språk. få – get

Why don't we get a microwave? Varför skaffar vi oss inte en mikrovågsugn?

FrameNet (Charles Fillmore)

- Ramar (frames) för verben
- Koppling fraskategorier+syntaktisk funktion och tematiska roller

Engelska, tyska, japanska och spanska

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

Tematiska roller

AGENT	Kalle öppnar dörren
UPPLEVAREN	Kalle såg en vit mus
DET UPPLEVDA	Kalle såg en vit mus
INSTRUMENTET	Kalle skar brödet med kniven
KÄLLA	Kalle reste från Oslo
MÅL	Kalle bilade till Uppsala

Kreativitet och lexikon

Lexikal semantik bygger ofta på ett statiskt lexikon.
Vi människor skapar hela tiden nya betydelser. Inte obegränsat eller osystematiskt.

Metaforer: ett uttryck tolkas, upplevs eller förstås genom ett annat uttryck. *Han iklädde sig med förtjusning fadersrollen. På samhällets botten. Marginalskatten äter upp inkomstökningar.*

Metonymi = tydligt samband mellan uttryckets vanliga referent och det vid yttrandet avsedda referenten. *Fyran – gästerna vid bord fyra. Växeln – personalen som sköter telefonväxeln*