

PLIN021 Sémantická analýza v praxi

OP VK Mezi bohemistikou a informatikou
www.projekt-inova.cz

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

8. března 2012

PLIN021 Sémantická analýza v praxi

Předmět PLIN021 Sémantická analýza v praxi je podpořen projektem OP VK Mezi bohemistikou a informatikou. Inovace vysokoškolské výuky češtiny v kontextu počítačového zpracování přirozeného jazyka (INOVA.CZ).

www.projekt-inova.cz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Význam a sémantika

Organizace semestru

Úvod do studia významu

Sémantika

20. století nebylo zlatým věkem sémantiky, protože mu vévodily dvě osobnosti: Leonard Bloomfield, který věřil, že význam leží mimo vědecké zkoumání jazyka, a Noam Chomsky, který se zabýval téměř výhradně abstraktní formální syntaxí. S postupem nového století je jasné, že časy se mění...

[Goddard, 2011] (překl. ZN)

Sémantika

- nauka o významu **znaků**
- znak = něco, co něco zastupuje (reprezentant)
- v našem případě nauka o významu morfémů, slov, slovních spojení, vět, promluv

Jak souvisí sémantika s počítačovou lingvistikou?

co je význam?

- *na sílu u výjezu ma* **na sílu**
- *z ná k = sílu, co sílu zastupuje* | *na pozici sílu*
- *v našem případě násila u výjezu ma možná, doo, dovíc k vneseš, sílu, pomek*

Je k souvisí sémantika s počítačovou lingvistikou?

co je význam?

Největší prostor dostanou samozřejmě slova. V průběhu kurzu ale uvidíme, že jednak přesně nevíme, co je jedno jednotlivé slovo, jednak existuje silné hnutí, které samostatná slova vůbec za nositele významu nepovažuje.

Jak souvisí sémantika s počítačovou lingvistikou? Chceme přece, aby počítače přirozenému jazyku rozuměly. Nevíme sice přesně, co znamená něčemu rozumět, ale víme, jaké jsou projevy porozumění. Také víme, jaké jsou výsledky počítačové analýzy přirozeného jazyka *bez* nějaké simulace porozumění.

Pro počítače musí být všechno přesně formálně definováno. Zkusme ale definovat formálně význam. . . daleko se nedostaneme. Na „význam významu“ budeme přicházet v průběhu kurzu (viz dál).

Význam (meaning)

- nemá klasickou definici
- pokusy o popis významu vedou k ostrým a věčným polemikám
- víme ale, že význam souvisí s porozuměním
- tušíme také, že jazyk je zajímavý právě tím, že jeho části mají význam
- během tohoto kurzu se dostaneme k tomu, co je to význam

Význam (meaning): co není

- **význam \neq reference**
příklady a protipříklady: B. Obama, já, jednorožec, nikdo, ahoj
vynálezce joja, vynálezce parkovacího automatu
někdy mluvíme o **smyslu (sense)**, aby bylo jasné, že nemíníme odkazovaný objekt
- **význam \neq (vědecká) znalost**
příklady: sůl = chlorid sodný NaCl
protipříklady: Bůh, pondělí, sodík
- **význam \neq užití**
význam znaku nelze určit přesněji než výčtem užití tohoto znaku
- **význam \neq pravdivost**
promluva je pravdivá, pokud jsou její části pravdivé a zároveň správně poskládané (pravdivostní podmínky)
jak ale popsat pravdivostní podmínky?

└ Význam a sémantika

└ Význam (meaning): co není

- **význam ≠ slovo**
příklad: a poutník šel B. Obama, (š, je šroubovat, nikdo, a ho)
výsledce: jít, vyřadit a šroubovat a šroubovat
výsledky: a poutník šel B. Obama, a by šel jít, š, není šroubovat
- **význam ≠ jméno / zkratka**
příklad: a il = c. Hlavě šel vy Na Cl
poutník šel B. Obama, poutník
- **význam ≠ síla**
význam má to slovo a čít poutník šel vyřadit šroubovat
a š
- **význam ≠ pravidlo**
poutník je pravidlo, pokud (na) čít pravidlo a šroubovat
a šroubovat poutník (poutník) poutník
je šroubovat poutník poutník

„B. Obama“ je odkaz na stále tutéž osobu, „já“ ne. „Jednorozec“ je odkaz na jakou bytost? „Nikdo“ odkazuje na koho? Vynálezce joja a parkovacího automatu je tatáž osoba (D. F. Duncan)

Můžeme význam soli rozložit na význam sodíku a chloru, ale dál se nedostaneme. . .

S užitím je taky problém, lidé totiž běžně užívají slova novými způsoby (ať už omylem, nebo jako výraz své kreativity).

Význam (meaning): co by mohl být

- konceptualistický přístup: význam slova je myšlenka v hlavě
 - v dětství všichni dostaneme do hlavy významy slov a také rutinu, jak nové významy přidávat a odvozovat
 - díky sdíleným významům také můžeme komunikovat
 - je-li význam myšlenka, kde se vzala? empiricismus vs. racionalismus
- strukturalismus: pochopit význam slova znamená pochopit, jak slovo interaguje s ostatními slovy (význam slova je určený souhrnem vazeb s jinými slovy)
- extensionalismus: rozdíly ve významu slov lze nejlépe zkoumat ve vztahu k extenzím (referencím) (e. pokrývá omezený rozsah)
- v současnosti existuje řada přístupů *mezi* uvedenými

└ Význam a sémantika

└ Význam (meaning): co by mohl být

- které přičítat? být přičítat: význam slova je vykládáno v slově
 a v různých větvích slovnice, ale i když význam slova a také
 tudíž, jak se od významu přičítat a dovozovat
 a díky od významu slova také se kládne ke značce, ale
 a (od významu slova) také se značí? například: co,
 co je slovo
- struktura významu: jak byk význam slova znamená, jak byk,
 slovo intencje a ostatní slovy
 (význam slova je a čí je význam slova k a jiným slovy)
- extensivní význam: což by se význam slova může a jiným slovy
 ve větě ke k významu (jako význam)
- jaký je význam slova k
- v soudu mají význam slova přičítat mezi sebou význam

Je důležité se někam zařadit? To záleží. Z pohledu CL to není důležité, navíc málokdo má takový odstup od své práce, aby si sebezařazení mohl dovolit. Takže typicky v oblasti poč. lingvistiky nastávají tyto případy:

- vědci neví, kam se zařadit
- vědce nezajímá, kam se zařadit
- vědci chtějí praktické výsledky

My se zatím zařadíme mezi tu třetí skupinu. Abychom dosáhli praktických výsledků v CL, musíme se usilovně věnovat formalizaci všech aspektů významu. O tom bude celý kurz. Nejvíc prostoru ponecháme formálním metodám pro práci s (kognitivním) významem slov.

Náplň kurzu

- zkoumat význam z pohledu počítačové lingvistiky
- jak formálně uchopit význam, když nemáme formální definici
- seznámit se s tradičními i současnými přístupy k sémantické analýze
- přemýšlet o aplikaci různých přístupů na český jazyk

Organizace semestru, informace

- přednášky
- studijní materiály v ISu: tyto prezentace
- odkazy na literaturu na konci prezentací
- interaktivní osnova v ISu: program přednášek + důležité informace
- v průběhu semestru narazíme na témata zajímavá pro bakalářskou diplomovou práci
- hodnocení: projekt **nebo** písemný test

Hodnocení: Projekt

- téma vztahující se k libovolné části kurzu
- možné okruhy: prozkoumání nějakého software; aplikace určitého software na český jazyk; srovnání dvou programů a recenze; nalezení určitého jevu v korpusu a např. pravidel, jak takový jev hledat; představení zajímavého článku (časopisecký n. sborník z konference, ne starší než 10 let)...
- zpracování formou písemné práce na 2–6 stran A4, PDF
- volba tématu v průběhu semestru, téma musí schválit ZN
- na jednom projektu mohou pracovat až 3 lidé společně, rozsah musí být odpovídající
- průběžné konzultace podle potřeby
- termín odevzdání: **23. 5. 2012**

Hodnocení: Písemný test

- ve zkuškovém období
- varianty pro studenty PLIN a ČJL



Studium významu: východiska 20. stol.

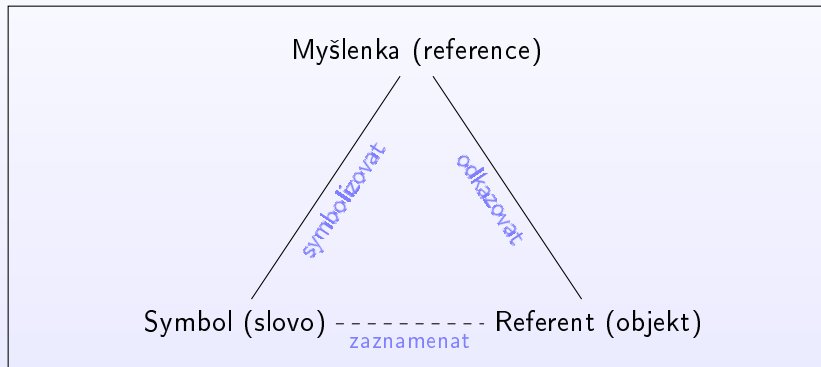
princip kompozicionality (Frege): význam celku lze určit z významů částí

rychlý + auto = rychlé auto

porušení principu kompozicionality – idiomy, frazémy

slaměný + vdovec \neq slaměný vdovec

Studium významu: východiska 20. stol.



C. K. Ogden, I. A. Richards: The Meaning of Meaning, 1923

Typy významu [Leech, 1981]

1. konceptuální (kognitivní, logický, denotační)
 - kontrast (muž vs. žena)
 - princip kompozicionality (žena = adult, human, feminine)
2. konotační (žena = upovídaná, citlivá ...)
3. sociální (mladík vs. ogar)
4. emocionální (kůň vs. herka)
5. reflektovaný (je obtížné užít slova v jednom významu, aby se neasocioval jiný, např. kozy)
6. kolokační (typické pro adjektiva – černý = bez placení v kolokaci „černý pasažér“)

Jak zkoumat význam

- význam je v hlavě, ale chceme o něm mluvit \Rightarrow používáme jazyk
- jakým jazykem popsat jazyk?
- z jazyka nemůžeme uniknout, přesto ale můžeme být přesní a systematictí – užitím **metajazyka**
- jak popsat význam? pomocí **sémantické reprezentace**
- sémantická reprezentace se vytváří z výrazů, které tvoří metajazyk sém. reprezentace (čili **sémantický metajazyk**)



Goddard, C. (2011).

Semantic Analysis: A Practical Introduction.

Oxford Textbooks in Linguistics. Oxford University Press.



Leech, G. N. (1981).

Semantics.

Hardsworth: Penguin.