

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

2. října 2012

Vágnost a subjektivita

Víceznačnost

Je víceznačnost problém?

Zkoumání významu

Význam zkoumáme z pohledu počítačové lingvistiky. Chceme se dobrat nějakého formálního modelu významu. Co nám v tom brání?

- význam je těžké definovat
- lidé jazyk používají „nepořádně“

Zkoumání významu

Řeč je jakési bludiště cest. Přijdeš z jedné strany a vyznáš se tu; přijdeš na totéž místo z jiné strany, a už se tu nevyznáš.

[Wittgenstein, 1953]

Vágnost a subjektivita významů

muž = dospělý člověk mužského pohlaví

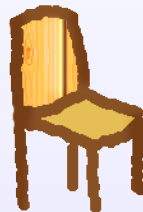
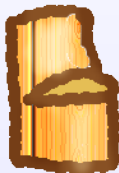
Je 18letý člověk mužského pohlaví muž?

Je 17letý člověk mužského pohlaví muž?

Je 16letý člověk mužského pohlaví muž?

Vágnost a subjektivita významů

špalek židle



Jsou významy diskrétní, nebo spojitý?

└ Vágnost a subjektivita

└ Vágnost a subjektivita významů

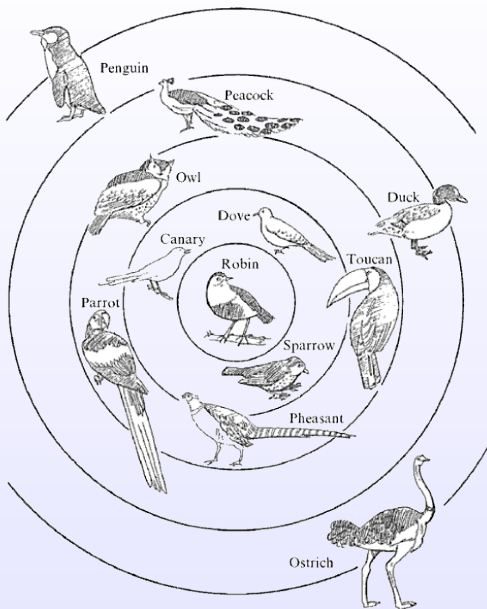
špalek židli



Jaká je na my židli židli, nebo špalek?

Mezi špalkem a židlí, mezi mužem a klukem je dost mezistupňů, na kterých se neshodneme. Jak to, že se vůbec domluvíme?

Tradičně vidíme významy odděleně (diskrétně), jsme na to zvyklí ze slovníků, kde jsou významy očíslovány a seřazeny (podle čeho vlastně?).
Není to chyba? Není rozumnější vidět otesaný špalek jako 10% židli a 90% špalek?



Aitchison,
2003 in
[Goddard, 2011]



Práce se spojitými prostory je složitější a méně intuitivní (méně přirozená?) než s diskrétními prostory.

Řada vědců (Aitchison) se přiklání k nějakému odstupňování významu. Odpovídá tomu i psychologická zkušenost: když řeknu „pták“, představíte si kosa nebo pěnkavu, těžko si někdo ihned vybaví pštrosa nebo kiviho. Kos nebo pěnkava jsou u nás „v centru“ (jsou prototypem ptáka). Tohle centrum ale asi bude kulturně nebo geograficky závislé. . .

Vágnost a subjektivita významů

Wittgensteinův příklad: hra – deskové hry, karetní hry, míčové hry, Olympijské hry, dětské hry ...

Co mají společného?

zábava, soutěž, vítězství, dovednost ...

rodinná podobnost – stavba těla, barva očí, temperament ...

E. Rosch ukázala, že kategorie jsou často organizovány okolo typických reprezentantů.[Langacker, 1987]



└ Vágnost a subjektivita

└ Vágnost a subjektivita významů

Wagnersteinův příklad: hra – ésslové ky, kuzeni ky, mítow ky
Olympjeid ky, áttuid ky...

Co mají společného?

od hra, soustě, věřstě, ésswé mst ...

nísswá joúúú mst – stáwá tllá, káwa úúú, te mje umest ...

E. Rosch a úúúú, Je kategóúú [úúú úúúú úúúúúúúú úúúú
úúúúúúúú úúúúúúúú. [úúúúúúúú 1987]

W. připodobnil významovou podobnost k rodové podobnosti. Existuje, ale přesně a obecně ji popsat je velký problém.

Víceznačnost, homonymie, polysémie

existuje na různých úrovních: přípony, koncovky, slova, slovní spojení, věty

- homonymie – náhodný jev: úplná h. (líčit, kolej) a částečná h. (stát, los ...)
- polysémie – přirozený jev: kohout, strom, kulhat ...

Víceznačnost, homonymie, polysémie

Jak rozeznat polysémii od úplné homonymie?

Nevíme.

Záleží na tom?

└ Víceznačnost

└ Víceznačnost, homonymie, polysémie

Jestli rozumet pokříd mii ot íjítit homonymie?

Nevíme.

Základ na tom?

Někdy je velice snadné homonymii a polysémii rozeznat. Tam, kde k přenesení významu došlo dávno, nebo se přenesl „daleko“, h. a p. může rozeznat jen expert. Pro počítačové zpracování není rozdíl mezi h. a p. důležitý, proto často mluvíme o víceznačnosti (ambiguity), aniž bychom specifikovali, jak vznikla. Později si ukážeme, že můžeme „vzdálenost“ významů docela dobře spočítat.

Je víceznačnost problém v NLP?

rychlý \Rightarrow fast

auto \Rightarrow car

rychlé auto \Rightarrow fast car

vysoký \Rightarrow high

škola \Rightarrow school

vysoká škola \Rightarrow university?

└ Je víceznačnost problém?

└ Je víceznačnost problém v NLP?

rychlý ⇒ fast
auto ⇒ car

rychlé auto ⇒ fast car

vysoký ⇒ high
škola ⇒ school

vysoká škola ⇒ university?

rychlé auto – auto, které zrovna jede rychle; auto, které může jet velmi rychle; auto, na které můžeme rychle vydělat peníze?

vysoká škola – univerzita; škola, jejíž budova je vysoká?

Je víceznačnost problém v NLP?

Odpověď: jak kdy, záleží na aplikaci . . .

└ Je víceznačnost problém?

└ Je víceznačnost problém v NLP?

Bohužel nedokážeme ani přesně stanovit, kdy je víceznačnost problém. Přesněji řečeno dokážeme to až pro konkrétní výrazy, ne obecně.

Jak rozlišovat významy (sense)?

Kolik významů má slovo kočka?

- SSJČ: 7
- SSČ: 2
- PSJČ: 10
- Slovník českých synonym: 4
- Český WordNet: 3

PLIN021 Sémantická analýza v praxi

└─ Je víceznačnost problém?

└─ Jak rozlišovat významy (sense)?

Kolik významů má slovo tuča?

- SSJČ: 7
- SSC: 2
- PSJČ: 3
- Slovník českých synonym: 4
- Český WordNet: 3

ukázka z DebDictu: kohoutek

Jak rozlišovat významy (sense)?

Praktické problémy: **granularita** a **užívání**

Jakou granularitu vlastně po aplikacích *požadujeme*?

Požadujeme, aby aplikace „znaly“ všechny významy, nebo jen ty, které se běžně užívají?

└ Je víceznačnost problém?

└ Jak rozlišovat významy (sense)?

Pastikův psůk my: *go sledovat a sledovat*

Je to a go sledovat sledovat po aplikaci go sledovat?

Pozdávame, a by a dlece „sady“ všech významy, nebo jen ty
které se sledovat sledovat?

Nabízí se možnost úplně vynechat významy, které v korpusu nenajdeme. Dostáváme se tady ale k věčnému problému korpusové lingvistiky – jak velkou část jazyka korpusy pokrývají? Neriskovali bychom vynecháním nepoužívaného významu situaci, že něco důležitého přehlédneme?

Granularita významu (sense): kočka

- 1. malá kočkovitá šelma, chovaná v domácnostech
- 2. malá n. středně velká šelma s hustým kožichem; zool. rod Felis
- 3. samice kočkovité šelmy vůbec
- 4. ob. kožišina na límci, kolem krku n. ramen
- 5. kocovina (Haš.)
- 6. věc připomínající někt. vlastnost kočky
- 7. druh důtek

Granularita významu (sense): kočka

- 2. malá n. středně velká šelma s hustým kožichem; zool. rod Felis
 - 1. malá kočkovitá šelma, chovaná v domácnostech
- 3. samice kočkovité šelmy vůbec
- 4. ob. kožišina na límci, kolem krku n. ramen
- 5. kocovina (Haš.)
- 6. věc připomínající někt. vlastnost kočky
- 7. druh dūtek

Granularita významu (sense): kočka

- 2. malá n. středně velká šelma s hustým kožichem; zool. rod Felis
 - 1. malá kočkovitá šelma, chovaná v domácnostech
 - 3. samice kočkovité šelmy vůbec
- 4. ob. kožišina na límci, kolem krku n. ramen
- 5. kocovina (Haš.)
- 6. věc připomínající někt. vlastnost kočky
- 7. druh dětek

Granularita významu (sense): kočka

- 2. malá n. středně velká šelma s hustým kožichem; zool. rod Felis
 - 1. malá kočkovitá šelma, chovaná v domácnostech
 - 3. samice kočkovité šelmy vůbec
- 4. ob. kožišina na límci, kolem krku n. ramen
- 6. věc připomínající někt. vlastnost kočky

- 5. kocovina (Haš.)

- 7. druh důtek

└ Je víceznačnost problém?

└ Granularita významu (sense): kočka

- 1. mačká se, utíká si volně kolem a křaví kočkovitě; zool. rod Felis
 - 2. mačká kočku nebo třeba, chozová v domácnostech
 - 3. samice kočky dle svého vůdce
- 4. ob. koččím se (mač, kolem třeba s. u mač)
- 5. včt. přejímající dítě, dle matky kočky
- 5. kosačička (Háč.)
- 7. dřev. dřevák

Zdá se, že významy netvoří seznam (jak jsme zvyklí ze slovníků), ale hierarchii. Návrh a tvorba hierarchického slovníku je zajímavé téma na závěrečnou práci.

Granularita: hierarchie významů





podle syntaktických kritérií:

- Lord *zanechal* v závěti všechn svůj majetek místnímu sirotčinci.
- Student *zanechal* studia podáním písemné žádosti.

Granularita: hierarchie významů

podle sémantických kritérií:

- abstraktní × konkrétní
- životný × neživotný
- člověk × zvíře
- emoce
- doména

-  Erk, K., McCarthy, D., and Gaylord, N. (2009).
Investigations on Word Senses and Word Usages.
In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 10–18, Suntec, Singapore. Association for Computational Linguistics.
-  Goddard, C. (2011).
Semantic Analysis: A Practical Introduction.
Oxford Textbooks in Linguistics. Oxford University Press.
-  Langacker, R. W. (1987).
Foundations of cognitive grammar: Theoretical Prerequisites.
Stanford University Press, Stanford, CA.
Vol 1, 1987(Hardcover), 1999(Paperback).
-  Wittgenstein, L. (1953).
Philosophical Investigations.
Basil Blackwell, Oxford.

