

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

30. března 2012

Formalismy pro reprezentaci znalostí

Sémantické rysy

Výběrová omezení

Sémantické role

Sémantické role a slovesa

Sémantické rámce

Formalismy pro reprezentaci znalostí

WSD je zajímavá a „klasická“ disciplína, řada vědců ale WSD odmítá kvůli uvedeným slabinám.

Navíc, WSD je kvantitativní analýza, neříká nic o významu.

Při studiu významu se chceme posunout dál, skutečně k jádru věci. Chceme pracovat nejen se slovy, ale i se znalostmi: jazykovými i obecnými. . .

└ Formalismy pro reprezentaci znalostí

└ Formalismy pro reprezentaci znalostí

WSD je zejména „deskriptivní“ disciplína, která věnuje sk WSD
odmítavě i velmi složitým otázkám.

Navíc, WSD je kvantitativní analýza, měří se v objemu.

Přítalio výje me u chceme pummatál, statečtí k jít a věci
C ha me pucovat nejen se dovo, ak i se znalostí: jazykovými i
obecnými...

Může to vypadat, že dost bylo matematiky. Na chvíli si od ní odpočineme a osvěžíme termíny z lingvistiky.

Časem se k matematice zase vrátíme...

Sémantické rysy (semantic features)

matka = FEMALE + ADULT + HAS CHILD

batole = HUMAN – ADULT

- sémantické rysy jsou „atomy“ významu
- sémantické rysy jsou distinktivní rysy
- význam je definován pomocí s. rysů a pravdivostních podmínek

[Croft and Cruse, 2004]

Sémantické rysy: synonymie

výraz	<i>jít</i>	<i>procházet</i>
rysy	MOTION ON FOOT SELF-PROPELLED MEDIUM VELOCITY	MOTION ON FOOT SELF-PROPELLED MEDIUM VELOCITY

Sémantické rysy: antonymie

výraz	<i>jít</i>	<i>běžet</i>
rysy	MOTION ON FOOT SELF-PROPELLED MEDIUM VELOCITY	MOTION ON FOOT SELF-PROPELLED HIGH VELOCITY

Sémantické rysy: problém s antonymií

výraz	<i>jít</i>	<i>letět střemhlav</i>
rysy	MOTION ON FOOT SELF-PROPELLED MEDIUM VELOCITY	MOTION ON WINGS GRAVITY-PROPELLED HIGH VELOCITY

Sémantické rysy: problém s antonymií

výraz	<i>jít</i>	<i>loudat se</i>
rysy	MOTION	MOTION
	ON FOOT	ON FOOT
	SELF-PROPELLED	SELF-PROPELLED
	MEDIUM VELOCITY	LOW VELOCITY

PLIN021 Sémantická analýza v praxi

└ Sémantické rysy

└ Sémantické rysy: problém s antonymi

výraz	ry	ry
	MOTION	MOTION
	ON FOOT	ON FOOT
	SELF-PROPELLED	SELF-PROPELLED
	MEDIUM VELOCITY	LOW VELOCITY

Sémantické rysy (jako všechny ostatní teorie) mohou být užitečné. Při jejich formálním uchopení narazíme na problémy: kolik rysů má každý výraz? Jak poznáme, které jsou podstatné a které ne? Trošku to připomíná problém, jak najít vhodnou otázku ve hře Myslím si zvíře. . .

Výběrová omezení (selectional restrictions)

slouží pro desambiguaci závislosti větných členů [Allen, 1995, kap. 10.1]

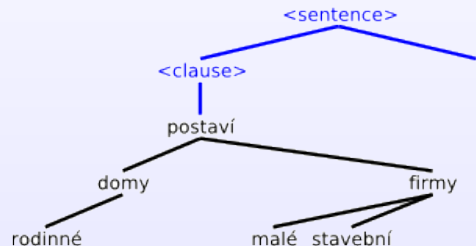
- Koupila jsem si pletenou čepici a šálu.
- Koupila jsem si nealkoholické pivo a křupky.

pletená šála – OK, nealkoholické křupky – NOT OK

Výběrová omezení (selectional restrictions)

slouží pro desambiguaci závislosti větných členů [Allen, 1995, kap. 10.1]

Rodinné domy postaví malé stavební firmy.



AGENT = rodinné
domy
THEME = malé
stavební firmy

AGENT = malé
stavební firmy
THEME = rodinné
domy

(AGENT postavit PERSON | INSTITUTION)
(THEME postavit BUILDING)

PLIN021 Sémantická analýza v praxi

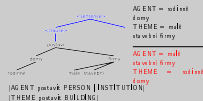
└ Výběrová omezení

└ Výběrová omezení (selectional restrictions)

Výběrová omezení (selectional restrictions)

Abstr. pro časopisový příspěvek o výskytu české [PLIN, 2015, kv. 11.]

Rodinné domy postavil malá stavební firma.



Obrázek je ze syntaktického analyzátoru SET

Sémantické role (semantic role, thematic relation, theta-role, deep case)

(AGENT postavit PERSON | INSTITUTION)
(THEME postavit BUILDING)

AGENT, EXPERIENCER, THEME, PATIENT, INSTRUMENT,
FORCE/NATURAL CAUSE, LOCATION, DIRECTION/GOAL,
RECIPIENT, SOURCE/ORIGIN, TIME, BENEFICIARY,
MANNER, PURPOSE, CAUSE

Jména rolí jsou u různých autorů mírně různá (např. AGENT/ACTOR). Přiřazení rolí je někdy zřejmé (Karel/AGENT rozbil okno), jindy zřejmé není (Kladivo/AGENT?INSTRUMENT? rozbilo okno).

[Fillmore, 1968]

└ Sémantické role

└ Sémantické role (semantic role, thematic relation, theta-role, deep case)

Sémantické role (semantic role, thematic relation, theta-role, deep case)

[AGENT postavě PERSON] [INSTITUTION]
[THEME postavě BUILDING]

AGENT, EXPERIENCER, THEME, PATIENT, INSTRUMENT,
FORCE/NATURAL CAUSE, LOCATION, DIRECTION/GOAL,
RECIPIENT, SOURCE/OBJECT, TIME, BENEFICIARY,
MANNER, PURPOSE, CAUSE

Jedna věta jazyka má vždy a totokórně několik smysl.
AGENT/ACTOR. Předmět věty má vždy nějaký smysl [AGENT
smysl věty], který závisí na smyslu [AGENT/INSTRUMENT
smysl věty].

[Rilmore, 1988]

dobry přehled sémantických rolí je kupodivu na Wikipedii

Sémantické role (semantic role, thematic relation, theta-role, deep case)

sémantické role [Jackendoff, 1992, kap. 2.2]

- jsou nositelé širšího významu
- jsou součástí sémantické struktury promluvy (conceptual structure), ne syntaxe
- typicky však odpovídají větným členům
- objevují se v sémantických rámcích (o těch později)

└ Sémantické role

└ Sémantické role (semantic role, thematic relation, theta-role, deep case)

sémantické role [Jackendoff, 1987, kap. 2.2]

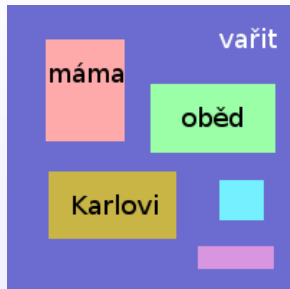
- jsou součástí většího výrazu
- jsou součástí sémantické struktury podobny (come from structure), se systémem
- typicky vázá na pověsti věty (deep case)
- odlišují se v sémantické kódu (kódu) (je třeba poznamenat)

Terminologie je tady dost zmatená. Je to způsobeno faktem, že syntaktické struktury jsou často těsně spojeny s těmi sémantickými, typicky je např. podmět ACTOR/AGENT. Př.: Karel (AGENT) rozbil okno (THEME/BENEFICIARY). Existují ale i protipříklady (různá pasiva): Okno se rozbilo. Okno bylo rozbito. Někteří lingvisté striktně rozlišují mezi syntaxí a sémantikou, tím pádem nemůžou slučovat theta role s tématickými relacemi.

Fillmorův Case for case se samozřejmě týká angličtiny. V češtině je situace zase trochu jiná: 1. kromě předložek máme i pády, 2. vzhledem k celkem volnému slovosledu uplatňujeme aktuální členění větné.

Sémantické role ve slovesných valencích

sloveso je centrem věty – na slovese „visí“
celý zbylý význam
sloveso si lze představit jako rámec, do
kterého zapadnou ostatní dílky



slovník slovesných valencí VerbaLex [Hlaváčková, 2007]

└ Sémantické role a slovesa

└ Sémantické role ve slovesných valencích

slava je se se m m o by - se slovesa, jdiť
 celý zbylý výraz m
 sloveso si lex přidává je in ú me, to
 lexi ho zapá toa osaz n íky



sloves k slovesných valencí Verba Lex [Hlaváčková, 2007]

Ukázka. Potom můžeme opět zkusit vzít libovolný skutečný text (třeba titulku z iDnes) a najít patřičný rámec. Uvidíme, že je to těžké. Když najdeme rámec, není zaplněný (je velmi těžké najít zaplněný). Co zbyde z věty kromě zaplněného rámce?

Ve VerbaLexu nejsou většinou uvedena příslovečná určení. Co ale ještě zbyde? Zbydou další větné členy, které zpravidla slouží jako prostředky koherence textu. Které to jsou? Co znamenají? Proč se jimi skoro nikdo nezabývá?

Námět na referát/BP (zabývá se jimi prof. Joshi

<http://www.cis.upenn.edu/~joshi/>)

Rámce – stereotypická informace [Minsky, 1974]

objekty, vlastnosti, vztahy mezi objekty, odvozovací pravidla

pro každý objekt jsou v rámci rubriky *sloty (slots)*, každá rubrika má položky (*links, facets*) jako např. aktuální hodnotu, implicitní hodnotu, rozsah možných hodnot ...

kočka domácí:

- je __ druhem zvíře
- má __ nepřítele pes
- má __ za __ potravu myš
- nachází __ se (blízko __ domu, doma)
- velikost malé __ zvíře

Rámce – skutečné projekty





FrameNet, ConceptNet

slovesa – VerbNet, VerbaLex

Ukázky FrameNetu Hrají v popisu rámců roli místní realie? Lze nějak hromadně určit případy, ve kterých ano? Další námět na referát nebo BP. Zásadní rozdíl mezi VerbaLexem a FrameNetem je v propojení rámců: ve F. jsou zachyceny vztahy mezi slovesnými rámci (např. příčina–následek). Ve V. je zase více informací lingvistických (gramatické kategorie, slovesné třídy). Má cenu zmiňovat Levinovou a slovesné třídy?

VerbNet a VerbaLex také neobsahují implicitní hodnoty (proč? není to námět na BP?), rozsah hodnot apod.

FrameNet a ConceptNet jsou spíš hnány potřebou aplikace. VerbNet a VerbaLex jsou dílem lingvistů.

-  Allen, J. (1995).
Natural Language Understanding (2nd ed.).
Benjamin-Cummings Publishing Co., Inc., Redwood City, CA,
USA.
-  Croft, W. and Cruse, D. (2004).
Cognitive linguistics.
Cambridge textbooks in linguistics. Cambridge University Press.
-  Fillmore, C. (1968).
The case for case.
In Bach, E. and Harms, R., editors, *Universals in Linguistic
Theory*. Holt, Rinehart, and Winston, New York.
-  Hlaváčková, D. (2007).
Databáze slovesných valenčních rámců VerbaLex.
PhD thesis, Masarykova univerzita, Filozofická fakulta, Ústav
českého jazyka.



Jackendoff, R. (1992).

Semantic Structures.

Current Studies in Linguistics. MIT Press.



Minsky, M. (1974).

A framework for representing knowledge.

Technical report, Massachusetts Institute of Technology,

Cambridge, MA, USA.