

PA153 Počítačové zpracování přirozeného jazyka

05 – Sémantika II

(logická reprezentace, od věty k diskurzu)

Karel Pala, Zuzana Nevěřilová

Centrum ZPJ, FI MU, Brno

14. října 2013

- 1 Lexikální význam a kontext
- 2 Kontext
- 3 Větná sémantika, logická sémantika
- 4 Sloveso jako predikát

Lexikální význam a kontext

lexikální význam: izolovaný význam slov

plnovýznamové (autosémantické) lexikální jednotky (LU):

- substantiva
- adjektiva
- verba
- adverbia

černý, moci, nechat, být, tak, vlastně

pomocné (syntaktické) LU: z, proč, jak, ten

ostatní: kykyryký

└ Lexikální význam a kontext

└ Lexikální význam a kontext

lexikální význam: izolovaný význam slov
přívýznamové (autosemantické) lexikální jednotky (LU):

- substantiva
- adjektiva
- verba
- adverbia

černý, moci, řešat, být, tak, vlastně
pomocné (symantické) LU: z, prož, jak, ten
ostatní: kykyrykj

Některé LU mají izolovaný význam, uvedené LU jsou do jisté míry protipříklady. Např. „černý“ znamená „neplatící“, ale jen v kolokaci „černý pasažér“. Pomocná a způsobová (modální) slovesa mají význam oslabený. U některých LU má smysl mluvit o významu skutečně jen v kontextu.

Lexikální význam a kontext

Krakutel z jejich mrusy se ploc blutkal, načež potom tražil také všechny své stěvače.

Vyšetřovatel jopuz hrych vlády tre moc naštval, bruvěž slekym rozzuřil vičké kruký moré posluchače.

Vyšetřovatel z jejich vlády se moc naštval, načež potom rozzuřil také všechny své posluchače.

Krakatel z jejich mravy se ploc blurkal, nalez potom trafil také vřechy své stěvače.

Vyšetřovatel jopuz hrych vlády tre moc našelval, bruvřil sládkym rozzuřil vičké kruky moe posluchače.

Vyšetřovatel z jejich vlády se moc našelval, nalez potom rozzuřil také vřechy své posluchače.

U těchto vět zkusíme určit slovní druhy a větné členy. U té první věty by to mělo být snazší než u té druhé. V první větě jsou plnovýznamová slova nahrazena nesmyslnými slovy, ale koncovka je zachována. V druhé větě jsou nahrazena všechna neplnovýznamová slova, konec slova je opět zachován.

Jde o ukázkou důležitosti funkčních slov pro pochopení významu věty. Význam těchto slov si uvědomíme právě jen v kontextu věty.

Kontext

- **verbální kontext (co bylo řečeno, co bude následovat)**
- situační kontext (místo, čas, počet komunikačních partnerů, jejich vzájemný vztah, presupozice mluvčího)
komunikační situace: $KS = s, a, o_1, \dots, o_n, p, t$,
kde s – mluvčí, a – adresát, o_1, \dots, o_n – promluvové objekty, p – místo komunikace, t – čas komunikace
presupozice mluvčího: společná báze znalostí (to, co není třeba zmiňovat)
- sociální kontext (vzdělání, zkušenost, životní podmínky, status sociální skupiny)

Ale sakra himl hergot, já nikdy neřekl, že na mě nikdo nemá

Kontext

Jak zkoumat verbální kontext:

- v době „předkorpusové“
- konkordance v korpusu (středně velké korpusy)
- word sketch (velké korpusy)

kandidát

czTenTen12 [Majka] frekvence = 213578 (39.3 v milionu)

<u>a_modifier</u>	<u>90268</u>	<u>-1.4</u>	<u>post_na</u>	<u>45490</u>	<u>-7.3</u>	<u>gen_2</u>	<u>22232</u>	<u>-0.9</u>
prezidentský	<u>5105</u>	10.02	post	<u>2878</u>	8.8	hodnost	<u>258</u>	7.01
žhavý	<u>2404</u>	9.23	primátor	<u>1617</u>	8.14	zvolení	<u>161</u>	6.62
závislý	<u>4647</u>	8.74	prezident	<u>3946</u>	7.52	pětice	<u>139</u>	6.49
vhodný	<u>9792</u>	8.43	senátor	<u>701</u>	7.35	nominace	<u>272</u>	6.31
republikánský	<u>1055</u>	8.34	eurokomisara	<u>216</u>	7.27	slyšení	<u>109</u>	6.28
navržený	<u>1516</u>	8.28	pozice	<u>4181</u>	7.2	výběr	<u>1822</u>	6.11
horký	<u>2315</u>	8.19	děkan	<u>382</u>	7.15	představování	<u>62</u>	6.11
opoziční	<u>745</u>	7.42	rektor	<u>321</u>	7.03	navrhování	<u>91</u>	6.08

Kontext a word sketch

Jak vypočítat *word sketch*?

word sketch grammar

multi word sketch



význam věty: význam slov + syntaktické vztahy mezi větnými složkami
(princip kompozicionality)

Compositionality Principle:

- The meaning of the whole is a function of the meaning of the parts and the mode of combining them.
- The meaning of a complex expression is uniquely determined by the meaning of its constituents and the syntactic construction used to combine them.

Logická sémantika

- redukuje lexikální význam LU na logický typ (individuum, čas, ...)
- zůstává predikátová struktura u sloves, příp. deverbativ
- propozice mají pravdivostní hodnotu

propozice¹: informační obsah věty vyjádřený větou pronesenou v určitém kontextu

více různých vět může vyjadřovat tutéž propozici:

The Earth is round. Země je kulatá.

¹<http://plato.stanford.edu/entries/propositions-structured/>

Logická sémantika a predikátová logika

Predikátová logika 1. řádu (First Order Predicate Logic) [Mendelson, 1997]

- termy: proměnné (x), funkce ($f(x)$)
- predikátové symboly ($P(x)$)
- logické spojky ($\vee, \wedge, \neg, \Rightarrow, \Leftrightarrow$)
- kvantifikátory (\forall, \exists)
- symbol rovnosti ($=$)
- mimologické symboly (aritmetické symboly, řetězce atd.)

Logická sémantika a predikátová logika

Kdo nebude znát malou násobilku, dostane hodnocení F.

$\forall x : \neg Z(x, \text{"malá násobilka"}) \Rightarrow H(x, \text{"F"})$

$(Z(x, y) - x \text{ zná } y, H(x, y) - x \text{ dostal hodnocení } y)$

Petr nezná malou násobilku.

$\neg Z(\text{"Petr"}, \text{"malá násobilka"})$

Úsudek:

$\neg Z(\text{"Petr"}, \text{"malá násobilka"}) \Rightarrow H(\text{"Petr"}, \text{"F"})$

je pravdivý

Logická sémantika a predikátová logika

Vyzkoušejte:

Komu se nelení, tomu se zelení.

$$\forall x : N(x) \Rightarrow Z(x)$$

Honza se odrazil od podlahy a vyskočil do dvou metrů.

$O(\text{"Honza"}, \text{"podlaha"})$

$V(\text{"Honza"}, \text{"2 metry"})$

Tahle vláda není ani ryba ani rak.

$\neg(\text{"tahle vláda"} = \text{"ryba"}) \wedge \neg(\text{"tahle vláda"} = \text{"rak"})$

Limity predikátové logiky

- ne všechny konstrukce v přirozeném jazyce jsou propozice
 - ▶ Dobrý den. Děkuji vám.
 - ▶ Kdybych tak měl milión ...
 - ▶ Nedá se nic dělat.
- ne všechny propozice jsou 1. řádu
Všichni lidé mají společné vlastnosti.
 $\exists V \forall x : V(x)$
- v přirozeném jazyce je mnohem víc kvantifikátorů
 - ▶ většina
 - ▶ velká část
 - ▶ kdekdo
 - ▶ pár lidí, skoro nikdo
 - ▶ pokud vůbec někdo

Predikátová logika, typy argumentů

$Z(x, y)$ – x žije v y

Honza žil v Brně.

$Z(\text{"Honza"}, \text{"Brno"})$

Honza žil ještě v sobotu.

$Z(\text{"Honza"}, \text{"sobota"})$

$Z(x, y)$ – x žije v y a y je místo

typované logiky

└ Větná sémantika, logická sémantika

└ Predikátová logika, typy argumentů

$Z(x, y)$ = x žije v y
Honzá žil v Brně.
 $Z("Honzá", "Brno")$
Honzá žil ještě v sobotu.
 $Z("Honzá", "sobota")$
 $Z(x, y)$ = x žije v y a y je místo
typované logiky

O tom bude mluvit Vojta K. v semináři č. 9.

Sloveso jako predikát

sloveso (příp. deverbativum) jako predikát, větné členy jako argumenty predikátu

valence

Sloveso jako predikát

Kluk rozbil okno.

Kámen vletěl do okna a rozbil ho.

Okno se rozbilo.

rozbít: AG(osoba) ART(výrobek) INS(nástroj)

kluk osoba

kámen nástroj

okno výrobek

Valenční slovníky

české:

- Vallex: In VALLEX 2.x, there are roughly 2,730 lexeme entries containing together around 6,460 lexical units (“senses”).²
- VerbaLex:³
 - ▶ 21032 literálů (sloveso + význam)
 - ▶ 10469 slovesných lemmat

anglické:

- VerbNet: 8537 total verbs represented⁴

²<http://ufal.mff.cuni.cz/vallex/2.6/doc/home.html>

³<http://nlp.fi.muni.cz/cs/VerbaLex>

⁴<http://verbs.colorado.edu/verb-index/index.php>

Valenční slovníky: Vallex

VALLEX 2.6

alphabet

class

functors

forms

aspect

control

reflex.

recipr.

complexity

VALEVAL

- F (10)
- G**
- H (51)
- CH (22)
- I (17)
- J (13)
- K (73)
- L (37)
- M (53)
- N (133)

- zlobit, zlobivat
- zlobit se, zlobívat se
- zlomit se, zlámat se
- zmáčknout, zmačkat
- zmáčknout se, zmačkat se
- zmáhat, zmocí/zmoci
- zmáhat se, zmoci se/zmoci se
- zmapovat

zmařit^{pf}

1 ≈ **zkazit; zničit**

-frame: **ACT**₁^{obl} **PAT**₄^{obl} **BEN**₃^{typ} **MEANS**₇^{typ}

-example: zmařil celé jednání svou nezodpovědností; zmařil mu život

-rfi: pass: jeho podvatrné plány se naštěstí dopředu zmařily

Valenční slovníky: VerbaLex

Verb classes

- admit-64 (65)
- adopt-91 (4)
- allow-63 (69)
- animal_sounds-38 (60)
- approve-75 (91)
- assessment-34 (50)
- **avoid-52 (51)**
- banish-10.2 (55)
- battle-36.3 (8)
- bodyinternalmotion-49 (131)
- build-26.1-4 (7)

Verb class

"destroy-44"

- babrat₁
- bořit₁
- bořit₂
- bourat₁
- bourat₄
- brakovat₃
- brát₂₇
- břídit₁
- demolovat₁

zmařit₃^{pf} **zničit**₅^{pf} **rozbít**₃^{pf}

mařit₃^{impf} **ničit**₅^{impf} **rozbíjet**₃^{impf}

1 zmařit₃, mařit₃ ≈

2 rozbít₃, rozbíjet₃, zničit₅, ničit₅ ≈

-frame: **GROUP** <institution:1>_{i1} **VERB** ^{obl} **GROUP** <institution:1>_{i4} ^{obl}

-example: *policie rozbila zločinecký gang (pf)*

Valenční slovníky: VerbNet

ROLES

- AGENT [+INT_CONTROL]
- PATIENT [+CONCRETE]
- INSTRUMENT [+CONCRETE]

FRAMES

NP V NP

EXAMPLE "The Romans destroyed the city."

SYNTAX AGENT V PATIENT

SEMANTICS CAUSE(AGENT, E) DESTROYED(RERESULT(E), PATIENT)

NP V NP PP.INSTRUMENT

EXAMPLE "The builders destroyed the warehouse with explosives."

SYNTAX AGENT V PATIENT {WITH} INSTRUMENT

SEMANTICS CAUSE(AGENT, E) USE(DURING(E), AGENT, INSTRUMENT) DESTROYED(RERESULT(E), PATIENT)

NP.INSTRUMENT V NP

EXAMPLE "The explosives destroyed the warehouse."

SYNTAX INSTRUMENT V PATIENT

SEMANTICS CAUSE(?AGENT, E) USE(DURING(E), ?AGENT, INSTRUMENT) DESTROYED(RERESULT(E), PATIENT)

Zachycení významu ve valenčním rámci

Verb classes	Verb class "destroy-44"	
<ul style="list-style-type: none">• admit-64 (65)• adopt-91 (4)• allow-63 (69)• animal_sounds-38 (60)• approve-75 (91)• assessment-34 (50)• avoid-52 (51)• banish-10.2 (55)• battle-36.3 (8)• bodyinternalmotion-49 (131)• build-26.1-4 (7)	<ul style="list-style-type: none">• babrat₁• bořit₁• bořit₂• bourat₁• bourat₄• brakovat₃• brát₂₇• břídit₁• demolovat₁	<p>zmařit^{pf}₃ zničit^{pf}₅ rozbít^{pf}₃</p> <p>mařit^{impf}₃ ničit^{impf}₅ rozbíjet^{impf}₃</p> <p>1 zmařit₃, mařit₃ ≈</p> <p>2 rozbít₃, rozbíjet₃, zničit₅, ničit₅ ≈</p> <p>-frame: GROUP <institution:1>_{i1} VERB ^{obl} GROUP <institution:1>_{i4}</p> <p>-example: policie <i>rozbila</i> zločinecký gang (pf)</p>

- sloveso a jeho synonyma (synset)
- překlad
- slovesná třída (slovesa komunikace, slovesa ničení ... [Wu and Palmer, 1994])
- slovesné rámce (jednotlivé případy užití slovesného synsetu)
- počet argumentů (slotů)
- syntaktické informace (větné členy: pořadí, pád, příp. předložka)
- sémantické role (role)
- výběrová omezení (typický reprezentant)

Propojení valenčních rámců: od slovníků k sémantické síti

popisy typických situací: skripty (scénáře), rámce: nakupovat, měřit, dolovat, zemřít

- návrhy od 70. let: Schank, Abelson, Minsky
- aplikace (datové zdroje): od 90. let FrameNet ⁵

⁵<https://framenet.icsi.berkeley.edu/fndrupal/>

Propojení valenčních rámců: FrameNet

Definition:

This transparent noun frame is concerned with **Units** for measuring the **Area** of regions.
Hillary lives in the middle of **10 ACRES** of wheat.

Paul owns **20 HECTARES** of land.

Semantic Type: Transparent Noun

FEs:

Core:

Area □

Excludes: Occupant

The region whose surface is being measured.

Klaas has a 10 **ACRE** peach **orchard**

Count □

The number of **Units**.

Smiley owns **fifteen** **HECTARES** of prime real estate.

Odkazy I



Mendelson, E. (1997).

Introduction to Mathematical Logic.

Discrete Mathematics and Its Applications Series. Chapman & Hall.



Wu, Z. and Palmer, M. (1994).

Verbs semantics and lexical selection.

In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, ACL '94, pages 133–138, Stroudsburg, PA, USA. Association for Computational Linguistics.