

# PA153 Počítačové zpracování přirozeného jazyka

## 05 – Sémantika II

(logická reprezentace, od věty k diskurzu)

Karel Pala, Zuzana Nevěřilová

Centrum ZPJ, FI MU, Brno

15. listopadu 2020

- 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

- substantiva
- adjektiva
- verba
- adverbia

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.

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.

- **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)

# Kontext

Jak zkoumat verbální kontext:

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

modifiers of "kandidát"	verbs with "kandidát" as subject	"kandidát" and/or ...
<b>prezidentský</b> ... prezidentského kandidáta	<b>kandidovat</b> ... kandidát kandiduje	<b>kandidátka</b> ... kandidátů a kandidátek
<b>perfektní</b> ... perfektní kandidát	<b>uspět</b> ... kandidát uspěje	<b>kandidátek</b> ... kandidáty a kandidátky na
<b>republikánský</b> ... republikánský kandidát	<b>postoupit</b> ... kandidáti postoupí do	<b>lídr</b> ... lídrem a kandidátem
<b>vhodný</b> ... vhodného kandidáta na pozici	<b>ucházet</b> ... se kandidát uchází	<b>koncipient</b> ... koncipienti a kandidáti
<b>žhavý</b> ... žhavým kandidátem	<b>jevit</b> ... kandidát se jeví	<b>kandidát</b> ... kandidáta či kandidátů



# Kontext a word sketch

Jak vypočítat *word sketch*?

word sketch grammar

multi word sketch

# Kontext a word embeddings

Reprezentace slova pomocí n-rozměrného vektoru<sup>1</sup>.

Vektor se vypočítá na základě kontextu:

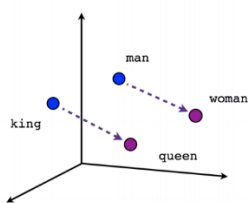
- continuous bag-of-words (CBOW):

výpočet pravděpodobnosti dalšího slova  $w$  za předpokladu kontextu  $c$

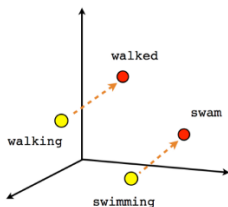
$P(w|c) = 0.9$ , kde  $w = \text{můj}$ ,  $c = \{\text{Kde, domov}\}$

- skip-gram: pravděpodobnost kontextu  $c$  za předpokladu slova  $w$

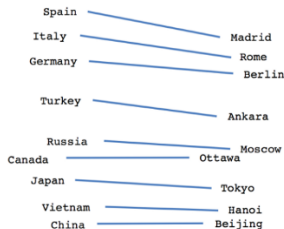
word2vec, GloVe, FastText, NumberBatch, ...



Male-Female



Verb tense



Country-Capital

<sup>1</sup>[https://towardsdatascience.com/](https://towardsdatascience.com/creating-word-embeddings-coding-the-word2vec-algorithm-in-python-using-dee)

## Kontext a word embeddings

Kontextové word embeddings (ELMo, BERT, CoVe, ...):

I **left** my phone on the **left** side of the table.



I left my phone on the table.

I left when he was cleaning the windows.

Turn left than continue straight on.



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

propozice<sup>2</sup>: informační obsah věty vyjádřený větou pronesenou v určitém kontextu

více různých vět může vyjadřovat tutéž propozici:

The Earth is round. Země je kulatá.

---

<sup>2</sup><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:

forall, exists, not,  $\rightarrow$ , and, or

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  
Lhát se nemá.  
 $N(L)$
- v přirozeném jazyce je mnohem víc kvantifikátorů
  - ▶ většina
  - ▶ velká část
  - ▶ kdekdo
  - ▶ pár lidí, skoro nikdo
  - ▶ pokud vůbec někdo

řešení: jiné logiky, rozšířená FOPL, intenzionální logiky (TIL)

# Predikátová logika, typy argumentů

$Z(x, y)$  –  $x$  žije v  $y$

Honza žil ještě 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

# Sloveso jako predikát

sloveso (příp. deverbativum) jako predikát, ostatní 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”).<sup>3</sup>
- VerbaLex:<sup>4</sup>
  - ▶ 21032 literálů (sloveso + význam)
  - ▶ 10469 slovesných lemmat

anglické:

- VerbNet: 8537 total verbs represented<sup>5</sup>

---

<sup>3</sup><http://ufal.mff.cuni.cz/vallex/2.6/doc/home.html>

<sup>4</sup><http://nlp.fi.muni.cz/cs/VerbaLex>

<sup>5</sup><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
- zmačknout, zmačkat
- zmačknout se, zmačkat se
- zmahat, zmoci/zmoci
- zmahat se, zmoci se/zmoci se
- zmapovat

**zmařit**<sup>pf</sup>

**1** ≈ **zkazit; zničit**

-frame: **ACT**<sub>1</sub><sup>obl</sup> **PAT**<sub>4</sub><sup>obl</sup> **BEN**<sub>3</sub><sup>typ</sup> **MEANS**<sub>7</sub><sup>typ</sup>

-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"

- babrat<sub>1</sub>
- bořit<sub>1</sub>
- bořit<sub>2</sub>
- bourat<sub>1</sub>
- bourat<sub>4</sub>
- brakovat<sub>3</sub>
- brát<sub>27</sub>
- břídit<sub>1</sub>
- demolovat<sub>1</sub>

**zmařit**<sub>3</sub><sup>pf</sup> **zničit**<sub>5</sub><sup>pf</sup> **rozbít**<sub>3</sub><sup>pf</sup>

**mařit**<sub>3</sub><sup>impf</sup> **ničit**<sub>5</sub><sup>impf</sup> **rozbíjet**<sub>3</sub><sup>impf</sup>

1 zmařit<sub>3</sub>, mařit<sub>3</sub> ≈

2 rozbít<sub>3</sub>, rozbíjet<sub>3</sub>, zničit<sub>5</sub>, ničit<sub>5</sub> ≈

-frame: **GROUP** <institution:1><sub>i1</sub> **VERB** <sup>obl</sup> **GROUP** <institution:1><sub>i4</sub> <sup>obl</sup>

-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"><li>• admit-64 (65)</li><li>• adopt-91 (4)</li><li>• allow-63 (69)</li><li>• animal_sounds-38 (60)</li><li>• approve-75 (91)</li><li>• assessment-34 (50)</li><li>• <b>avoid-52 (51)</b></li><li>• banish-10.2 (55)</li><li>• battle-36.3 (8)</li><li>• bodyinternalmotion-49 (131)</li><li>• build-26.1-4 (7)</li></ul>	<ul style="list-style-type: none"><li>• babrat<sub>1</sub></li><li>• bořit<sub>1</sub></li><li>• bořit<sub>2</sub></li><li>• bourat<sub>1</sub></li><li>• bourat<sub>4</sub></li><li>• brakovat<sub>3</sub></li><li>• brát<sub>27</sub></li><li>• břídit<sub>1</sub></li><li>• demolovat<sub>1</sub></li></ul>	<p><b>zmařit</b><sub>3</sub><sup>pf</sup> <b>zničit</b><sub>5</sub><sup>pf</sup> <b>rozbít</b><sub>3</sub><sup>pf</sup></p> <p><b>mařit</b><sub>3</sub><sup>impf</sup> <b>ničit</b><sub>5</sub><sup>impf</sup> <b>rozbíjet</b><sub>3</sub><sup>impf</sup></p> <p>1 zmařit<sub>3</sub>, mařit<sub>3</sub> ≈</p> <p>2 rozbít<sub>3</sub>, rozbíjet<sub>3</sub>, zničit<sub>5</sub>, ničit<sub>5</sub> ≈</p> <p>-frame: <b>GROUP</b> &lt;institution:1&gt;<sub>i1</sub> <b>VERB</b> <sup>obl</sup> <b>GROUP</b> &lt;institution:1&gt;<sub>i4</sub> <sup>obl</sup></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 <sup>6</sup>

---

<sup>6</sup><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.