Semantics II: logical representation, from sentence to discourse

Karel Pala, Zuzana Nevěřilová, Vojtěch Kovář

Natural Language Processing Centre
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno

{pala, xpopelk, xkovar3}@fi.muni.cz

PA153 Natural Language Processing

Outline

Lexical meaning and context

- Lexical meaning (lexical semantics)
  - meaning of isolated words
- Autosemantic lexical units
  - they carry meaning
  - nouns, adjectives, verbs, adverbs
  - black, high, run, let, actually
  - (but e.g. “high school”)
- Synsemantic lexical units
  - have meaning only in combination with autosemantics
  - from, why, how, the

Lexical meaning and context

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

Context

- Verbal context
  - previous/following words, sentences, ...
- Situation context
  - place, time
  - number of communication partners and their relationships
  - knowledge of communication partners
  - *righarrow* communication situation
- Social context
  - education of the audience, status of the social group
  - "what the f*ck are you doing here?"

Verbal context: Corpora

- concordance lines (small corpora)
- word sketch (large corpora)
  - "sketch grammar" + statistics

kandidát

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

a_modifier	90268	-1.4
prezidentský	5105	10.02
žhavý	2404	9.23
závislý	4647	8.74
vhodný	9792	8.43
republikánský	1055	8.34
navržený	1516	8.28
horký	2315	8.19
opoziční	745	7.42

post_na	45490	-7.3
post	2878	8.8
primátor	1617	8.14
prezident	3946	7.52
senátor	701	7.35
eurokomisara	216	7.27
pozice	4181	7.2
děkan	382	7.15
rektor	321	7.03

gen_2	22232	-0.9
hodnost	258	7.01
zvolení	161	6.62
pětice	139	6.49
nominace	272	6.31
slyšení	109	6.28
výběr	1822	6.11
představování	62	6.11
navrhování	91	6.08

Sentence semantics

- Meaning of a sentence
  - meaning of the words + syntactic relationships between them
- 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.

Logical analysis of the sentence

- Sentence → logical formula
  - predicate structure for verbs
  - propositions have truth value (true, false)
- Propositions
  - information content of the sentence in a certain context
  - different sentences can lead to the same proposition
  - "The Earth is round."
  - "The Earth has a round shape."
  - "Země je kulatá."

First order predicate logic

- "Those who do not know statistics, will be classified F"
- $\forall x : \neg knows(x, "statistics") \Rightarrow classif(x, "F")$
- "Peter does not know statistics"
- $\neg knows("Peter", "statistics")$
- We can deduce
  - $classif("Petr", "F")$
  - "Peter is classified F."

First order predicate logic

Komu se nelení, tomu se zelení.

Honza se odrazil od podlahy a vyskočil do dvou metrů.

Tahle vláda není ani ryba ani rak.

First order predicate logic

Komu se nelení, tomu se zelení.
$\forall x : N(x) \Rightarrow Z(x)$

Honza se odrazil od podlahy a vyskočil do dvou metrů.

Tahle vláda není ani ryba ani rak.

First order predicate logic

Komu se nelení, tomu se zelení.
$\forall x : N(x) \Rightarrow Z(x)$

Honza se odrazil od podlahy a vyskočil do dvou metrů.
$O(\text{"}Honza\text{"}, \text{"}podlaha\text{"})$
$V(\text{"}Honza\text{"}, \text{"}2\ metry\text{"})$

Tahle vláda není ani ryba ani rak.

First order predicate logic

Komu se nelení, tomu se zelení.
$\forall x : N(x) \Rightarrow Z(x)$

Honza se odrazil od podlahy a vyskočil do dvou metrů.
$O(\text{"}Honza\text{"}, \text{"}podlaha\text{"})$
$V(\text{"}Honza\text{"}, \text{"}2\ metry\text{"})$

Tahle vláda není ani ryba ani rak.
$\neg(\text{"}tahle\ vláda\text{"}=\text{"}ryba\text{"}) \wedge \neg(\text{"}tahle\ vláda\text{"}=\text{"}rak\text{"})$

First order predicate logic: limits

- Some things are not propositions
  - Good morning. Thank you
  - If only I had money...
- Some propositions are not first order
  - All people have some common features.
  - $\exists F \forall x : F(x)$
- More quantifiers
  - somebody
  - big part of
  - many people

First order predicate logic: limits

- Types of arguments
  - Peter lived on the first floor.
  - Peter stil lived on Sunday
- $\rightarrow$ logic formalisms with types

Verb as a predicate

- Verb = predicate
  - other sentence parts = arguments
- → verb valencies
- Break: AG(person) ART(product) <by> INS(instrument)
  - A boy break a window by a stone.
  - The window was broken.
  - A stone broke the window.
  - A boy broke the window.

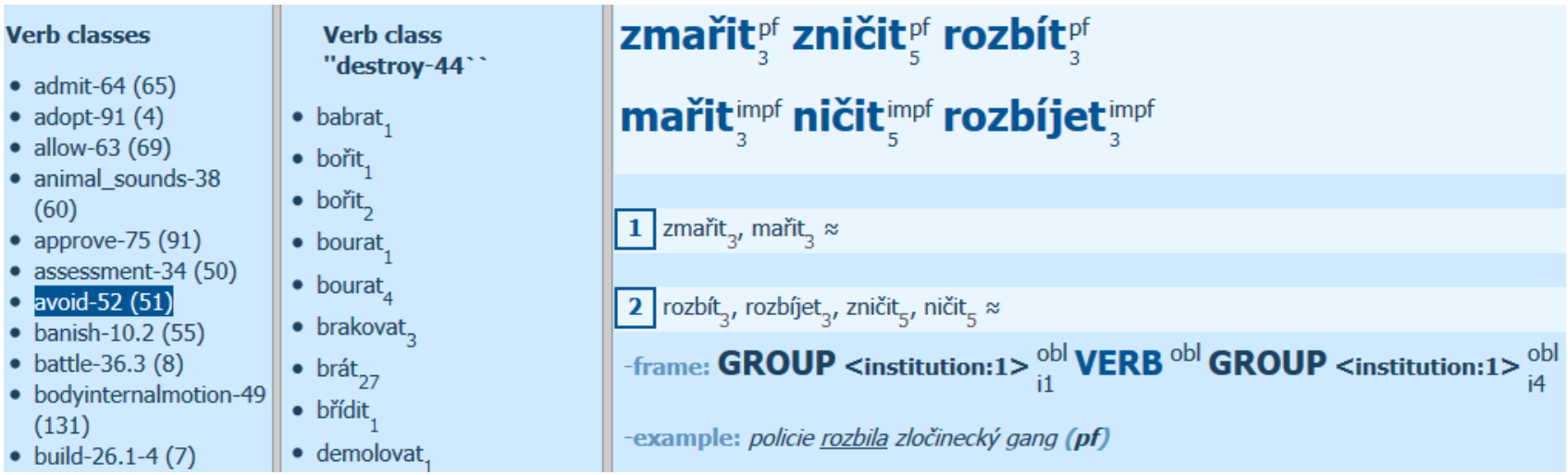
Verb valency lexicons

- VALLEX (Czech)
  - Institute of Formal and Applied Linguistics
  - 2,730 entries, 6,460 senses
- VerbaLex (Czech)
  - Natural Language Processing Centre at FI
  - 10,469 lemmas, 21,032 senses (literals)
- VerbNet (English)
  - University of Colorado (Martha Palmer)
  - 8,537 verbs

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, zlobívat
zlobit se, zlobívat se
zlomit se, zlámat se
zmáčknout, zmačkat
zmáčknout se, zmačkat se
zmáhat, zmoci/zmoct
zmáhat se, zmoci se/zmoct se
zmapovat
zmařit pf
1 ≈ zkazit; zničit
-frame: ACT obl 1 PAT obl 4 BEN typ 3 MEANS typ 7
-example: zmařil celé jednání svou nezodpovědností; zmařil mu život
-rfl: pass: jeho podvratné plány se naštěstí dopředu zmařily

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$_1$
- bořit$_1$
- bořit$_2$
- bourat$_1$
- bourat$_4$
- brakovat$_3$
- brát$_{27}$
- břídit$_1$
- demolovat$_1$

zmařit$^{pf}_{3}$ **zničit**$^{pf}_{5}$ **rozbít**$^{pf}_{3}$

mařit$^{impf}_{3}$ **ničit**$^{impf}_{5}$ **rozbíjet**$^{impf}_{3}$

1 zmařit$_3$, mařit$_3$ ≈

2 rozbít$_3$, rozbíjet$_3$, zničit$_5$, ničit$_5$ ≈

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

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

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(RESULT(E), PATIENT)

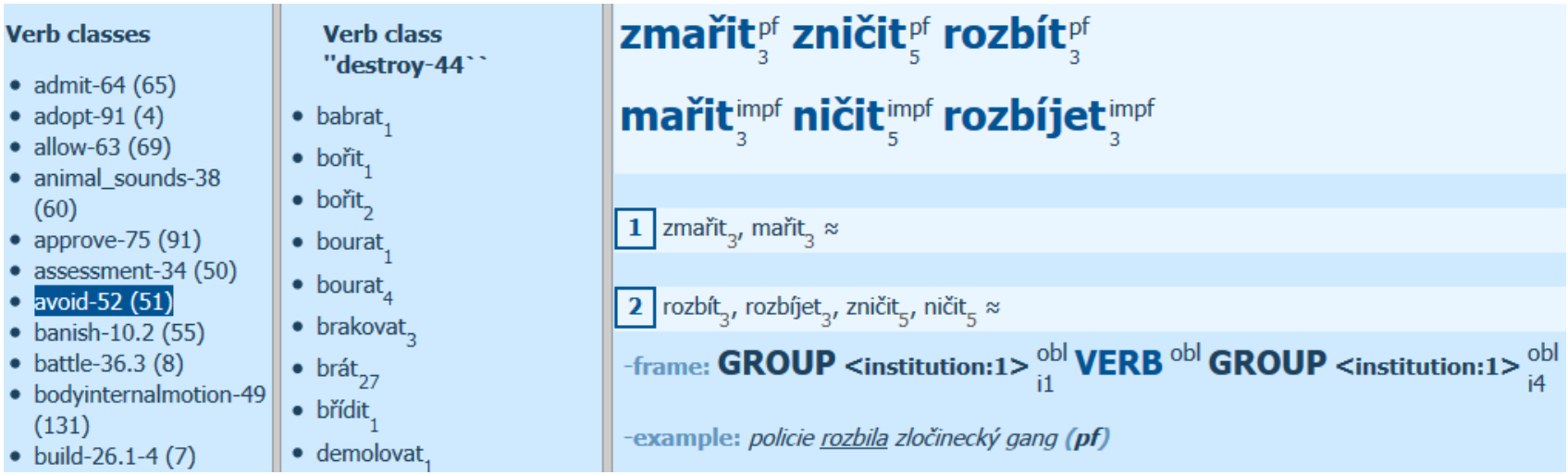
NP V NP PP.INSTRUMENT

EXAMPLE	"The builders destroyed the warehouse with explosives."
SYNTAX	AGENT V PATIENT {WITH} INSTRUMENT
SEMANTICS	CAUSE(AGENT, E) USE(DURING(E), AGENT, INSTRUMENT) DESTROYED(RESULT(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(RESULT(E), PATIENT)

Meaning in the valency frame



- verb + synonyms = synset (synonymical set)
- translation
- verb class (communicate, destroy) based on Palmer's work
- sub-frames: the particular frames for the words from the synset
- syntactic information: number and order of arguments, cases, prepositions
- semantic roles + selectional preferences

Framenet

- lexicon of general valency frames
- interconnection of frames

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.

References


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.