

Slovník a význam

PLIN035: Počítačová lexikografie

František Kovařík

Opáčko

Cross-reference

Hyperlink

Opáčko

Slovník dosáhne „nejdál na sémantiku“.

Nemůžeme zachytit všechna slova/významy.

Experiment – závěr

Všimli jste si netradičních slovních užití
nezachytitelných ve slovníku?

Experiment – závěr

Našli jste část pragmatiky
zachytitelnou ve slovníku?

Jazyková ekonomie

**Efektivní systém,
málokdy problém.**

Jazyková ekonomie

**Efektivní systém,
málokdy problém.**

Některá slova až desítky významů!

Problém

Co je význam?

Co už je nějaký význam?

Co když se dá jakékoli slovo použít jakkoli?

Adam Kilgarriff: I Don't Believe in Word
Senses (1997)

Co je slovní význam?

**Nejbližše slovnímu významu
je jeho užití.**

„Řešení“

Patrick Hanks:

Významové potenciály.

Významové potenciály

Hypoteční banka

Krevní banka

Mama hotel

WSI: word-sense induction

(WS discrimination)

Jaké významy slovo má?

=> vytvoření kontrastních skupin

WSD: word-sense disambiguation

Který význam má slovo v tomto kontextu?

Jak rozlišovat významy?

Jak rozlišovat významy?

Kontext.

Morfologie.

Významové potenciály

The doctor treats people with medicine.

We treated the floor with preservative.

The doctor treated my mother with kindness.

Five people were treated for gunshot wounds.

Významové potenciály

*The doctor treats people **with** medicine.*

*We treated the floor **with** preservative.*

*The doctor treated my mother **with** kindness.*

*Five people were treated **for** gunshot wounds.*

Opáčko – regulární polysémie

(Vypil x Vyhodil) jsem tři lahve.

I like coffee. x Three coffees, please.

Dubový (les x nábytek).

Rozšiřování polysémie

Marvelization

Greenwashing

meerkatted



Vznik nových významů

Metafora, metonymie, synekdocha,
oxymóron, accismus, eufemismus,
dysfemismus, hyperbola, ironie, sarkasmus,
pun, odkaz, metalepsis, ...

Vznik nových významů

<https://en.wikipedia.org/wiki/>

[Figure of speech](#)

Druh textů zaměřený na ohýbání významových potenciálů?

Druh textů zaměřený na ohýbání významových potenciálů?

POEZIE.

WSI – word sense induction

*„You shall know a word by the company it
keeps.“*

(Firth, 1957)

Za vším...

... hledej kolokaci.

Klastrování

Klastry (clusters) ~ „významové skupiny“.

Počítáme podle kolokace (Word Sketch).

WSI – princip

1 význam na dokument

1 význam na kolokaci

1 význam = 1 klastr

Zjednodušený algoritmus

POKUD: slovo má v dokumentu dostatek souvškytů s kolokací v existujícím klastru,

→ přidej slovo do klastru;

JINAK: vytvoř nový klastr s tímto slovem a kolokací.

Jiná metoda

- podle embeddings.

Word Sense Disambiguation

Rozlišování významů.

CPA – corpus pattern analysis

Vzorce v konkordancích...

... se spoustou výjimek.

=> fuzzy (prototypy)

CPA – corpus pattern analysis

Norms and Exploitations

How clustering works: *treat*

In fact they'd treated him rather badly, although he...
He treated his father with such disdain, ...
He treats them as if they're his personal friends...
...our ability to treat others with respect...
That doesn't give you the right to treat me like a child
...ensure that the people here are treated fairly

Cluster 1: recurrent patterns

To treat a person...

- + adverb (badly, fairly, considerately...)
- + with + noun (with contempt, with kindness)
- + like or as if (like a child, as if you're an idiot)

Cluster 2: recurrent patterns

To treat a person...

- + with + noun (with antibiotics, with steroids)
- + for + noun (for smoke inhalation, for gunshot wounds...)

To treat ...

- + an illness (used in treating cancer, diabetes etc)
- + a person (all patients are treated within 2 hours)

Two other women were being treated for gunshot wounds.
A young girl who died after being treated with steroids...
...a drug that is widely used to treat breast cancer...
Middle-ear infections are frequently treated with antibiotics
and...
...and a smaller rise in the number of patients treated in hospital wards...
...were taken to hospital and treated for smoke inhalation

Koligace

Vztah mezi

lexikální jednotkou

a gramatickou kategorií / textovou

strukturou.

Koligace

Preferovaná pozice ve větě.

Preference určitého tvaru (pasivum, množné číslo...) nebo negativního kontextu.

Charakteristický ko-text.

Koligace v češtině

Tři psi x pět psů.

Pořád skáče x *skočí.

Definice

Definuje něco?

Definice

... definuje něco?

=> *explanation* (vysvětlení)

Definice – např.

Genus proximum + differentia specifica.

Synonymum.

Ustálené formulace.

- **pedantic** of, relating to, or being a pedant
- **importance** the quality or state of being important
- **realization** the action or realizing: the state of being realized

a book, optical disc, mobile device, or online lexical resource (such as Dictionary.com) containing a selection of the words of a language, giving information about their meanings, pronunciations, etymologies, inflected forms, derived forms, etc., expressed in either the same or another language; lexicon; glossary. Print dictionaries of various sizes, ranging from small pocket dictionaries to multivolume books, usually sort entries alphabetically, as do typical CD or DVD dictionary applications, allowing one to browse through the terms in sequence. All electronic dictionaries, whether online or installed on a device, can provide immediate, direct access to a search term, its meanings, and ancillary information:

Definition of *dictionary*

absolute (noun)

Something that is absolute (*American Heritage Dictionary*)

absolute (noun)

Something that is absolute (A

ary)



Jak na to?

Popis prototypu.

Konec snahy popsat vše.

Pochopení

Závisí i na adresátovi.

Volba metajazyk.

Rekurzivita, cyklení významů.

Typy definic

„Klasické definice“.

FSD = full-sentence definition.

Vyjádření pragmatiky

shack up: to live with another person as sexual partners without being married. This word shows that you do not approve of people who do this.

(macmillandictionary.com)

Vyjádření pragmatiky (subjektivní názor)

school An institution originally non-compulsory, now a forced hellhole where otherwise reasonably intelligent people are forced to go for 14 years. (Urban Dictionary)

Pastiňáky

Citlivá témata aka PARSNIP


P – Politics 

A – Alcohol 

R – Religion 

S – Sex 

N – Narcotics 

I – “-isms” (e.g. communism, atheism...) 

P – Pork (i.e. cultural taboos) 

- slova vysvětlovat, ale nehodnotit, neprojektovat svoje (kulturní) očekávání (jenže... nakolik jsme si jich *u sebe* vědomi?)

Přehled

Slova mají významové potenciály.

Nejblíže významu je užívání slova.

Při tvorbě definic se řídíme selským
rozumem.