

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

23. dubna 2012

Zkoumání kontextu

Formalismy

Domény

Významové potenciály

Kontext

The context of a piece of language (...) is its surrounding environment. But this can include as little as the articulatory movements immediately before and after it, or as much as the whole universe, with its past and future.

(Werth, 1999 in [Requejo, 2007])

└ Zkoumání kontextu

└ Kontext

The context of a piece of language (...) is its surrounding environment. But this can include as little as a sentence or two, or as many as the whole of society.

(Wether, 1999 in [Rogalska, 2007])

Kontext už jsme zmiňovali. Bohužel jsme jej zatím přesně nevymezili. Proč? Protože je to těžké. Zdá se, že každé slovo má v promluvě jinou váhu. Podobně i „váha“ kontextu nijak zvlášť nesouvisí s jeho „velikostí“.

Kontext v kontextu kognitivní lingvistiky

- klasický přístup = začneme zkoumat kontext, teprve až selžou všechny ostatní pokusy
- „nový“ přístup = kontext určuje význam, začněme tedy kontextem. . .
- existence kontextu je důvod, proč nelze zcela oddělit sémantiku od pragmatiky [Requejo, 2007]
- porozumění promluvě závisí na jazykové znalosti, ale také na znalosti světa (Bransford et Johnson, 1972)

└ Zkoumání kontextu

└ Kontext v kontextu kognitivní lingvistiky

- klasický přístup = za česme zkusmat kontext, teprve až vidíme všechny ostatní jazyky
- „nový“ přístup = kontext a č. je výsledek, za č. se tedy kontextem...
- vzhledem k kontextu je č. jednoduché, protože zcela odlišné od ostatních jazyků [Bresnan, 2007]
- poznámka: srovnání č. s jazyky z oblasti, ak. tak se z. v knize světa [Bresnan et Johnson, 2012]

Klasický přístup je silně ovlivněný Chomského generativní gramatikou.
Ch. zkoumal různé jazykové roviny odděleně.

Kontext: rozdělení

- jazykový
- situační (objektivní i subjektivní)

└ Zkoumání kontextu

└ Kontext: rozdělení

- jazykový
- situační (subjektivní i objektivní)

Jazykový kontext: co bylo/bude řečeno, odkazování v promluvě (anafora, antecedent, katafora). Situační kontext: kdo/co se vyskytuje v místě a čase promluvy, např. kdo je *já*, kde je *tady*? Subjektivní jazykový kontext je hůř postižitelný než objektivní, jsou to naše nálady, ale i temperament. Příklad: ta samá zpráva u nás může vyvolat rozdílnou reakci podle našeho vnitřního rozpoložení. Ano, je řeč o emocích. Vrátime se k nim v závěru kurzu.

Kontext a kognitivní lingvistika

- slova nenesou významy, naopak slova jsou instrukce k výstavbě významu jazykového výrazu
- význam promluvy není funkcí významu jednotlivých částí
- víceznačnost je norma
- nemá cenu zkoumat menší jednotky (slova, izolované věty) bez znalosti kontextu

- slova se mohou významy, mohou i slova jsou kontextu k výstavě významu jazykové ho výzva
- význam promluvy není funkce významu je složky k částí
- významnost je sama
- nemá cenu zkusit měřit jednotky slova, kolovat výhy bez zkusit kontext

Kognitivní l. nemá na naši problematiku vysloveně revoluční pohled (nic starého neplatí), dívá se spíš „z druhé strany“ (to staré platí, ale dosud jsme se dívali ze špatného úhlu). Ona druhá strana vychází z hypotéz o naší mysli. Nezajímá nás jen výsledek (jazyková analýza promluvy), ale i cesta k němu. Modely k. l. mají (snad) blízko k tomu, jak zpracováváme jazykové vstupy my, lidé.

Klasický přístup se studuje déle, a proto má víc a známějších formalismů. Na druhou stranu kognitivním lingvistům je jasné, že bez nějaké formalizace se dál nedostanou. Proto formální modely existují a podíváme se na ně. . .

Kontext: formalismy a studium

- sémantické a asociativní sítě, rámce (Fillmore, 1985), skripty (Abelson), domény (Langacker 1987), mentální prostory (Fauconnier, 1994)
- významové potenciály (Hanks)
- kontextové vektory

└ Formalismy

└ Kontext: formalismy a studium

- sémantická a asociativní síť, síťce (Fillmore, 1985), sítě (Aksentijevič, 1987) (Langeheine, 2007), mentální prototypy (Fillmore, 1984)
- výrazové prototypy (Rasnik)
- kontextové sítě

O sém. sítích, rámcích a skriptech jsme už mluvili. Domény a mentální prostory zmíníme dnes. Jsou to novější formalismy, které se v praxi spíš nepoužívají (možná jen v prototypch). Inspirativní otázka je, jestli s tím nemáme začít.

Kontext: formalismy

sémantické a asociativní sítě, rámce (Fillmore, 1985), skripty (Abelson)

- zachycení kontextu pomocí relací: isa, has a, located in . . .
- typický (mnohdy nezmíněný) kontext
- částečně kulturně a geograficky závislé
- některé části jsou nezávislé, ale nevíme, které to jsou
- jak reagovat na změny stereotypů (např. telefonování)?

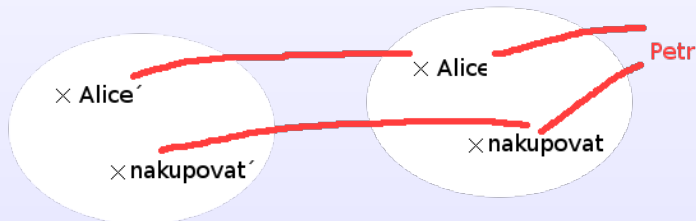
- zachycení kontextu a formální sémantika, has a, located in ...
- typický (mnohdy seznámí) koncept
- částí síťového a geografického síťového
- síťové části jsou seznámí, ale seznámí, síťové síťové
- jak seznámí seznámí síťové síťové [zapř. telefonová]

Kulturní závislost rámců či skriptů by byla dobrým tématem BP nebo aspoň referátu. Dalším zajímavým tématem může být rychlost změny stereotypů – zatímco obchod je víceméně stejný jako před 100 lety (snad kromě platebních karet), telefonování se změnilo, telegramy zanikly, sociální sítě se rozmohly ...

Mentální prostory (Fauconnier, 1994)

pracují s rámci, ale umožňují zrcadlení

Petr si myslí, že Alice šla nakupovat.



Domény (Langacker 1987)

What one finds in language depends in large measure on what one expects to find. [Langacker, 1987]

Domény (Langacker 1987)

význam je kognitivní jev

cognitive grammar – vychází z lingvistiky, ale inspiruje se v kognitivní psychologii a umělé inteligenci

1. sémantika není univerzální, ale jazykově závislá
2. nelze oddělit syntax a sémantiku
3. nelze oddělit syntax a lexikon

Domény (Langacker 1987)

Každá jazyková jednotka je v určité míře závislá na kontextu. Kontext, který charakterizuje sémantickou jednotku, nazýváme doména.

Základní domény vs. abstraktní domény

Koncept NEHET předpokládá existenci konceptu PRST. Koncept PRST předpokládá existenci konceptu RUKA. . . předpokládá existenci konceptu TĚLO. TĚLO je **základní doména**, předpokládá sice existenci trojrozměrného prostoru, není tím ale nijak zvláštní. Doménu, která není základní, nazýváme **abstraktní**.

Význam slova odpovídá konfiguraci (umístění) v jedné nebo více doménách.

Domény (Langacker 1987)

Domény mají dimenze, např. TEPLOTA je jednorozměrná. Díky dimenzím je možné významy uspořádat.

Domény jsou uzavřené (konečný rozsah BAREV, které můžeme rozeznat), nebo otevřené (ČAS).

Hodnoty rozměru v doménách mohou být spojité nebo diskrétní, základní jsou většinou spojité, abstraktní jsou někdy diskrétní (RODINNÉ VAZBY).

Většina významů je definovaných ve více doménách zároveň: např. BANÁN je definovaný v doménách BARVA, TVAR, JÍDLO, STROM, CHUŤ.

Domény mají různé us, např. TEPLOTA je jednodoménová. Díky širšímu je možné vyjádřit například:

Domény jsou uzavřené (konečný seznam BAREV, které můžeme nazvat), nebo otevřené (ČAS).

Naše metody v doménách mohou být spjaté nebo částečně odlišné (jako v případě spojité a bodu kde jsou střídy částí těla) (ROVNĚŽ VÁZBY).

Všechny tyto mají je číselnou sítí ve více doménách (zároveň: např. BANÁN je číselnou sítí BARVA, TVAR, JÍDLO, STROM, CHŮT.

Trochu to připomíná sémantické rysy. Na rozdíl od nich (rys u pojmu buď je, nebo není přítomen) je však definován rozsah hodnot (např. teplota živého těla). Domény rozhodně stojí za prozkoumání a jsou dobrým námětem pro referát nebo BP. V BP by mohlo být zajímavé např. určit domény slovesným rámcům z VerbaLexu.

Významové potenciály (Hanks)

- slovníky pro lidi jsou pro počítače nevyhovující
- slovníky kombinované a hierarchické (WordNet) mají lepší výsledky v počítačovém zpracování (Nirenburg 2007)
- význam není jedno políčko v „kontrolním seznamu“
- významy lze nejlépe interpretovat pomocí pravděpodobnosti, s jakou se **užití** blíží prototypu (Fillmore)
- významy jsou spojeny se vzory (patterns, Hunston a Francis, 2000), konstrukcemi (Goldberg) či „frazémy“ (Melčuk), ale i se slovy – korpusová lingvistika dokáže tyto **vzory** (preference užití) zjistit a studovat
- významy vně kontextu neexistují, existují jen **významové potenciály**, které se kontextem **aktivují** [Hanks, 2000]

Významové potenciály (Hanks)

Corpus Patterns Analysis

PATTERN: [[Human]] translate ([[Document]]) (from [[Language 1]]) (into [[Language 2]])

IMPLICATURE: [[Human]] expresses the meaning of [[Document]] in [[Language 1]] in the words and phraseology of [[Language 2]]

└ Významové potenciály

└ Významové potenciály (Hanks)

Corpus Pattern Analysis

```
PATTERN: [[Hans]] translate [[Document]] from [[language 1]] into [[language 2]]
```

```
IMPLICATION: [[Hans]] express the meaning of [[Document]] in [[language 1]] in the words and physiology of [[language 2]]
```

Patrick Hanks je korpusový lingvista, který v letech 2005–2008 působil na FI, teď působí na Karlově univerzitě. Jeho přístup je velmi blízký přístupu kognitivní lingvistiky, i když myslím, že k. l. není příliš v jeho oblasti zájmu. Korpusová lingvistika se rozvíjí s tím, jak nám rostou korpusy a také software pro jejich zpracování.

Významové potenciály (Hanks): normy (norm) [Hanks, 2010]

- **norma** = užití slova podle určitého syntagmatického vzoru
- slova se užívají jednak **v souladu** s normou (očekáváním), jednak mluvčí normu **porušují**
- **vzory** se skládají z užitého slova a z **lexikálních množin**, se kterými se slovo užívá
- lexikální množiny mohou být obrovské, např. [[Human]]
- čím menší lexikální množina, tím silnější je její vliv na význam vzoru

└ Významové potenciály

└ Významové potenciály (Hanks): normy
(norm) [Hanks, 2010]

- *norma* = síli slovo *zauk* a čít ho systematické ho vzora
- slovo se užívá j *ěsa k v sa daj* a normu j *ěsa k v sa daj*,
j *ěsa k v sa daj* normu *sa daj*
- *vzory* se užívá j *ěsa k v sa daj* a *ležišlech moudra*, se
kterými se slovo užívá
- *ležišlech moudra* máho a být uková, např. [Hanks]
- Čím se síli *ležišlech moudra*, tím síliší je j *ěsa k v sa daj* se výzorem
vzora

Konečně se dostáváme k nějakému tvrzení, které se vyjadřuje k velikosti kontextu. Zpočátku to vypadalo, že prostě kontext uchopit nemůžeme (přestože s konkrétní situací se dokážeme vyrovnat).

Významové potenciály (Hanks): porušení normy (exploitation)

- kreativní užití jazyka
- porušení normy nepřesahuje 10 % případů v korpusu
- pokud ano, je to nejspíš dosud neobjevená norma
- i porušení normy má jistá pravidla „dvojitá šroubovice“ systémů pravidel: pravidla, jak slova používat normálně, a pravidla, jak normu porušit
- často studovaným vzorem je **valence**

Významové potenciály (Hanks): valence

...možná bych prodloužila poslední verš...	nesedí	mi tam nejen vizuálně, že je příliš krátký
Whitesun a stále nedokončeným C-3PO. Cliegg	sedí	na létajícím křesle, protože mu chybí jedna
atraktivní módní přehlídce. Tvůrkyně ošacení	seděla	na premiéře v první řadě v atraktivní róbě
zablokovali oba směry magistrály. Někteří	sedí	na schodech k opeře. Mávají rudými vlajkami
centrem byl zdemolován vůz TV Nova. Jinak zde	sedí	ještě menší skupina demonstrantů. Jde většinou
hráli v sále hospody a hodně lidí zůstalo	sedět	venku, protože tentokrát počasí přálo.
jsem byl naprosto unešena.Jsou perfektní.	Sedí	jako ulité.Jen délka ,ale tu jsem za 15
psala, dorazil a věcičky jsou nádherný Vše	sedí	tak jak má a já jsem mooc spokojená Moc
dnes mi s podprsenkami dorazil... Skvěle	sedí	. Jsem moc spokojená! Budu se těšit zase
moc děkuji, těhotenské rifle jsou super -	sedí	úplně perfektně!!! A taky děkuju za dáreček

[[Human | Group of Humans]] sedět [[Location]]

[[Garment]] sedět [[Human, dative]] [[Manner]]

[[Art]] sedět [[Human, dative]] [[Location]] [[Manner]]

Významové potenciály (Hanks): elipsa

...možná bych prodloužila poslední verš...	nesedí	mi tam nejen vizuálně, že je příliš krátký
Whitesun a stále nedokončeným C-3PO. Cliegg	sedí	na létajícím křesle, protože mu chybí jedna
atraktivní módní přehlídce. Tvůrkyně ošacení	seděla	na premiéře v první řadě v atraktivní róbě
zablokovali oba směry magistrály. Někteří	sedí	na schodech k opeře. Mávají rudými vlajkami
centrem byl zdemolován vůz TV Nova. Jinak zde	sedí	ještě menší skupina demonstrantů. Jde většinou
hráli v sále hospody a hodně lidí zůstalo	sedět	venku, protože tentokrát počasí přálo.
jsem byl naprosto unešena. Jsou perfektní.	Sedí	jako ulité. Jen délka ,ale tu jsem za 15
psala, dorazil a věcičky jsou nádherný Vše	sedí	tak jak má a já jsem mooc spokojená Moc
dnes mi s podprsenkami dorazil... Skvěle	sedí	. Jsem moc spokojená! Budu se těšit zase
moc děkuji, těhotenské rifle jsou super -	sedí	úplně perfektně!!! A taky děkuju za dáreček

[[Human | Group of Humans]] sedět [[Location]]

[[Garment]] sedět [[Human, dative]] [[Manner]]

[[Art]] sedět [[Human, dative]] [[Location]] [[Manner]]

Významové potenciály (Hanks): elipsa

Elipsy mají také svá pravidla. Co může být vypuštěno?

- podmět
- předmět
- příslovečné určení

Za jakých okolností se může vyskytnout elipsa? Vypuštěno může být jen to, co je zřejmé.

Elipsa jako porušení normy? Je to otázka frekvence.

└ Významové potenciály

└ Významové potenciály (Hanks): elipsa

Elipsy mají také své pravidla. Co může být vynecháno?

- psát mět
- před mět
- předložit síce si

Za je být k okdantí se může vynechat elipsa? Vypačí se může být je tto, co je ztojt.

Elipsa jako porušení normy? Je to otázka frakce.

Elipsa jako porušení normy: svíčková na smetaně je dnes nepochybně normou, původně jde o elipsu.

Existence a pravidla elipsy jsou dobrý námět pro referát nebo BP.



Hanks, P. (2000).

Do word meanings exist?

Computers and the Humanities, 34:205–215.

10.1023/A:1002471322828.



Hanks, P. (2010).

Elliptical arguments: a problem in relating meaning to use.

In Paquot, M. and Granger, S., editors, *eLexicography in the 21st century : New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, volume 7 of *Cahiers du Cental*, pages 109–124, Louvain-la-Neuve, Belgium. Université Catholique de Louvain, Presses universitaires de Louvain.



Langacker, R. W. (1987).

Foundations of cognitive grammar: Theoretical Prerequisites.

Stanford University Press, Stanford, CA.

Vol 1, 1987(Hardcover), 1999(Paperback).



Requejo, M. D. P. (2007).

The role of context in word meaning construction: A case study.

International Journal of English Studies, 7.