

Databáze slovesných valenčních rámců

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

Mgr. Jakub Machura, Ph.D.

Slovesná valence

- **slovesná valence** – schopnost slovesa vázat k sobě další lexikální jednotky (substantiva, adjektiva, adverbia, infinitiv jiného slovesa, vedlejší větu)
- *jíst jídlo, stát se slavným, vidět rudě, jít běhat, říct, aby přišli*
- schopnost je dána (2 přístupy):
 - gramaticky (gramatickými vlastnostmi slovesa) – *Vallex*
 - sémanticky (významem slovesa) – *VerbaLex*
- pozn. existuje také valence substantiv, adjektiv či adverbí

Slovesná syntax – teoretická východiska

- **závislostní syntax** (L. Tesnière)
 - doplnění slovesa – aktanty a cirkumstanty (= volná doplnění)
- **pádová gramatika** (Case Grammar, Ch. Fillmore)
 - zavádí tzv. hloubkové pády (Semantic Roles – Agent, Patient ad.), Frame semantics
- v ČR zejména František Daneš (oba přístupy)
 - gramatické větné vzorce (GVV)
 - sémantické větné vzorce (SVV)
- později Petr Karlík (ÚČJ FF MU, směr Fillmore) a Jarmila Panevová (ÚFAL MFF UK, směr Tesnière a FGP Petra Sgalla)

Valenční rámce

- formální záznam slovesné valence
- zachycují valenci
 - levostrannou (subjekt/podmět)
 - pravostrannou (objekt/předmět a další aktanty/argumenty/doplnění)
- bývají doplněny o sémantické informace – hloubkové pády / sémantické role / tématické role / funktory, příp. o další informace o slovesech
- *Petr jí jídlo příborem v jídelně.*

Valenční slovníky a databáze

- Svozilová N., Prouzová H., Jirsová A. – **Slovesa pro praxi. Valenční slovník nejčastějších českých sloves**. Praha: Academia, 1997. (tištěná publikace, elektronicky na webu ÚJČ AV ČR)
- **Vallex** – ÚFAL MFF UK, od r. 2002, webová aplikace a tištěná publikace
 - Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., Žabokrtský, Z.: **Valenční slovník českých sloves VALLEX**. Praha: Karolinum, 2016.
 - popis viz <http://ufal.mff.cuni.cz/vallex>
 - prohlížení Vallexu 4.5
<https://ufal.mff.cuni.cz/vallex/4.5/about.html>

Valenční slovníky a databáze

- **VerbaLex** – CZPJ FI MU (Hlaváčková, Horák)
 - webová aplikace
 - od r. 2004
 - proti Vallexu více zaměřen na sémantickou rovinu
 - bližší popis viz dizertační práce *Hlaváčková, D. Databáze slovesných valenčních rámců VerbaLex, 2008* (v ISu)

VerbaLex – cíl

- vytvořit rozsáhlou elektronickou databázi českých slovesných valenčních rámců
- zachytit valenci sloves na syntaktické i sémantické úrovni
- doplnit další relevantní informace o chování sloves v přirozeném kontextu

VerbaLex – stručná charakteristika

- začátek práce v roce 2004
- 15 anotátorů + 6 pracovníků CZPJ (softwarová podpora)
- zpracováno **10 596** slovesných lemmat
- **výchozí zdroje**
- **BRIEF** (1997, FI MU), 15 000 sloves, 50 000 povrchových valenčních rámců
- **Czech WordNet** (2002–2004, Balkanet, FI MU), 1 359 valenčních rámců k 824 synsetům
- **Vallex** (od 2002, UFAL MFF UK), Vallex 2.0 – 4 250 lemmat

VerbaLex – stručná charakteristika

- **typické rysy VerbaLexu**
- částečně inspirace z WordNetu
- slovesné lemma (variantní lemma) s číslem významu
- synonymické řady (tvořeny posloupností lemmat, např.: *jíst:1, požit(požívat):2*)
- syntaktická a sémantická rovina rámce
- základní a komplexní valenční rámeček

Základní valenční rámec

- valenční doplnění na syntaktické úrovni (přímé a předložkové pády)
- valenční doplnění na sémantické úrovni (sémantické role)
- nejfrekventovanější idiomatická doplnění

Komplexní valenční rámec

- synonymie, číslování významů polysémních sloves
- definice významu synonymické řady
- homonymie (číslování odlišných významů, př. *sladit, stát*)
- možnost tvoření pasiva + tranzitivnost, intranzitivnost
- slovesný vid (slovesa dokonavá, nedokonavá, obouvidá)
- sémantické třídy sloves
- způsob užití slovesa (základní, přenesené, idiomatické)
- reflexivita (refl. tantum, reciprocita, syntakt. refl., absol. synonym.)
- doloženo konkrétními příklady

Základní valenční rámec

Synset: *jíst:1, požit(požívat):2*

Def: *přijímat potravu*

Subsynset: *jíst:1*

Sémantická role 1. úrovně

AG – agens

INS - instrument

obligatorní

Pozice slovesa

fakultativní

AG(kdo1;<person:1>;obl) **VERB SUBS**(co4;<food:1>;obl) **INS**(čím7;<cutlery:2>;opt)

*zájmenný výraz
s číslem pádu*

Sémantická role 2. úrovně

SUBS – substance

Komplexní valenční rámec – příklad

jíst:1(impf), požit:2(pf), požívat:2(impf)

definition: *přijímat potravu*

class: eat-39.1

passive: yes

jíst:1 ≈

-frame: **AG**(kdo1;<person:1>;obl) **VERB SUBS**(co4;<food:1>;obl) **INS**(čím7;<cutlery:2>;opt)

-example: *synovec jedl zmrzlinu (impf)*

-example: *dcera jí polévku lžící (impf)*

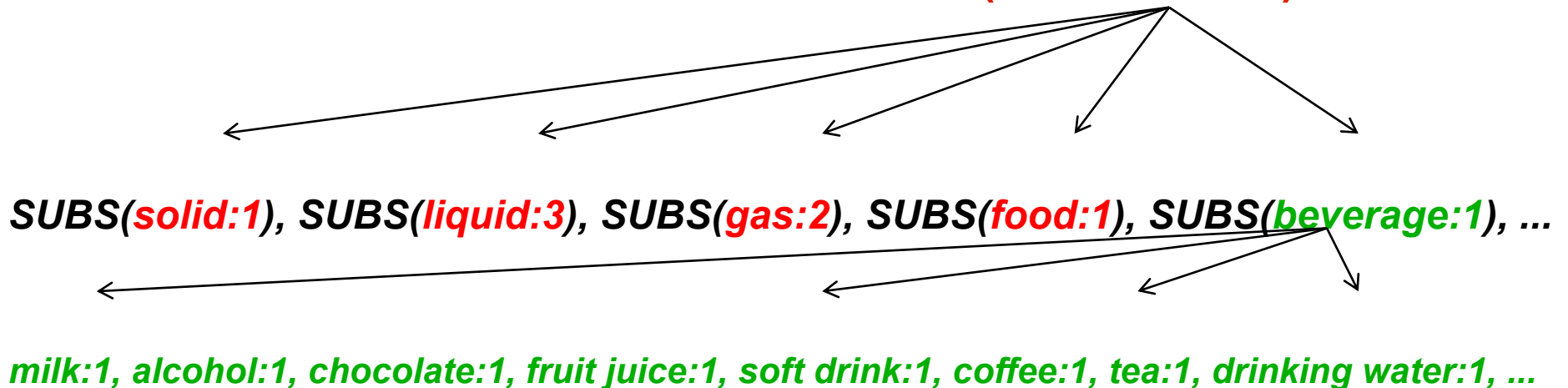
-synonym: **požit:2, požívat:2**

-use: prim

-reflexivity: no

Dvouúrovňové sémantické role

- **EuroWN vrcholová ontologie (Top-Ontology)** ▶ vrcholová hyperonyma EuroWN (63 prvků), entity 1., 2. a 3. řádu
- **Např.:** EuroWN Top-Ontology: entita 1. řádu ▶
- **Substance** – ve VerbaLexu sémantická role **1. úrovně** ▶ **SUBS**
- **2. úroveň** – hyperonymum ▶ **substance:1** *definition - that which has mass and occupies space*
- **dvouúrovňová sémantická role** ▶ **SUBS(substance:1)**



• **hyponymické lexikální jednotky specifikující význam**

(beverage:1 – definition – any liquid suitable for drinking)

Sémantické třídy sloves

- **Motivace:** sémantická klasifikace predikátů (Daneš, Grepl, Karlík), vztah sémantických rolí ve VerbaLexu k sémantickým třídám sloves
- **Východisko:**
- Beth Levin, *English Verb Classes and Alternations* –
 - 48 základních sémantických tříd
- Martha Palmer, *VerbNet* – 82 základních sémantických tříd, celkem 395 podtříd
- české sémantické třídy – modifikovaný překlad, doplnění o další synonyma, vidové protějšky a prefigovaná slovesa,
 - 82 základních sémantických tříd, celkem 258 podtříd, aktuálně seznam zahrnuje 11 241 sloves, z toho 6 393 různých lemmat

Použité softwarové nástroje

- **Existující nástroje**

- **DEBDict** – slovníky *SSČ*, *SSJČ*, *SČFI*, *SČS*

- synonymie, definice významu synonymické řady, homonymie, možnost tvoření pasiva + tranzitivnost, intranzitivnost, slovesný vid, způsob užití slovesa, reflexivita

- **Webové rozhraní Bonito2, Word Sketch Engine**

- korpusy SYN2000, ALL

- ověřování valenčních doplnění a nejfrekventovanějších idiomatických doplnění, konkrétní příklady užití ve větách

- **DEBVisDic**

- dodržení návaznosti VerbaLexu na Czech WordNet a Princeton WordNet, číslování významů

Použité softwarové nástroje

- **Nástroje upravené pro potřeby VerbaLexu**
- **Editor gVIM** – editace databáze, kontrola formálních chyb
- **převody do formátů xml, pdf a html** – formální podoba datové struktury databáze
- v současné době možnost zadávat data přes webový formulář

Dostupnost

- dostupnost:
 - <https://nlp.fi.muni.cz/verbalex/html2> (starší webové rozhraní, úplné informace)
 - <https://nlp.fi.muni.cz/verbalex/html3/> (novější webové rozhraní, neúplné informace)
 - v obou případech login: verbalex, heslo: cjbb85