

# Databáze slovesných valenčních rámců VerbaLex

Dana Hlaváčková

# Úvod – zadání práce (valence)

- vytvořit rozsáhlou elektronickou databázi českých slovesných valenčních rámců
- **valence** – významem determinovaná schopnost slovesa vázat na sebe další slova
- zachytit valenci sloves na syntaktické i sémantické úrovni
- doplnit další relevantní informace o chování sloves v přirozeném kontextu
- popsat teoretický základ databáze VerbaLex

# VerbaLex - stručná charakteristika

- začátek práce v roce 2004
- 15 anotátorek, 6 pracovníků CZPJ se podílelo na softwarové podpoře
- aktuálně 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

## Typické rysy VerbaLexu

- slovesné lemma (variantní lemma) s číslem významu
- synonymické řady (tvořeny posloupností lemmat, př.: *jíst:1, požit(požívat):2*)
- syntaktická a sémantická rovina rámce
- základní a komplexní valenční rámeček

# Základní valenční rámec

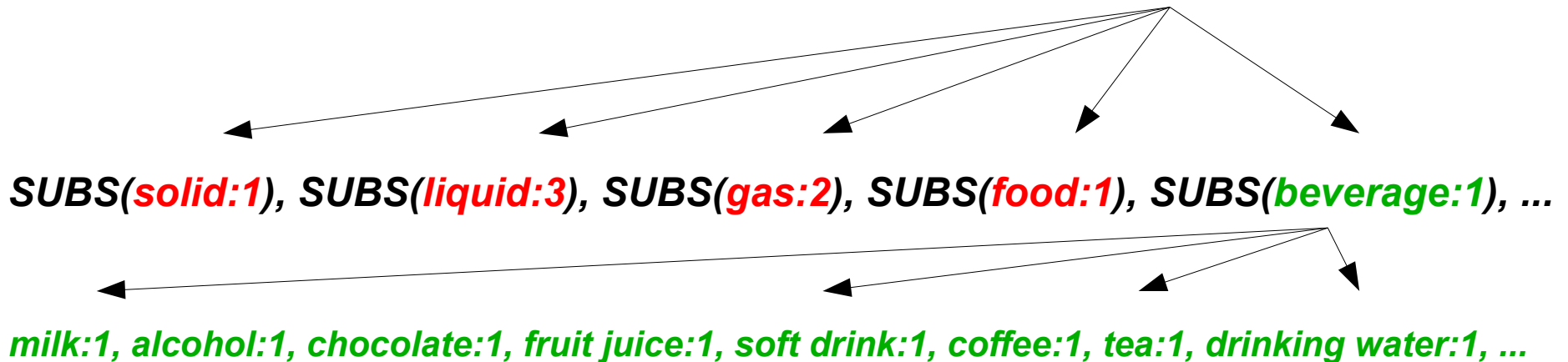
- valenční doplnění na syntaktické úrovni (přímé a předložkové pády)
- valenční doplnění na sémantické úrovni (sémantické role)
- nejfrekventovanější idiomatická doplnění

# 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

# 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)*

# 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

# 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

## Nástroje upravené pro potřeby VerbaLexu

**Editor gVIM** – editace databáze, kontrola formálních chyb

**převody do formátů xml, pdf a html** – formální podoba datové struktury databáze