

Linking Czech Verbal Synsets to English WordNet

Vašek Němčík, Dana Hlaváčková, Aleš Horák, Karel Pala

Centre of Natural Language Processing
FI MU Brno

WordNet

- Large lexical database, semantic net
- Nouns, verbs, adjectives and adverbs in semantic relations
- Synsets (sets of synonyms) in hyper-hyponymic relations
- PrincetonWN (PWN) – from 1995
- CzechWN (CzWN) – from 1998
- Interlingual index (ILI)
- 5158 verb synsets

VerbaLex

- Database of Czech verb valency frames
- CZPJ FI MU Brno, from 2005
- Headwords = lemmata in synsets (Czech verbs with sense numbers from CzWN + new added verbs and synsets)
- Basic valency frame – morpho-syntactic level + semantic level (semantic roles, deep cases) + obligatory/optional verb complements
- 1st level semantic roles (list of 30 labels)
- 2nd level semantic roles (literals from PWN)
- Complex valency frame – verb aspect, definition, verb passive, semantic class, use, reflexivity, example
- 6 360 synsets, 21 193 literals, 10 482 lemmas, 19 556 frames

Basic Valency Frame

Synset: *jíst:1, požit(požívat):2*

Def: *přijímat potravu*

Subsynset: *jíst:1*

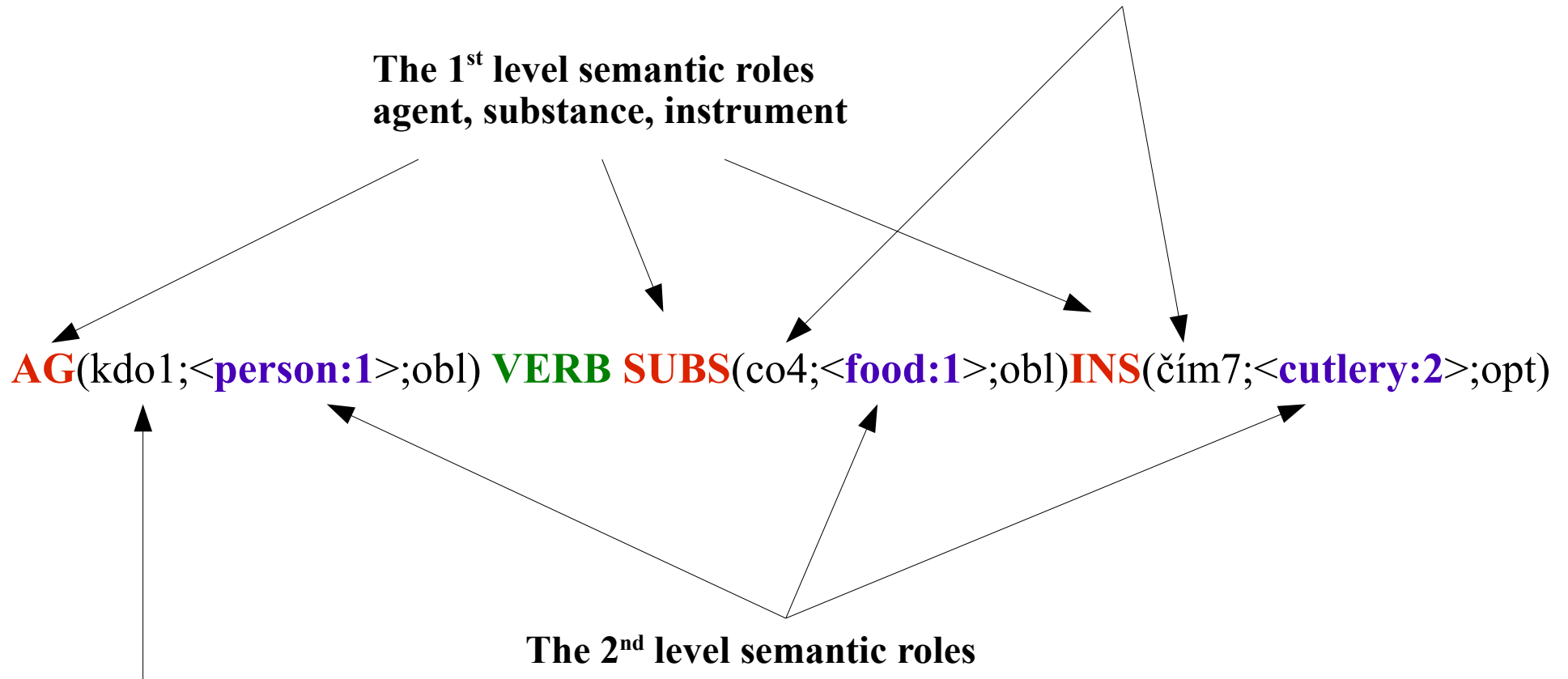
*The pronominal term
indicating number of case
what – accusative, by sth
– instrumental*

**The 1st level semantic roles
agent, substance, instrument**

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

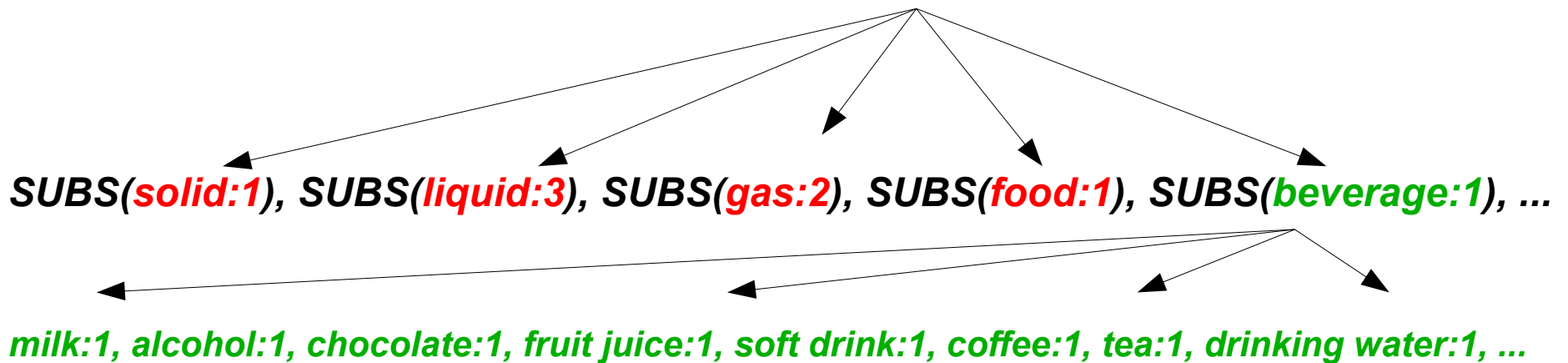
The 2nd level semantic roles

*The pronominal term
indicating number of case
who - nominative*



Two level semantic roles

- **EuroWN Top-Ontology** ▶ top hypernyms EuroWN (63 items), 1., 2. and 3. order entity
- **Example:** EuroWN Top-Ontology: 1. order entity ▶
- **Substance** – 1st level semantic role in VerbaLex ▶ **SUBS**
- **2nd level** – hypernym ▶ **substance:1** *definition - that which has mass and occupies space*
- **two level semantic role** ▶ **SUBS(substance:1)**



- **hyponymic literals from PWN**

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

Complex Valency Frames

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

VerbaLex – New Synsets

- 3 686 newly added synsets
- for each new synset, find:
 - a link to an English counterpart in PWN
 - a link to the Czech hypernym
- **WordNet Assistant**
- automates humdrum tasks
- helps human lexicographers to locate relevant synsets

WordNet Assistant – Details

- dictionary look-up
 - GNU/FDL English-Czech dictionary, ZČU Plzeň
- WN interface queries
 - DEB server service providing access to WN data
- offering candidates to the human user
 - heuristic ordering according to relevance
 - the goal is not to offer just one candidate
 - comparison with similar/related synsets is crucial

Problems in Linking to PWN

- **Unlinkable Czech synsets (~15 %)**
- Perfective verbs – usually end of action
doletět, dočesat, dokrmit ...
- Reflexive verbs
*naběhat se, nalítat se, načekat se,
maskovat se ...*
- Metaphoric expressions
 - *nalinkovat* (to plan something for somebody)
 - *žrát* (to nag at somebody)

Problems in Linking II

- Expressive
 - *ňafat se* (to argue with somebody)
- Czech verbs with no English equivalent
 - *přistavit* (park/stop a vehicle somewhere)
- Czech verbs with no equivalent in PWN
 - *přepólovat* (change polarity)

Data Check

- more Czech synsets linked to one PWN synset
- can be hardly prevented
- helps to detect duplicate synsets
- some synsets need to be merged
- some synsets need finer description
- *round:8, flesh out:3, fill out:6*
 - *baculatět:n1, buclatět:n1*
 - *kulatět se:n1, kulatit se:n1*

Choose file: 

Load!

Show definitions: Show examples: (672) **logaritmovat:n2****Example:** *logaritmoval součin svou čísel*(673) **lokalizovat:n1****Example:** *autor lokalizoval děj románu na venkov***Example:** *restaurátor lokalizoval sochu nad průčelím**Národního divadla***Example:** *autor lokalizoval děj knihy do 15. století*(674) **lokat:n3, loknout si:n3****Example:** *lokal krásu místní krajiny***Example:** *lokal svobodu plnými doušky***Example:** *lokl si volnosti*(675) **loknout:n2(lokat), loknout si:n2****Example:** *lokl čerstvý vzduch***Example:** *lokl si čerstvého vzduchu*(676) **loknout si:n1, napít se:n1, cvaknout si:n3****Example:** *zedník se napil přímo z láhve***Example:** *přátelé si cvaknuli slivovice*(677) **lombardovat:n1****Example:** *obchodník lombardoval svými cennými papíry***Example:** *banka lombardovala svými aktivy***SYNSETS FOR LOKALIZOVAT:N1 (27)**

- [ENG20-01662808-v] **set:9, localize:4, localise:4, place:10** (2/30)
Definition: locate
Example: *The film is set in Africa*
No equivalent
- [ENG20-02266215-v] **locate:3, place:4, site:1** (2/30)
Definition: assign a location to
Example: *The company located some of their agents in Los Angeles*
Equivalent: nastavět:2, lokalizovat:1
- [ENG20-02614695-v] **place:12, localize:1, localise:1** (2/30)
Definition: identify the location or place of
Example: *We localized the source of the infection*
Equivalent: lokalizovat:2
- [ENG20-01452069-v] **put:1, set:1, place:1, pose:5, position:2, lay:1** (1/30)
Definition: put into a certain place or abstract location
Example: *Put your things here*

Choose file: WNNNEW.txt



Load!

Show definitions: Show examples: (2405) **rektifikovat:n1****Example:** ředitel rektifikoval své rozhodnutí**Example:** dělník rektifikoval měřící přístroj**Example:** kritik rektifikoval utkvělou představu

čtenářů o autorovi

Example: úřad rektifikoval nedopatření(2406) **rektifikovat:n2****Example:** továrna rektifikovala kapalinu redestilací(2407) **rektifikovat:n3****Example:** geometr rektifikoval délku oblouku(2408) **rekultivovat:n1****Example:** zemědělci rekultivovali půdu**Example:** dědeček rekultivoval louku**Example:** odborná firma rekultivovala bývalou skládku(2409) **rekurovat:n1****Example:** poddaní rekurovali proti nařízení**Example:** rekuroval k nejvyššímu soudu(2410) **rekvalifikovat:n1****Example:** školy rekvalifikují uchazeče o zaměstnání(2411) **rekvalifikovat se:n1****Example:** pracovník se rekvalifikoval na novou profesi**SYNSETS FOR REKULTIVOVAT:N1 (5)**

- [ENG20-00373052-v] **reform:2, reclaim:3, regenerate:3, rectify:3** (1/5)

Definition: bring, lead, or force to abandon a wrong or evil course of life, conduct, and adopt a right one**Example:** The Church reformed me**Example:** reform your conduct**No equivalent**

- [ENG20-00372303-v] **reclaim:4** (1/5)

Definition: make useful again; transform from a useless or uncultivated state**Example:** The people reclaimed the marshes**No equivalent**

- [ENG20-00292619-v] **domesticate:2, domesticate:1, domestice:1, reclaim:5, tame:4** (1/5)

Definition: overcome the wildness of; make docile and tractable**Example:** He tames lions for the circus**Example:** reclaim falcons