

# Towards a Corpus-based Valency Lexicon of Czech Nouns

Jana Klímová, Veronika Kolářová, Anna Vernerová

Institute of Formal and Applied Linguistics  
Faculty of Mathematics and Physics  
Charles University in Prague



# Outline

- Starting point
  - Czech valency lexicons Vallex and PDT-Vallex
  - Valency theory of Functional Generative Description
- Goals
  - Corpus-based valency lexicon of Czech nouns: NomVallex
  - Comprehensive lexicographic treatment of noun valency
  - Deeper understanding of issues of noun valency

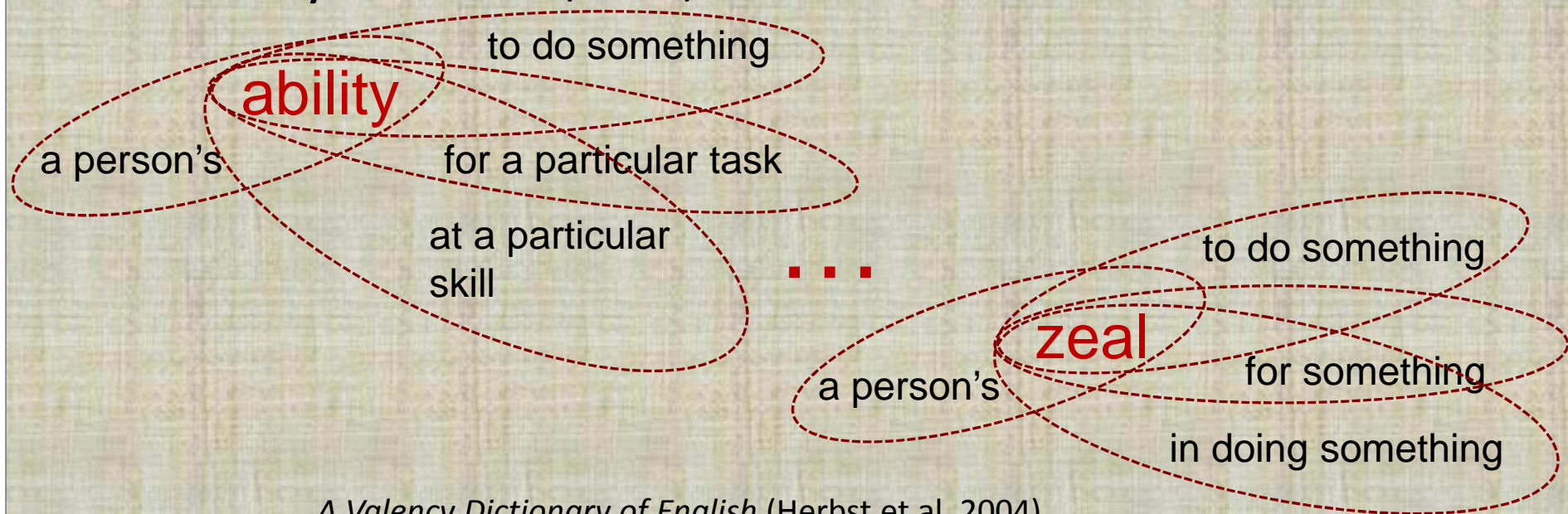


# Valency of nouns

- Valency: the range of syntactic elements either required or specifically permitted by a verb or other lexical unit

*Concise Oxford Dictionary of Linguistics*

- Valency lexicons (a ... z)



*A Valency Dictionary of English* (Herbst et al. 2004)

– complementation patterns of English verbs, nouns and adjectives



# Corpus-based valency lexicon of Czech nouns (NomVallex)

- Supported by the Czech Science Foundation (2016-2018; 16-02196S)
- Team: Veronika Kolářová, Jarmila Panevová, Jana Klímová, Anna Vernerová
- Theoretical framework: valency theory of Functional Generative Description (Panevová)
  - semantic roles
    - participants: Actor (ACT), Patient (PAT), Addressee (ADDR), Effect (EFF), Origin (ORIG)
    - obligatory free modifications: Direction (DIR), Location (LOC), Manner (MANN)
  - valency frames
- Goal: a lexicon available as machine readable data in a format suitable for NLP applications
- Valency of deverbal nouns
- Extension of two valency lexicons, PDT-Vallex and VALLEX



# Vallex, PDT-Vallex, NomVallex

- Vallex
  - valency structure of Czech **verbs** in their particular senses, characterized by glosses and examples (based on data of the Czech National Corpus, CNC)
    - Vallex 3.0: almost 5 000 Czech verbs in more than 10 800 lexical units, i.e., given verbs in the given senses
- PDT-Vallex
  - a lexicon built during the tectogrammatical annotation of the Prague Dependency Treebank (PDT)
  - verbs, **nouns**, adjectives



# Vallex, PDT-Vallex, NomVallex

- Vallex
  - valency structure of Czech **verbs** in their particular senses, characterized by glosses and examples (based on data of the Czech National Corpus, CNC)
    - Vallex 3.0: almost 5 000 Czech verbs in more than 10 800 lexical units, i.e., given verbs in the given senses
- PDT-Vallex
  - a lexicon built during the tectogrammatical annotation of the Prague Dependency Treebank (PDT)
  - verbs, **nouns**, adjectives
- **NomVallex**
  - Extends VALLEX: adding noun entries
  - Extends PDT-Vallex: comprehensive lexicographic treatment



# Slovník slovesných, substantivních a adjektivních vazeb a spojení

(Svozilová, Prouzová, Jirsová 2005)

- Lexicon of verbal, noun and adjectival valency modifications
- So far the only printed lexicon capturing valency of Czech nouns and adjectives
- No semantic roles, no valency frames
- Noun entries provided selectively
- Only so-called right valency

**zpráva** o něčem, o někom; odněkud; pro někoho: *každodenní z-y o teroristických útocích // z. o hrozících přívalových deštích // ...*

**message** about something, about somebody; from somewhere; for somebody: *everyday messages about terrorist attacks // ...*



# Valency of Czech nouns

- Czech: highly inflectional language
- Forms of adnominal participants
  - not only infinite and prepositional phrases but also prepositionless cases (GEN, DAT, INS)
  - typical as well as special shifts in surface forms of participants
- Polyvalent nouns
  - various combinations of participants
- Various factors influencing valency of nouns
  - meaning of the noun
  - word-formation process
  - syntactic homonymy of some adnominal forms
  - semantic class membership
  - being a part of a complex predicate
  - ...



# Valency of Czech nouns

- Czech: highly inflectional language
- Forms of adnominal participants
  - not only infinite and prepositional phrases but also prepositionless cases (GEN, DAT, INS)
  - typical as well as special shifts in surface forms of participants
- Polyvalent nouns
  - various combinations of participants
- Various factors influencing valency of nouns
  - **meaning of the noun**
  - word-formation process
  - syntactic homonymy of some adnominal forms
  - semantic class membership
  - being a part of a complex predicate
  - ...



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)
- Noun: **vyznamenání** ‘honor’
  - derived from the verb *vyznamenat* ‘to honor’

|  |   |   |
|--|---|---|
| Action / state                           | Abstract result of an action                      | Concrete result of an action / physical object / place / person |
| <b>‘the process of honoring someone’</b> | <b>‘honor, award’</b>                             | <b>‘badge, medal, order’</b>                                    |
| Syntactic derivation                     | Boundary between syntactic and lexical derivation | Lexical derivation  |



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)
- Noun: **vyznamenání** ‘honor’
  - derived from the verb *vyznamenat* ‘to honor’

|  |   |   |
|--|---|---|
| Action / state                           | Abstract result of an action                      | Concrete result of an action / physical object / place / person |
| <b>‘the process of honoring someone’</b> | <b>‘honor, award’</b>                             | <b>‘badge, medal, order’</b>                                    |
| Syntactic derivation                     | Boundary between syntactic and lexical derivation | Lexical derivation  |



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)
- Noun: **vyznamenání** ‘honor’
  - derived from the verb *vyznamenat* ‘to honor’

|  |   |   |
|--|---|---|
| Action / state                           | Abstract result of an action                      | Concrete result of an action / physical object / place / person |
| <b>‘the process of honoring someone’</b> | <b>‘honor, award’</b>                             | <b>‘badge, medal, order’</b>                                    |
| Syntactic derivation                     | Boundary between syntactic and lexical derivation | Lexical derivation  |



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)
- Noun: **vyznamenání** ‘honor’
  - derived from the verb *vyznamenat* ‘to honor’

|  |   |   |
|--|---|---|
| Action / state                           | Abstract result of an action                      | Concrete result of an action / physical object / place / person |
| <b>‘the process of honoring someone’</b> | <b>‘honor, award’</b>                             | <b>‘badge, medal, order’</b>                                    |
| Syntactic derivation                     | Boundary between syntactic and lexical derivation | Lexical derivation  |



# Polysemy of deverbal nouns

- What does the noun denote?
- Syntactic vs. lexical derivation (Kuryłowicz 1936)
- Noun: **vyznamenání** ‘honor’
  - derived from the verb *vyznamenat* ‘to honor’

|  |   |   |
|--|---|---|
| Action / state                           | Abstract result of an action                      | Concrete result of an action / physical object / place / person |
| <b>‘the process of honoring someone’</b> | <b>‘honor, award’</b>                             | <b>‘badge, medal, order’</b>                                    |
| Syntactic derivation                     | Boundary between syntactic and lexical derivation | Lexical derivation  |

- Different valency behavior in different meanings?



# Typical vs. special valency behavior of nouns

- Typical valency behavior (Panevová 2000, Karlík 2000)
  - Nouns denote an action or a state as their source verbs do
  - Nouns inherit the valency frame from their source verbs
  - Their forms undergo typical shifts
    - e.g. ACC → GEN, *varování cestujících.ADDR* ‘warning passengers-GEN’
- Special valency behavior (Kolářová 2010, Kolářová 2014)
  - Tied with changes in meaning (e.g., an action → a figurative sense)
  - Characteristic properties of valency complementation
    - special forms (corresponding to special shifts)
      - e.g. ACC → DAT, *varování cestujícím.ADDR* ‘warning passengers-DAT’
    - reduction of the number of slots in the valency frame
      - pure reduction
      - incorporation of a participant (e.g. actor nouns incorporate Actor)
    - change of the character of a valency complementation to exclusively nominal
      - e.g. Patiens → Material (MAT) / Partitive
      - *balení másla.PAT do obalu brigádníky.ACT* → *jedno balení másla.MAT* (“container”)  
‘packing butter.PAT into the wrap by volunteers .ACT’ → ‘one pack of butter.MAT’



# Nouns in PDT-Vallex

- Valency lexicon built during the annotation of the PDT:  
<https://ufal.mff.cuni.cz/pdt2.0/visual-data/pdt-vallex/vallex.html>
- PDT-Vallex in PDT2.0 (2006): 3727 noun entries
  - Alevtina Bémová; nominal parts of complex predicates: Veronika Kolářová
- Only nouns and their senses (including corresponding valency frames) that occurred in the PDT data

## \* balení ‘pack(ing)’

ACT(.2,.7,.u) PAT (.2,.u) ?EFF(na-1[.4],v-1[.4],do-1[.2]) v-w112f1 Used: 3x  
(balit ‘to pack’) *balení másla PAT na jednotlivé porce.EFF*  
‘packing butter.PAT to particular portions.EFF’

?MAT (.2) v-w112f2 Used: 2x  
*jedno balení másla.MAT*  
‘one pack of butter.MAT’

EMPTY v-w112f3 Used: 2x  
(způsob provedení) *kniha v brožurkovém balení*  
(design) ‘a book in a paperback binding’



# Nouns in PDT-Vallex

- Valency lexicon built during the annotation of the PDT:  
<https://ufal.mff.cuni.cz/pdt2.0/visual-data/pdt-vallex/vallex.html>
- PDT-Vallex in PDT2.0 (2006): 3727 noun entries
  - Alevtina Bémová; nominal parts of complex predicates: Veronika Kolářová
- Only nouns and their senses (including corresponding valency frames) that occurred in the PDT data

Syntactic derivation, inherited valency frame, typical valency behavior

## \* **balení** ‘pack(ing)’

ACT(.2,.7,.u) PAT (.2,.u) ?EFF(na-1[.4],v-1[.4],do-1[.2]) v-w112f1 Used: 3x

(balit ‘to pack’) *balení másla PAT na jednotlivé porce.EFF*

‘packing butter.PAT to particular portions.EFF’

?MAT (.2) v-w112f2 Used: 2x

*jedno balení másla.MAT*

‘one pack of butter.MAT’

EMPTY v-w112f3 Used: 2x

(způsob provedení) *kniha v brožurkovém balení*

(design) ‘a book in a paperback binding’



# Nouns in PDT-Vallex

- Valency lexicon built during the annotation of the PDT:  
<https://ufal.mff.cuni.cz/pdt2.0/visual-data/pdt-vallex/vallex.html>
- PDT-Vallex in PDT2.0 (2006): 3727 noun entries
  - Alevtina Bémová; nominal parts of complex predicates: Veronika Kolářová
- Only nouns and their senses (including corresponding valency frames) that occurred in the PDT data

Syntactic derivation, inherited valency frame, typical valency behavior

## \* **balení 'pack(ing)'**

ACT(.2,.7,.u) PAT (.2,.u) ?EFF(na-1[.4],v-1[.4],do-1[.2]) v-w112f1 Used: 3x

(balit 'to pack') *balení másla PAT na jednotlivé porce.EFF*

'packing butter.PAT to particular portions.EFF'

?MAT (.2) v-w112f2 Used: 2x

*jedno balení másla.MAT*

'one pack of butter.MAT'

Lexical derivation, "container",

special valency behavior:

Reduction of the number of slots, MAT participant

EMPTY v-w112f3 Used: 2x

(způsob provedení) *kniha v brožurkovém balení*

(design) 'a book in a paperback binding'



# Nouns in PDT-Vallex

- Valency lexicon built during the annotation of the PDT:  
<https://ufal.mff.cuni.cz/pdt2.0/visual-data/pdt-vallex/vallex.html>
- PDT-Vallex in PDT2.0 (2006): 3727 noun entries
  - Alevtina Bémová; nominal parts of complex predicates: Veronika Kolářová
- Only nouns and their senses (including corresponding valency frames) that occurred in the PDT data

Syntactic derivation, inherited valency frame, typical valency behavior

## \* balení 'pack(ing)'

ACT(.2,.7,.u) PAT (.2,.u) ?EFF(na-1[.4],v-1[.4],do-1[.2]) v-w112f1 Used: 3x

(balit 'to pack') *balení másla PAT na jednotlivé porce.EFF*

'packing butter.PAT to particular portions.EFF'

?MAT (.2) v-w112f2 Used: 2x

*jedno balení másla.MAT*

'one pack of butter.MAT'

Lexical derivation, "container",  
special valency behavior:  
Reduction of the number of slots, MAT participant

EMPTY v-w112f3 Used: 2x

(způsob provedení) *kniha v brožurkovém balení*

(design) 'a book in a paperback binding'

Lexical derivation, EMPTY valency frame



# NomVallex: Corpus-based valency lexicon of Czech nouns

- Focus on valency of deverbal nouns
  - especially nouns that in one of their meanings can denote an action
    - nouns derived by productive means, suffixes –(e)ní/tí
      - e.g. honění ‘hunting’
    - nouns derived by non-productive means or by the zero suffix
      - e.g. honba ‘hunt’, hon ‘hunt’
- Exploitation and extension of PDT-Vallex
- Conception of lexicographic treatment: closer to Vallex
  - more corpus data
  - semantic class membership
  - all documented senses of treated nouns



# Corpus data

- Not only PDT data (as in PDT-Vallex) but also SYN family of the CNC and Araneum Bohemicum corpus
  - More data ⇒ more information
  - Verification of usage of some adnominal forms stated in PDT-Vallex
    - especially the instrumental form of Actor modifying non-productively derived nouns, e.g. *?četba kým.ACT* ‘reading by somebody’
    - *Dnes je na programu festivalu ... četba liturgické poezie.PAT rabínem.ACT Karolem Sidonem.* (SYN2009PUB)
    - ‘Reading of liturgical poetry.PAT by rabbi.ACT Karol Sidon is on the today’s program of the festival.’



# Manual searching through the CNC

- Valency of nouns → Searching for syntactic relations
  - Query: a linear notation of dependency relations, CQL query type
  - A noun modified by a participant in prepositionless DAT:
  - `([lemma="pomoc"] [!(tag="[Z|R|V|J].*")]{0,2} [tag="N...3.*"])`

Optional positions

We exclude punctuation, prepositions, verbs and conjunctions

A noun in DAT

- A query can often cover various dependency relations that do not match our intention ⇒ All found occurrences have to be manually checked and evaluated
  - precise but VERY time-consuming
- A recent improvement: an automatic syntactic analysis of the CNC subcorpus SYN2015
  - `[tag="N...3.*" & p_lemma="pomoc"]`

Lemma of the noun's parent



# Sketch Engine: Word Sketch

- One-page, automatic, corpus-derived summary of a word's grammatical and collocational behavior
- Collocates are grouped according to grammatical relations in which they occur
  - Word Sketch provides a simple grammar - a collection of definitions that allow the system to automatically identify possible relations of words to the keyword
- Czech Word Sketch
  - Good: adnominal prepositional phrases and prepositionless genitive
  - So far not covered: e.g. adnominal prepositionless dative and instrumental, combinations of participants



# Sketch Engine

**POKUS** syn2009pub frekvence = 84,385 (99.9 v milionu)

**pokus 'attempt'**

| coord       | <u>1,923</u> | 0.6  | <u>post inf</u> | <u>8,103</u> | 2.3  | prec verb    | <u>2,025</u> | 0.9  | post verb    | <u>1,238</u> | 0.5  | a modifier   | <u>28,681</u> | 1.5  |
|-------------|--------------|------|-----------------|--------------|------|--------------|--------------|------|--------------|--------------|------|--------------|---------------|------|
| vražd       |              | 5    | získat          | <u>199</u>   | 3.4  | vyšetřovat   | <u>60</u>    | 5.11 | dojit        | <u>70</u>    | 3.5  | úspěšný      | <u>2,318</u>  | 8.89 |
| omyl        |              | 8    | dostat          | <u>192</u>   | 3.05 | čekat        | <u>60</u>    | 1.74 | stát         | <u>45</u>    | 0.22 | další        | <u>2,217</u>  | 5.96 |
| výtržnictví | <u>184</u>   | 9.11 | vytvořit        | <u>125</u>   | 4.55 | představovat | <u>59</u>    | 3.13 | končit       | <u>34</u>    | 3.02 | střelecký    | <u>1,737</u>  | 9.87 |
| krádež      | <u>61</u>    | 4.72 | prosadit        | <u>114</u>   | 5.38 | dělat        | <u>45</u>    | 2.15 | vyjít        | <u>27</u>    | 2.81 | poslední     | <u>1,392</u>  | 6.25 |
| znásilnění  | <u>48</u>    | 7.36 | najít           | <u>83</u>    | 2.86 | šetřit       | <u>42</u>    | 4.44 | lze          | <u>21</u>    | 2.02 | trojkový     | <u>652</u>    | 9.48 |
| loupež      | <u>48</u>    | 6.16 | prodat          | <u>63</u>    | 4.2  | stát         | <u>41</u>    | 0.08 | podařit      | <u>20</u>    | 3.15 | marný        | <u>542</u>    | 8.98 |
| podvod      | <u>41</u>    | 5.11 | vrátit          | <u>60</u>    | 3.07 | provádět     | <u>37</u>    | 3.93 | skončit      | <u>16</u>    | 0.7  | oživovací    | <u>449</u>    | 8.97 |
| svoboda     | <u>40</u>    | 3.53 | omezit          | <u>58</u>    | 4.88 | připomínat   | <u>31</u>    | 2.88 | ukazovat     | <u>13</u>    | 2.4  | podobný      | <u>395</u>    | 5.93 |
| věc         | <u>38</u>    | 1.94 | změnit          | <u>58</u>    | 3.51 | probíhat     | <u>29</u>    | 2.85 | provádět     | <u>12</u>    | 2.32 | sebevražedný | <u>354</u>    | 8.39 |
| vydírání    | <u>23</u>    | 6.09 | proniknout      | <u>57</u>    | 6.14 | existovat    | <u>29</u>    | 2.39 | vycházet     | <u>12</u>    | 1.95 | jediný       | <u>321</u>    | 5.35 |
| sebevražda  | <u>21</u>    | 5.22 | zachránit       | <u>57</u>    | 4.8  | uskutečnit   | <u>29</u>    | 2.22 | probíhat     | <u>12</u>    | 1.59 | předchozí    | <u>294</u>    | 6.53 |
| ozbrojování | <u>19</u>    | 6.86 | zavést          | <u>56</u>    | 4.86 | znamenat     | <u>28</u>    | 2.25 | selhat       | <u>11</u>    | 5.3  | zoufalý      | <u>291</u>    | 7.96 |
| zdraví      | <u>16</u>    | 2.62 | ovlivnit        | <u>56</u>    | 4.57 | odmítnat     | <u>27</u>    | 3.5  | pověst       | <u>10</u>    | 1.84 | nezdařený    | <u>287</u>    | 8.3  |
| útok        | <u>14</u>    | 1.16 | oživit          | <u>54</u>    | 5.72 | objevovat    | <u>26</u>    | 3.67 | představovat | <u>10</u>    | 0.58 | tříbodový    | <u>279</u>    | 7.99 |
| napadení    | <u>12</u>    | 4.65 | založit         | <u>54</u>    | 4.43 | vnímat       | <u>25</u>    | 3.99 | zopakovat    | <u>9</u>     | 3.55 | nový         | <u>276</u>    | 2.64 |
| vloupání    | <u>12</u>    | 4.22 | překonat        | <u>52</u>    | 4.75 | sledovat     | <u>25</u>    | 2.49 | vyplývat     | <u>9</u>     | 3.35 | dlouhý       | <u>272</u>    | 5.12 |
| jednotlivec | <u>12</u>    | 3.75 | obnovit         | <u>51</u>    | 5.08 | vidět        | <u>25</u>    | 1.37 | uskutečnit   | <u>9</u>     | 0.54 | platný       | <u>257</u>    | 7.1  |
| velezrada   | <u>11</u>    | 6.97 | zbavit          | <u>51</u>    | 4.97 | učinit       | <u>23</u>    | 4.22 | považovat    | <u>9</u>     | 0.34 | opakovaný    | <u>230</u>    | 7.47 |
| únos        | <u>11</u>    | 4.97 | zastavit        | <u>50</u>    | 3.95 | popisovat    | <u>20</u>    | 3.17 | opakovat     | <u>8</u>     | 2.78 | bezpečný     | <u>230</u>    | 5.94 |
| přepadení   | <u>10</u>    | 3.77 | propašovat      | <u>48</u>    | 7.22 | chápat       | <u>19</u>    | 4.02 | absolvovat   | <u>8</u>     | 1.62 | trestný      | <u>198</u>    | 5.19 |
| experiment  | <u>9</u>     | 4.46 | udělat          | <u>48</u>    | 2.75 | hodnotit     | <u>19</u>    | 2.82 | hodlat       | <u>8</u>     | 1.44 | vyšetřovací  | <u>195</u>    | 7.41 |
| výzkum      | <u>9</u>     | 2.2  | zabránit        | <u>46</u>    | 4.33 | činit        | <u>18</u>    | 1.96 | krýt         | <u>7</u>     | 3.67 | údajný       | <u>194</u>    | 7.01 |
| snaha       | <u>9</u>     | 1.73 | dát             | <u>44</u>    | 1.66 | považovat    | <u>18</u>    | 1.34 | proběhnout   | <u>7</u>     | 1.22 | vydařený     | <u>179</u>    | 7.05 |
| poškození   | <u>8</u>     | 3.83 | sestavit        | <u>43</u>    | 5.43 | začínat      | <u>18</u>    | 1.0  | vypadat      | <u>7</u>     | 1.09 | přesný       | <u>179</u>    | 6.05 |
| zneužívání  | <u>8</u>     | 3.69 | postavit        | <u>43</u>    | 2.91 | podnikat     | <u>17</u>    | 4.37 | přicházet    | <u>7</u>     | 1.06 | dosavadní    | <u>162</u>    | 5.72 |

valency



## valency

| prec  | prep | 13,635       | 1.2  | post       | z | 718        | 0.8  | post       | o | 24,423       | 45.5 | post        | na | 1,244      | 0.7  | post        | v | 1,080     | 0.4  |
|-------|------|--------------|------|------------|---|------------|------|------------|---|--------------|------|-------------|----|------------|------|-------------|---|-----------|------|
| z     |      | <u>3,188</u> | 4.5  | dálka      |   | <u>102</u> | 7.56 | vražda     |   | <u>2,872</u> | 10.1 | zvíře       |    | <u>287</u> | 6.53 | soutěž      |   | <u>33</u> | 0.57 |
| při   |      | <u>2,779</u> | 6.66 | vzdálenost |   | <u>70</u>  | 6.0  | sebevražda |   | <u>883</u>   | 9.46 | člověk      |    | <u>107</u> | 1.57 | řada        |   | <u>25</u> | 0.87 |
| za    |      | <u>2,405</u> | 5.32 | pole       |   | <u>61</u>  | 4.45 | atentát    |   | <u>700</u>   | 8.91 | výška       |    | <u>93</u>  | 6.01 | nadhoz      |   | <u>23</u> | 8.52 |
| o     |      | <u>1,012</u> | 3.17 | úhel       |   | <u>45</u>  | 6.51 | ublížení   |   | <u>631</u>   | 8.6  | myš         |    | <u>69</u>  | 7.91 | dálka       |   | <u>20</u> | 5.18 |
| k     |      | <u>852</u>   | 3.39 | kop        |   | <u>45</u>  | 3.69 | rekord     |   | <u>585</u>   | 7.25 | roh         |    | <u>25</u>  | 4.29 | trh         |   | <u>20</u> | 0.81 |
| na    |      | <u>700</u>   | 1.0  | hranice    |   | <u>32</u>  | 2.42 | krádež     |   | <u>532</u>   | 7.47 | opice       |    | <u>19</u>  | 6.48 | hod         |   | <u>15</u> | 4.19 |
| s     |      | <u>574</u>   | 2.05 | otočka     |   | <u>17</u>  | 7.57 | znásilnění |   | <u>497</u>   | 8.99 | ukončení    |    | <u>13</u>  | 3.36 | laboratoř   |   | <u>14</u> | 4.25 |
| pro   |      | <u>466</u>   | 3.01 | vápno      |   | <u>17</u>  | 4.31 | podvod     |   | <u>440</u>   | 7.87 | mistrovství |    | <u>13</u>  | 0.49 | finále      |   | <u>14</u> | 1.9  |
| po    |      | <u>282</u>   | 2.31 | blížkost   |   | <u>15</u>  | 3.32 | útěk       |   | <u>432</u>   | 8.32 | zajatec     |    | <u>10</u>  | 5.43 | disk        |   | <u>13</u> | 4.77 |
| proti |      | <u>165</u>   | 3.65 | čára       |   | <u>12</u>  | 3.14 | překonání  |   | <u>309</u>   | 8.39 | dobrovolník |    | <u>10</u>  | 4.07 | závěr       |   | <u>13</u> | 1.52 |
| před  |      | <u>142</u>   | 2.3  | pozice     |   | <u>12</u>  | 1.6  | převrat    |   | <u>295</u>   | 8.26 | branka      |    | <u>10</u>  | 0.35 | kvalifikace |   | <u>11</u> | 2.63 |
| kvůli |      | <u>134</u>   | 4.02 | polovina   |   | <u>10</u>  | 0.38 | vytvoření  |   | <u>258</u>   | 7.42 | chvíle      |    | <u>9</u>   | 1.2  | oblast      |   | <u>11</u> | 0.25 |
| od    |      | <u>80</u>    | 0.87 | volej      |   | <u>9</u>   | 5.82 | vloupání   |   | <u>256</u>   | 7.59 | hřiště      |    | <u>9</u>   | 0.17 | historie    |   | <u>10</u> | 0.96 |
| během |      | <u>77</u>    | 3.77 | nájezd     |   | <u>8</u>   | 3.24 | záchrana   |   | <u>225</u>   | 6.53 | potkan      |    | <u>8</u>   | 6.12 | útok        |   | <u>10</u> | 0.68 |
| u     |      | <u>71</u>    | 1.31 | počátek    |   | <u>8</u>   | 1.5  | návrat     |   | <u>220</u>   | 5.96 | sestavení   |    | <u>8</u>   | 5.15 | sezona      |   | <u>10</u> | 0.31 |



# Sketch Engine

## Valency (PAT vs. ACT)

|                           |                        |                        |                             |                           |
|---------------------------|------------------------|------------------------|-----------------------------|---------------------------|
| <b>prec proti</b> 154 2.1 | <b>gen 1</b> 7,908 1.2 | <b>gen 2</b> 3,400 0.5 | <b>is subj of</b> 5,558 1.2 | <b>is obj2 of</b> 258 1.5 |
| protestovat 18 4.64       | vražda 1,278 9.24      | čin 507 6.14           | vyjít 277 6.12              | dopustit 46 5.62          |
| protest 13 3.46           | krádež 449 7.49        | stadium 306 9.53       | skončit 240 4.59            | uniknout 24 4.93          |
| zasahovat 7 2.56          | ubližení 346 8.27      | případ 195 4.07        | získat 68 1.86              | zúčastnit 22 2.86         |
|                           | host 163 4.17          | stádium 124 9.38       | jit 66 1.8                  | vzdát 19 4.27             |
| <b>prec před</b> 139 1.1  | soupeř 110 3.85        | výsledek 117 3.31      | ztroskotat 64 7.61          | týkat 14 1.77             |
| varovat 34 4.25           | znásilnění 103 7.76    | počet 117 3.19         | učinit 63 5.56              | přibývat 12 3.34          |
|                           | domáci 97 2.61         | řada 102 2.89          | ukázat 61 4.25              | zanechat 10 4.33          |
| <b>prec kolem</b> 22 1.0  | vláda 71 2.43          | účastník 91 4.48       | dostat 58 1.33              | účastnit 10 3.09          |
| aféra 11 3.6              | vloupání 57 6.11       | metoda 84 5.65         | selhat 57 7.12              | zabránit 10 2.29          |
|                           | sebevražda 52 6.12     | většina 61 2.86        | stačit 57 3.79              |                           |
| <b>prec kvůli</b> 112 3.2 | podvod 52 5.24         | svědek 53 4.9          | uskutečnit 53 3.07          | <b>is obj3 of</b> 378 1.2 |
| vyšetřovat 9 2.42         | den 32 0.42            | série 52 3.91          | zdařit 49 7.81              | čelit 95 5.41             |
|                           | poškození 31 5.38      | průběh 41 3.21         | stát 49 0.34                | zabránit 67 5.03          |
| <b>byt adj</b> 400 1.5    | loupež 31 5.18         | cesta 37 1.43          | vytvořit 48 3.19            | odolávat 18 6.2           |
| úspěšný 98 4.6            | Jan 28 1.9             | základ 27 2.63         | čekat 45 1.32               | odolat 15 5.64            |
| marný 73 8.44             | vydírání 26 5.64       | fáze 26 3.9            | dopadnout 41 4.4            | bránit 15 2.37            |
| přesný 9 2.29             | ministr 25 1.23        | zákaz 26 3.5           | minout 40 5.85              | věnovat 13 0.97           |
| platný 8 2.98             | opozice 23 4.12        | desítka 25 2.11        | probíhat 40 3.28            | vyhnout 11 3.23           |
| nutný 8 2.41              | skupina 23 0.51        | doba 25 0.02           | miřit 38 3.97               | uniknout 10 3.66          |
| nadějný 7 4.01            | lékař 22 1.87          | spáchání 23 5.61       | vypadat 38 3.47             | vzdorovat 9 5.44          |
| jasný 7 1.68              | ministerstvo 22 0.97   | rámec 23 2.08          | končit 38 3.14              | přihlížet 9 4.59          |
| bezpečný 7 1.27           | hráč 22 0.46           | stovka 22 2.17         | prosadit 37 3.8             | zamezit 8 4.4             |
|                           | pan 20 2.91            | pachatel 21 2.46       | pověst 37 3.66              |                           |
|                           | útočník 20 2.03        | cíl 21 2.04            | provést 37 3.66             |                           |
|                           | Petr 20 1.66           | důvod 21 1.66          | vyrazit 32 3.66             |                           |



# Semantic classes in NomVallex

- Providing information about semantic class membership
  - in contrast to PDT-Vallex
- Following semantic classification applied in VALLEX
- Focusing on selected semantic classes
  - communication, exchange, contact, psychological nouns, mental action
- Consistency across a whole semantic class

## \* **zmínka** ‘mention, notice, remark’

## **PDT-Vallex entry**

ACT(.2,.u) PAT(o-1[.6]) ?EFF(že[v],.s) v-w9626f1 Used: 1x

Svobodova.ACT zmínka o tom, že nelze.PAT zbohatnout ‘Svoboda’s remark about...’

zmínka o rozdělení.PAT unie ‘remark on the Union division’

jediná samostatná zmínka o České republice.PAT ‘the only remark about the CzRep’

Noun of communication → missing **ADDR(.3)** should be added

zmínka kamarádovi.ADDR Jiřímu Ulrychovi ... že je.EFF delegován k ... (SYN2013PUB)

‘remark to a friend.ADDR Jiří Ulrych ... that he is.EFF delegated to...’



# Polysemy treatment in NomVallex

- PDT-Vallex: only the senses of nouns encountered during annotation of the PDT
- NomVallex:
  - valency patterns for all senses of the treated nouns, as documented in the CNC and monolingual dictionaries
  - focusing on senses with valency potential

## \* **mluvení** ‘talking, speaking’

**PDT-Vallex entry**

ACT(.2,.u) v-w1867f1 **Used: 1x**

*(umět mluvit) mluvení cizím jazykem.MANN*

(to be able to speak) ‘speaking a foreign language’

Missing:

ACT(.2,.u) ADDR(s+7) PAT(o+6)

*process of talking with somebody about something*

- *její.ACT mluvení o knihách.PAT (SYN2010)* ‘her.ACT talking about books.PAT’

- *mluvení s novináři.ADDR (SYN2009PUB)* ‘talking to/with journalists.ADDR’



# Additional information in NomVallex

- All documented combinations of participants in various forms, including word order variants
  - especially trivalent nouns constitute rather complex patterns
- Information about the type of special valency behavior
  - the form corresponding to special shifts, reduction of the number of slots in the valency frame, the special nominal participant MAT
- Frequency and stylistic evaluation of a pattern
  - scale: very frequent – frequent – rare
  - grammatical but rare
    - e.g. double post-nominal genitives
  - legal language
    - e.g. Patient in prepositionless genitive modifying the noun *pokus* 'attempt'
      - *pokus vraždy*.PAT, lit. attempt murder-GEN.SG, 'attempted murder'
- Links to the PDT-Vallex and Vallex entries of the verb from which the noun is derived and to other nouns derived from the same verb



# Nominal Entry Example: *vyznamenání* 'honor'

Semantic class: evaluation

Source verb: *vyznamenat*<sup>pf</sup> 'to honor', Nouns: *vyznamenávání*;

## 1. proces vyznamenání někoho 'the process of honoring someone'

Frame: ACT(.2,.7,.u) PAT(.2,.u)

Example: *vyznamenání veterána.PAT premiérem*.ACT 'honoring of the veteran by the prime minister'; *vyznamenání premiérem*.ACT 'honoring by the prime minister'; *premiérovo*.ACT *vyznamenání veterána*.PAT 'the prime minister's honoring of the veteran'; *vyznamenání veterána*.PAT 'honoring of the veteran'; *?premiérovo*.ACT *vyznamenání* 'the prime minister's honoring'; *veteránovo*.PAT *vyznamenání premiérem*.ACT 'the veteran's honoring by the prime minister'; *veteránovo*.PAT *vyznamenání* 'the veteran's honoring'

## 2. pocta, vyznamenání udělené někomu 'honor, award'

Frame: ACT(.2,.u) PAT(.3)

The type of special valency behavior: special form of PAT (special shift ACC → DAT)

Example: *vyznamenání veteránovi*.PAT 'honor addressed to the veteran'; *?premiérovo*.ACT *vyznamenání veteránovi*.PAT 'the prime minister's honor addressed to the veteran'; *?vyznamenání premiéra*.ACT *veteránovi*.PAT 'honor of the prime minister addressed to the veteran'

## 3. odznak, medaile, řád 'badge, medal, order'

Frame: EMPTY

Example: *ověnčený vyznamenáními* 'decked with medals'

## 4. nejvyšší stupeň celkového prospěchu 'honors'

Frame: EMPTY

Example: *studovat s vyznamenáním* 'to study with honors'



# Nominal Entry Example: *vyznamenání* 'honor'

Semantic class: evaluation

Source verb: *vyznamenat*<sup>pf</sup> 'to honor', Nouns: *vyznamenávání*;

## 1. proces vyznamenání někoho 'the process of honoring someone'

Frame: ACT(.2,.7,.u) PAT(.2,.u) Syntactic derivation, inherited valency frame, typical valency behavior

Example: *vyznamenání veterána.PAT premiérem*.ACT 'honoring of the veteran by the prime minister'; *vyznamenání premiérem*.ACT 'honoring by the prime minister'; *premiérovo*.ACT *vyznamenání veterána*.PAT 'the prime minister's honoring of the veteran'; *vyznamenání veterána*.PAT 'honoring of the veteran'; *?premiérovo*.ACT *vyznamenání* 'the prime minister's honoring'; *veteránovo*.PAT *vyznamenání premiérem*.ACT 'the veteran's honoring by the prime minister'; *veteránovo*.PAT *vyznamenání* 'the veteran's honoring'

## 2. pocta, vyznamenání udělené někomu 'honor, award'

Frame: ACT(.2,.u) PAT(.3)

The type of special valency behavior: special form of PAT (special shift ACC → DAT)

Example: *vyznamenání veteránovi*.PAT 'honor addressed to the veteran'; *?premiérovo*.ACT *vyznamenání veteránovi*.PAT 'the prime minister's honor addressed to the veteran'; *?vyznamenání premiéra*.ACT *veteránovi*.PAT 'honor of the prime minister addressed to the veteran'

## 3. odznak, medaile, řád 'badge, medal, order'

Frame: EMPTY

Example: *ověnčený vyznamenáními* 'decked with medals'

## 4. nejvyšší stupeň celkového prospěchu 'honors'

Frame: EMPTY

Example: *studovat s vyznamenáním* 'to study with honors'



# Nominal Entry Example: *vyznamenání* 'honor'

Semantic class: evaluation

Source verb: *vyznamenat*<sup>pf</sup> 'to honor', Nouns: *vyznamenávání*;

## 1. proces vyznamenání někoho 'the process of honoring someone'

Frame: ACT(.2,.7,.u) PAT(.2,.u) Syntactic derivation, inherited valency frame, typical valency behavior

Example: *vyznamenání veterána.PAT premiérem*.ACT 'honoring of the veteran by the prime minister'; *vyznamenání premiérem*.ACT 'honoring by the prime minister'; *premiérovo*.ACT *vyznamenání veterána*.PAT 'the prime minister's honoring of the veteran'; *vyznamenání veterána*.PAT 'honoring of the veteran'; *?premiérovo*.ACT *vyznamenání* 'the prime minister's honoring'; *veteránovo*.PAT *vyznamenání premiérem*.ACT 'the veteran's honoring by the prime minister'; *veteránovo*.PAT *vyznamenání* 'the veteran's honoring'

## 2. pocta, vyznamenání udělené někomu 'honor, award'

Frame: ACT(.2,.u) PAT(.3)

Boundary between syntactic and lexical derivation, special valency behavior: special form of PAT

The type of special valency behavior: special form of PAT (special shift ACC → DAT)

Example: *vyznamenání veteránovi*.PAT 'honor addressed to the veteran'; *?premiérovo*.ACT *vyznamenání veteránovi*.PAT 'the prime minister's honor addressed to the veteran'; *?vyznamenání premiéra*.ACT *veteránovi*.PAT 'honor of the prime minister addressed to the veteran'

## 3. odznak, medaile, řád 'badge, medal, order'

Frame: EMPTY

Example: *ověnčený vyznamenáními* 'decked with medals'

## 4. nejvyšší stupeň celkového prospěchu 'honors'

Frame: EMPTY

Example: *studovat s vyznamenáním* 'to study with honors'



# Nominal Entry Example: *vyznamenání* 'honor'

Semantic class: evaluation

Source verb: *vyznamenat*<sup>pf</sup> 'to honor', Nouns: *vyznamenávání*;

## 1. proces vyznamenání někoho 'the process of honoring someone'

Frame: ACT(.2,.7,.u) PAT(.2,.u) Syntactic derivation, inherited valency frame, typical valency behavior

Example: *vyznamenání veterána.PAT premiérem*.ACT 'honoring of the veteran by the prime minister'; *vyznamenání premiérem*.ACT 'honoring by the prime minister'; *premiérovo*.ACT *vyznamenání veterána*.PAT 'the prime minister's honoring of the veteran'; *vyznamenání veterána*.PAT 'honoring of the veteran'; *?premiérovo*.ACT *vyznamenání* 'the prime minister's honoring'; *veteránovo*.PAT *vyznamenání premiérem*.ACT 'the veteran's honoring by the prime minister'; *veteránovo*.PAT *vyznamenání* 'the veteran's honoring'

## 2. pocta, vyznamenání udělené někomu 'honor, award'

Frame: ACT(.2,.u) PAT(.3)

Boundary between syntactic and lexical derivation, special valency behavior: special form of PAT

The type of special valency behavior: special form of PAT (special shift ACC → DAT)

Example: *vyznamenání veteránovi*.PAT 'honor addressed to the veteran'; *?premiérovo*.ACT *vyznamenání veteránovi*.PAT 'the prime minister's honor addressed to the veteran'; *?vyznamenání premiéra*.ACT *veteránovi*.PAT 'honor of the prime minister addressed to the veteran'

## 3. odznak, medaile, řád 'badge, medal, order'

Frame: EMPTY

Lexical derivation, EMPTY valency frame

Example: *ověnčený vyznamenáními* 'decked with medals'

## 4. nejvyšší stupeň celkového prospěchu 'honors'

Frame: EMPTY

Example: *studovat s vyznamenáním* 'to study with honors'



# Nominal Entry Example: *vyznamenání* 'honor'

Semantic class: evaluation

Source verb: *vyznamenat*<sup>pf</sup> 'to honor', Nouns: *vyznamenávání*;

## 1. proces vyznamenání někoho 'the process of honoring someone'

Frame: ACT(.2,.7,.u) PAT(.2,.u) Syntactic derivation, inherited valency frame, typical valency behavior

Example: *vyznamenání veterána.PAT premiérem*.ACT 'honoring of the veteran by the prime minister'; *vyznamenání premiérem*.ACT 'honoring by the prime minister'; *premiérovo*.ACT *vyznamenání veterána*.PAT 'the prime minister's honoring of the veteran'; *vyznamenání veterána*.PAT 'honoring of the veteran'; *?premiérovo*.ACT *vyznamenání* 'the prime minister's honoring'; *veteránovo*.PAT *vyznamenání premiérem*.ACT 'the veteran's honoring by the prime minister'; *veteránovo*.PAT *vyznamenání* 'the veteran's honoring'

## 2. pocta, vyznamenání udělené někomu 'honor, award'

Frame: ACT(.2,.u) PAT(.3)

Boundary between syntactic and lexical derivation, special valency behavior: special form of PAT

The type of special valency behavior: special form of PAT (special shift ACC → DAT)

Example: *vyznamenání veteránovi*.PAT 'honor addressed to the veteran'; *?premiérovo*.ACT *vyznamenání veteránovi*.PAT 'the prime minister's honor addressed to the veteran'; *?vyznamenání premiéra*.ACT *veteránovi*.PAT 'honor of the prime minister addressed to the veteran'

## 3. odznak, medaile, řád 'badge, medal, order'

Frame: EMPTY

Lexical derivation, EMPTY valency frame

Example: *ověnčený vyznamenáními* 'decked with medals'

## 4. nejvyšší stupeň celkového prospěchu 'honors'

Frame: EMPTY

Lexical derivation, EMPTY valency frame

Example: *studovat s vyznamenáním* 'to study with honors'



# Current work

- Preliminary list of entries
  - starting point
    - source verbs contained in Vallex
    - comparison of the list of derived nouns with entries contained in PDT-Vallex
    - adding missing nouns (lemmas)
      - *hovor* ‘talk’, *komunikování* ‘communicating’...
  - initial phase: more than 1000 entries
  - resulting lexicon: 400-500 representative entries
  - nouns of communication, contact, exchange and psychological nouns
- Adjustment of Word Sketch grammar to specifically Czech adnominal forms
  - adnominal prepositionless dative and instrumental
  - adnominal expression of a proposition
  - “left” valency (possessive adjectives and pronouns)
  - combinations of participants



# Future work

- Identification of all documented meanings of the treated nouns
  - corpus data, monolingual dictionaries, valency lexicons
- Searching for noun valency through the CNC and the Araneum Bohemicum corpus
- Finding a solution to - so far - unanswered questions
  - idioms' treatment
    - *mluvení do duše* 'giving somebody a talking-to'
  - nominalizations of complex predicates
    - *dávání přednosti* 'giving way', *ztráta kontroly* 'loss of control'
- More semantic classes
- New theoretical findings in the field of noun valency



# Conclusion

- NomVallex: starting project building upon solid theoretical foundations
- Exploitation and extension of valency lexicons PDT-Vallex and Vallex
- Valency of nouns based on corpus data
  - PDT, CNC, Araneum Bohemicum
- Comprehensive description of valency patterns of Czech deverbal nouns
  - valency frames
  - semantic class membership, sense identification, additional information (frequency, special valency behavior etc.)



# References

- Čermáková, A. (2009). *Valence českých substantiv*. Praha: Lidové noviny. ISBN 978-80-7106-426-800.
- Herbst, T. et al. (2004): *A valency dictionary of English. A corpus-based analysis of the complementation patterns of English verbs, nouns, and adjectives*. Berlin: Walter de Gruyter.
- Kolářová, V. (2010). *Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valencí)*. Praha: Karolinum. ISBN 978-80-246-1828-9.
- Kolářová, V. (2014). Special valency behavior of Czech deverbal nouns. In O. Spevak (Ed.) *Noun Valency*, Amsterdam: John Benjamins, pp. 19--60. ISBN 9789027259233.
- Kuryłowicz, J. (1936). Dérivation lexicale et dérivation syntaxique. *Bulletin de la Société de Linguistique de Paris*. 1936, 37, pp. 79--92.
- Lopatková, M.; Kettnerová, V.; Bejček, E.; Vernerová, A. and Žabokrtský, Z. (2015). *VALLEX 3.0 - Valenční slovník českých sloves*. Charles University in Prague, [online] <http://ufal.mff.cuni.cz/vallex/3.0/>.
- Matthews, P. H. (2007): *Concise Oxford Dictionary of Linguistics*. Oxford University Press.
- Panevová, J. (1974 and 1975). On verbal frames in functional generative description. *Prague Bulletin of Mathematical Linguistics*. Part I: 1974, 22, pp. 3--40. Part II: 1975, 23, pp. 17--37.
- Panevová, J. (2000). Poznámky k valenci podstatných jmen. In Z. Hladká, P. Karlík (Ed.) *Čeština – univerzálie a specifika 2*. Brno: Vydavatelství MU, pp. 173--180. ISBN 80-210-2262-0.
- Przepiórkowski, A. et al. (2014). Walenty: Towards a comprehensive valence dictionary of Polish. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk and S. Piperidis (Eds) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavík, Iceland: ELRA, pp 2785—2792.
- Svozilová, N.; Prouzová, H. and Jirsová, A. (2005). *Slovník slovesných, substantivních a adjektivních vazeb a spojení*. Praha: Academia.