Valence českých deverbativních substantiv reprezentujících vybrané sémantické třídy

Veronika Kolářová

Ústav formální a aplikované lingvistiky
Matematicko-fyzikální fakulta
Univerzita Karlova v Praze

Osnova

- Valenční slovník NomVallex
- Vybrané sémantické třídy
- Specifické valenční chování substantiv
  - Formy participantů
  - Redukce počtu participantů
- Kvantitativní analýza valenčního chování
  - Český národní korpus
  - Pražský závislostní korpus

NomVallex

Valenční slovník českých substantiv založený na korpusu

- GAČR projekt (2016-2018; GA16-02196S)
- Tým: Veronika Kolářová, Jarmila Panevová, Jana Klímová, Anna Vernerová
- Pracoviště: Ústav formální a aplikované lingvistiky, MFF UK, Praha
- Teoretický rámec: valenční teorie funkčního generativního popisu
  - sémantické role (aktanty a volná doplnění; Panevová), valenční rámce
- Valence deverbativních substantiv
  - verbální substantiva (VS): *-ní/-tí*, např. *vyrobení / vyrábění*
  - dějová substantiva (DS): různé přípony, včetně nulového sufixu, např. *výroba*
- Korpusová data: PDT, ČNK a Araneum Bohemicum
- Sémantické třídy

Vallex, PDT-Vallex, NomVallex

- Vallex
  - Valenční slovník zachycující valenční strukturu českých sloves v jejich jednotlivých významech, které charakterizuje pomocí glos a příkladů
    - 4 598 českých sloves, která odpovídají 10 701 lexikálním jednotkám, tedy vždy „danému slovesu v daném významu“
- PDT-Vallex
  - Valenční slovník budovaný při anotacích taktogramatické roviny Pražského závislostního korpusu (PDT)
  - slovesa, substantiva (3727 hesel), adjektiva
- NomVallex
  - Rozšiřuje Vallex: dodání substantivních hesel
  - Rozšiřuje PDT-Vallex: detailnější způsob zpracování

NomVallex
Valence substantiv a sémantické třídy

- Vybrané sémantické třídy
  - návaznost na sémantické třídy ve Vallexu
  - Communication, Contact, Exchange, Mental action, Psychological nouns
- Co máme od valenčního chování jednotlivých sémantických tříd očekávat?
  - vliv posunu ve významu na valenci
  - formy participantů
  - redukce počtu participantů ve valenčním rámci
  - preference v souvýskytu participantů
  - verbální substantiva vs. dějová substantiva
  - centrum – periférie
- Charakteristické vlastnosti jednotlivých sémantických tříd
- Zajímavé rozdíly mezi jednotlivými sémantickými třídami

Typické vs. specifické valenční chování substantiv

- Typické valenční chování (Panevová 2000, Karlík 2000)
  - označují děj nebo stav jako základové sloveso
  - dědí valenční rámec slovesa
  - formy podléhají typickým posunům
    - např. Ak → Gen, *varování cestujících*.ADDR
- Specifické valenční chování (Kolářová 2010, Kolářová 2014)
  - úzce vázáno na posun ve významu substantiva
    - už nikoli děj nebo stav jako u základového slovesa
  - specifické vlastnosti valenčních doplnění
    - specifické formy (odpovídají specifickým posunům)
      - např. Ak → Dat, *varování cestujícím*.ADDR
    - redukce počtu valenčních doplnění
      - prostá redukce (*dovtípit se čeho* x **důvtip čeho*)
      - zabudování aktantu (např. činitelská jména zabudovávají ACT, *učitel – kdo učí*)
    - změna charakteru valenčního doplnění ze slovesného na substantivní
      - např. Patiens → Material (MAT) / Partitiv
      - *balení másla*.PAT *do obalu brigádníky*.ACT → *jedno balení másla*.MAT („kontejner“)

Typické vs. specifické valenční chování substantiv

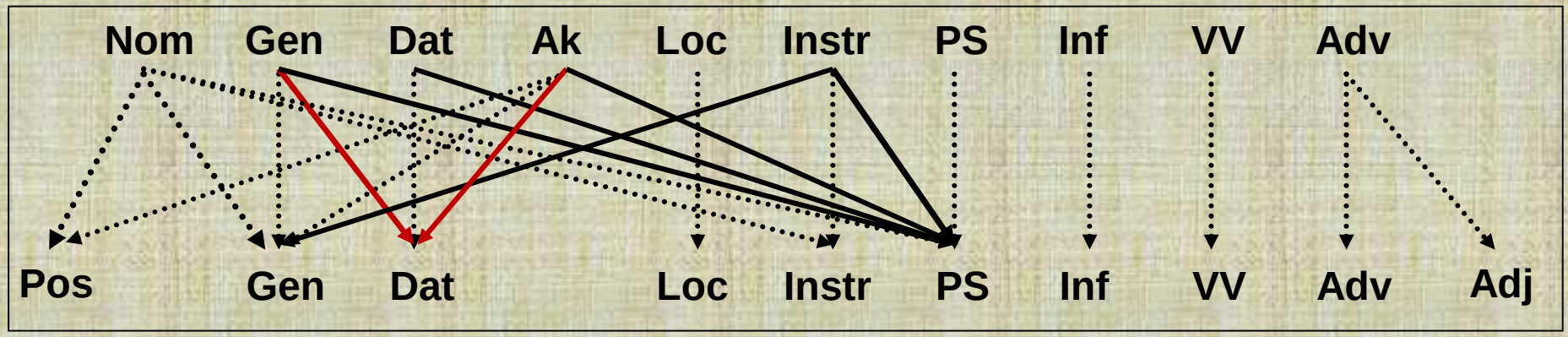
- Typické valenční chování (Panevová 2000, Karlík 2000)
  - označují děj nebo stav jako základové sloveso
  - dědí valenční rámec slovesa
  - formy podléhají typickým posunům
    - např. Ak → Gen, *varování cestujících*.ADDR
- Specifické valenční chování (Kolářová 2010, Kolářová 2014)
  - úzce vázáno na posun ve významu substantiva
    - už nikoli děj nebo stav jako u základového slovesa
  - specifické vlastnosti valenčních doplnění
    - **specifické formy (odpovídají specifickým posunům)**
      - např. Ak → Dat, *varování cestujícím*.ADDR
    - redukce počtu valenčních doplnění
      - prostá redukce (*dovtípit se čeho* x **důvtip čeho*)
      - zabudování aktantu (např. činitelská jména zabudovávají ACT, *učitel – kdo učí*)
    - změna charakteru valenčního doplnění ze slovesného na substantivní
      - např. Patiens → Material (MAT) / Partitiv
      - *balení másla*.PAT *do obalu brigádníky*.ACT → *jedno balení másla*.MAT („kontejner“)

Specifické posuny v povrchových realizacích participantů

Ak → Dat a Gen → Dat

- *dotázat se Petra → dotaz Petrovi*
- *varovat řidiče → varování řidičům*
- *poctít umělce → pocta umělci*
- *udeřit poslance → úder poslanci*

Povrchové realizace participantů sloves



Povrchové realizace participantů substantiv

Specifický posun Ak → Dat

- Tři sémantické třídy: Communication, Evaluation, Contact
- Životnost participantu, u něhož k posunům dochází
- Communication: ADDR
  - *dotaz, informace, instrukce, otázka, pozvánka, prosba, upozornění, varování, výstraha, výzva, zpráva, žádost*
- Evaluation: PAT
  - pozitivní: *pocta, podpora, pochvala, ocenění, uznání, vyznamenání*
  - negativní: *napomenutí, pokuta, trest, urážka, veto*
  - životnost vs. neživotnost PAT
    - *pochvala české vlády / české vládě* vs. *pochvala kravaty / *kravatě*
- Contact: PAT
  - *políbení, polibek, rána, úder, zásah*
  - někdy mění význam: *zásah cíle* (Contact) vs. *zásah právnímu vědomí občanů* (Psychological noun)

Typické vs. specifické valenční chování substantiv

- Typické valenční chování (Panevová 2000, Karlík 2000)
  - označují děj nebo stav jako základové sloveso
  - dědí valenční rámec slovesa
  - formy podléhají typickým posunům
    - např. Ak → Gen, *varování cestujících*.ADDR
- Specifické valenční chování (Kolářová 2010, Kolářová 2014)
  - úzce vázáno na posun ve významu substantiva
    - už nikoli děj nebo stav jako u základového slovesa
  - specifické vlastnosti valenčních doplnění
    - specifické formy (odpovídají specifickým posunům)
      - např. Ak → Dat, *varování cestujícím*.ADDR
    - **redukce počtu valenčních doplnění**
      - prostá redukce (*dovtípit se čeho* x **důvtip čeho*)
      - **zabudování aktantu** (např. činitelská jména zabudovávají ACT, *učitel – kdo učí*)
    - změna charakteru valenčního doplnění ze slovesného na substantivní
      - např. Patiens → Material (MAT) / Partitiv
      - *balení másla*.PAT *do obalu brigádníky*.ACT → *jedno balení másla*.MAT („kontejner“)

Zabudovaný Patiens

- Výsledky činnosti „předávání věci / informace“
- Výsledky předávání / nabytí věci (Exchange)
  - co bylo předáno / dodáno / darováno / vyplaceno…
  - *dar, dárek, dodávka, nákup, věno; podání*
  - peněžní obnosy (*dluh, náhrada, plat, půjčka, výplata*)
- Výsledky předávání informace (Communication)
  - co bylo sděleno / nabídnuto / vzkázáno…
  - *sdělení, nabídka, vzkaz*
- Patiens ve valenčním rámci substantiva s daným významem chybí

*** výplata**

ACT(.2,.7,.u) PAT(.2,.u) ADDR(.3) v-w8191f1 Used: 1x

***výplaty mezd*.PAT**

***výplata dividend*.PAT / *náhrad*.PAT *Gen*.ADDR**

EMPTY v-w8191f2 Used: 0x

(plat) měsíční výplata

Kvantitativní analýza

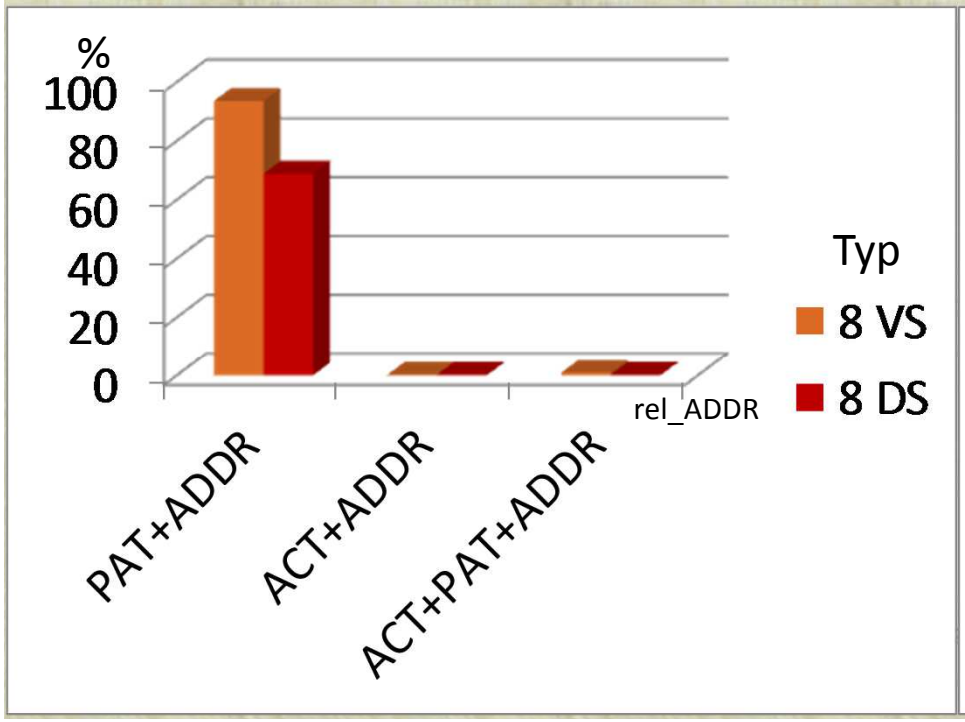
- Srovnání valenčního chování vybraných sémantických tříd
  - preference v souvýskytu participantů
  - formy participantů
- 1. sonda: Data Českého národního korpusu (ČNK), řada SYN
  - pilotní srovnání dvou sémantických tříd
  - Communication, Exchange
  - Bonito / KonText
  - dotazovací jazyk: CQL
    - [lemma="... "] [!(tag="[Z|R|V|J].*")]{0,4} [tag="N…3.*"]
- 2. sonda: Data Pražského závislostního korpusu (PDT 3.0)
  - závislostní struktura, anotace valence, aktanty
  - příprava na zpracování valence substantiv v NomVallexu
  - Communication, Exchange, Contact, Psychological nouns, (Mental action)
  - dotazovací jazyk: PML-TQ

Pilotní srovnání dvou sémantických tříd

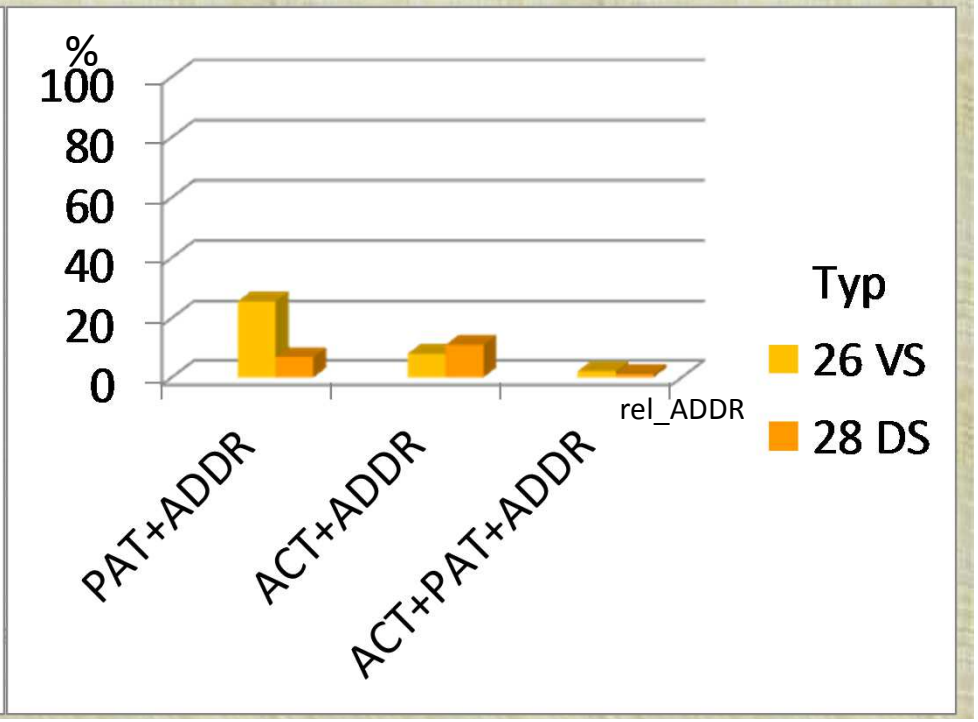
Data ČNK: SYN2000 (Kolářová 2010, 2014)

- Kombinace ADDR a dalších aktantů
- Manuální vyhledávání, ruční kontrola a pročištění výsledků
- Rozdílné preference v kombinacích aktantů
- Rozdíly mezi verbálními a dějovými substantivy

Exchange



Communication

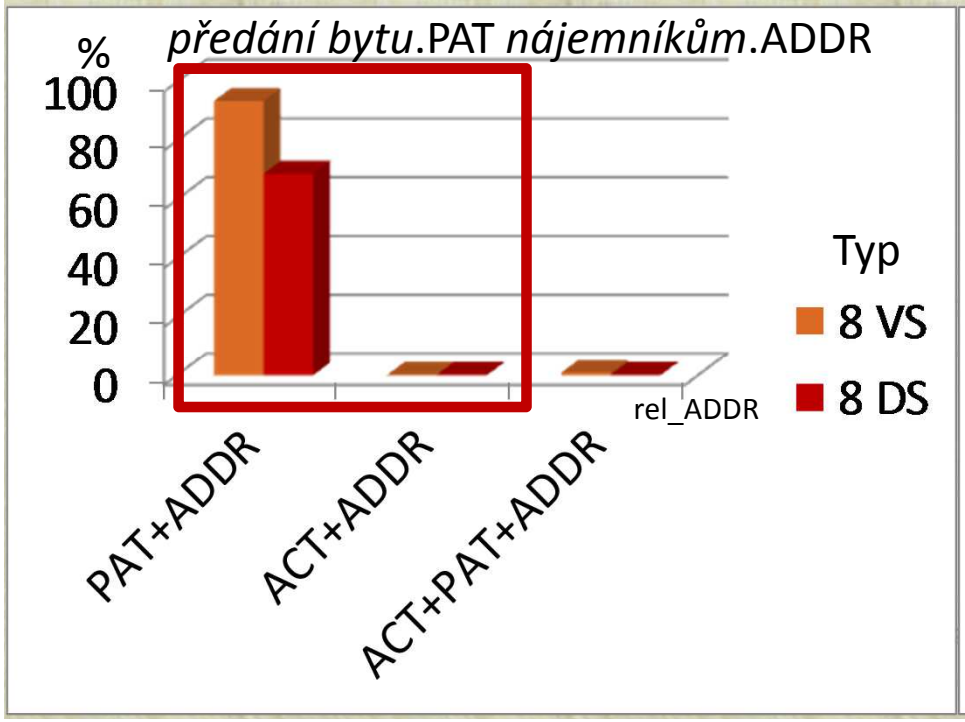


Pilotní srovnání dvou sémantických tříd

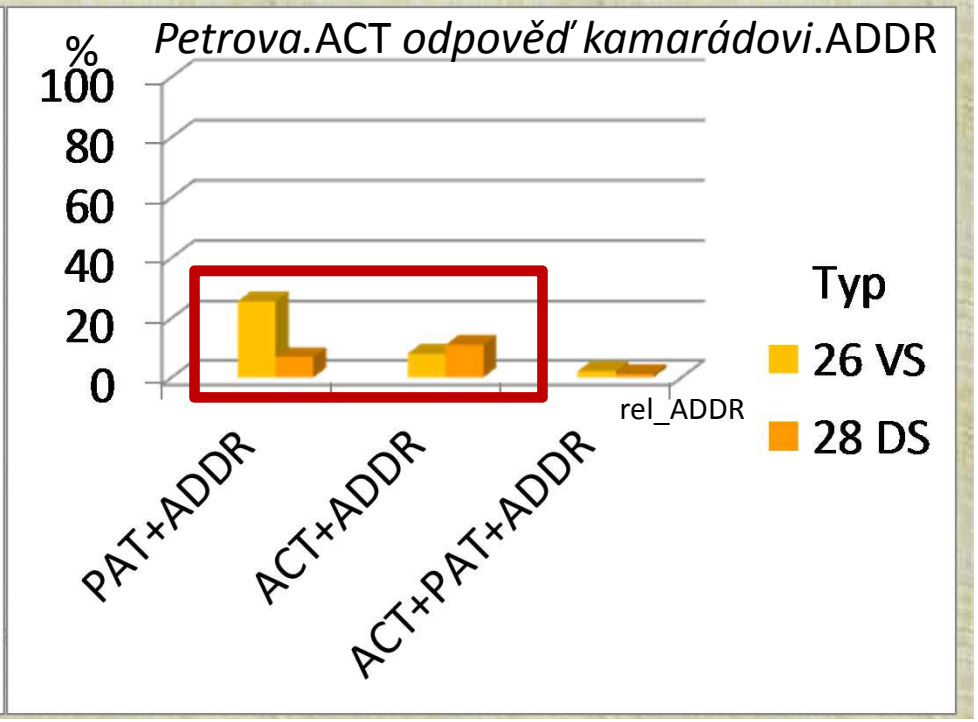
Data ČNK: SYN2000 (Kolářová 2010, 2014)

- Kombinace ADDR a dalších aktantů
- Manuální vyhledávání, ruční kontrola a pročištění výsledků
- Rozdílné preference v kombinacích aktantů
- Rozdíly mezi verbálními a dějovými substantivy

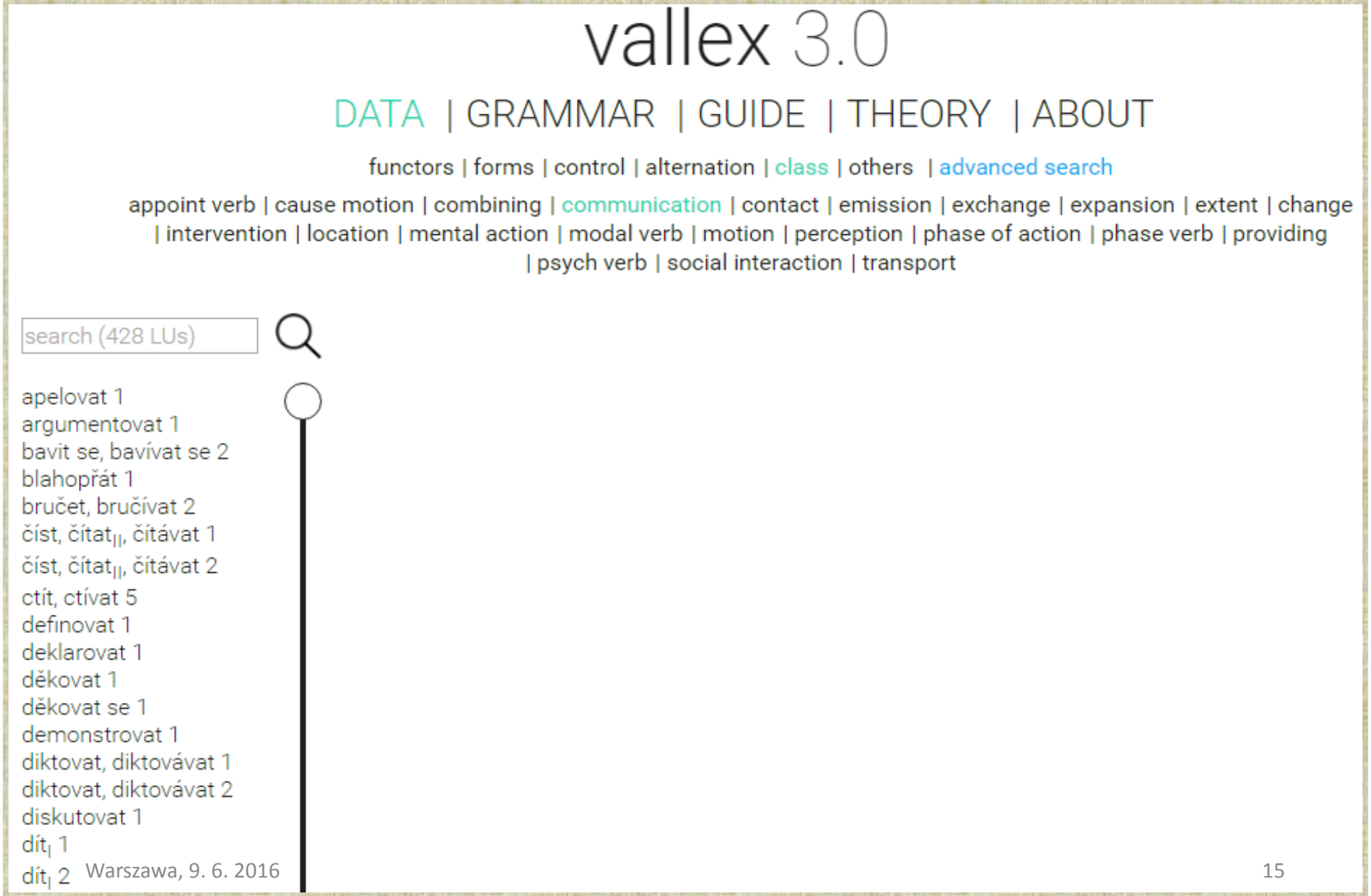
Exchange



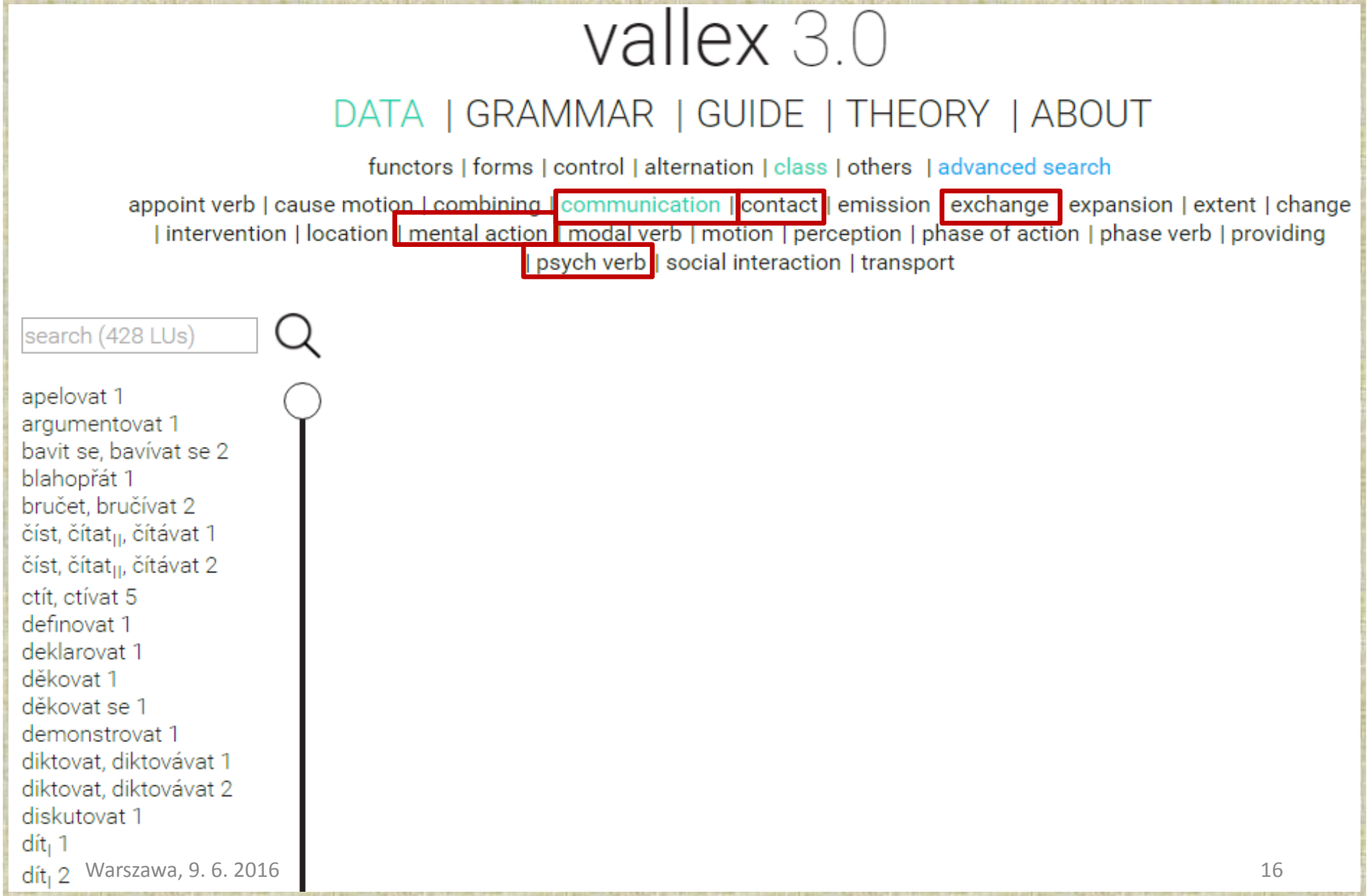
Communication



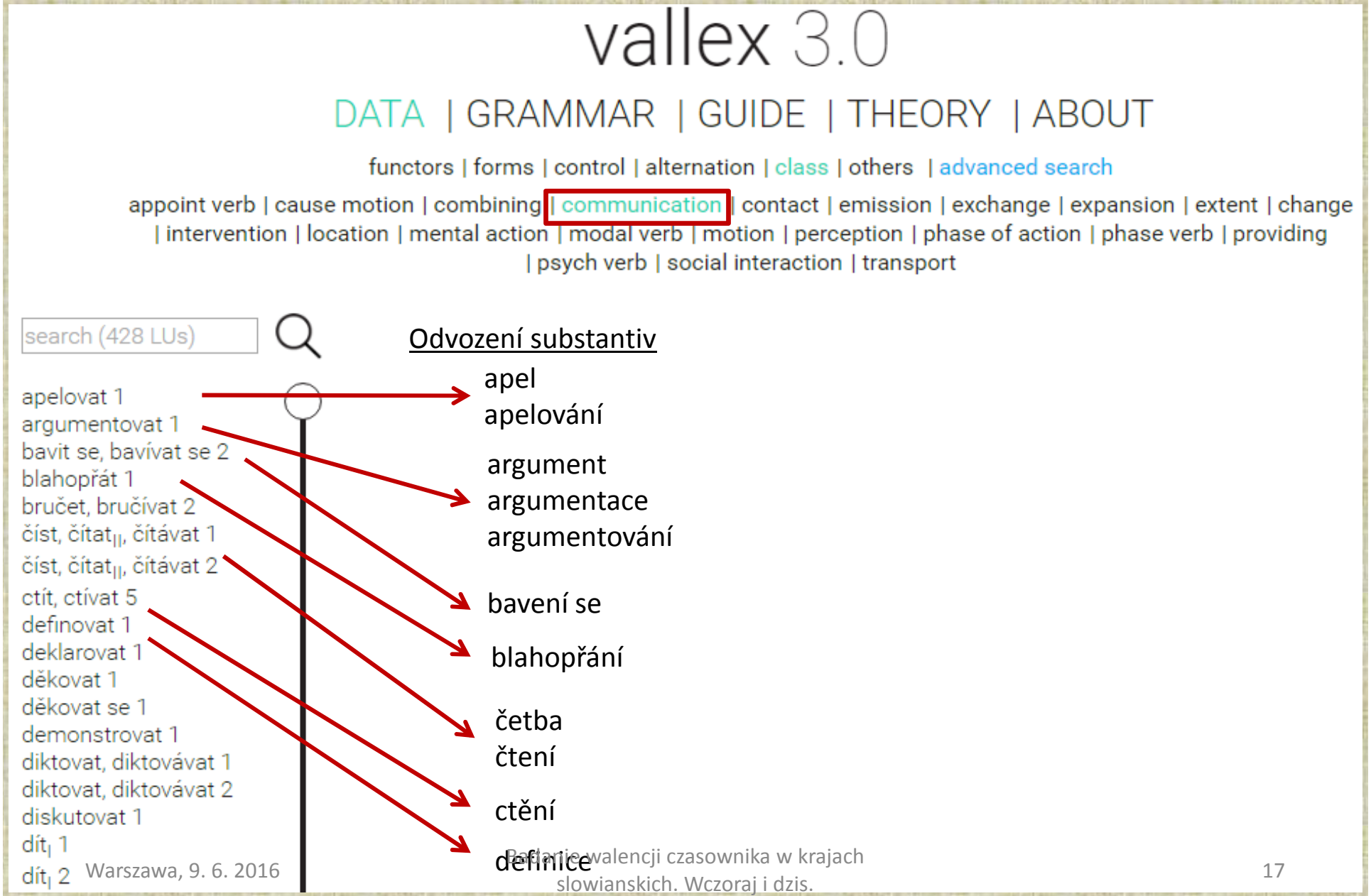
NomVallex: Výběr substantiv



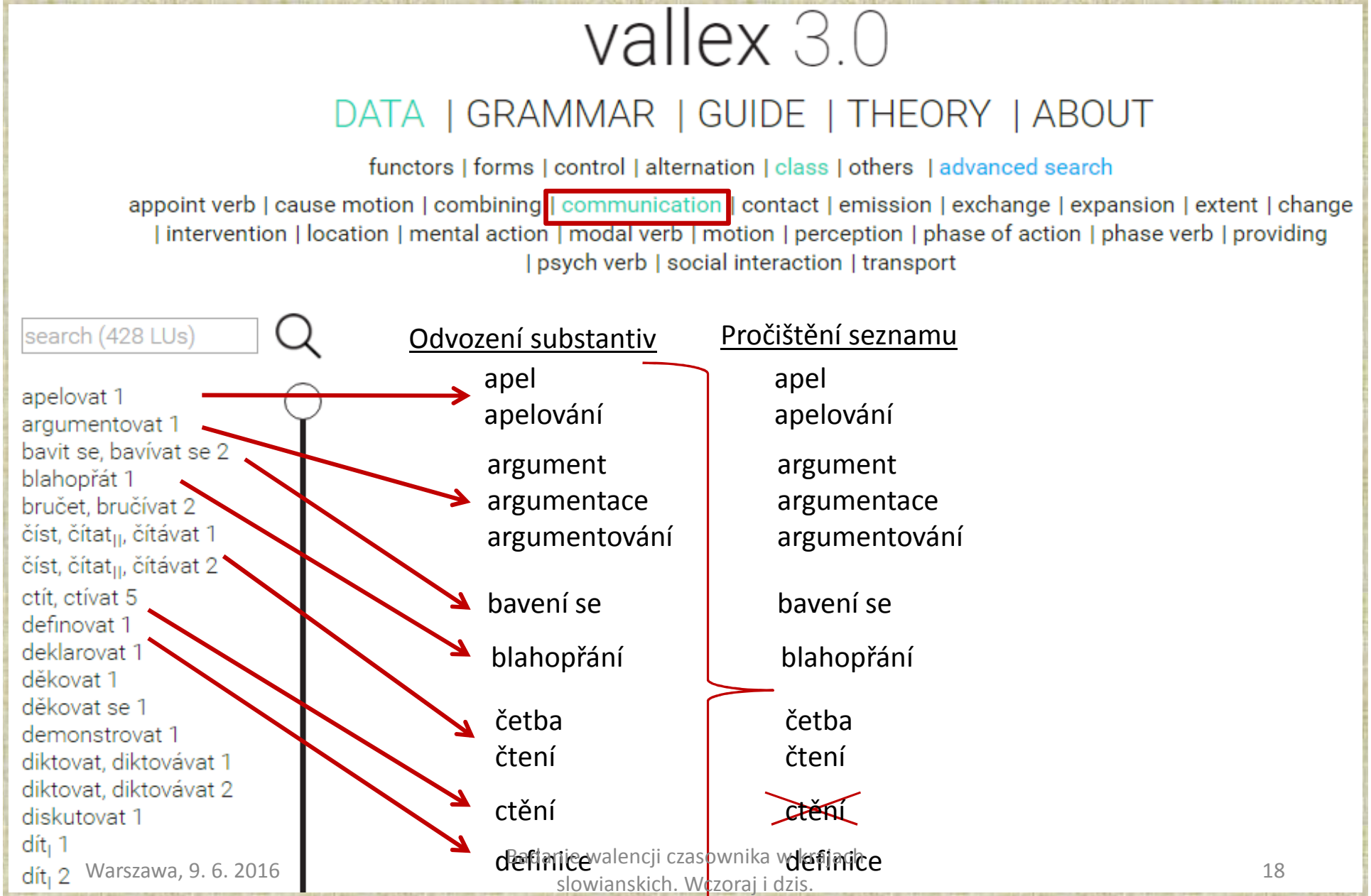
NomVallex: Výběr substantiv



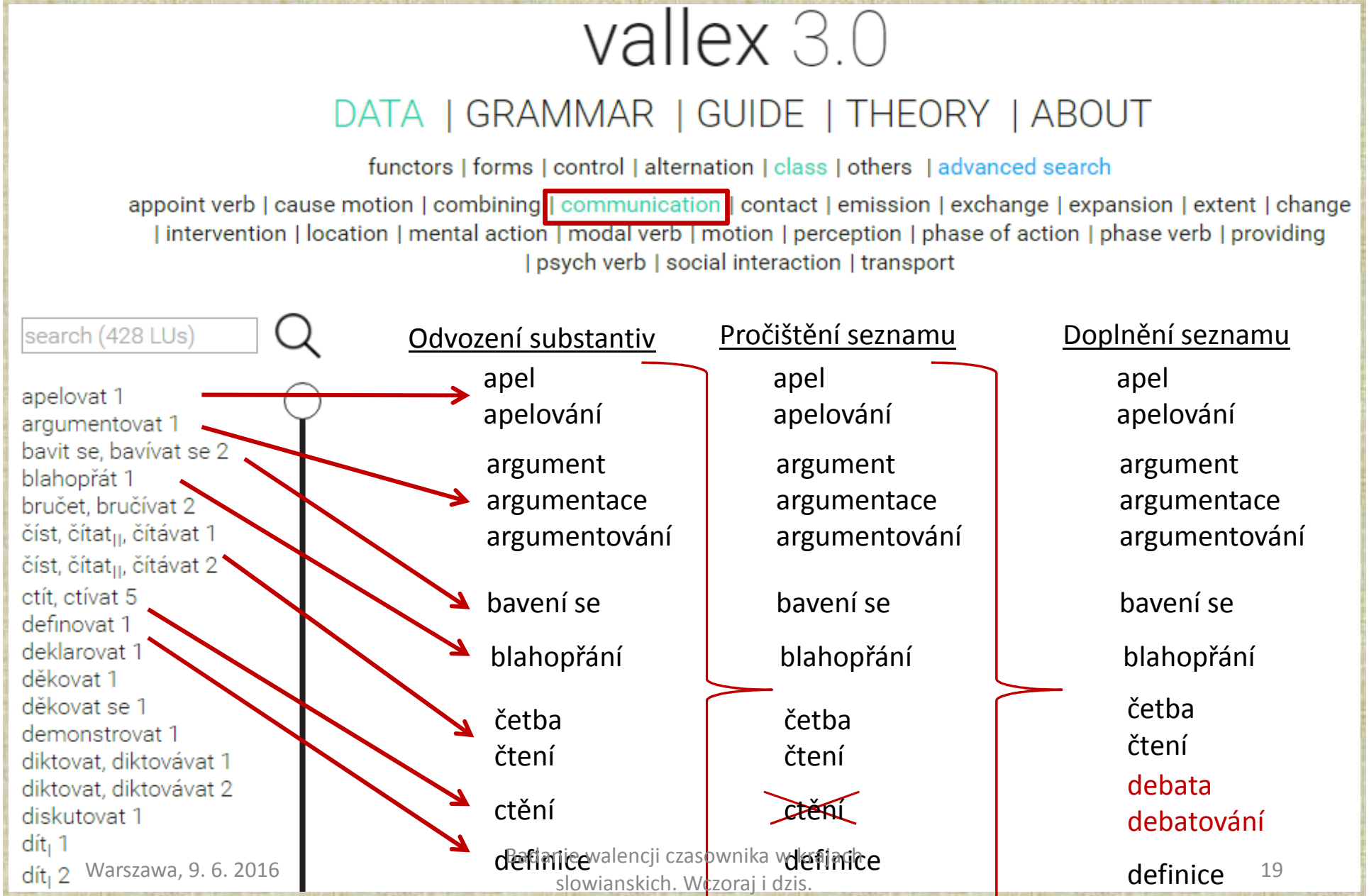
NomVallex: Výběr substantiv



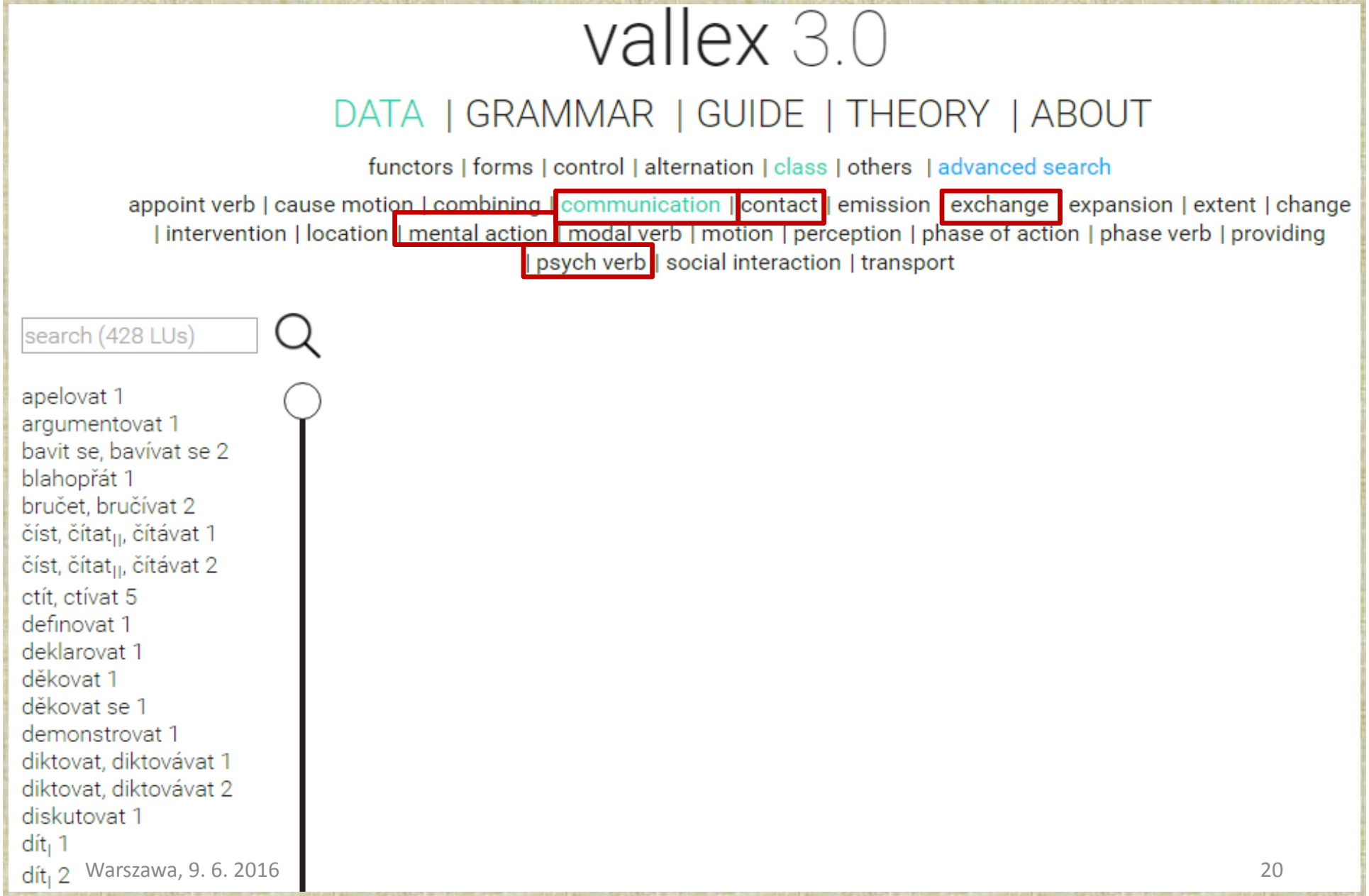
NomVallex: Výběr substantiv



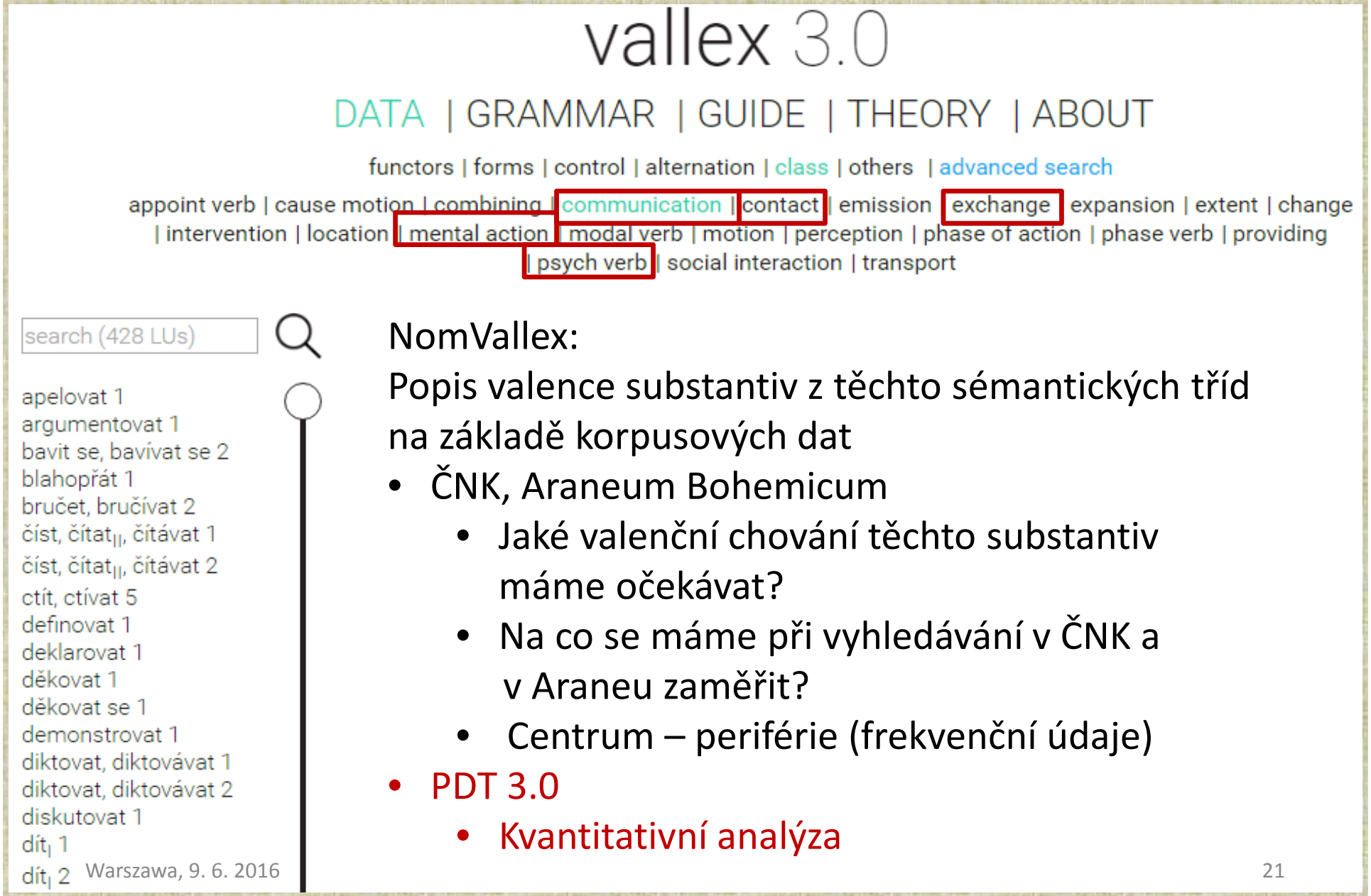
NomVallex: Výběr substantiv



NomVallex: Výběr substantiv



NomVallex: Výběr substantiv



NomVallex:

Popis valence substantiv z těchto sémantických tříd na základě korpusových dat

- ČNK, Araneum Bohemicum
  - Jaké valenční chování těchto substantiv máme očekávat?
  - Na co se máme při vyhledávání v ČNK a v Araneu zaměřit?
  - Centrum – periférie (frekvenční údaje)
- PDT 3.0
  - Kvantitativní analýza

Kvantitativní analýza: Sonda do PDT 3.0

- Substantiva, která byla v datech PDT 3.0 rozvita nějakým aktantem
  - povrchově vyjádřený nebo nevyjádřený aktant
  - zanedbává se polysémie substantiv
- Kombinace aktantů
- Formy aktantů

Typ substantiva	**Communication**	**Exchange**	**Contact**	**Psychological nouns**
Verbální substantivum	145	94	31	29
Dějové substantivum	102	34	10	18
Celkem (lemmata)	247	128	41	47

Kombinace aktantů: DS Exchange Dotaz v PML-TQ (PDT 3.0)

(Jan Kolář)

```
t-node $a:=[
   a/lex.rf a-node [ m/tag ~ "^N" ],
    t_lemma="dar" or t_lemma="dluh" or t_lemma="dodávka"
    or t_lemma="doprava" or t_lemma="koupě" or ……..
   , t-node $b:= [ functor ~ "ACT|PAT|ADDR|EFF|ORIG" ],
];
>> give $a, $b.functor & " " & first_defined($b.is_generated,
0) & " "
>> distinct $1, concat($2, " " over $1 sort by $2)
>> distinct $2, count(over $2)
>> $1, $2, "= " & percnt($2 div sum($2 over all),2) & " % of "
&sum($2 over all) , "rank " & rank(over all sort by $2 desc)
```

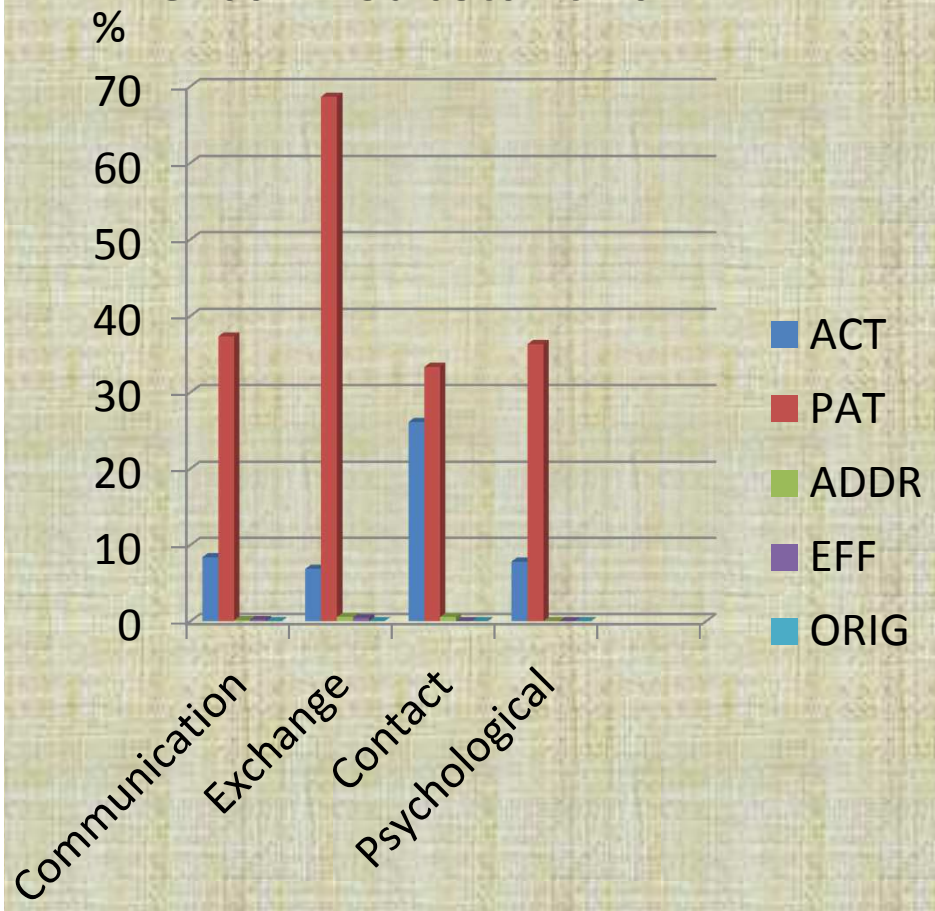
Dějová substantiva: Exchange
Souhrnná tabulka pro kombinace aktantů

PAT 0	348	= 64.44 % of 540	rank 1
ACT 0	97	= 17.96 % of 540	rank 2
ADDR 0	16	= 2.96 % of 540	rank 3
EFF 0	14	= 2.59 % of 540	rank 4
ORIG 0	13	= 2.41 % of 540	rank 5
ADDR 0 PAT 0	11	= 2.04 % of 540	rank 6
ACT 1 PAT 0	7	= 1.30 % of 540	rank 7
ACT 1 ADDR 1 PAT 1	5	= 0.93 % of 540	rank 8
ACT 1	5	= 0.93 % of 540	rank 8
ACT 1 PAT 1	5	= 0.93 % of 540	rank 8
ACT 0 PAT 0	3	= 0.56 % of 540	rank 11
ACT 1 ADDR 1 PAT 0	3	= 0.56 % of 540	rank 11
EFF 0 PAT 0	3	= 0.56 % of 540	rank 11
ORIG 0 PAT 0	3	= 0.56 % of 540	rank 11
PAT 1	2	= 0.37 % of 540	rank 15
PAT 0 PAT 0	1	= 0.19 % of 540	rank 16
ACT 1 EFF 0	1	= 0.19 % of 540	rank 16
ACT 0 ADDR 0	1	= 0.19 % of 540	rank 16
ADDR 1	1	= 0.19 % of 540	rank 16
ACT 0 EFF 0	1	= 0.19 % of 540	rank 16

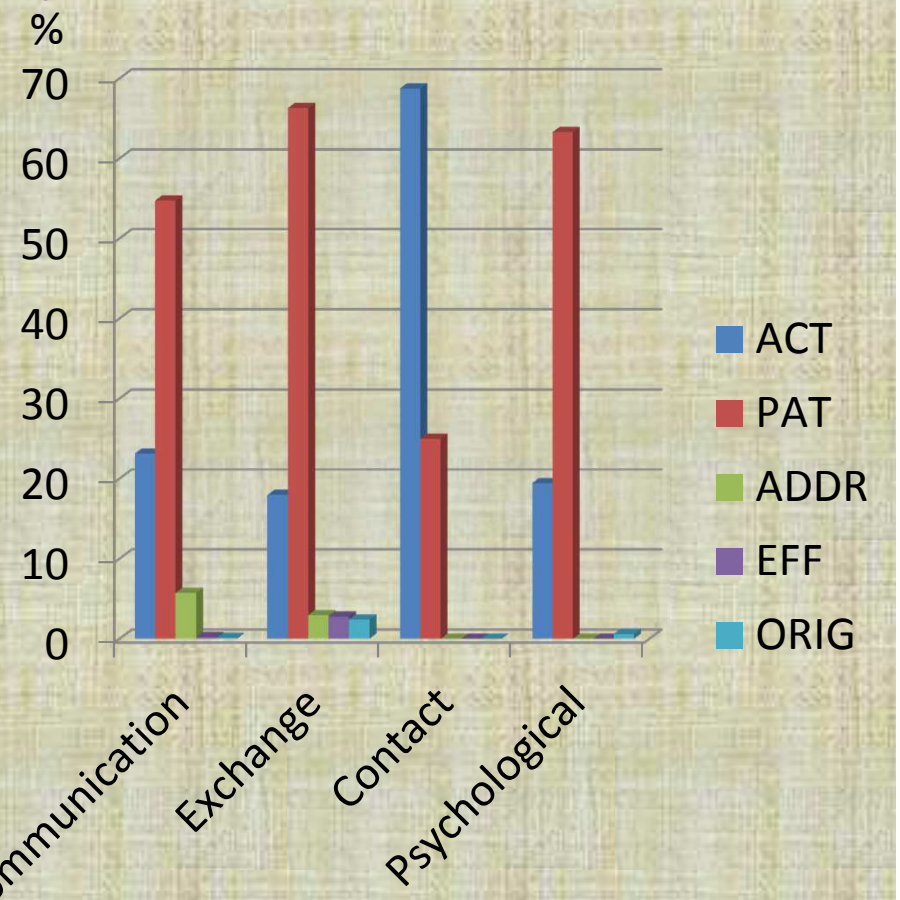
PDT 3.0: Povrchově vyjádřen pouze 1 aktant

- Převažuje PAT
- Výjimka: třída Contact

Verbální substantiva

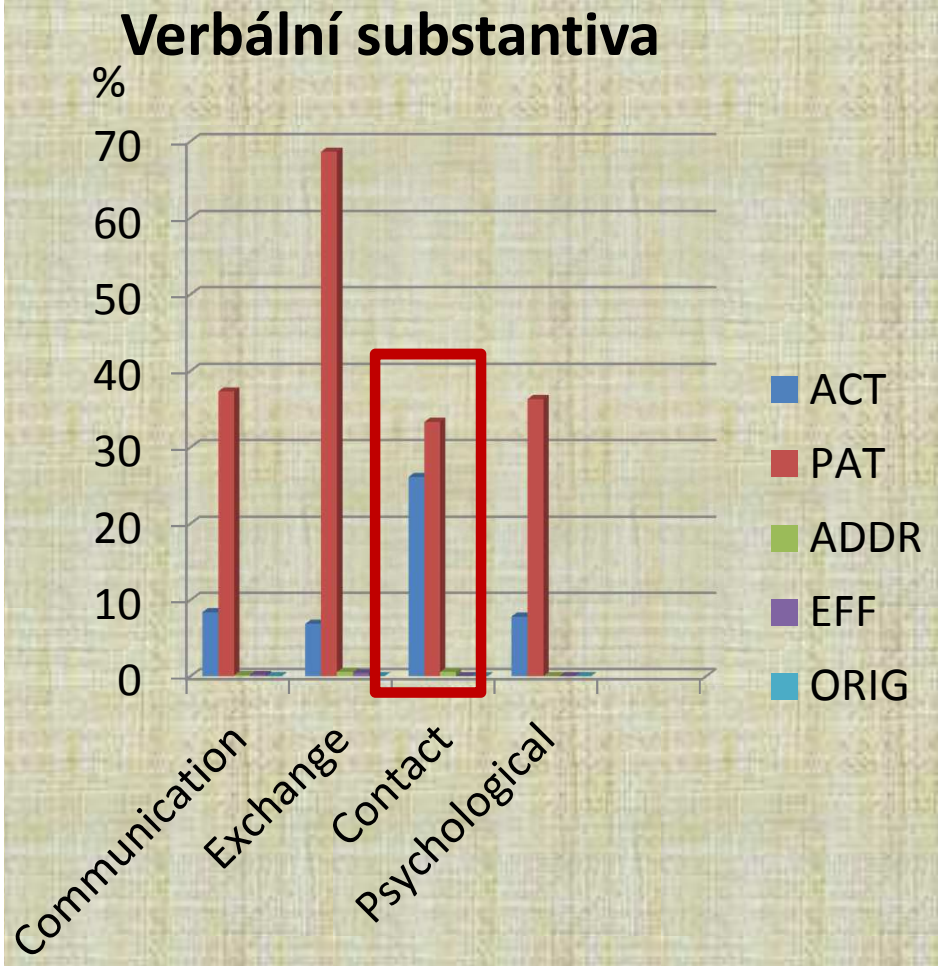


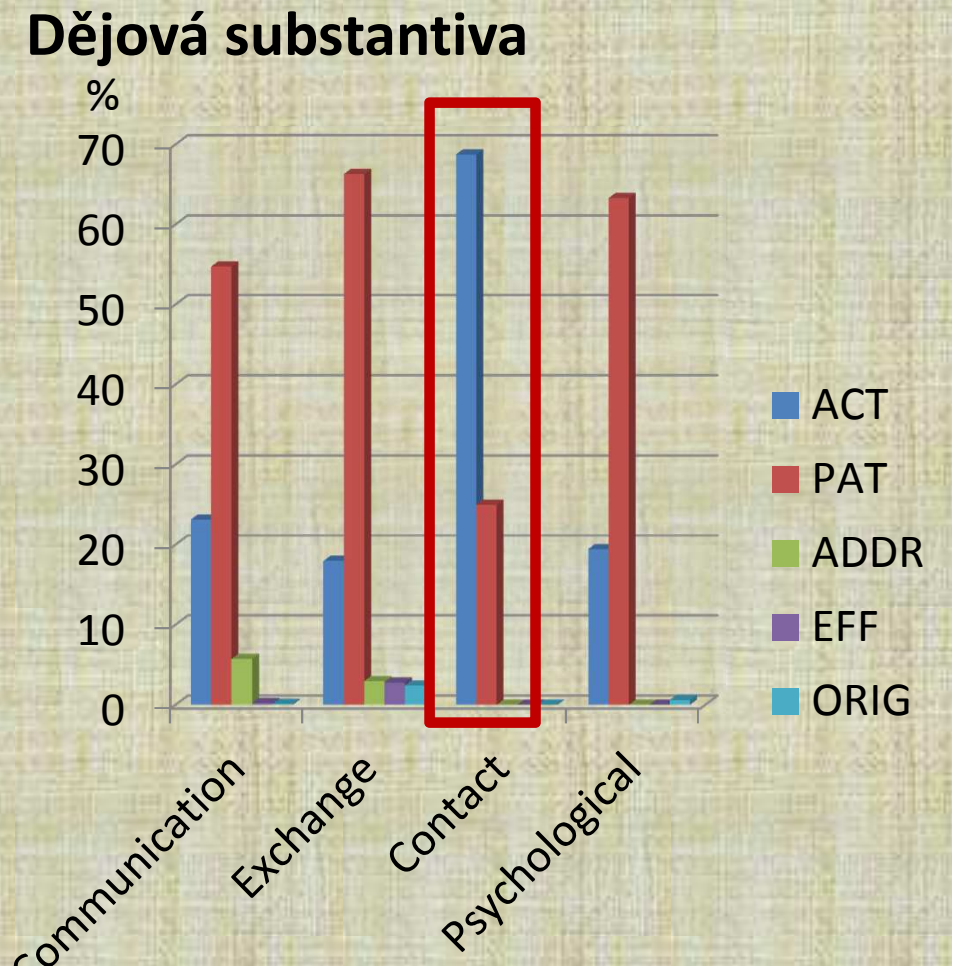
Dějová substantiva



PDT 3.0: Povrchově vyjádřen pouze 1 aktant

- Převažuje PAT
- Výjimka: třída Contact

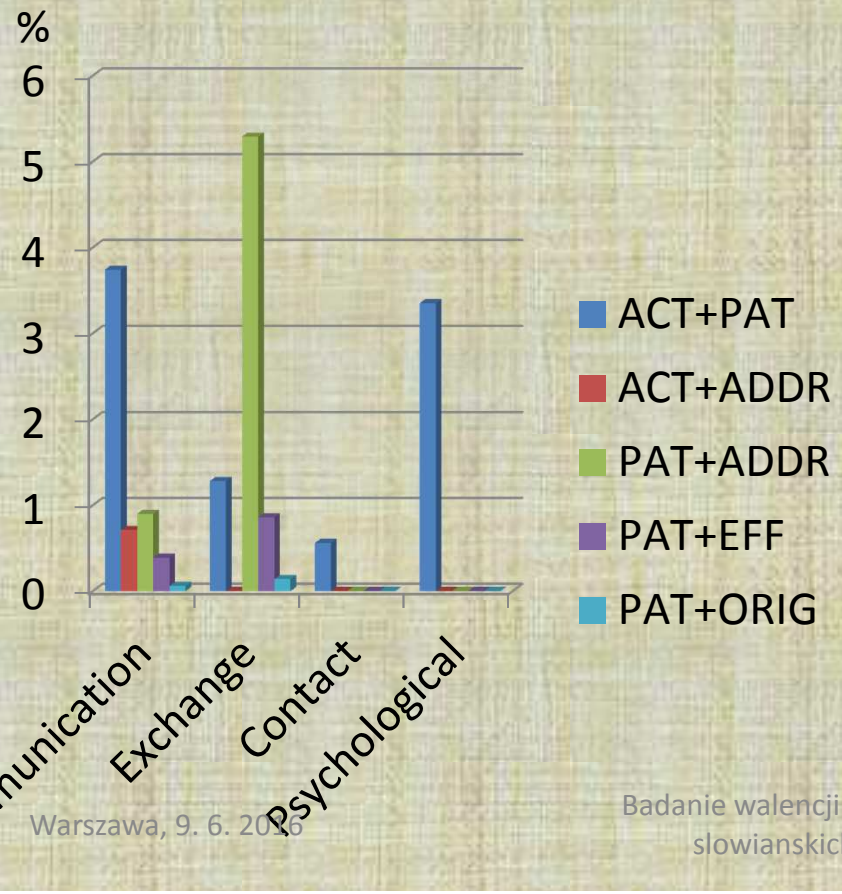




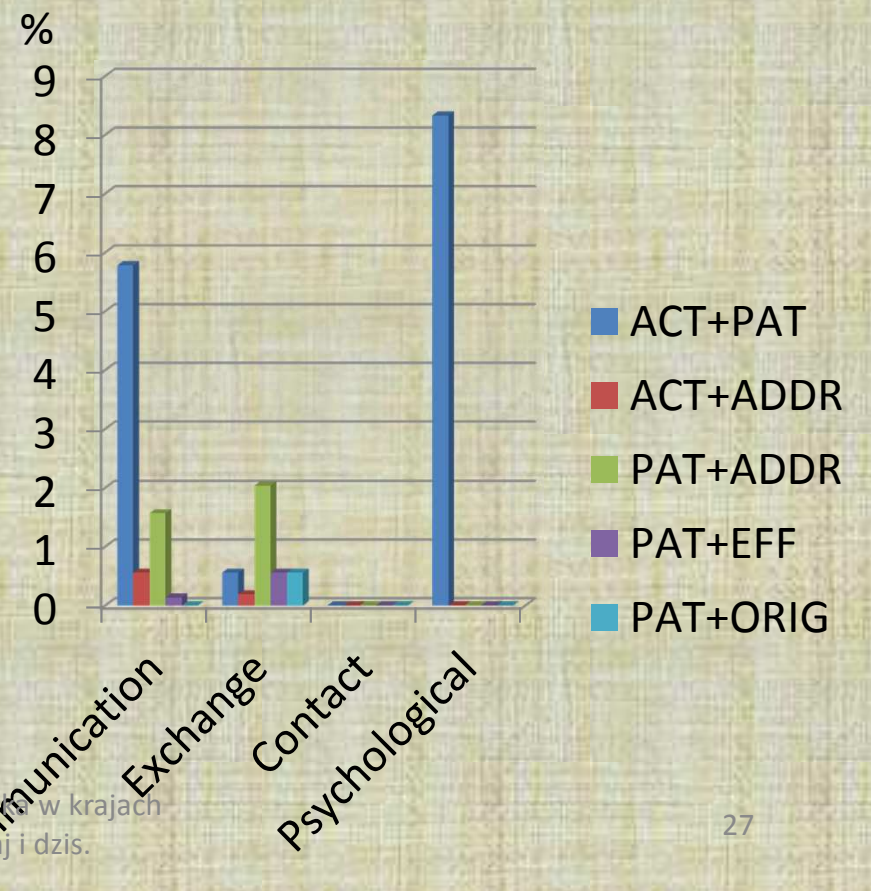
PDT 3.0: Kombinace dvou povrchově vyjádřených aktantů

- Potvrzuje výsledky dřívější sondy do ČNK SYN2000
  - rozdíly mezi třídami Exchange a Communication
- Třída Contact: nízká / nulová frekvence kombinace aktantů

Verbální substantiva



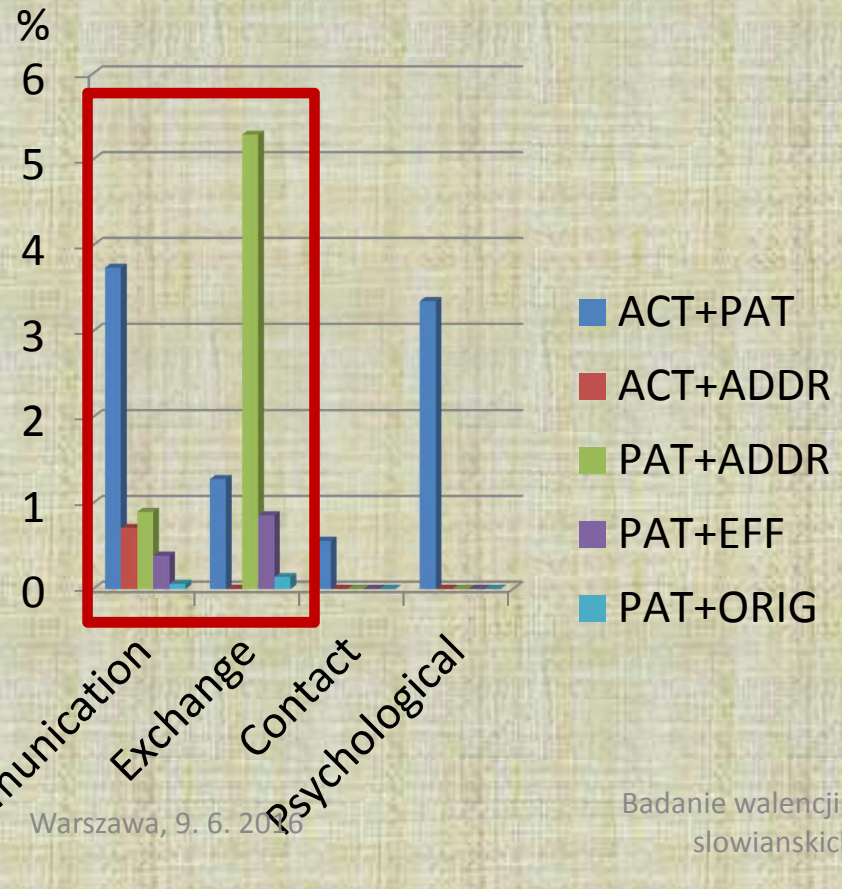
Dějová substantiva



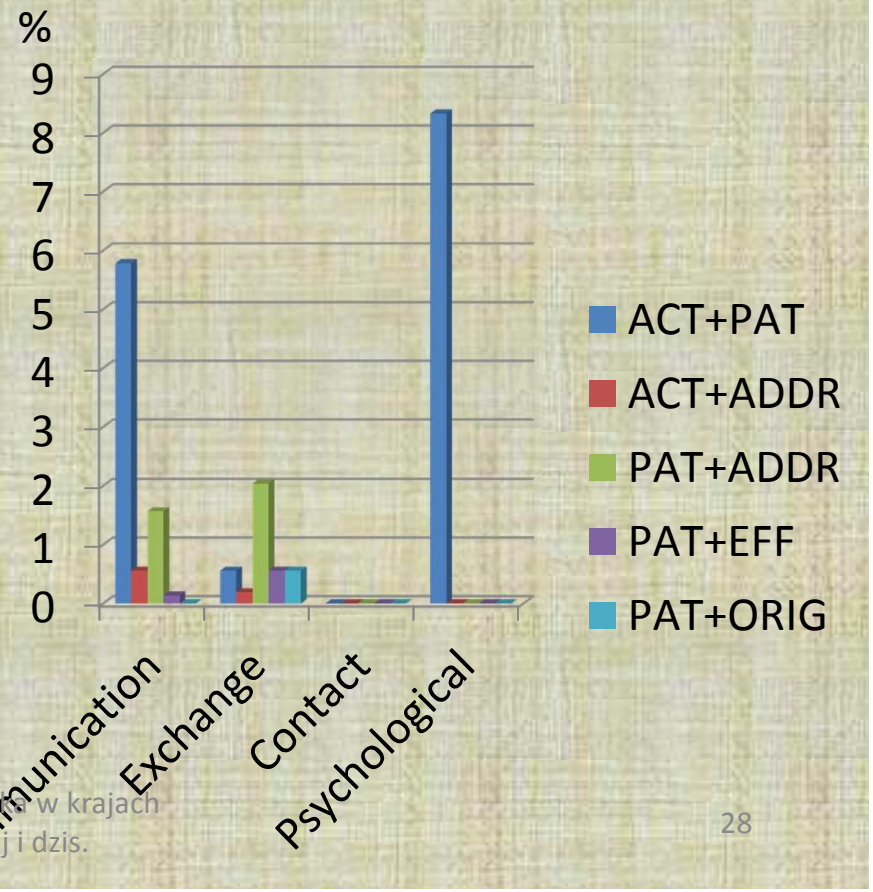
PDT 3.0: Kombinace dvou povrchově vyjádřených aktantů

- Potvrzuje výsledky dřívější sondy do ČNK SYN2000
  - rozdíly mezi třídami Exchange a Communication
- Třída Contact: nízká / nulová frekvence kombinace aktantů

Verbální substantiva



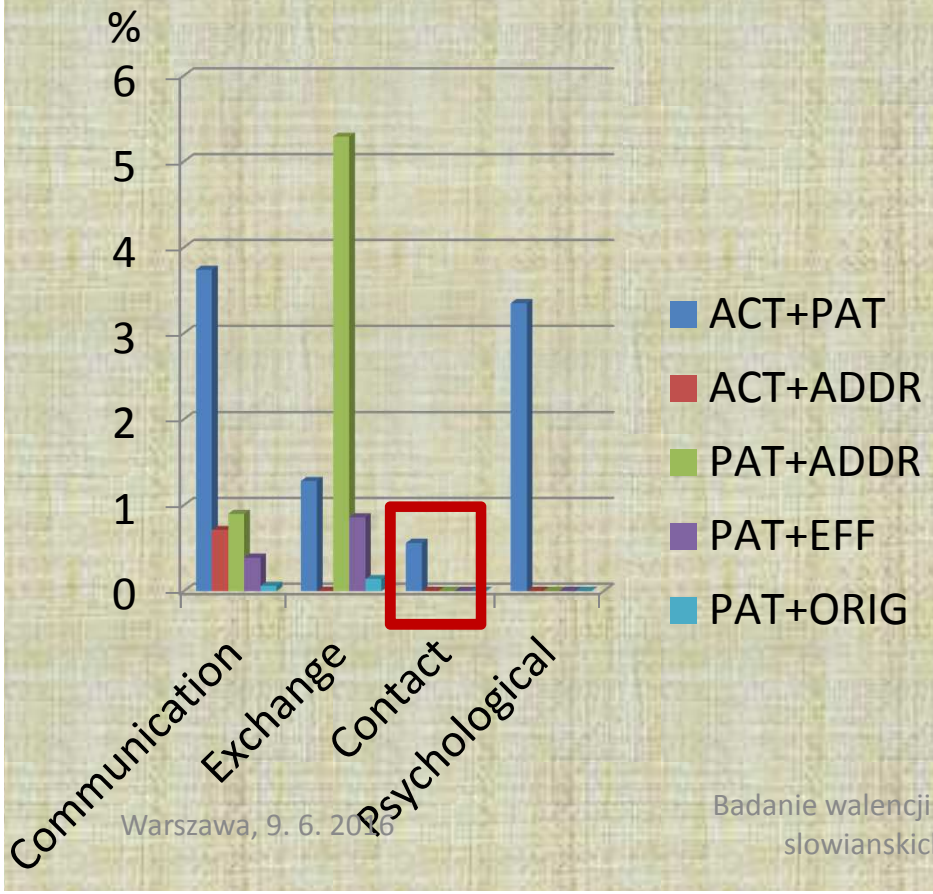
Dějová substantiva



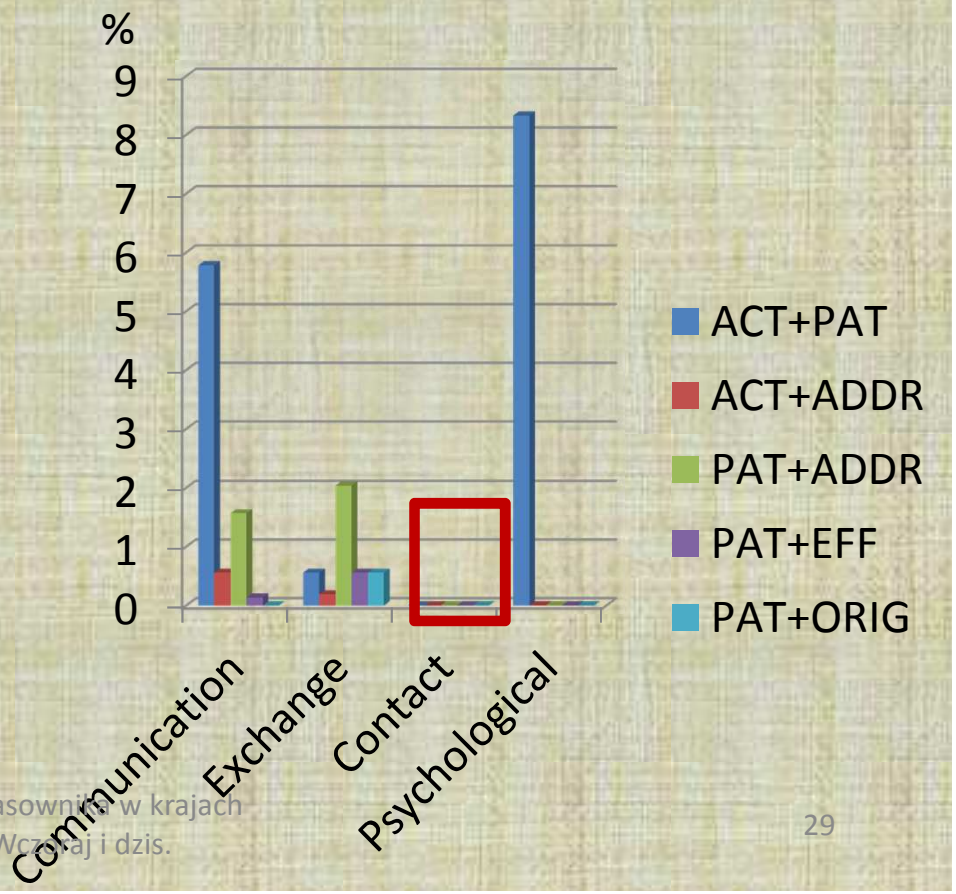
PDT 3.0: Kombinace dvou povrchově vyjádřených aktantů

- Potvrzuje výsledky dřívější sondy do ČNK SYN2000
  - rozdíly mezi třídami Exchange a Communication
- Třída Contact: nízká / nulová frekvence kombinace aktantů

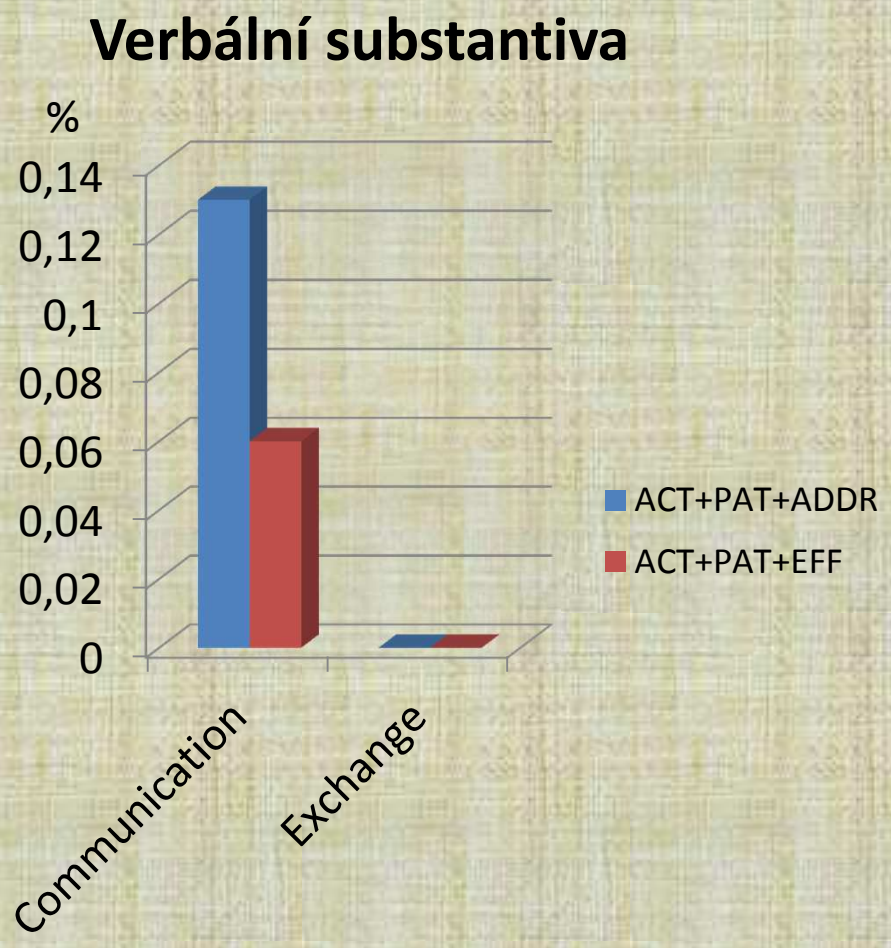
Verbální substantiva

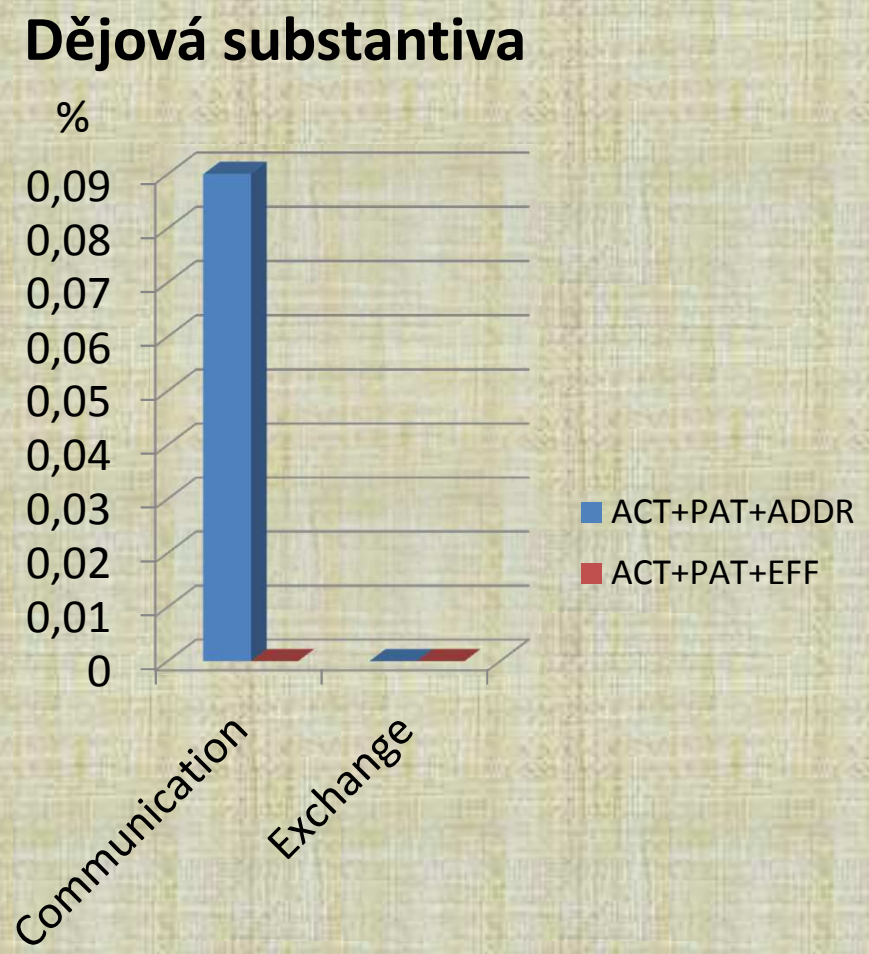


Dějová substantiva



PDT 3.0: Kombinace tří povrchově vyjádřených aktantů

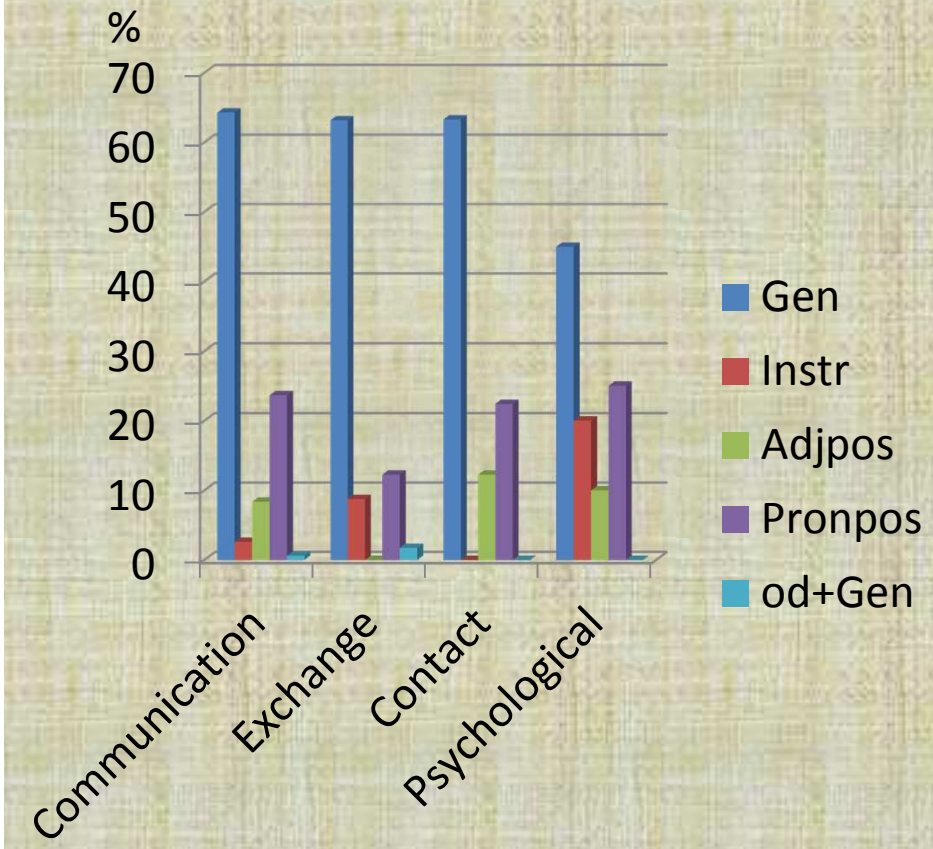




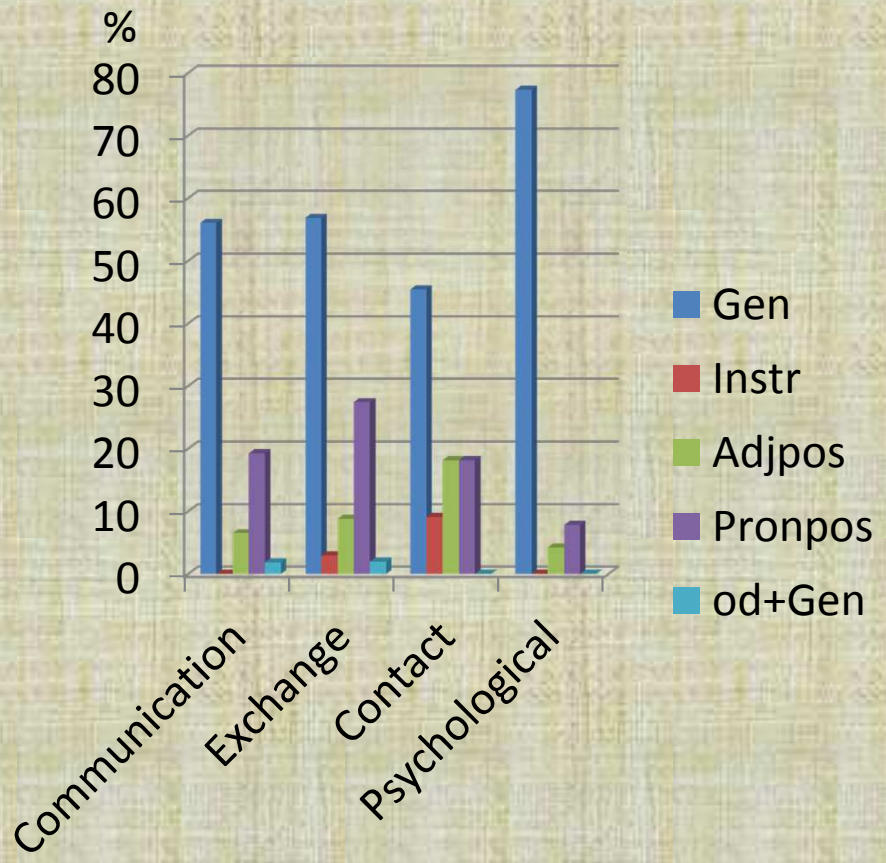
Forma vyjádření konatele (Aktora)

- Gen > $Pron_{pos}$ > Adj_{pos} / Instr > od+Gen
- Forma Instr se u některých tříd vůbec nevyskytuje

Verbální substantiva



Dějová substantiva



Závěr

- Specifické valenční chování substantiv
  - posun Ak → Dat u sémantických tříd Communication, Evaluation a Contact
  - redukce počtu valenčních doplnění (zabudovaný PAT) u tříd Communication a Exchange
- Kvantitativní analýza sémantických tříd Communication, Exchange, Contact a Psychological nouns
  - shodné rysy napříč třídami, u verbálních i dějových substantiv
    - převažuje vyjádření samotného PAT
    - při vyjádření ACT převažuje Gen, následovaný posesivy
  - rozdílné rysy
    - rozdílné preference v souvýskytu aktantů
    - vyjádření ACT formou Instr
- Valenční chování substantiv souvisí s příslušností k sémantické třídě
  - zachycení ve valenčním slovníku NomVallex

Literatura

- Čermáková, A. (2009). *Valence českých substantiv*. Praha: Lidové noviny. ISBN 978-80-7106-426-800.
- Kolářová, V. (2010). *Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valencí)*. Praha: Karolinum.
- Kolářová, V. (2014). Special valency behavior of Czech deverbal nouns. In O. Spevak (Ed.) *Noun Valency,* Amsterdam: John Benjamins, pp. 19--60.
- Kolářová, V. (2014): Preference v souvýskytu aktantů u českých substantiv mluvení. *Korpus – gramatika – axiologie*, Vol. 5, No. 10. Hradec Králové: Gaudeamus, pp. 23--40.
- 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/.
- 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 – univerzália 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.