EXEMPLIFIKACE SUBSTANTIV V LEXIKÁLNÍ DATABÁZI PRALEX (ZPRACOVÁNÍ PŘÍKLADOVÉ ČÁSTI HESLA)

Marta Koutová, Ústav pro jazyk český AV ČR, v. v. i., oddělení současné lexikologie a lexikografie

ÚVOD

Podrobné zpracování exemplifikace je těžištěm naší práce v rámci výzkumného záměru *Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století.*

Základní materiálové východisko: Český národní korpus ÚČNK

Používaný korpus: ORIG_SYN (dříve pod názvem SYN – obsahuje korpusy SYN2000, SYN2005 a SYN2006PUB). Zahrnuje současné psané texty z oblasti publicistiky, beletrie i odborné literatury.

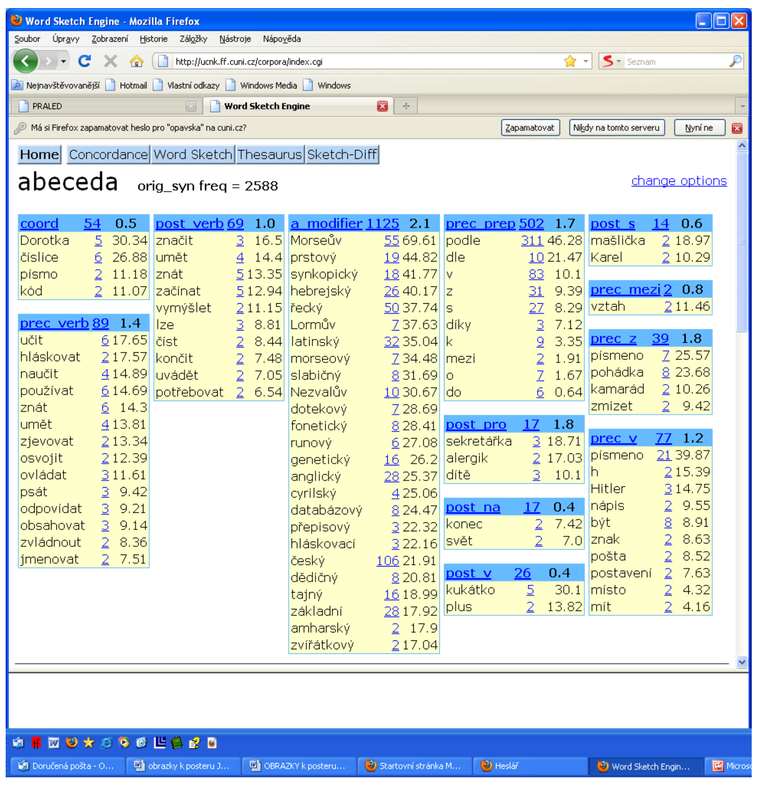
Cíl exemplifikace v LDB Pralex: zachycení spojitelnosti heslových slov v příkladové části hesla – výběr relevantních dokladů, tj. příkladů na konkrétní užití k předem založeným významům podle SSJČ, resp. dalších výkladových slovníků + zachycení dokladů k novým významům.

1 ZACHYCENÍ SPOJITELNOSTI HESLOVÝCH SLOV

- U substantiv se zaměřujeme na syntaktickou i sémantickou spojitelnost.
- Celou škálu spojitelnosti ilustrujeme primárně pomocí necitátových, tzv. upravených dokladů, tj. příkladových spojení ve formě syntagmat, která představují běžné kolokace zpracovávaných substantiv.
- V menší míře uvádíme i větné doklady (citátové).

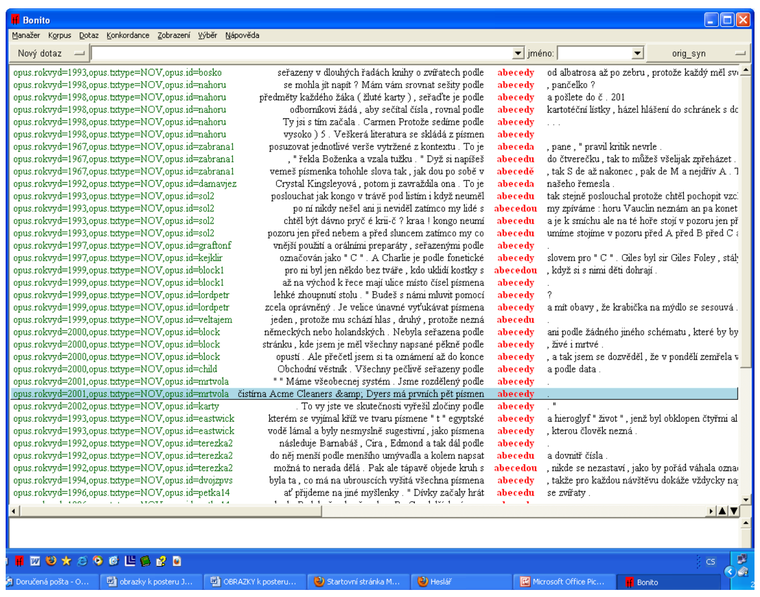
2 PRÁCE S WSE A BONITEM

- Pro třídění dokladů z korpusu využíváme nástroje Word Sketch Engine (WSE) a Bonito.
- WSE používáme pro výběr a podrobné třídění minimálních kontextů k jednotlivým významům.



obr. 1: nástroj WSE

- Z Bonita vybíráme tzv. citátové doklady – v průměru 3–5 větných příkladů s uvedením zdroje.



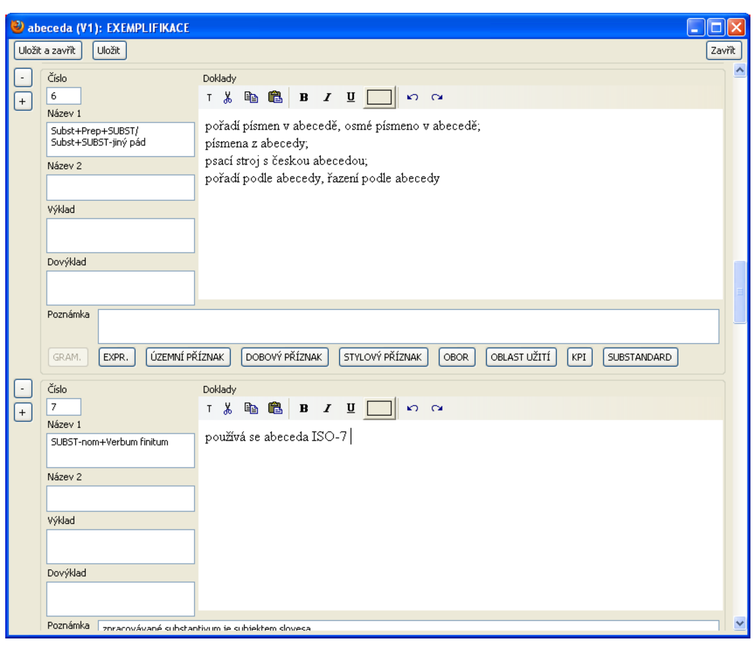
obr. 2: Bonito

3 PRÁCE S KORPUSOVÝMI DOKLADY

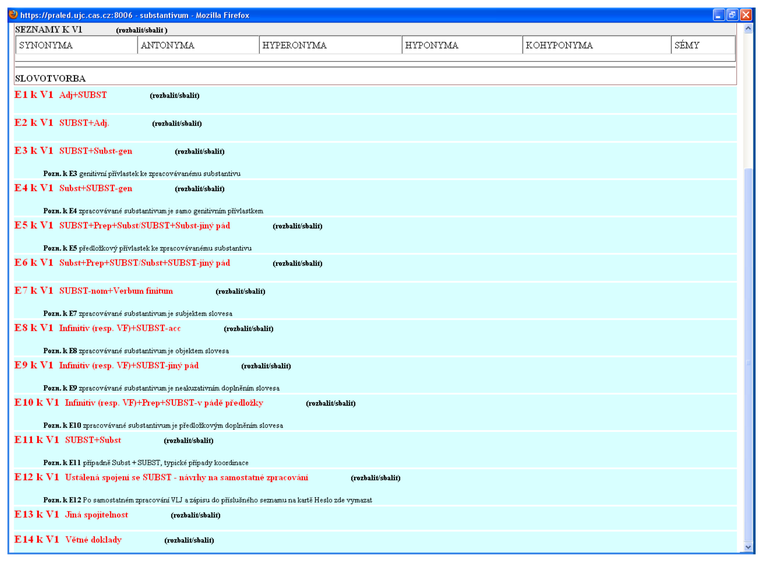
- Vybíráme nejen spojení zcela typická, ale abychom dostali širší obraz sémantické spojitelnosti daného lexému, vybíráme i spojení běžná, v úzu hojně zastoupená.
- V rámci příslušného bloku postupujeme částečně na základě frekvence, částečně na základě sémantiky tak, aby byly vedle sebe uvedeny kolokace z jedné lexikálněsémantické skupiny. Pokud nenajdeme žádné relevantní doklady, příslušný blok zůstane prázdný.
- Rozvití minimálních syntagmat: v některých případech nestačí uvádět minimální syntagmata, ale je nutné uvádět rozvitější doklady, aby příklad dával smysl, popř. aby se přesně vymezilo, ke kterému významu daný příklad patří, srov. např.:
  autodílna: *autodílna v **Mariánské** ulici* (nikoli pouze „autodílna v ulici“);
  houslista: *houslista orchestru **Národního divadla*** (nikoli pouze „houslista orchestru“);
  kachna: *bránit se **tiskovým** kachnám* (k V3: ‚nepravdivá, vymyšlená, senzační zpráva‘).
- Někdy sice není nezbytně nutné rozvití uvádět, avšak pokud vhodně dokresluje dané spojení a v úzu se běžně vyskytuje, můžeme uvést i rozvitější doklady tohoto typu:
  kontrola: *jít na kontrolu **k lékaři***
  auto: *slabost pro **rychlá** auta; stříet s **protijedoucím** autem*

4 CELKOVÁ STRUKTURA EXEMPLIFIKACE U SUBSTANTIV

- Příkladovou část hesla rozdělujeme do 14–15 exemplifikačních bloků. V členění do bloků se projevuje syntaktická spojitelnost daného substantiva, uvnitř jednotlivých bloků je zachycena spojitelnost sémantická.

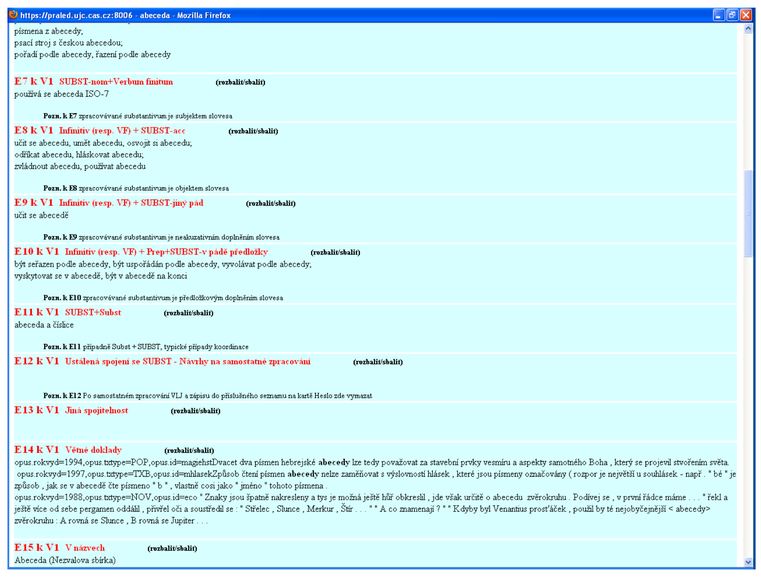


obr. 3: Nástroj Exemplifikace v LDB PRALEX (část formuláře pro zápis dat)



obr. 4: Náhled Exemplifikace v LDB PRALEX (před vyplněním)

- Do exemplifikačních bloků E1–E13 uvádíme výhradně tzv. upravené doklady, tj. minimální nebo někdy i rozvitější syntagmata, která získáváme především pomocí WSE. Pro větné doklady slouží blok E14. Fakultativní blok E15 slouží pro doklady na lemma užité v názvech a přidáváme ho jen v případě potřeby.



obr. 5: Exemplifikace (příkladová část) hesla ***abeceda*** v LDB PRALEX

5 VNITŘNÍ ČLENĚNÍ V RÁMCI JEDNOHO EXEMPLIFIKAČNÍHO BLOKU

- V každém exemplifikačním bloku uvádíme nejprve doklady na běžné, nepříznakové užití. Pokud u nějakého spojení převládá užití v plurálu, uvádíme doklad v této podobě, srov. např.:
  automobil: *majitel automobilu, řidič automobilu*
  obyvatel: *místní obyvatelé, zdejší obyvatelé, tamní obyvatelé*
  podpatek: *vysoký podpatek* i *vysoké podpatky*
- Pokud v korpusu najdeme i doklady na užití odborné, přenesené apod., uvádíme je až po „neutrálních“ dokladech.
- Doklady na odborné užití signalizujeme titulkem „odb.“ (zatím nejde o oborové kvalifikátory, ale jen o vyčlenění jasných případů zvlášť), srov. např. u hesla ***mezera***:
  vzduchová mezera;
  větrací mezera;
  volná mezera, zející mezera;
  úzká mezera, široká mezera;
  dvoucentimetrová mezera;

  odb. *dilatační mezera*

  Ve výjimečných případech používáme v titulku i konkrétnější označení dané oblasti, např. „sport.“ aj.
- Doklady na přenesené užití registrujeme pod titulkem „přen.“, srov. např. u hesla ***četník***:
  přen. *role četníka Balkánu*

  Podobně u hesla ***kometa***:
  přen. *zářivá kometa zápasu; vyhoupnout se mezi komety letošního roku.*
- Někdy může být určité spojení užíváno v přímém i přeneseném významu. V těchto případech dáváme za takový doklad do závorky komentář „(i přen.)“.
- Je-li spojení se substantivem užíváno jako neustálené (v přímém významu) i jako frazém, případně sousloví, pak uvádíme v závorce komentář „(i fraz.)“, případně „(i sousl.)“:
  podpatek: *srazit podpatky* (i fraz.), *sklapnout podpatky* (i fraz.), *scvaknout podpatky* (i fraz.);
  kůň: *bílý kůň* (i fraz.), *černý kůň* (i fraz.);
  skříňka: *černá skříňka* (i sousl.).

6 SPECIFIKA A PŘÍKLADY ZPRACOVÁNÍ JEDNOTLIVÝCH BLOKŮ EXEMPLIFIKACE U SUBSTANTIVNÍCH HESEL

Blok E1 (Adj+SUBST)

- První blok zachycuje spojení zpracovávaného substantiva s anteponovaným adjektivním přívlastkem. V tomto bloku zachycujeme odděleně doklady na singulárové a plurálové užití zpracovávaného substantiva podle toho, co u daných spojení převažuje.
  Jako příklad uvedeme heslo ***čeština***:
  spisovná čeština, obecná čeština, hovorová čeština;
  stará čeština, archaická čeština, obrozenecká čeština, veleslavínská čeština, středověká čeština;
  současná čeština, dnešní čeština, novodobá čeština;
  lámaná čeština, špatná čeština; dobrá čeština, plynná čeština, perfektní čeština, dokonalá čeština;
  dobře srozumitelná čeština, slušná čeština, bezchybná čeština; krásná čeština, kultivovaná čeština;
  jadrná čeština;
  mluvená čeština, psaná čeština;
  znakovaná čeština

Blok E2 (SUBST+Adj)

- Do tohoto bloku dáváme výhradně odborná spojení zpracovávaného substantiva s adjektivem v postpozici typu *konvalinka vonná, kyselina sírová, atribut verbální.*
  Nepatří sem případy, kdy je postpozice ovlivněna stylisticky nebo aktuálním větným členěním, srov. např. u hesla ***ozon***: *troposférický ozon má vůči UV-záření stejné chování jako **ozon stratosférický***.

Blok E3 (SUBST+Subst-gen)

- Do bloku E3 vybíráme příklady na spojení daného substantiva s jiným substantivem v genitivu. Snažíme se přitom o oddělení objektových a subjektových genitivů:
  autocenzura textů x *autocenzura redaktorů;*
  makrofotografie hmyzu x *makrofotografie Jiřího Štreita*

  Příklad zpracování – heslo ***čeština***:
  čeština emigrantů, čeština cizinců;
  čeština Karla Hynka Máchy

Blok E4 (Subst+SUBST-gen)

- v bloku E4 je zpracovávané substantivum samo genitivním přívlastkem.

  Příklad zpracování – heslo ***čeština***:
  znalost češtiny; výuka češtiny, kurs češtiny, studium češtiny, hodina češtiny;
  učitel češtiny, profesor češtiny, lektor češtiny pro cizince, uživatelé češtiny;
  užívání češtiny; výslovnost češtiny;
  slovník češtiny, mluvnice češtiny, učebnice češtiny

Blok E5 (SUBST+Prep+Subst / SUBST+Subst-jiný pád)

- Blok E5 zachycuje předložkové přívlastky ke zpracovávanému substantivu. Do tohoto bloku patří rovněž doklady na spojení zpracovávaného substantiva se substantivem v jiném pádě, srov. např. u hesla ***dárek***: *dárek ženě, dárky dětem.*
- Vazby se stejnou předložkou a stejným pádem dáváme k sobě, převažuje zde tedy formální kritérium nad sémantickým.
- Pokud se však u některých dokladů předložka pojí se stejným pádem, ale sémantika je odlišná, pak rozdělujeme dané doklady na více řádků. Srov. např. u hesla ***automat***:
  automat na nápoje, automat na kávu, automat na jízdenky;
  automat na mince, automat na karty

  Příklad zpracování – heslo ***čeština***:
  čeština pro cizince, čeština pro samouky;
  čeština na univerzitě, čeština na gymnáziu;
  čeština ve škole;
  čeština se silným přízvukem

Blok E6 (Subst+Prep+SUBST / Subst+SUBST-jiný pád)

- Do bloku E6 dáváme spojení, kde se zpracovávané substantivum vyskytuje jako předložkový přívlastek k jiným substantivům nebo se s jinými substantivy pojí jako přívlastek dativní či instrumentální (např. ***arterie***: *průřez arterií*).

  Příklad zpracování – heslo ***čeština***:
  překlad do češtiny, převod do češtiny;
  překladatel do češtiny;
  překladatel z češtiny; překlad z češtiny do němčiny;
  zkouška z češtiny, maturita z češtiny, písemka z češtiny, diktát z češtiny, známka z češtiny;
  učitelka na češtinu;
  návod v češtině, text v češtině, nápis v češtině, vysílání v češtině;
  vztah k češtině, láska k češtině;
  prohřešek proti češtině;
  zájem o češtinu;
  cit pro češtinu;
  problémy s češtinou, zacházení s češtinou

Blok E7 (SUBST-nom+Verbum finitum)

- V bloku E7 je zpracovávané substantivum subjektem slovesa.

  Příklad zpracování – heslo ***čeština***:
  ze všech stran zněla čeština;
  úředním jazykem je čeština

Blok E8 (Infinitiv (resp. VF)+SUBST-acc)

- V bloku E8 je dané substantivum objektem slovesa.

  Příklad zpracování – heslo ***čeština***:
  učit češtinu;
  studovat češtinu, učit se češtinu;
  ovládat češtinu, užívat češtinu;
  slyšet češtinu;
  komolit češtinu, lámat češtinu, prznit češtinu

Blok E9 (Infinitiv (resp. VF)+SUBST-jiný pád)

- Blok E9 zachycuje případy, kdy zpracovávané substantivum je neakuzativním doplněním slovesa, tj. vyskytuje se v genitivu, dativu či instrumentálu.

  Příklad zpracování – heslo ***čeština***:
  učit se češtině, rozumět češtině;
  mluvit spisovnou češtinou, hovořit lámanou češtinou, vládnout krásnou češtinou;
  kniha je psána krásnou češtinou

Blok E10 (Infinitiv (resp. VF)+Prep+SUBST-v pádě předložky)

- V bloku E10 je zpracovávané substantivum předložkovým doplněním slovesa.
- Slovesa v blocích E8–E10 uvádíme většinou v infinitivním tvaru. Pokud je ale důležitý původce děje, je nutné uvádět daná slovesa ve finitním tvaru, srov. např.:
  veřej: *vrata vrzala ve veřejích, dveře skřípou ve veřejích* (tj. nikoli pouze *skřípat ve veřejích, vrzat ve veřejích*).

  Příklad zpracování – heslo ***čeština***:
  přeložit knihu do češtiny, překládat do češtiny, tlumočit do češtiny, převést text do češtiny;
  překládat z češtiny, přecházet z češtiny do angličtiny, maturovat z češtiny;
  kniha vyšla v češtině;
  zápasit s češtinou, válčit s češtinou, být s češtinou na štíru

Blok E11 (SUBST+Subst, případně Subst + SUBST)

- V bloku E11 je naším úkolem uvádět typické příklady koordinace. Zapisujeme sem především koordinace se spojkami *a* a *nebo* (protože jiné nám WS nezobrazí), popř. též se spojkou *i* nebo s předložkou *s*, jako např. u lemmatu ***otec***: *otec s matkou; matka s otcem.*

  Příklad zpracování – heslo ***čeština***:
  čeština a angličtina, čeština a slovenština; směsice češtiny a němčiny

Blok E12 (Ustálená spojení se SUBST – návrhy na samostatné zpracování)

- Tento blok není na rozdíl od bloků E1–E11 a E13 vymezen syntakticky. Jde o blok, do kterého zpracovatelé substantiv zatím pracovně zapisují ta ustálená spojení se zpracovávaným substantivem, která identifikují v korpusu a na základě lexikografického kritéria potřebnosti výkladu navrhují k samostatnému zpracování.

Blok E13 (Jiná spojitelnost)

- Tento blok zatím slouží jako sběrný koš pro případy, které se syntakticky nehodí do žádného jiného bloku. Ukázky různých spojení, která patří do E13, lze nalézt např. u hesla ***puchýř***:
  nohy plné puchýřů; puchýře naplněné vodou; látka vyvolávající na kůži puchýře; opar ve stadiu puchýře;
  mít dlaně samý puchýř

Blok E14 (Větné doklady)

- Na rozdíl od předchozích bloků zachycuje blok E14 celé větné doklady z korpusu, tj. příklady rozvitější, často i takové, které díky širšímu kontextu vysvětlují a ilustrují méně jasná spojení z ostatních exemplifikačních bloků.

  Příklad zpracování – heslo ***čeština***:
  opus.rokvyd=2003,opus.txtype=SCI,opus.id=ener0302 Konference bude vedena v angličtině a < češtině>.
  opus.rokvyd=1993,opus.txtype=PUB,opus.id=lnd93161 V Tatrách zní letos < čeština> sporadicky.
  opus.rokvyd=1995,opus.txtype=PUB,opus.id=mf950102 V < češtině> zní vaše jméno podobně jako kuře.

Blok E15 (V názvech)

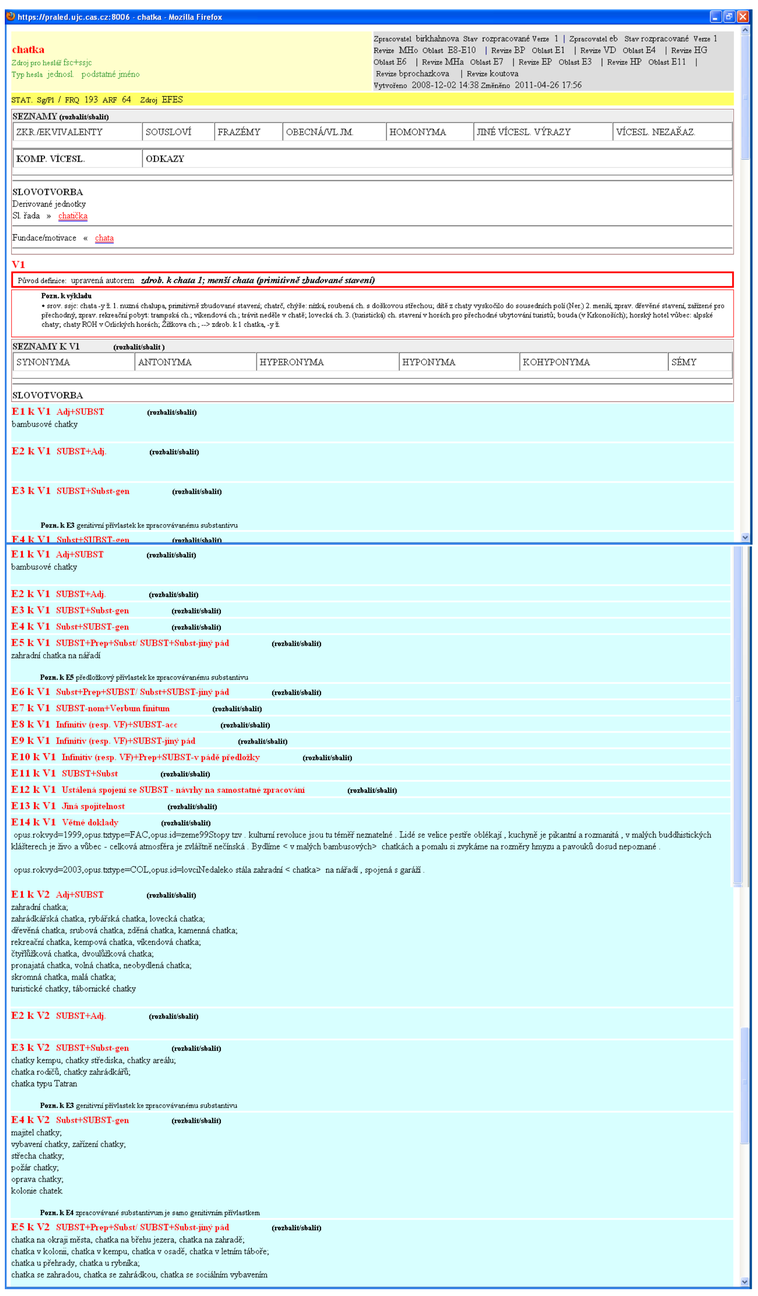
- Do bloku E15 dáváme příklady na užití daného substantiva v nejrůznějších názvech. Tento blok v základní nabídce exemplifikační struktury substantiv není, přidáváme jej fakultativně – pouze v případě, že nějaké relevantní doklady v korpusu nalezneme.

  Příklad zpracování – heslo ***čeština***:
  Nová slova v češtině: slovník neologizmů;
  Příruční mluvnice češtiny;
  Slovník spisovné češtiny pro školu a veřejnost

7 ČLENĚNÍ NA VÝZNAMY NA ZÁKLADĚ PRIMÁRNÍCH A REFERENČNÍCH ZDROJŮ

Vyhodnocení celé sémantické struktury polysémních hesel a nové formulace výkladových definic budou až úkolem nového výkladového slovníku; pro potřeby LDB Pralex zatím pracovně používáme výklady ze Slovníku spisovného jazyka českého (SSJČ) a případně i z dalších výkladových slovníků. Pokud v korpusu najdeme doklady pro významy nezachycené ve slovnících, nebo zaznamenáme nějaké změny vůči předchozímu zpracování v nich, založíme další význam. Polysémní hesla tedy primárně členíme na významy ve shodě se SSJČ, při práci s korpusem a dalšími slovníky někdy doplňujeme další významy.

- Např. u substantiva ***delegát*** jsme zachytili na základě výskytu v korpusu ORIG_SYN další dva významy:
  V2: ‚pracovník cestovní kanceláře pečující o rekreanty v místě jejich (zahraničního) pobytu‘ (*delegát cestovní kanceláře, delegát Čedoku*);
  V3: ‚osoba vyslaná fotbalovým svazem a pověřená dohledem nad regulérností hodnocení utkání‘ (*fotbalový delegát; svazový delegát; delegát zápasu, delegát utkání; delegát svazu*).
- Zpracování v SSJČ porovnáváme se zpracováním v SSČ, ASCS aj., a pokud tyto slovníky zachycují oproti SSJČ další významy, v relevantních případech na základě těchto novějších výkladových slovníků a doloženosti v korpusu významovou strukturu zpracovávaného hesla doplňujeme.
- Další význam někdy vymezujeme na základě dokončeného procesu lexikalizace, která byla v SSJČ jen naznačena. Pokud bylo registrováno v SSJČ nějaké přenesené užití nebo významový odstín a dnes je tento významový posun v korpusu hojně doložen, může to být signálem, že se ustaluje další samostatný význam. Srov. např. u hesla ***buňka*** – v SSJČ uvedeno u 2. významu jako „zř. přen.“ ‚malá místnost, jedna z mnoha stejných‘; v LDB vyčleněno jako další význam (V5): ‚malá místnost, jedna z mnoha stejných; dočasná stavba menších rozměrů sloužící konkrétním účelům‘:
  stavební buňka, pracovní buňka, prodejní buňka; *montážní buňky;*
  rodinná buňka, obytná buňka; *mobilní buňky;*
  sanitární buňka, servisní buňka; *panelové buňky*
  sklepní buňka, skladovací buňka;
  provizorní buňka;
  prefabrikovaná buňka, montovaná buňka, smontovaná buňka;
- Pokud najdeme jen několik dokladů, které signalizují nové významové posuny, na jejichž základě se zatím nedá nový význam přesně vymezit, uvádíme je v poli „Doklady pro nové významy“, aby bylo možné je později ověřit dohledáním v dalších zdrojích a vyhodnotit, zda jde opravdu o další význam. Všechny tyto signály budou využity při práci na novém výkladovém slovníku.



obr. 6: Příklad polysémního hesla – ***chatka***