

Jazyková potencialita: studium na bázi hapaxů legomenon

Renata Novotná
Ústav Českého národního korpusu FF UK,
renata.novotna@ff.cuni.cz

ABSTRACT: The paper deals with word forms occurring in the SYN corpus (1.2 billion words) only once, twice or three times, the so-called hapax legomena, which provide a basis for the study of potentiality in language. As the material was very large, only 20 samples were chosen, each containing 3 000 forms, i. e. 60 000 forms overall. Approximately 50 % of word forms were various mistakes, especially typing errors, or words from other languages. Therefore only the remaining 30 000 word forms were selected as the basis for this study. The analysis showed that the most significant suffixes for hapax legomena are *-ovský* (e. g. *jimmyreedovský*), *-ák* (e. g. *medvěďák*), *-ista* (e. g. *havlista*), *-ing/-ink* (e. g. *gardening*, *dancink*), *-ovitý* (e. g. *kladivovitý*), type *po vojensku*, diminutives derived from abstracts (e.g. *minulůstka*) and names of women professions (e. g. *meduprodavačka*). Moreover, compound words with the first parts *dlouho-* (e. g. *dlouhorožec*), *gala-* (e. g. *galamenu*), *jino-* (e. g. *jinomluva*), *kino-* (e. g. *kinofajnšmekr*), *mega-* (e. g. *megakatastrofa*), *nízko-* (e. g. *nízkohlučný*), *polo-* (e. g. *poločitelný*), *video-* (e. g. *videokomentář*) were typical for new words.

1 Hapax legomena a jazyková periferie

Tento příspěvek se orientuje na oblast tzv. hapaxů legomenon, tedy jazykových jednotek, které mají v korpusu jeden výskyt (řec. hapax – jednou, legomenon – řečené). Na rozdíl od běžného korpusového přístupu, který se zaměřuje na jednotky frekventované, z centra jazykového systému, tu jde o opačný přístup zabývající se jazykovou periferií. Jde ovšem i o poněkud rehabilitované chápání tohoto pojmu na rozdíl od tradice nastolené diachronní lexikografií. M. Dokulil užívá starší termín produktivnost a uvádí, že stanovení produktivity může být dvojitý: a) přímé – zkoumá se tvořivá potence slovo tvorného prostředku nebo typu, b) nepřímé – zkoumá se realizace této potence (Dokulil, 1962; 79).

1.1 Výchozí materiál

Materiálově příspěvek vychází ze vzorku tvarů, které byly v korpusu SYN o rozsahu 1.2 miliardy slov zachyceny 1-3 výskyty. Tento rozsah byl zvolen z toho důvodu, že záměrem bylo podchytit tvary s nízkými frekvencemi obecně, nejen pouze s jedním výskytem, tj. tzv. hapaxy a semihapaxy (Čermák, 2010; 163). V celém korpusu SYN je celkem 3008111 hapaxů, tj. tvarů, které se vyskytly jen jednou. Jde téměř o 49.4 % z celkového počtu různých slov v korpusu SYN, kterých je 6088605¹. Tento počet zahrnuje jak tvary složené z písmen, tak nejruznější tvary obsahující číslice a další znaky. Z tohoto důvodu byl pro účely výzkumu jazykové potenciality materiál pročištěn tak, aby jeho základem byla pouze slova obsahující písmena. Tímto postupem bylo získáno celkem 1.3 milionu těchto tvarů, tzv. semihapaxů (tj. zhruba jedna tisícina výchozího korpusu SYN). S ohledem na tak velké množství, které nelze manuálně prozkoumat, byly provedeny pouze dílčí sondy, konkrétně

¹ Tak např. Primož Jakopin (2007) uvádí, že "Hapax legomena amount to 49,3%, but it is nevertheless probably the most complete source of neologisms in Slovenian".

20 sond o rozsahu 3000 tvarů, celkem tedy bylo podchyceno 60000 tvarů. Z tohoto počtu tvořily necelých 50 % chybné tvary (zvl. překlepy), cizojazyčné citátové tvary apod. - celkem 29986 tvarů. Základem pro studium jazykové potenciality, které hapaxy často představují, se proto stalo 30014 tvarů, které odpovídají 15840 lemmatům (tato lemmata byla hapaxům přiřazována ručně, nejde tedy o lemmata přiřazená hapaxům v korpusu), ve slovníku ovšem spíše obecně nedoloženým. Je však třeba upozornit, že v korpusu SYN se sledovaly pouze tvary (nikoli lemmata) o 1-3 výskytech, tj. odpovídající lemma může mít více výskytů než 3, tak např. lemma *kinový* je zachyceno celkem 5x, a to ve tvarech *kinovém* (1x), *kinovému* (1x), *kinovou* (1x), *kinový* (1x) a *kinovým* (1x).

1.2 Tvary zařazené do materiálu

Do materiálu byly zahrnuty jen smysluplné lexikální jednotky, a to následující skupiny:

- a) nově utvořená slova, zejména termíny, např. *divakcína*, *rudokřídly*, popř. běžná slova – *klackovatět*
- b) cizí slova, která se skloňují nebo časují, popř. jsou jinak začleněna do české slovní zásoby – např. *kinges* (*pozvat dva kingesy na fet*), *swordfish* (*na britských lodích pokládali swordfishe za německé bombardéry*)
- c) pravopisné varianty typu *kinderčokoláda* – *kindrčokoláda*
- d) varianty interjekcí – např. *hehé*, *he-hé*, *hehééé*

1.3 Tvary vyloučené z materiálu

Jak bylo naznačeno už výše, řada tvarů nebyla do výsledného materiálu zahrnuta. Šlo o následující skupiny:

- a) překlepy – např. *miohou* místo *mohou*, *vícmělákala* – *víc mě lákala*
- b) chybný pravopis – např. *rozhavené* místo *rozžhavené*, *heidrichiáda* místo *heydrichiáda*
- c) odchylky od výslovnosti, např. *svítáááá* místo *svítá*, *klasaveček* místo *krasaveček*
- d) přepis cizí výslovnosti – např. *vičtaun* (tj. which town), *Palé ruajal* (tj. Palais Royal)
- e) citátová cizí slova (včetně názvů cizí reality) – např. *jedná se o mláďata ovce vřesovištní* – *heideschnucke*; *calliphora vicina*; *en viden kjent fjord*; *Sarajevski svijet*; *Specification for classification system for vulcanized rubbers*; *kinsmarauli* (gruzínské červené víno)
- f) nářeční slova – např. *povřiuslo*, *odmlúvat*, *minota*
- g) staročeská slova – např. *ruoznicě*, *klapý*, *hejzúk*
- h) slova z e-mailových a internetových adres, např. *autoradce* – www.autoradce.cz, *heimat-saaz* – www.heimat-saaz.de aj.

2 Hapax legomena a jejich studium

Hapax legomena jsou definována jako slova, která se vyskytují jen jednou v daném textu nebo v souboru textů. V korpusové lingvistice jde o slovo, které je zachyceno jen jednou v určitém korpusu. Hapax legomena byla většinou zkoumána v diachronním pohledu, a to jako zpravidla obtížně přeložitelná slova, vyskytující se jen jednou ve všech literárních památkách určitého, zejména mrtvého jazyka. Zkoumání mrtvých jazyků je omezeno na zachované literární památky a významy jednotlivých slov se pak dají určovat jen podle jejich různých kontextů a použití, případně podle etymologie a slov podobných. Zkoumání hapax legomen je z tohoto hlediska důležité například v biblistice, při textové kritice a

překladech biblických textů. V korpusové lingvistice se využívá relativní frekvence hapaxů, na jejímž základě se zjišťuje variantnost slovníku. Existují i jiná využití počtu hapaxů v kvantitativní lingvistice, např. Good-Turingův odhad apod. Analýza hapaxů může být využita rovněž ve forenzní lingvistice, zejména při určování autorství nebo v případech podezření z plagiátorství aj.

3 Hapax legomena v české lexikografii

Pokud jde o lexikografii obecně, Hartmann-James (1998; 67) uvádějí, že jde o slovo, které se v korpusu objevuje jen jednou a které v důsledku toho není zahrnuto do slovníku. Zajímavý je přístup k tzv. hapaxům v českých výkladových slovnících, zejména v Příručním slovníku jazyka českého, který vycházel v letech 1935-1957. A. Získal (1938) uvádí, že "kritické stanovisko nutno zaujmouti i k slovům ojedinele v materiále doloženým (hapax legomena). Přitom však je potřebí rozeznávat, zdali hapax legomenon je slovo odvozené od běžného základu a běžnou odvozovací příponou, tedy slovo sice málo doložené, ale významu každému hned zřejmého, či zdali je to slovo bez těchto vlastností. V tomto druhém případě je potřebí větší opatrnosti již také proto, že může jíti o chybu tiskovou nebo o omyl spisovatelův." V předmluvě k tomuto slovníku na str. VIII, (díl I, 1935-1937) se zdůrazňuje, že "tak zv. hapax legomena, tj. slova jen jednou v našich sbírkách doložená, se zpravidla vynechávají". Slovník spisovného jazyka českého však hapaxy (mají-li 1-3 výskyty v kartotéce lexikálního archivu ÚJČ) uvádí a označuje je hvězdičkou, např. *peřatý*, *podlesk*, *ropotavý*, *rozkrajovač*, *skřehotavec* aj.

4 Hapax legomena a jejich zastoupení v textech

Jak uvádí M. Těšitelová, "u plnovýznamových slov, u nichž se výrazně uplatňují, tj. u substantiv, adjektiv, sloves a adverbií, představují minimálně 39 % - maximálně 66 % všech příslušných lexikálních jednotek v textu" (1974; 67). Počet hapaxů je však závislý na celkové velikosti korpusu, v daném případě šlo o korpus, který byl podkladem frekvenčního slovníku Frekvence slov, slovních druhů a tvarů v českém jazyce, který měl rozsah cca 1.6 milionu slov. Dále Těšitelová píše, že v uměleckém stylu je počet slov s frekvencí 1 vždy vyšší než v slovníku textů naučných. Pokud jde o lexikální jednotky s frekvencí 1-10, ty jsou těžištěm tzv. rozptýlení slovníku, resp. bohatství slovníku v užším slova smyslu. Toto pásmo představuje u plnovýznamových druhů slov maximálně 97 %, minimálně 77 % slovníku. Plnovýznamová slova s frekvencí 1-10 mají nejvýznamnější postavení ve slovníku kteréhokoli textu bez zřetele k funkčnímu stylu i k autorovi.

5 Hapax legomena v korpusu SYN

Na základě výše uvedených 20 sond z 1.3 miliónu tvarů (viz 0.1) se ukázalo, že centrem vzniku nových pojmenování v rámci jazykové potenciality jsou především některé frekventované sufixy, např. *-ovský*, *-ovitý*, *-ing/-ink* apod. a dále některé části kompozit, např. *dlouho-*, *mega-*, *video-* aj.

5.1 Zastoupení frekventovaných sufixů

Abychom však mohli posoudit zastoupení sufixů typických pro tzv. hapaxy, je třeba nejprve zjistit, jak jsou procentuálně u hapaxů zastoupeny frekventované sufixy z centra slovní zásoby. V daném případě jde zvláště o sufixy *-ový*, *-ka* a *-ovat*:

Sufix	Počet tvarů	Procenta	Počet lemmat	Procenta
<i>-ový</i>	2328	7.7	1067	6.7
<i>-ka</i>	1688	5.6	800	5
<i>-ovat</i>	724	2.4	367	2.3

Tabulka 1: Zastoupení frekventovaných sufixů

Ukazuje se tedy, že frekventované sufixy se pohybují v pásmu 2 – 8 % z celkového materiálu, jak uvidíme dále, u sufixů typických pro oblast tzv. hapaxů jde zpravidla o méně než 2 %.

5.2 Suffixy typické pro hapax legomena

Jak bylo uvedeno už výše, pro tzv. hapaxy jsou charakteristické následující sufixy:

Sufix	Počet tvarů	Procenta	Počet lemmat	Procenta
<i>-ovský</i>	636	2.1	321	2.0
<i>-ák</i>	321	1.0	127	1.0
<i>-ista</i>	281	0.9	147	0.9
<i>-ing</i>	227	0.7	124	0.8
<i>-ovitý</i>	147	0.5	98	0.6

Tabulka 2: Zastoupení sufixů typických pro hapax legomena

5.2.1 Sufix *-ovský*

Adjektiva s tímto sufixem jsou v rámci nashromážděného materiálu tvořena převážně od proprií. Celkem lze odlišit dva typy, jeden tvořený od toponym, např. *stodůlkovský*, jeden od antroponym, např. *jimmyreedovský*. V některých případech je proprium, které tvoří odvozovací základ adjektiva, doplněno ještě prefixem, např. *antibriandovský*, *powitoldovský*, resp. jde o kompozita složená z adjektiv stejného typu, tj. např. *klausovskomachovský*, *havlíčkovsko-peroutkovsko-vaculíkovský* apod. Pokud jde o adjektiva s tímto sufixem neodvozená od proprií, patří k nim např. *dívkovský*, *poutníkovský*, *divousovský*, *mission-impossiblovský* apod. V případě některých toponym však jde pouze o sufix *-ský* připojený ke kmeni zakončenému na sufix *-ov*, např. *Světlov – světlovský*.

5.2.2 Sufix *-ák*

Tento sufix je typický pro obecněčeskou slovní zásobu. Jako příklady lze uvést *mirindák* (utvořeno od nápoje Mirinda), *medvěďák* (několikadenní spánek), *odkladák* (domov důchodců), *klafák* (velký zub), *avéhák* (utvořeno od zkratky ÁVH Államvédelmi Hatóság – obdoba Státní bezpečnosti v 50. letech v Maďarsku) aj. Některé hapaxy jsou sémanticky nejasné, protože jediný kontext nemůže poskytnout dostatečnou informaci o významu daného lexému, např. *renčák* (zřejmě peníze)

(1) Jurka dostal za pušku deset tisíc renčáků,
rundák

(2) Tvé a mé klíma, tedy pomyslná krajina či rundák, před kterým jsme se producírovali, obsahuje zároveň jihočesko-sázavskou dumavost, moravskou buclatost, slovanské furiantství, pražsko-pepickou skepsi - leč nadto naši společnou latinskost...
stoleďák

(3) Střeleckého klání se zúčastnily klub vojenských důchodců a vojenské útvary z jihu Čech, Policie ČR a písecký klub vojáků v záloze. Hodnocení bylo obdobné, jako u stoleďáků. Hranice 55 let, co rok mladší, to minus jeden bod a naopak starším se bod připisoval...

5.2.3 Sufix *-ista*

Také v případě sufixu *-ista* nacházíme v materiálu slova odvozená od proprií, např. *antibushista*, *rumlista*, *havlista*, *havlofašista*, *antibenešista*, *klausooptimista* aj. Často se v materiálu objevují kompozita, např. *autor-kurzista*, *dárce-oportunista*, *poutník-turista* apod. Většina materiálu rozšiřuje počet substantiv označujících názvy profesí, popř. dalších činností, např. *deathmetalista*, *galérista*, *heligónista*, *kinopianista*, *klapkostrójista*, *rumunista*, *videoartista* aj. Častý je prefix *anti-*, který je v jednom případě užít zdvojeně, tj. *antiantikomunista*.

5.2.4 Sufix *-ing/-ink*

Doložený materiál v rámci hapaxů ukazuje, že počestěný pravopis *-ink* je v menšině (15 lemmat ku 120), ačkoliv ve většině českých výkladových slovníků převládá právě tento počestěný pravopis, tedy *-ink*. S počestěným pravopisem se objevují např. lexémy *autotrénink*, *galamítink*, k dalším lexikálním jednotkám patří např. *dancink*, *darlink*, *reportink*, *stopink* aj. S původním anglickým pravopisem byly mezi hapaxy zachyceny jednotky *autoscaling* (užito i v gen. *autoscalingu*), *gardening* (zachycen i gen. *gardeningu*), *kiteboarding* (kromě nominativu se objevily i tvary gen. *kiteboardingu* a instr. *kiteboardingem*), *repackaging* (zachycen gen. *repackigingu*), *videomonitoring* (zachycen gen. *videomonitoringu*). V Pravidlech českého pravopisu (1993) je např. uveden lexém *tramping* s pravopisem *trepink* i *tramping*. V korpusu SYN je však doložen pravopis *tramping* 771 výskyty, zatímco pravopis *trepink*, popř. *trampink* pouze 41 výskyty. Takových případů najdeme více. Pravidla pravopisu tento typ tedy poměrně uměle a zbytečně dělí na dvě skupiny. M. Jelínek (Daneš-Dokulil-Kuchař 1967; 640) uvádí, že nejvíce substantiv tohoto typu bylo přejato pro pojmenování různých sportovních činností, v současné době už je však škála slov s tímto sufixem velmi široká, zejména v ekonomii, např. *timesharing*, *daytrading*, *headhunting*, v technice, např. *klastering*, *repowering*, *autotuning* aj.

5.2.5 Sufix *-ovitý*

Tento sufix vyjadřuje podobnost se substantivem v základu adjektiva a je rovněž typický pro oblast tzv. hapaxů. Adjektiva se sufixem *-ovitý* jsou odvozena jednak od názvů věcí, např. *kladivovitý*, *rukavicovitý*, *utěrkovitý* aj., dále od názvů rostlin, např. *jiřinkovitý*, *meduňkovitý*, *resedovitý*, *rulíkovitý* aj., od názvů zvířat, např. *polokoryšovitý*, *meduzovitý*, *stonožkovitý* aj. Objevují se i kompozita typu *jílovcovito-pískovcovitý*.

5.3 Kategorie modifikační

Vedle konkrétních sufixů z kategorie mutační je pro oblast hapaxů legomenon příznačný nárůst pojmenování ve dvou modifikačních slovtvorných kategoriích – u zdrobnělin a slov přechýlených. Zdrobněliny, např. *haveloček* mají 743 tvarů (2.5 %) a 377 lemmat (2.4 %), slova přechýlená, např. *heligonkárka* mají 328 tvarů (1 %), 238 lemmat (1.5 %).

5.3.1 Zdrobněliny

Jak uvádí L. Doležel (Daneš-Dokulil-Kuchař 1967; 494), "počet deminutiv je neuzavřený, existuje možnost prakticky od každého substantiva utvořit deminutivum". Zdrobněliny v rámci hapaxů jsou dost často doplňovány prefixem *mini-*, např. *minitovárnička*, *minitraktůrek*, *minitrapásek*, *miniturnajík*, *minivázička*, *minivláček*, *minizahrádečka* aj. Zajímavé je tvoření zdrobnělin od abstrakt, např. *anketka*, *avantgardička*, *avantýrka*, *jistotka*, *jízlivůstka*, *jobíček*, *minulůstka*, *nitřičko*, *odměnka*, *polonápadíček*, *poutička*, *povodnička*, *rentička*, *reportážka*, *reprizička*, *růček*, *stoletíčko*, *sviňárnička*, *svobodička*, *swingůvka*, *utěšeníčko*, *utrpeníčko*, *úvodek*, *úvažka* aj. K méně obvyklým zdrobnělinám patří dále názvy

profesí, např. *autůrek*, popř. *autoreček*, *hejtmánek*, *jogínek*, *klauniček*, názvy staveb a místností, např. *garážička*, *klášteříček*, *stolárnička*, *svobodárnička* a některé další, např. *angreštík*, *automátek*, *automobílek*, *divadlíčko*, *gangsteřík*, *helikoptérka*, *jilmeček*, *jogurtíček*, *kinečko*, *kláseček*, *klauňátko*, *medík*, *medvídeček*, *megahrníček*, *megasrdéčko*, *mistráček*, *poušťka*, *poutíčko*, *površek*, *reproduktůrek*, *ručinka*, *ruinka*, *rumík*, *vietnamčátka* aj. Podle daných kontextů se zdá, že tu převládá pragmatický, evaluativní význam nad kvantitativním (tj. *hejtmánek* není malý *hejtman*).

5.3.2 Slova přechýlená

V rámci přechýlených slov upoutají pozornost zejména netypické názvy profesí, např. *angioložka*, *automechanika*, *autozávodnice*, *aviatička*, *galéristka*, *jiřinářka*, *kinařka*, *klaunka*, *meduprodavačka*, *místohejtmanka*, *místorychtářka*, *nimrodka*, *odhřebíkovávačka*, *odlišťovatelka*, *politručka*, *stomaterapeutka*, *stomatochirurgka*, *svíčkářka*, *swingerka*, *utišovatelka*, *úvěrářka*, *videodýdžejka*, *videoumělkyně*, *vietnamistka*, *vikariátka*, *viktimoložka* aj. Vedle domácích přechýlenin je tu významná složka i u termínů cizího původu.

5.4 Typ *povojsku*

Tento typ je častěji psaný zvlášť, tj. např. *po jiráskovsku*. V rámci tvarů (resp. tokenů) vyskytl 56x (0.2 %) a v rámci lemmat (resp. typů) 41x (0.25 %). Jeho procentuální zastoupení sice není vysoké, ale vykazuje velký nárůst v porovnání s centrem slovní zásoby. Je typicky tvořen od proprií, a to jak od antroponym, např. *po dantovsku*, *po divišovsku*, *po gandalfovsku*, *po gándhiovsku*, *po havlíčkovsku*, *po hawkingo-blumenfeldovsku*, *po jimovsku*, *po jiránkovsku*, *po jiráskovsku*, *po jirkovsku*, *po kingovsku*, *po kiplingovsku*, *po kissovsku*, *po renčínovsku*, *po rumcajsovsku*, *po rumlovsku*, *po stoneovsku*, *po svorovsku* aj., tak od toponym, např. *po avignonsku*, *po galicijsku*, *po hebrejsku*, *po nižněgorodsku* aj. K nepropriálním tvarům patří *po gádžovsku*, *po gangstersku*, *po havířsku*, *po házenkářsku*, *po klaunsku* aj., vždy se však týká osob.

5.5 Kompozita

Pokud jde o kompozita, charakteristická jsou především kompozita typu *útočník-střelec*, *divadlo-dokument*, které představují 629 tvarů (2 %) a 485 lemmat (3 %). Z dalších kompozit vybíráme ta, která byla v rámci sond zachycena jako celek, tj. např. kompozita s první částí *dlouho-* od *dlohobrtnický* po *dlouhoživotný*:

polo-, např. *poločíselný*: 2285 tvarů (7.6 %), 1270 lemmat (8.0 %)

Tato kompozita jsou jak adjektivní, např. *poloalkoholický*, *polodementní*, *polokřesťanský* apod., tak substantivní, např. *polodivoch*, *polohladovka*, *polokůň*, ojediněle tvořená i od adverbii, např. *poloasociálně*, *pololežmo-polosedmo*, *polomaradonovsky*, popř. od verb, např. *pololikvidovat*, *polonaplnit*, *polonést* aj.

mega-, např. *megakatastrofa*: 1320 tvarů (4.4 %), 669 lemmat (4.2 %)

U kompozit s první částí *mega-* se vyskytují jen adjektiva a substantiva, např. *megahambinec*, *megaklokan*, *megaohňostroj* apod., *megabohatý*, *megahvězdný*, *megapopový* aj. Objevují se i kompozita složená z více částí, např. *megahypersuperultrafilm*, *megahyperultrasuperninja*.

video-, např. *videokomentář*: 1164 tvarů (3.8 %), 595 lemmat (3.7 %)

Kompozita s první částí *video-* jsou typická především pro substantiva, např. *videoantologie*, *videofibrogastroskop*, *videomaniak* aj., v některých případech je sémantická souvislost části *video-* s druhou částí kompozita poněkud nejasná, např. *videomašinka*,

(4) *Všechny ty bary a putyky, rozhozené v obrovském údolí, mají čtyři věci – televizi, kulečnick, gramofon a videomašinky.*

videonáhrobek

(5) *Barrows si už nechal zaregistrovat patent videonáhrobku. Podle časopisu New Scientist ale není prvním vynálezcem, který přišel s elektronickým náhrobním kamenem.*

videoskaut:

(6) *Na práci s videem po zápasech a různé rozbory jsem byl zvyklý z Evropy. Ale v Golden State máme speciálního videoskauta. Když jsme přišli o poločase do šatny, už jsme měli připravený sestřih z první půle.*

nízko-, např. *nízkohlučný*: 411 tvarů (1.4 %), 212 lemmat (1.3 %)

Kompozita s první částí *nízko-* jsou převážně adjektivní, např. *nízkokořenící, nízkoobsažný, nízkotukový*, substantiva jsou zpravidla odvozena od adjektiv, např. *nízkopodlažní – nízkopodlažnost*.

kino-, např. *kinofajnšmekr*: 272 tvarů (0.9 %), 142 lemmat (0.9 %)

Hapaxy s první částí *kino-* mají vedle dokladů typu *kinoarchív, kinobar, kinoherec, kinopokladna, kinotýdeník* aj. i některá méně obvyklá kompozita, např. *kinohltavec, kinokrál, kinozámek* apod.

dlouho-, např. *dlouhorožec*: 222 tvarů (0.7 %), 110 lemmat (0.7 %)

Podobně jako u kompozit s první částí *nízko-* i v tomto případě převládají adjektiva, např. *dlouhočnělečný, dlouhokvetoucí, dlouhotažný*, substantiva jsou v menšině a jde zejména o termíny, např. *dlouhohrozen, dlouhonosec, dlouhošijka* aj.

gala-, např. *galamenu*: 130 tvarů (0.4 %) , 70 lemmat (0.4 %)

Kompozita s první částí *gala-* jsou především substantivní, např. *galalivrej, galapokrm, galavolejbal, galavystoupení*, z adjektivních byly zachyceny *galaindustriální, galaoperetní, galavečerní* aj.

jino-, např. *jinomluva*: 111 tvarů (0.37 %), 55 lemmat (0.35 %).

Ke kompozitům s první částí *jino-* patří jak substantiva, např. *jinokrajec, jinořeč, jinověrství* aj., tak adjektiva, např. *jinobytný, jinosvětský, jinožánrový* aj.

Vedle uvedených kompozit se v materiálu objevila i kompozita adjektivní typu *ruskojaponský* nebo *rudo-zeleno-bílý* - celkem 751 lemmat, tj. 4.7 % materiálu, 1338 tvarů, tj. 4.5 % materiálu. Nejvíce kořenů mělo kompozitum *hebrejsko-řecko-latinsko-anglicko-německo-maďarsko-český*.

6 Shrnutí

Studium tzv. hapaxů, byť i takto omezené, ukázalo, že velká část jazykové potenciality tvořící další slovo na základě už existujících prostředků a modelů je soustředěna kolem odvozenin od proprií, srov. např. *jimmyreedovský, stodůlkovský, po renčínovsku* apod.

Původně propriální kořen má i řada slov s dalšími sufixy, popř. kompozita, jejichž tvorba je vyvolána aktuálním jevem nebo osobou, srov. např. propriální kořen **klaus-**: *klausák, klausboard, klausenský, klausiáda, klausián, klausiánský, klausiánství, klausicismus, klausicismus-topolánkismus, klausismus, klausista, klausiště, klauskalouskizace, klausniáda, klausobijce, klausofil, klausofilní, klausofob, klausofobie, klausofobní, klausokracie,*

klausonomika, klausooptimista, klausopakt, klausoprdelovlezectví, klausoskeptik, klausospeak, klausování, klausovat, klausovatět, klausovatý, klausovice, klausovina, klausovitý, klausovka, klausovláda, klausovládí, klausovsko-grebeničkovský, klausovsko-mackovský, klausovskomachovský, klausovsko-ódéesovský, klausovsko-zahradilovsko-nagrovský, klausovskozemanovský, klausovství, klausovština, klauszemácký, klauszemanismus, po klauszemansku, klauszemansky, klauszemanský, klauszémie.

Z českých sufixů jsou pro nová pojmenování využívány sufixy *-ák* a *-ovitý*, z cizích sufixů mezinárodní sufix *-ista* a z angličtiny přejatý sufix *-ing/-ink* (v tomto případě jsou však slova s daným sufixem přejímána jako celek). Typickým rysem vzniku nových pojmenování je také tvoření zdvojnásobení, zejména od abstrakt, a přechýlených slov. Modifikační kategorie tvoří v rámci shromážděného materiálu 3.5 % tvarů a 3.9 % lemmat. Velké procento materiálu představují rovněž kompozita, jednak kompozita typu *dívka-kuřačka, svíčka-hromnička, reportérka-fotografka* apod., jednak kompozita s prvními částmi *dlouho-, gala-, jino-, kino-, mega-, nízko-, polo-, video-* aj. (celkem 19.5 % tvarů a 19.5 % lemmat).

LITERATURA

CVRČEK, V. et al. (2010): *Mluvnice současné češtiny*. Praha: Nakladatelství Karolinum.

ČERMÁK, F. (2010): *Lexikon a sémantika*. Praha: Nakladatelství Lidové noviny.

DANEŠ, F. – DOKULIL, M. – KUCHAR, J. eds. (1967): *Tvoření slov v češtině II*. Praha: Academia.

DOKULIL, M. (1962): *Tvoření slov v češtině I*. Praha: Nakladatelství Československé akademie věd.

HARTMANN, R.R.K. – JAMES, G. (1998): *Dictionary of Lexicography*. London-NewYork: Routledge.

JAKOPIN, P. (2002): *Extraction of lemmas from a web index wordlist*. VII. seminář TELRI (Trans-European LanguageResources Infrastructure), dostupný z <http://www.hnk.ffzg.hr/telri7/lop.htm#jakopin>

McENERY T. - BAKER P. - HARDIE A. (2006): *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

TĚŠITELOVÁ, M. (1974): *Otázky lexikální statistiky*. Praha: Academia.

ZÍSKAL, A. (1938): Závěrečný příspěvek k lexikografické teorii a praxi (Podrobný slovník současného spisovného jazyka a uspořádání hesel v něm). *Slovo a slovesnost*, roč. 4, 19-27, dostupný z <http://sas.ujc.cas.cz/archiv.php?art=266>

Pravidla českého pravopisu (1993). Praha: Academia.

Průruční slovník jazyka českého I. (1937).

Slovník spisovného jazyka českého I. (1988).

wikipedia: http://en.wikipedia.org/wiki/Hapax_legomenon