

Tvoření slov – poznámky ke korpusově podložené studii¹



Miloslav Vondráček

ABSTRAKT:

V tomto příspěvku sumarizuji zkušenosti z čtyřletého projektu GA ČR *Slovník afixů užívaných v češtině*. Uvádím teoretická východiska i praktické problémy, které bylo potřeba řešit. Nejvíce prostoru věnuji heslům mnou zpracovávaným, která se ukázala jako problematická z hlediska jednoznačnosti vyslovených závěrů. Nejčastěji šlo o nejistý slovní druh základového slova a v důsledku toho i o nejistotu ohledně užitého slovotvorného prostředku.

KLÍČOVÁ SLOVA:

korpus, tvoření slov, afixy, afixoidy, prefixace, suffixace, cirkumfixace

ABSTRACT:

In this paper, I summarize the experience of the four-year project *The Dictionary of Czech Affixes*. I present the theoretical basis and the practical problems that were necessary to solve. The crucial chapters proved to be problematic from the point of view of unambiguously expressed conclusions. Most often, problems arose from the word-type of base words, and the consequent uncertainty regarding applied word-forming elements.

KEYWORDS:

corpora, wordformation, affixes, afixoides, prefixation, suffixation, cirkumffixation

1. Na úvod o projektu *Slovník afixů užívaných v češtině* (GA ČR, reg. č. 13-07138S): Publikace, která vyšla tiskem počátkem roku 2017 (s vročením 2016) a elektronicky je dostupná na webu ÚJČ, je výsledkem čtyřleté práce (2013–2016) vícečlenné skupiny lidí. V přípravné fázi bylo nezbytné vyřešit řadu teoretických, především metodologických, ale i praktických otázek. Ačkoli příprava byla pečlivá, některé problémy, více či méně očekávané, se naplno objevily při zpracování konkrétních hesel. Bylo proto potřebné v mezích možností sjednotit postup členů týmu.

Nejprve bylo třeba rozhodnout, jaký jazykový materiál bude zpracován. Popsána je většina slovotvorných afixů, včetně afixů tvořících pomezí prostředků slovotvorných a tvarotvorných, ale i tzv. koncovkové tvoření. Otázkou se stalo nalezení hranice mezi vznikem slovotvorné varianty typu *Hradečan/Hradečák* a varianty morfologické (*prostor/prostora*), stejně jako rozhodnutí, zda je příslušný afix prostředkem relativně samostatným, nebo jde o alomorfy (např. přejatý prefix *ko-/kon-/kom* nebo český sufix *-ejš(i)/-ějš(i)* (MČ1, Dokulil a kol. 1986: 181). Zohledněno nebylo tvoření kompozičně-sufixální, nebo bylo zmíněno jen okrajově. Prostor je naopak věnován tzv. afixoidům. Pokud nebylo možno jednoznačně hovořit o afixech, užíváme neutrální označení sta-

1 Tento příspěvek vznikl při řešení GA ČR *Slovník afixů užívaných v češtině*, reg. č. 13-07138S.



vební prvek, popř. řetězec. Některé afixy se ukázaly jako téměř monofunkční/jednovýznamové, jiné jako funkčně/významově velmi pestré. Omezené množství afixů uváděných v dosavadní literatuře se nepodařilo jednoznačně prokázat, např. *-ůvk(a)* (TSČz, rejstřík /Daneš–Dokulil–Kuchař 1967/). V rámci projektu bylo popsáno cca 1400 afixů.

Dále byl dohodnut pracovním postup. Využili jsme referenčního žánrově vyváženého korpusu SYN2010, v době začátku grantu nejnovějšího. Dotaz byl zadáván zpravidla prostřednictvím regulárních výrazů ve formě [lemma="*.itý"] pro sufikální tvoření, analogicky [lemma="na.*"] pro tvoření prefixální, popř. [lemma="zá.*ka"] pro cirkumfixaci. Protože lemmatizace je vytvořena tak, že lemma zahrnuje i negované formy, komparativy a superlativy (Vondráček 2013), které jsme zpracovávali samostatně, bylo potřeba pro očištění vzorku užít komplikovanější zadání typu [(lemma="*.atý") & (tag="A.....1A.*")]. Získávali jsme různě rozsáhlý jazykový materiál, který byl převeden na lemmata. Takový výčet obsahoval v různé míře relevantní doklady, bylo tedy nutno materiál ručně vyhodnotit a odložit lemmata, v nichž nešlo o afix. Pokud bylo lemmat do 250, byl nálezný hodnocen v plném rozsahu manuálně. Jestliže byl nálezný výrazně rozsáhlejší, byl redukován tak, aby vznikl vzorek o rozsahu 250–300 lemmat. Z něj byla dovozena míra relevance a prostřednictvím získaného indexu bylo vypočteno odhadnuté množství relevantních lemmat v celém vzorku (Králík 2001). Potom následovala funkční/sémantická klasifikace nálezu.

Struktura hesla obvykle začíná obecnou informací o tom, z jakého typu základových slov deriváty vznikají — zda z jednoduchých, nebo i složených, zda z českých, nebo i cizích, případně jen z cizích. Následuje charakteristika jednotlivých funkcí/významů se vzorovými doklady. Poté je uvedena frekvenční analýza obsahující údaje o počtu celkově získaných lemmat, údaj o míře jejich relevance, u rozsáhlých nálezů i údaj o celkovém odhadnutém počtu validních lemmat v kompletním nálezu. Následuje výčet dvaceti nejfrekventovanějších lexémů s odkazem k sémantické/funkční skupině a s údajem o frekvenci dokladů v celkovém nálezu. Pod frekvenční analýzou jsou uvedeny příklady nezohledněné z důvodu synchronní slovtvorné neprůhlednosti, stejně jako příklady slov, v nichž se nejedná o afix, ale o řetězec zcela náhodně s afixem shodný.

Stavba hesla s přejatým afixem a obvykle i základem je v závěru odlišná. Nemáme dostatek dokladů pro úsudek, zda byly do češtiny různé slovnědruhové deriváty převzaty souběžně, nebo byly tvořeny slovtvorné řady a svazky v našem jazykovém prostředí (a v jakém pořadí derivátů). Proto statistické údaje spočívají ve výčtu 10 nejčastějších substantiv, 10 nejčastějších adjektiv atd. (Šimandl 2016)

2. K několika konkrétním otázkám, před než jsem byl postaven při formulování hesel, jež mi byla svěřena, a naráželi na ně i ostatní členové řešitelského týmu:

2.a Jako první jsem zpracovával sufix *-k(a)*, mimořádně funkčně a sémanticky bohatý — vyčlenil jsem 27 sémanticko-funkčních skupin (srov. Bílková 2014). Úskalím nezbytné kategorizace je, že vedle sebe vznikají skupiny vnitřně členité (výsledky přechylování zahrnují u mě přechýlené názvy osob i zvířat, ačkoli bylo možno vytvořit dvě podskupiny) a naopak úzce vymezené skupiny zahrnující omezený počet dokladů (názvy oděvů a obuvi).

Za výsledky přechylování považuji i názvy osob typu *dojička, sběračka, šička, třídička*, které bývají interpretovány jako odvozené přímo ze sloves (nejen MČz, Petr



a kol. 1986: 33) a provázené poznámkou, že mužské protějšky k nim neexistují, nebo jsou sekundární; k tomu srov. i rozdíl mezi výsledky desubstantivní derivace *hráčka* ← *hráč* a odvozování deverbálního *hračka* ← *hrát* (*si*). Pro všechna jména ženského rodu lze však doložit mužské protějšky obdobného významu, které nepůsobí jako výsledky desufixace: *dojič ovcí*, *sběrač kovů*, *třídíč odpadu*. Ty lze v češtině považovat za motivačně i fundálně výchozí. Tím nevyklučuji, že se odpojením přechylovací přípony nahodile maskulina tvoří (*hostes*, *fanyln*, *servír*, snad i *prostitut*) — včetně druhotného, a přesto neutrálního generické označení *rodič*.

Názvy oděvů a obuvi vyvolávají další otázku: relativně bezproblémový je lexém *kopačka*, u nějž je doložen singulárový tvar, byť výrazně méně zastoupený. Podstatnější otázkou je, zda základ může být adjektivní a -č- může být výsledkem alternace -c-, s čímž souvisí dílem přetrvávající spor, zda takové případy řešit jako resufixace typu -c(i)/-k(a).²

Spornější je, zda pod heslo -k(a) řadit i pluralia typu *podvlékačky*, *oteplovačky*, nebo komponent -k(y) považovat za samostatný sufix, či přinejmenším za alomorf. To se týká i dalších typů objektů, kde je motivující/fundující slovo zřejmé, ale výsledný derivát označuje více či méně odlišnou skutečnost (*kukačky* — hodiny, *sirky* — balení). Pokud není motivující/fundující slovo, a tedy ani povaha sufixu jednoznačná, odkazují k další možné příponě jako prostředku slovotvorné derivace: *kuličky* — hra i sada hracích nástrojů ← *koule? koulet?* ► -ičk(a) (Vondráček 2014b).

2.b V sémantickém poli názvy dokladů mnohdy nelze jednoznačně rozhodnout, jestli název vznikl ze základu adjektivního, substantivního, nebo verbálního (a sufixu ► -enk(a)/-ěnk(a)). Názvy typu *jízdenka*, *loděnka*, *omluvenka*, *účtenka* k adjektivnímu původu přímo neukazují, ale nevyklučují jej, protože k nim víceslovné/neuniverbizované konkurenty obvykle dosud existují: *jízdní doklad*, *lodní lístek*, *omluvný list*, *účtetní doklad* (Vondráček 2014a).

2.c Problémem dalších skupin je jejich značná významová nesourodost. Sufixem -k(a) jsou tvořeny některé názvy polévek, nemocí, popř. částí těla typu *čočkovka*, *česnečka* ← *česn-ek-ový?*, *horečka* (Černá 2009), převládá ale tvoření sufixy složenými ► -čk(a), -ačk(a), -ečk(a), -ičk(a), -ovk(a): *zelňačka*, *bramboračka*, *levačka*, *pravačka*.

Podobně obtížné je přesněji charakterizovat názvy lidských produktů. Jsou tvořeny z adjektiv odvozených od vlastních jmen osob sufixy -ovsk(y), -ův(), nebo přímo z těchto vlastních jmen (pak by v některých případech šlo o sufix ► -ovk(a))?: *bondovka*, *holmesovka* (detektivní příběh s J. Bondem a jemu podobné), *bekovka*, *masaryčka* (čepice vyráběná A. Bekem a jí podobné / jako nosil T. G. Masaryk), *morseovka*, *vincentka*, *becherovka*, *Škodovka* ← *Škodovy závody*, *Kolbenka* (továrny) aj.

2.d V mnoha případech nemůžeme vyloučit více slovnědruhově odlišných slov základových. V takových případech na nejednoznačnost alespoň upozorňuji: Názvy prostředků v širokém smyslu se odvozují ze základů sloves dokonavých i nedokonavých: *bruska*, *dýmka*, *chňapka*, *klapka*, *odměrka*, *obálka*, *nálevka*, *pájka*, *přenoska*, *skláádka*, *šlapka*, *škrabka*, *výhybka*, *výlevka*, *vývrtka*, *zásuvka*. Uvažovat lze také o odvo-

2 Sama motivace názvů obuvi je nejspíš různorodá: *tenis-ka* — bota/obuv na tenis? tenis-ová bota/obuv?; *cvička* — bota na cvič-ení? cvič-ební bota? v níž lze cvič-it?; *botas-ka/kecka* — bota značky Botas/Cat; *kopačka* — kopa-c-í bota? již lze kopa-t? na kopa-nou?



zování od adjektiv (*přepravka* ← *přepravní schrán(k)a*; *přenoska* ← *přenosný box*; *zásuvka* ← *zásuvná přihrádka*); i tato spojení jsou doložena.

U názvů dokladů typu *žádanka*, *povolenka*, *složenko*, *doručenka*, *navštivenka*, *sázenka*, *umístěnka*, *povolenka* lze připustit slovesný základ, přesněji základ spočívající v příčestí trpném: *žádán*, *složen*, *doručen*, *umístěn*, *navštíven*, *povolen*. Sémantika příčestí trpného není u řady odvozenin snadno zdůvodnitelná. Napětí mezi motivací a fundací je ostatně pro participiální odvozeniny obvyklé. Nabízí se tedy i tvoření z infinitivních kmenů sloves složenými sufixy ► *-enk(a)*, ► *-nk(a)*. Právě tak přichází v úvahu základ substantivní (*doklad o složení*, *doručení*, *sázce*, *umístění*). Nelze tudíž jednoznačně rozhodnout, zda jsou tyto názvy tvořeny formantem jednoduchým, nebo složeným. Srov. např. řešení zvolená v MČ1 a PMČ (Dokulil a kol. 1986; Grepl a kol. 1995; Štícha a kol. 2013; Štícha 2018).

3. Další okruh sporných jevů ukáží na sufixu *-avk(a)*. Jde zejm. o rozpor mezi synchronním pohledem a přístupem diachronním. Pro ověření předpokladů jsem tedy často nahlížel do PSJČ (PSJČ 1935–1957), Českého etymologického slovníku (Rejzek 2015), (SSJČ, Havránek 1989) a mnoha dalších zdrojů. Míra, v jaké jednotliví autoři hesel přihlíželi ke staršímu stavu slovní zásoby, se liší. Sám na možnou paralelní motivaci/fundaci alespoň upozorňuji.

3.a Sufixem *-avk(a)* se odvozují pojmenování objektů různého typu a pojmenování stavů. Otázkou je, od jakých základů jsou některé z nich tvořeny. Liší se pohled synchronní a diachronní. Synchronně se zdá, že jde převážně o deverbativa, ale historicky k řadě slov existovala i adjektiva (doložena jsou v PSJČ a jinde). Další teoretickou otázkou je, zda by se v případě deverbativ odděloval i kmenotvorný vokál, nebo jen infinitivní *-t* (pak by šlo o sufix ► *-vk(a)*). Př.: Synchronně se mohou zdát ze sloves odvozena slova *padavka* ← *padat* (ale dle PSJČ je doloženo i adj. *padavý*), *pijavka* ← *pije* (ale i *pijavý*), *prskavka* ← *prskat* (ale i *prskavý*), *měňavka* ← *měnit* (ale i *měňavý*, živé dosud), *hořavka* ← *hořet* (ale i *hořavý*, byť podle PSJČ již neživé), *kapavka* ← *kapat* (ale i *kapavý*); adjektivní specifikace chorob typu *horká nemoc* není historicky ojedinělá (Černá 2009), *sykavka* ← *sykat* (ale i *sykavý*), *třesavka/třasavka* ← *třesavý/třasavý* (PSJČ), ale vyloučit nelze z dnešního pohledu přijatelnější odvození z přítomného kmene *třese* (*se*). Dále přibývá problém s posloupností slovotvorných procesů. Př.: *lučavka* ← *lučba* (zastarale ‚chemie‘; adj. *Lučavý* PSJČ dokládá, ale zdá se být derivován od substantivního základu *lučavka*).

4. Jak bylo řečeno v úvodu, zpracování adaptovaných výrazů je pojato poněkud odlišně, případ od případu se liší i způsob zpracování frekvenčních údajů, např. podle možnosti sémanticky jazykový materiál klasifikovat. Problém je zejm. se stanovením posloupnosti/paralelnosti přejímání / autonomního tvoření různých slovnědruhových derivátů.

4.a Mnoho sufixů má původ v klasických jazycích, paralely v řadě jazyků moderních. České ekvivalenty přípon považujeme za adaptační (Uhlířová 1995; Šimandl 2016). Jsou-li základovými slovy výrazy s příponou *-ik()*, *-ik(a)*, *-ick(ý)*, uvažujeme o resuffixaci spíše než o příponě *-um* nebo palatalizaci *c/k* (*politikum* ← *politický*). U sloves se sufixy *-ick(ý)*, *-ik()*, *-ik(a)*, *-ikum*, *-i(e)* apod. nelze vždy jednoznačně říci, které slovo je základem pro tvoření slov dalších. Pravděpodobně byla většina z nich převzata souběžně, nebo byla v českém prostředí tvořena paralelně. Pokud slova ob-



sahují přední komponent *anti-*, *an-*, *ana-*, předpokládáme, že byla převzata již s tímto prvkem, a tedy že nejde o cirkumfixaci v českém prostředí, srov. *antimalarikum*, **malarikum* vs. *antipyretikum* ← *pyretikum* (ASCS, Kraus–Petráčková 1995).

4.b Obdobně prefix latinského původu *non-* (pův. *nōn*) se pojí výhradně s cizími základy, nejen románskými. Slova nevznikla v češtině, byla adaptována. České jsou slovnědruhovými deriváty. Předpona vyjadřuje význam ‚ne‘, ‚nikoli‘, ‚bez‘, ‚nic‘. Přítomen je díky slovtvornému potenciálu češtiny v mnoha slovních druzích. Není snadné rozhodnout, který slovní druh je základový, v češtině se jako výchozí jeví adjektivum (Kraus–Petráčková 2005).

5. Potíže jsou i se zpracováním cirkumfixů.³ Je třeba rozhodnout, zda výraz vznikl spojením předložkové vazby substantiva a souběžnou suffixací, nebo zda existuje možné motivující/fundující adjektivum (EŠČ, Karlík–Nekula–Pleskalová 2002). To se týká např. slova *podprsenka*. Primárně se zdá, že je tvořeno ze spojení prepozice *pod*, substantiva *prso/-a* a suffixu *-enk(a)*. Vedle toho však dodnes existuje spojení *podprsní pás*, jehož frekvenci musíme považovat v minulých obdobích za převažující, zejm. u chudších vrstev.

5.a Prefix *mimo-* vyjadřuje význam ‚míjet, minout, probíhat jinudy‘. Protože je také v lexikální podobě předložkou, přikláníme se k tomu, že deriváty jsou tvořeny z původních předložkových spojení se substantivem, tedy cirkumfixací v širším pojetí (*mimosoudní* ← *mimo soud*), spíše než předponovým odvozováním od adjektiv (*mimosoudní* ← *soudní*). Dalším argumentem je, že deriváty s komponentem *mimo-* v některých případech nevyjadřují mírnou negaci hypotetického základového adjektiva, ale nesou relativně samostatný věcný význam. Srov.: *řádný* = ‚správný, náležitý‘ vs. *mimořádný* = ‚výjimečný, neobvyklý, nepravidelný‘. V závislosti na významu mohou být diskutabilní výrazy typu *vegetační* (ve *vegetačním* a *mimovegetačním* období). Desubstantivní odvozenina ze spojení *mimo vegetaci* / *mimo vegetační prostor* přichází v úvahu ve významu ‚prostor bez jakékoli rostliny‘.

Slovtvorné varianty měkké deklinace mírně převažují nad variantami deklinace tvrdé (*mimořádný* ← *mimo řád*).

6. Souhrnně lze říci, že ve slovníku se ve většině případů otevřeně hovoří o sémantice slovtvorných morfémů, jejich polysémii či monosémii, resp. o distribuci funkcí. Obtížný se tento krok jeví u prefixace sloves (Uher 1987; Veselý 2010). Shromáždili jsme množství jazykového materiálu — nejen ze zmíněného korpusu SYN2010, ale i z korpusu SYN a mnoha dalších zdrojů, tištěných i elektronických (tyto doklady nebyly započítány do statistik). V důsledku toho byly přehodnoceny dosavadní seznamy afixů, uváděné ve slovtvorné literatuře.

Řešení některých teoretických problémů krystalizovalo v průběhu projektu, přičemž nemyslím, že byly všechny vyřešeny. Současně skutečnost, že v mnoha případech připouštíme možná variantní řešení, považuji za přednost slovníku. Osobním přínosem pro mne bylo, že jsem musel používat rozsáhlý okruh odborných materiálů a pronikl do hloubi výrazů zdánlivě slovtvorně jednoduchých. Posouzení výsledného díla je samozřejmě na lingvistické obci.

³ Za cirkumfixaci považujeme nejen souběžné připojení prefixu i suffixu, ale i změnu původní předložky v předponu probíhající při suffixaci.



BIBLIOGRAFIE:

- Bílková, J. (2014): Interpretace deminuce v češtině, in: Orgoňová, O. — Muziková, K. — Popovičová, Z. (eds.), *Jazyk a jazykoveda v interpretácii*, Bratislava: Univerzita Komenského, s. 270–276.
- Černá, A. M. (2009): *Staročeské názvy chorob*. Praha: Karolinum.
- Daneš, F. — Dokulil, M. — Kuchař, J. (1967): *Tvoření slov v češtině 2. Odvozování podstatných jmen*. Academia: Praha.
- Dokulil, M. a kol. (1986): *Mluvnice češtiny 1*. Praha: Academia.
- Grepl a kol. (1995): *Příruční mluvnice češtiny*. Praha: NLN.
- Havránek, B. (1989): *Slovník spisovného jazyka českého*. Praha: Academia.
- Karlík, P. — Nekula, M. — Pleskalová, J. (ed.) (2002): *Encyklopedický slovník češtiny*. Praha: NLN.
- Králík, J. (2001): Zamyšlení nad velkými výběry, in: Čermák, F. — Blatná R. (ed.), *Korpusová lingvistika: Stav a modelové přístupy (Studie z korpusové lingvistiky 1)*, Praha: NLN a ÚČNK, s. 38–53.
- Kraus, J. — Petráčková, V. (1995): *Akademický slovník cizích slov*. Praha: Academia.
- Kraus, J. — Petráčková, V. (2005): *Nový akademický slovník cizích slov*. Praha: Academia.
- Křen, M. — Bartoň, T. — Cvrček, V. — Hnátková, M. — Jelínek, T. — Kocek, J. — Kopřivová, M. — Novotná, R. — Petkevič, V. — Procházka, P. — Schmiedtová, V. — Skoumalová, H.: *SYN2010: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <<http://www.korpus.cz>>.
- Křen, M. — Čermák, F. — Hlaváčová, J. — Hnátková, M. — Jelínek, T. — Kocek, J. — Kopřivová, M. — Novotná, R. — Petkevič, V. — Procházka, P. — Schmiedtová, V. — Skoumalová, H. — Šulc, M.: *Korpus SYN, verze 3 z 27. 1. 2014*. Ústav Českého národního korpusu FF UK, Praha 2014. Dostupný z WWW: <<http://www.korpus.cz>>.
- Petr, J. a kol. (1986): *Mluvnice češtiny 2*. Praha: Academia.
- PSJČ (1935–1957): *Příruční slovník jazyka českého*. Praha: Státní nakladatelství.
- Rejzek, J. (2015): *Český etymologický slovník*. Praha: Leda.
- Šimandl, J. (2016): Předmluva, in: Šimandl, J. (ed.), *Slovník afixů užívaných v češtině*, Praha: Karolinum.
- Šimandl, J. (v tisku): Přejatá substantiva na *-er/-ér*: tvary, výslovnost, gramatická informace pro slovník, in: Balowski, M. (ed.), *Chaos i ład w języku i literaturze zeczeskiej*, Instytut Filologii Słowiańskiej UAM, Wydawnictwo PRO, Poznań 2016.
- Štícha, F. (2018): *Velká akademická gramatika spisovné češtiny*. Praha: Academia.
- Štícha, F. a kol. (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.
- Uher, F. (1987): *Slovesné předpony*. Brno: Univerzita J. E. Purkyně.
- Uhlířová, L. (1995): Matematické termíny typu *jacobián, laplaceián*, *Naše řeč* 78. 4, s. 183–188.
- Veselý L. (2010): Ke slovesnému vidu v češtině, *Naše řeč* 93. 3, s. 113–124.
- Vondráček, M. (2013): Povaha stupňování příslovcí, in: Faltýnek, D. — Gvoždiak, V. (ed.), *Tygramatika*. Soubor studií věnovaných prof. Janu Kořenskému k 75. narozeninám, Praha: Dokořán, s. 131–151.
- Vondráček, M. (2014a): Sémantika a pragmatika gramatického čísla: aspekt slovtvorný, in: Orgoňová, O. — Muziková, K. — Popovičová, Z. (ed.), *Jazyk a jazykoveda v interpretácii*, Bratislava, Univerzita Komenského, s. 244–253.
- Vondráček, M. (2014b): Názvy dokladů — aspekt slovtvorný, *Korpus — gramatika — axiologie* 10/2014, s. 57–67.