

Synonymní/variantní tvary infinitivu jako problém konzistentní lemmatizace

Klára OSOLSOBĚ

Filozofická fakulta Masarykovy univerzity, Brno
osolsobe@phil.muni.cz

Synonymous/variant forms of infinitive as a problem of consistent lemmatization

ABSTRACT: The NovaMorf project (see more Hlaváčová 2009; Osolsobě et al. 2017; Petkevič et al. 2019) provided us with two following proposals which have been accepted as our point of departure. The first introduces a concept of the multiple lemma. The second expresses the so-called golden rule of morphology, i.e. the requirement for systematic processing of variants, so that each variant would differ by a pair of lemma + tag. These two requirements lead to the systematic corpus-based study of the variations in the Czech morphology. The aim of our text is to give a comprehensive picture of the consistent lemmatization of verb forms. If we have two variants of an infinitive such, that they can be considered synonymous (*myslet/myslit*), then, let the concept of *multiple lemma* be applied. Here, we give a complete list containing the synonymous lemmas of Czech verbs extracted from grammars, dictionaries and the large language corpora.

KEYWORDS: lemmatization, infinitive, variant

1. Úvod

Cílem tohoto textu je podat ucelený obraz jedné oblasti automatického zpracování přirozeného jazyka (češtiny), a sice lemmatizace slovesných tvarů. Do slovesného tvarosloví zahrnujeme: a) tvar(y) infinitivu, b) tvary l-ového přičestí a přechodníku minulého, c) tvary indikativu a přechodníku přítomného, d) tvary imperativu. K slovesnému paradigmatu volně patří i e) tvary pasivního přičestí a/nebo tvary dějového jména na *ní/tí*. Lemmatem slovesného tvaru je v rámci automatické morfologické analýzy zpravidla tvar infinitivu. Pod tvar infinitivu není zahrnut tvar verbálního substantiva na *ní/tí*, jeho lemmatem je nominativ singuláru příslušného substantiva. Pod vícenásobné lemma {tvar infinitivu, dlouhý/adjektivizovaný tvar pasivního přičestí na *ný/tý*} mají být podle návrhu NovaMorf (Petkevič et al. 2019) zahrnuty krátké (participiální) i dlouhé (adjektivizované) tvary pasivního přičestí.

Popis tvarosloví českého slovesa v synchronní podobě prezentovaný morfologickou anotací korpusů řady SYN vykazuje nekonzistentní řešení v případě, že variantní tvary má základní tvar slovesa – infinitiv. Zaměříme se tudíž na interpretace slovesných tvarů jednoho základu, které lze interpretovat jako tvary synonymní/variantní (např. *myslet/myslit, obléci/obléknout* atd.). Také se budeme snažit zaznamenat nedostatky pokrytí morfologického slovníku (Hajič – Hlaváčová 2013).

2. Zpracování variantního základního tvaru v rámci automatické morfologické analýzy

Existuje-li více než jedna varianta, která může představovat lemma/základní tvar, pak řešení na poli automatické analýzy nebyla dosud jednoznačná. Tak například paralelní

variantní tvary (*filozofie/filosofie, citron/citrón, blúza/blúza/blusa, být/bejt, mýt/mejt*) byly zpracovány nejednotně. V rámci projektu NovaMorf byly přijaty dva návrhy. První se týká variantnosti základního tvaru/lemmatu a zavádí koncept vícenásobného lemmatu. Druhý se týká tzv. zlatého pravidla morfologie, které zavádí požadavek systematického zpracování variant, a to tak, aby se každá varianta lišila dvojicí lemma+tag (viz více Hlaváčová 2009; Osolobě et al. 2017; Petkevič et al. 2019). Tyto dva požadavky vedly k systematickému korpusově podloženému studiu variantnosti českého tvarosloví.

3. Variantní tvary českých sloves

Pokud variantní tvary nejsou v infinitivu, pak přestože jednotlivé tvary slovesných subparadigmat lze vztahovat k různým třídám a vzorům, lemmatizace není problematická. Příklady ukazuje tabulka 1.

Tab. 1

ind. prez.	imp.	l-ové	pasivní/slovesné subst.	infinitiv
<i>probodne</i>	<i>probodni</i>	<i>probodnul/probodl</i>	<i>probodnut/proboden</i>	<i>probodnout¹</i>
<i>kope/kopá</i>	<i>kop/kopej</i>	<i>kopal</i>	<i>kopán</i>	<i>kopat</i>
<i>háže/hází</i>	<i>haž/házej</i>	<i>házal</i>	<i>házení</i>	<i>házet</i>
<i>uzřu²/uzřím</i>	<i>uzři</i>	<i>uzřel</i>	<i>uzření</i>	<i>uzřít</i>
<i>vlaje/vlá</i>	<i>vlej</i>	<i>vlál</i>	<i>vlání</i>	<i>vlát</i>
<i>nehraje/nehrá</i>	<i>nehraj/nehrej</i>	<i>nehrál</i>	<i>nehrán</i>	<i>nehrát</i>

3.1 Typologie variantních tvarů infinitivu

V následujícím přehledu se budeme zabývat pouze variantami kmenotvorné přípony infinitivního slovesného tvaru, a to i tehdy, kdy dochází k přehodnocení kmenotvorné přípony v důsledku historického vývoje. Nebudeme se zabývat teritoriálně ovlivněným krácením kmenotvorné přípony v infinitivu.³ Dále nás budou zajímat varianty infinitivu související s variantností kořenového vokálu *í/e/ě: sít/set*.⁴

Stranou ponecháme varianty kořenového vokálu související s tzv. úžením *é/[ý]* (*nést/nýst, lézt/lízt, prolézat/prolízat*), se změnou *[ýí]/ej* (*ryt/rejt, hýbat/hejbat, lít/lejt*) a krácením dlouhého kořenového vokálu (*hrát/hrat*)⁵ i další teritoriálně podmíněné alternace (*ryt/rét*).

¹ Diachronní infinitiv *probósti* (viz stč. banka: ... jehož nemohl nižádný přemoci ani <probósti> kopiem ani mečem ...) je patrně důvodem pro synchronní dublety (bez kmenotvorného -nu-) v l-ovém a pasivním přítčestí.

² *Jen vezmu ji do rukou, ucítím bolest a poté <uzřu> krev rudou jako růže.*

³ Doklady se najdou v korpusu SYNv7: (*Kdo ju včil postiská, kdo ju zahřeje, komu bude <prat> a vařit a fusekle látat? Madla v kuchyni se dívala v televizi na nějaký zas dokument o zločinech komunismu, jsem si sedl ke stroji v příjemné chuti něco pěkného si <psat>. Když sem si uvědomil, že by sem tu měl <zdechnut>, tak sem zpanikařil a bez rozmyslu sem vlez do te d'ury. „Neboj, děvcho, u nas se říka – kvuli dvuch piv je škoda boty <dřít> ...“ smál se Ostravák. „To mu musíš <řeknut>, že si sprostě mluvil?“)*

⁴ Varianty s krátkým [eě] jsou různě hodnoceny kodifikačními příručkami, nicméně v úzu prezentovaném korpusy jsou doloženy. Patrně zde působí tlak analogie sloves, která mají pouze krátké [eě] v kořeni infinitivu (*pět, pomět se, spět*), viz Štícha et al. (2013).

⁵ *V Riu se budu především snažit <hrat> svůj tenis, a to co nejlíp.*

Tyto varianty jsou variantami hláskoslovnými, a týkají se tudíž všech sloves, jejichž hláskoslovná podoba alternaci připouští.

Nebudeme se zde zabývat variantami infinitivního tvaru typu *prosit/prositi* a *moci/moct* (variantní koncovka), neboť jde o univerzální variantnost českých sloves.

Také variantní infinitivy tvaru typu *sprát/seprat* (variantní prefix) ponecháme stranou (varianty existují u jednotlivých sloves a souvisejí s přehodnocením kmenotvorného vokálu na kmenový, viz např. *od-tí-0-t/ode-t-nou-t*, ale *roz-pj-a-l* i *roze-pj-a-l/roze-p-nu-l*).⁶

3.1.1 Typ 1

Existují dvě varianty infinitivu, které lze pokládat za synonymní. U jednotlivých tvarů slovesných subparadigmat lze vždy určit, ke kterému infinitivu/lemmatu se mají přiřadit. Každému lemmatu odpovídá a) kompletní soubor tvarů a b) nekompletní soubor tvarů. Příklady ukazuje tabulka 2.

Tab. 2

ind. prez.	imp.	l-ové	pasivní/slovesné subst.	infinitiv
<i>věje</i>	<i>věj</i>	<i>vál</i>	<i>vání</i>	<i>vát</i>
<i>vane</i>	<i>vaň</i>	<i>vanul</i>	<i>vanutí</i>	<i>vanout</i>
<i>plaje/plá</i>	<i>plaj</i>	<i>plál</i>	<i>plání</i>	<i>plát</i>
<i>plane</i>	<i>plaň</i>	<i>planul</i>	<i>planutí</i>	<i>planout</i>
–	–	<i>stál</i>	<i>stětí</i>	<i>stít</i>
<i>setne</i>	<i>setni</i>	<i>setnul/setl</i>	<i>setnutí</i>	<i>setnout</i>

3.1.2 Typ 2

Existují dvě varianty infinitivu, které lze pokládat za synonymní. U jednotlivých tvarů slovesných subparadigmat nelze někdy určit, ke kterému infinitivu/lemmatu se mají přiřadit. Některé tvary lze na základě analogie morfologického systému češtiny řadit k oběma variantním lemmatům (podbarvené). Příklady ukazuje tabulka 3.

⁶ Ziková (2016) uvádí, že alternace *e/0* v prefixech slovesných tvarů nelze spojovat s potřebami výslovnosti, nýbrž souvisejí s morfosyntaxí a závisí na morfologické interpretaci vokálu/slabičné sonory v následující slabice (viz např. *roze-stř-í-t* × *roz-stř-íh-a-t*, *ode-př-í-t* × *od-př-í-sáh-nou-t*, *ode-br-a-t* × *od-brzd-í-t*, *se-hn-a-t* × *s-hní-0-t*). Varianty se vyskytují u sloves typu *začít*, u kterých byl kořenový vokál (nosovka) v důsledku hláskoslovného vývoje (denazalizace) přehodnocen na kmenový vokál. Také se okrajově objevují u sloves **[bdpsž]rát/.*[bdpsž]eru a.*[mljē]t/.*[melu]*. Uvedená slovesa mají (v důsledku hláskoslovného vývoje – zániku/vokalizace) jerů a metateze likvid) v tvarech od kmene přítomného kořenový vokál nenulový a v tvarech od kmene infinitivního kořenový vokál nulový. Je tedy možné, že dublety jsou u tvarů od kmene minulého důsledkem působení analogie (přizpůsobení podoby prefixu tvarům od kmene přítomného a naopak). V korpusech jsme v infinitivu zaznamenali pouze varianty u vulgárního slovesa (*Někdo to totiž bude muset s odpuštěním <odsrat>, když se nebude hrát. Tohle si fakt někdo musí šeredně <odesrat> ...*) a okrajově i u sloves *odebrat* (*Za vysvětlování doktorky, že je potřeba Nikymu <odbrat> mozkomíšní tekutina, jsem se zatajeným dechem četla rizika, která jsou spojená s tímto zákrokem. Dal by se tady <odebrat> vzorek krve a poslat ho do Německa na ty LTT testy?*) a *seprat* (*Hledme, jak se stydí, a že ještě samou hanbou nepropad: vždyť kolik pecnů denně musí <sprat> ten žráč a na čí účet, že si jenom tak pouští krev žilou? A dovedl se o to <seprat> jak s ďáblem, tak s andělem ...*). PSJČ uvádí variantu *smlít/smlet/semlít*, SSSJČ pak *sedrat/sdrát, seprat/sprát*.

Tab. 3

ind. prez.	imp.	l-ové	pasivní/slovesné subst.	infinitiv
<i>obleče</i>	<i>obleč</i>	<i>oblékl</i>	<i>oblečen</i> ⁷	<i>obléci</i>
<i>oblékne</i>	<i>oblékni</i>	<i>oblékl/obléknul</i>	<i>oblečen</i> ⁷ /obléknut	<i>obléknout</i>
<i>chytí</i>	<i>chyt'</i>	<i>chytíl</i>	<i>chycen</i>	<i>chytit</i>
<i>chytne</i>	<i>chytňi</i>	<i>chytíl/chytnul</i>	<i>chycen</i> ⁸ /chytnut	<i>chytnout</i>
<i>vymyslím</i>	<i>vymysli</i>	<i>vymyslel</i>	<i>vymyšlen</i>	<i>vymyslet</i>
<i>vymyslím</i>	<i>vymysli</i>	<i>vymyslil</i>	<i>vymyšlen</i>	<i>vymyslit</i>

3.1.3 Typ 3

Existují dvě varianty infinitivu, které lze pokládat za synonymní. U jednotlivých tvarů slovesných subparadigmat nelze nikdy určit, ke kterému infinitivu/lemmatu se mají přiřadit. Variantní i nevariantní tvary lze na základě analogie morfologického systému češtiny řadit k oběma variantním lemmatům. Příklady ukazuje tabulka 4.

Tab. 4

ind. prez.	imp.	l-ové	pasivní/slovesné subst.	infinitiv
<i>tlí/tleje</i>	<i>tli/tlej</i>	<i>tlel</i>	<i>tlení</i>	<i>tlít/tlet</i>

3.1.4 Typ 4

Existují dvě varianty infinitivu, které lze pokládat za synonymní. K nim lze přiřadit pouze nevariantní tvary. Příklady ukazuje tabulka 5.

Tab. 5

ind. prez.	imp.	l-ové	pasivní/slovesné subst.	infinitiv
<i>čnί</i>	<i>čňi</i>	<i>čněl</i>	<i>čnění</i>	<i>čňit/čnět</i>
<i>mele</i>	<i>mel</i>	<i>mlel</i>	<i>mlet(ý)</i>	<i>mlít/mlet</i>

⁷ Zajisté je technicky vzato možné řadit tvary k lemmatům podle kmenotvorné přípony. Nicméně se domníváme, že tento přístup odporuje jak tradičním popisům (mluvnice, slovníky, dosavadní praxe automatických morfologických analyzátorů), tak úzu: *Musíte dodržovat dress code? Neřekla bych. Na obrazovce jsem v modelech, které mám připravené speciálně na vysílání, takže nikdo neřeší, jak přijdu do práce <oblečená>. Dokážu se ale <obléknout> adekvátně situaci. „Panenky mám už ze sedmnácti zemí světa, třeba i z Ekvádoru, Egypta, Japonska, Finska, Ruska nebo Číny. Ty nové jsou už <oblečené>, starší musíme <obléknout>, než je vystavíme,“ poznamenala Marie Roháčková. Jsem poslušná. Když mi řekli, že se tak mám <obléknout>, <oblékla> jsem se. Akademická gramatika spisovné češtiny (Štícha et al. 2013, s. 473) se o rozdílu pasivních přičestí na -en a na -t od sloves, která kolísají mezi třídami (1. a 2.), explicitně nezmiňuje.*

⁸ Totéž platí i pro slovesa, která kolísají mezi typem *prosit* a *tisknout*. Přestože tato synonyma v mluvnicích většinou nebývají zmiňována (výjimkou je zmínka o kolísání slovesa *chytit/chytnout* ve Štíchově Akademické gramatice spisovné češtiny), zaznamenávají je výkladové slovníky. Doklady z korpusu svědčí o tom, že dosavadní technické řešení uměle odděluje dvě různá lemmata, přičemž např. tvary pasivního přičestí *chycen** jsou důsledně řazeny pod lemma *chytit*, není v souladu s jazykovou intuící: *Je tu příroda, člověk si u vody nad pruty odpočine od všeho a sem tam <chytne> i rybu. <Chycené> ryby většinou rozdávám příbuzným, protože je nejím.*

4. Automatické nástroje

K analýze dat jsme použili automatický nástroj *Morfio* (Cvrček – Vondříčka 2009) a nástroj *Deriv* (Šmerk 2009). Oba nástroje pracují s korpusem,⁹ *Deriv* je navíc propojen s elektronickými verzemi tištěných slovníků.¹⁰ Synonymii jsme ověřovali v tištěných slovnících, v korpuse SYNV7 a ve webovém korpuse Araneum. Při analýze otevřené třídy sloves typu *myslit/myslet* jsme vycházeli z korpuse SYN2015.

5. Návrh zásad lemmatizace

Vzhledem k tomu, že korpusová analýza ukazuje, že variantní/synonymní infinitivy se ani zdaleka netýkají jen sloves, která jsou známá ze školských mluvnic, pokusíme se stanovit pravidla lemmatizace pro automatické nástroje a podáme pokud možno úplný korpusově a slovníkově podložený výčet lemmat.

5.1 Návrh k typu 1

Existují-li dvě varianty infinitivu, které lze pokládat za synonymní a u jednotlivých tvarů slovesných subparadigmat lze vždy určit, ke kterému infinitivu/lemmatu se mají přiřadit, pak tam, kde každému lemmatu odpovídá a) kompletní soubor tvarů, i tam, kde jednomu z lemmat odpovídá b) nekompletní soubor tvarů, necht' je využit koncept vícenásobného lemmatu.

K případu a) patří lemmata:¹¹

{*vát, vanout*}, {*vyvát, vyvanout*}, {*ovát, ovanout*}, {*zavát, zavanout*}, {*odvát, odvanout*}, {*přivát, přivanout*}, {*provát, provanout*}, {*vevát, vevanout*}, {*navát, navanout*}, {*svát, svanout*}, {*rozvát, rozvanout*}, {*vsout, vsunout*},¹² {(*pro*)*slout, (pro)*slynout}¹³, {*nadout, nadmout*}, {*vzdout/vzdemout*}, {*ožít, oživnout*}, {*snout*,¹⁴ *snovat*}.

Dvě různá slovesa jsou *klít* a *klnout* (komu) i *zaklít* (koho v co) a *zaklnout*, podobně *plout* a *plynout*, *dout* a *dmout se*, *kout* a *kovat*, *zasnout*¹⁵ a *zasnít*¹⁶ (*se*).

⁹ *Morfio* pracuje s korpusem SYN2005, SYN2010 a SYN2015. *Deriv* pracuje s webovými korpusem řady csTenTen.

¹⁰ Je možno přepnout na platformu DebDict. Více Osolsobě et al. 2009.

¹¹ Tento seznam je uzavřený. Přechody/kolísání mezi neproduktivní 3. třídou slovesnou typ *krýt* a produktivní 2. třídou slovesnou (popř. 3. třídou typem *kupovat*) jsou záležitostí historického vývoje češtiny. Do seznamu nezařazujeme slovesa, která v synchronním úzu prezentovaném korpusem řady SYN nejsou doložena. Jedná se sloveso *rouť* ve významu *rvát* (viz doklad z diakorp: *neb jich bieše taďy čěsta, bliz od té zdi toho města, přěd nímž boha umučili, v hlavu rujíc, v líce bili.*) a *trout* ve významu *otravovat* (srov. Šmilauer 1972, s. 215). V případě tvaru *rouť* se jedná v korpusech řady SYN převážně o variantní pravopis slova *raut*. Tvary jsou přiřazeny lemmatu *routa* a interpretovány jako tvary genitivu plurálu. Tvar *trout* je užit v angličtině (*trout* = angl. pstruh).

¹² Dvě různá slovesa jsou *sesout* a *sesunout*, *vysout* a *vysunout*, *zasout* a *zasunout*. Infinitiv *vsout* je doložen pouze v SSJČ, v synchronních korpusech se nevyskytuje. Jeden tvar vztažitelný k infinitivu *vsout* je doložen v SYN v7: ... *vyřítla se dvě děvčata, současně se chtěla naráz prodrat skrz úzké dveře, <vsula> se tam, a když se prorvaly ven, tak řvaly smíchy, ...* Doklady z korpuse Araneum jsou překlapy.

¹³ Infinitiv *proslout* uvádí SSJČ, v synchronních korpusech doložen není, na internetu se doklady najdou.

¹⁴ V korpuse SYNV7 jsou tvary *snout* řazené k lemmatu *snout/Vf.** vesměs propria (příjmení **Snout**) nebo jejich části (název mužstva **Lassie's snout**), popř. překlep (mezera navíc: <**Učí snout**> *vlásky* ...). Doklady na užití slovesa *snout* jsou ojedinělé: *Už slova děti uprchlíkův v nás mohou <snout> jen temný čoud.*

K případu b) patří lemmata:¹⁷

{tít, tnout}, {odtít, odetnout}, {podtít, podetnout}, {přetít, přetnout}, {stít, setnout}, {roztít, rozetnout}, {vtít, vetnout}, {zatít, zatnout}, {načít, načnout}, {počít, počnout}, {začít, začnout}, {započít, započnout}, {vzít, vezmout},¹⁸ {opomenout, opominout, opomnět}, {pozapomenout, pozapomnět}, {připomenout, připomnět}, {rozvzpomenout, rozvzpomnět}, {vzpomenout, vzpomnět}, {vyvzpomenout, vyvzpomnět}, {zapomenout, zapomnět}, {plát, planout}, {vzplát, vzplanout}, {zaplát, zaplanout}.

Sporné je, zda ve slovníku automatického analyzátoru zaznamenávat případy typu *žít* a *žnout*, konkrétně prefigované varianty *nažít* a *nažnout*, *sežít* a *sežnout*, a to s ohledem na homonymii se slovesem *žít* ve významu ‚být naživu‘.¹⁹ Velmi problematické se jeví také zachycení slovesa *ro(z)žnout/rozžít* (na Moravě běžně užívané ve významu *rozsvítit*),²⁰ a to jednak s ohledem na neustálenou grafickou podobu, jednak s ohledem na homonymii tvaru imperativu *rožni* s tvary substantiva *rožeň*.

5.2 Návrh k typu 2

Existují-li dvě varianty infinitivu, které lze pokládat za synonymní a u jednotlivých tvarů slovesných subparadigmat nelze někdy určit, ke kterému infinitivu/lemmatu se mají přiřadit (některé tvary lze na základě analogie morfologického systému češtiny řadit k oběma variantním lemmatům), pak necht' je využít koncept vícenásobného lemmatu.

Vícenásobná lemmata necht' mají slovesa kolísající mezi typy **prosit** – (**trpět**) – **sázet**:

Internetová jazyková příručka k nim uvádí, citují: „Ve 4. slovesné třídě najdeme několik sloves, která kolísají mezi vzory ‚prosit‘ a ‚sázet‘. Patří k nim frekventovaná slovesa *bydlit* – *bydlet*, *musit* – *muset*, *myslit* – *myslet*, z méně užívaných pak *kvílit* – *kvílet*, *šlit* – *šilet* a *hanit* – *hanět*. Historicky k nim můžeme řadit i slovesa *bujit* – *bujet*, *lačnit* – *lačnět*, *truchlit* – *truchlet*.“

¹⁵ Podle SSJČ má sloveso *zasnout* dva významy, a sice 1) ‚usnout‘ a 2) ‚upadnout do snění‘. U slovesa *zasnit* se oba významy objevují. V současném úzu prezentovaném korpuse řady SYN však v případech tvarů od lemmatu *zasnit* naprosto převažují tvary s reflexivním *se* (viz také výklad SSČ), okrajově se pak vyskytují ještě i tvary *zasnit si*. Sloveso *zasnout* není reflexivní. Nereflexivní doklady u slovesa *zasnit* v SSJČ jsou převážně z literatury 19. stol. V korpuse SYN jsme objevili ojedinělé doklady nereflexivního užití: „*To by bylo fajn, <zasnil> Čáslava nad vidinou startu na světovém šampionátu. „Obrázky nás probudí neb <zasní>, už nevíme dál ... tak zhasni!“ „Snad se mi to podaří, <zasnila> jednadvacetiletá Monika Řeháková.*

¹⁶ V SYNv7 jsou potenciálně homonymní tvary *zasni*, *zasněme*, *zasněte* desambiguovány oběma lemmaty, tedy *zasnit* nebo *zasnout*. Ve všech případech jde ovšem o tvary slovesa *zasnit se*, desambiguace by se dala zpřesnit právě použitím pravidel pro výskyt zvrtného *se*.

¹⁷ Tento seznam je uzavřený. 2. třída vzor *začít* je uzavřená třída sloves. Kromě sloves 2. třídy typu *začít* řadíme do tohoto seznamu sloveso *plát/plánout* a jeho prefigované varianty. Sloveso je zvláštní tím, že kromě paralelních tvarů podle 3. třídy typu *krýt* a 2. třídy, viz typ 1 případ a), má navíc ještě tvary od kmene přítomného podle 5. třídy typu *dělat* (tedy např. *Do samoty mi <pláš> jako chudokrevný měsíc. Vláš tává hřívá a svíce <pláš> a dohořívá až dohoří*). Bohužel většina tvarů **plám*, **pláš*, **plá*, **pláme*, **pláte*, jsou překlepy.

¹⁸ Doklady tohoto tvaru se v korpuse SYNv7 nevyskytují, objevují se ovšem ve webovém korpuse *Aranea*: „*Můžu si <vezmout> kamaše?*“

¹⁹ V případě infinitivu *nažít* jde ve většině dokladů v SYNv7 o sloveso *žít* ve významu ‚zažít‘: *Na druhou stranu je takový autorský typ sběratel, takže si určité věci potřebuje <nažít>*. Pouze jeden doklad je na význam *nažnout*: „*Dobře, přikývne, „ale musíme jít <nažít> trávu.*“ V případě infinitivu *sežít* jde v jednom případě o sloveso *sežnout*: „*Šla <sežít> ten kousek meze, co máme ještě u lesa,*“ ... Ve dvou případech jde o překlep.

²⁰ SYNv7: *Okamura strávil dětství v Bystřici. Dodnes proto říká „žufánek a <rožni>“.*

Rozsáhlé seznamy synonymních i nesynonymních dvojic sloves založené na analýze korpusu csTenTen12 sestavila N. Volková (2017). Tyto seznamy se opírají o analýzu dat z brněnského morfologického slovníku (Osolobě 1996), který je bohužel značně závislý na zdrojích, z nichž vzešel (heslář SSJČ), a obsahuje velké množství sloves, která v synchronním úzu nejsou doložena. Z toho důvodu jsme se rozhodli pro analýzu dat z korpusu SYN2015, a to prostřednictvím nástroje *Morfio*. Z automaticky vygenerovaného seznamu dvojic jsme ruční analýzou získali přehled o slovesech lišících se v infinitivu kmenotvorným [ieě]. Rozdělili jsme je na 1) případy synonymních lemmat a 2) případy dvou odlišných sloves.

Slovesa první skupiny necht' mají vícenásobné lemma:²¹ {bydlit, bydlet}, {dobydlit, dobydlet}, {obydlit, obydet}, {odbydlit, odbydlet}, {pobydlit, pobydlet}, {probydlit, probydlet}, {vybydlit, vybydlet}, {zabydlit, zabydlet}, {myslit, myslet}, {namyslit, namyslet}, {domyslit, domyslet}, {pomyslit, pomyslet}, {promyslit, promyslet}, {přimyslit, přimyslet}, {smyslit, smyslet}, {vymyslit, vymyslet}, {zamyšlit, zamyslet}, {musit, muset}, {bujit, bujet}, {blyštit, blyštět}, {hanit, hanět}, {chrapit, chraptět}, {chroptit, chroptět}, {kabonit, kabonět}, {kapalnit, kapalnět}, {krabatit, krabatět}, {kulatit, kulatět}, {kvílit, kvílet}, {lačnit, lačnět}, {nadstavit, nadstavět}, {ošklivit, ošklivět²²}, {oželit, oželet}, {pohanit, pohanět}, {posmutnit, posmutnět}, {potemnit, potemnět}, {poulit, poulet}, {prýštit, prýštět}, {pučit,²³ pučet}, {rozbulit, rozbulet}, {rozesmutnit, rozesmutnět}, {rozezvucit, rozezvucet}, {rozkošatit, rozkošatět}, {rozradostnit, rozradostnět}, {řeřavit, řeřavět}, {svědit, svědět}, {svraštit, svraštět}, {svrbit, svrbět}, {šílit, šílet}, {škaredit, škaredět}, {špičatit, špičatět}, {špinavit,²⁴ špinavět}, {temnit, temnět}, {truchlit, truchlet}, {utvářít, utvářet}, {vzít, vzet}, {vláčit, vláčet}, {vybulit, vybulet}, {vyhladovit, vyhladovět}, {vyhovít, vyhovět}, {vykoulit, vykoulit}, {vypučit, vypučet}, {vytvářít, vytvářet}, {vyzdravit, vyzdravět}, {zabahnit, zabahnět}, {zachroptit, zachroptět}, {zakvílit, zakvílet}, {zašpičatit, zašpičatět}, {zašustit, zašustět}, {zatemnit, zatemnět}, {zavrtit, zavrtět}, {zešílit, zešílet}, {zšeřit, zšeřet}, {zemnit, ztemnět}, {zvučit, zvučet}.

Slovesa druhé skupiny necht' mají lemmata dvě:²⁵ bělit a bělet (*se*), divočít a divočet, dostavit (*se*) a dostavět, dovážít a dovážet, hrbatit a hrbatět, hutnit a hutnět, chybit a chybět, jasnít (*se*) a jasnět, klidnit a klidnět, nastavit a nastavět, navážít a navážet, obestavit a obestavět, odšumit a odšumět, odvážít (*se*) a odvážet, otěhotnit a otěhotnět, otupit a otupět, ozdravit a ozdravět, ozřejmit a ozřejmět, pochybit a pochybět, praštit a praštět, proděřavit a proděřavět, přestavit a přestavět, převážít a převážet, přistavit a přistavět, přivážít a přivážet, rozestavit a rozestavět, rozněžnit (*se*) a rozněžnět, rozvážít a rozvážet, stavít (*se*) a stavět, svářít (*se*) a svářet, světlit a světlet, trčít a trčet, tvářít (*se*) a tvářet, vestavit a vestavět, vrásnit a vrásnět, vystavit a vystavět, vytrčít a vytrčet, vyvážít a vyvážet, zastavit a zastavět, zavážít a zavážet, zbělít a zbělet, zbídačit a zbídačet, zbystřít a zbystřet, zcivilnit a zcivilnět, zdomácnit a zdomácnět, zdrsnit a zdrsnět, zduchovnit a zduchovnět, zdůvěrnit a zdůvěrnět, zelenit a zelenět,

²¹ Tento seznam je z principu otevřený, měl by ovšem zachytit slovesa, která se v úzu prezentovaném korpusy skutečně objevují.

²² V korpusu SYNv7 jsou doklady jak na význam ‚mít v ošklivosti co‘ (*A přesto si to <ošklivě!> čím dál víc ...*), tak na význam ‚stávat se ošklivým‘ (*Lidé se věkem ohýbali a <ošklivěli> ...*).

²³ Bohužel je většina dokladů tvarů *puč[ii]*. * užita ve významu *půjčit*.

²⁴ Lemma *špinavět* je přiřazeno tvarům *špinavěj*, ve všech případech jde o překlepy (*špinavej*, *špinavější*).

²⁵ Tato skupina představuje velký problém pro automatickou morfologickou analýzu, a to s ohledem na homonymii všech tvarů kromě infinitivu a l-ového přičestí. Výsledky desambiguace nejsou uspokojivé. Významový rozdíl mezi deadjektivními slovesy bývá rozdíl mezi stávat se jakým/ činit jakým. Tento význam však v úzu bývá někdy potlačen (viz výše). Další pozorovaný významový rozdíl je +/- přítomnost externího činitele děje – např. *znervóznit* a *znervózňt*.

zestručnit a zestručňet, zesvětlit a zesvětlet, zesvětštit a zesvětštět, zešeřit a zešeřet, zešikmit a zešikmět, zeškaredit a zeškaredět, zeštíhlit a zeštíhlet, zevšednit a zevšednět, zhmotnit a zhmotnět, zhutnit a zhutnět, zchromit a zchromět, zintenzívnit a zintenzívnět, zjasnit a zjasnět, zjemnit a zjemnět, zjinačit a zjinačet, zkapatnit a zkapatnět, zklidnit a zklidnět, zkrabatit a zkrabatět, zkrásnit a zkrásnět, zkrvavit a zkrvavět, zkudrnatit a zkudrnatět, zlhostejnit a zlhostejnět, zlidštit a zlidštět, zmalátnit a zmalátnět, zmatnit a zmatnět, zmléčnit a zmléčnět, zmocnit a zmocnět, zmohutnit a zmohutnět, zmrvtit a zmrvtět, zmrzačit a zmrzačet, zmužnit a zmužnět, znečitlivit a znečitlivět, znehybnit a znehybnět, znejasnit a znejasnět, znejistit a znejistět, zneklidnit a zneklidnět, znervóznit a znervóznět, zněžnit a zněžnět, zobecnit a zobecnět, zostřit a zostřet, zošklivit a zošklivět, zpevnit a zpevnět, zplacatit a zplacatět, zploštit a zploštět, zpohodlnit a zpohodlnět, zpomalit a zpomalet, zpovrchnit a zpovrchnět, zpozornit a zpozornět, zpravidelnit a zpravidelnět, zprotivit a zprotivět, zprůhlednit a zprůhlednět, zprůsvitnit a zprůsvitnět, zprůzračnit a zprůzračnět, zpružnit a zpružnět, zpřesnit a zpřesnět, zpřísnit a zpřísnět, zruměnit a zruměnět, ztitěrnit a ztitěrnět, ztučnit a ztučnět, ztupit a ztupět, zubatit a zubatět, zuhelmatit a zuhelmatět, zútulnit a zútulnět, zvěcnit a zvěcnět, zveličit a zveličet, zveřejnit a zveřejnět, zvětšit a zvětšet, zvláčnit a zvláčnět, zvlčit a zvlčet, zvnitřnit a zvnitřnět, zvroucnit a zvroucnět, zvýraznit a zvýraznět, zvýznamnit a zvýznamnět, zženštit a zženštět.

Patrným přegenerováním morfologického slovníku MorFFlex (Hajič – Hlaváčová 2013) jsou lemmata *navelit*, *zamlžet*, *zaokrouhlet*, *zavazít*.

Vícenásobná lemmata necht' mají slovesa kolísající mezi typy *nést/péci/umřít – tisknout*.²⁶ {vybřít, vybřednout}, {zabřít, zabřednout}, {zábst, zábnout}, {zazábst, zazábnout}, {vlást,²⁷ vládnout}, {naklást, nakladnout²⁸}, {krást, kradnout²⁹}, {ukrást, ukradnout}, {vykrást, vykradnout³⁰}, {rozkrást, rozkradnout³¹}, {okrást, okradnout³²}, {vkrást, vkradnout³³}, {příkrást, příkradnout³⁴}, {zkrást, zkradnout³⁵}, {propást, propásnout}, {mást, mátnout}, {pomást, pomátnout}, {zmást, zkrátnout}, {nalézt, naleznout}, {vynalézt, vynaleznout}, {odvynalézt, odvynaleznout}, {znovuvynalézt, znovuvynaleznout}, {hrýzt, hryznout}, {odhrýzt, odhryznout}, {podhrýzt, podhryznout}, {zahřýzt, zahryznout}, {houst, hudnout³⁶}, {zahoust, zahudnout³⁷},

²⁶ Jedná se o uzavřený seznam sloves. Přechody mezi neproduktivními typy 1. třídy (*nést, péci*) a 2. třídou (typem *tisknout*) jsou popsány v českých mluvnicích (Komárek et al. 1986; Stícha et al. 2013). Přesto nejsou dosud beze zbytku zachyceny automatickými morfologickými slovníky. V korpusu SYNv7, ale například i ve webových korpusech se okrajově vyskytují jak tvary velmi archaické (např. *vlást*), ale i velmi progresivní (např. *kradnout, zmohnout, rostnout, hnětnout, ...*). S ohledem na budoucí údržbu slovníku je třeba zvážit, zda potenciální tvary nepřegenerovat.

²⁷ Doklady z korpusu SYNv7: ... *Zas budou příšery <vlást> kurvy horší než hovno ze sraček uplácaj si tvář svobodnymilovníků ... „Řekům sluší nad barbary <vlást>“, dodává, protože „barbar a otrok jest od přírody totěž.“*

²⁸ Doloženo pouze v korpusu csTenTen: *Vím, je to jen červík, ale vážně mi nedělá velkou radost to, že to mohlo <nakladnout> lary ...*

²⁹ *Takový člověk určitě <kradnout> nebude ...*

³⁰ *Favorita stačil <vykradnout> 24letý lapka na Klíši ...*

³¹ *... za to my nemůžeme, že náš stát politici nechali <rozkradnout>, ...*

³² *T-mobile se mě snaží <okradnout>.*

³³ *Jin a jang však mají v sobě tu vlastnost, že se umí do nás <vkradnout> pomalu, ...*

³⁴ *... jen slouhové a služebníci, jimž vrchnost dovolí si něco <příkradnout> ...*

³⁵ *Než aby si to alespoň museli <zkradnout>, tak jim to dejte.*

³⁶ V korpusu SYNv7 je jeden doklad: *Ti budou <hudnout> o zradě.* Ve druhém případě jde o překlep: *Dietologové je nedoporučují jíst ve velkých dávkách, když se snažíme <hudnout> ...*

³⁷ V korpusu SYNv7 je jeden doklad: *... co by si mohli <zahudnout> v hospodě ...*

{vmést, vmetnout³⁸}, {rozkvést, rozkvetnout³⁹}, {vykvést, vykvetnout⁴⁰}, {růst, rosnout⁴¹}, {srůst, srostnout⁴²}, {hníst, hnětnout⁴³}, {prohníst, prohnětnout⁴⁴}, {vyhníst, vyhnětnout⁴⁵}, {spočíst, spočtnout⁴⁶}, {započíst, započtnout⁴⁷}, {otřást, otrásnout⁴⁸}, {setřást, setřásnout⁴⁹}, {proplést, proplétnout⁵⁰}, {řít, řeknout}, {dořít, dořeknout}, {odřít, odřeknout}, {podřít, podřeknout}, {zřít, zřeknout}, {doobléci, doobléknout}, {dovléci, dovléknout}, {navléci, navléknout}, {obléci, obléknout}, {odvléci, odvléknout}, {podvléci, podvléknout}, {povléci, povléknout}, {provléci, provléknout}, {přivléci, přivléknout}, {převléci, převléknout}, {přivléci, přivléknout}, {svléci, svléknout}, {vléci, vléknout⁵¹}, {vyvléci, vyvléknout}, {zsvléci, zsvléknout}, {zavléci, zavléknout}, {střící, střehnout⁵²}, {dosíci, dosáhnout}, {zmoci, zmohnout⁵³}, {tlouci, tlouknout}, {natlouci, natlouknout}, {otlouci, otlouknout}, {přítlouci, přítlouknout}, {stlouci, stlouknout}, {vtlouci, vtlouknout}, {vytlouci, vytlouknout}, {zatlouci, zatlouknout}, {ztlouci, ztlouknout}, {otéci, oteknout}, {utéci, uteknout}, {vztéci,⁵⁴ vzteknout}, {ztéci, zteknout⁵⁵}, {posíci, poseknout⁵⁶}, {odemčít, odemknout}, {uzamčít, uzamknout}, {zamčít, zamknout}.

³⁸ V korpusu SYNv7 jsou dva doklady: *Zručně pracovali pendreký a já jsem dostal chuť <vmetnout> jim do štítů: I vás vezme čert! Leckterý z mužů, kteří sem za mnou přijdou, může mu plivnout, <vmetnout> do tváře hanbu jeho ženy, a tuto ošklivou slinu nesetře do nejdelší smrti.*

³⁹ Doloženo pouze v korpusu csTenTen: *Závěrečné verš „růže musí <rozkvetnout> / a ne uvadnout“ to celý zabíjí, protože takhle může rýmovat akorát zamilované deváták, ne čtyřicetiletý rocker.*

⁴⁰ Doloženo pouze v korpusu csTenTen: *Prima nechala <vykvetnout> jarní programové schéma ve znamení novinek.*

⁴¹ Doloženo pouze v korpusu csTenTen: *Mám děcko a vidím ho denně celý den <rostrnout>.*

⁴² Doloženo pouze v korpusu csTenTen: *... tedki musí <srostnout> kliční kost a bude to dobrý, ...*

⁴³ Pouze jediný doklad v SYNv7: *Když věděla, že bratři na víkend přijedou, začala už dny předtím vařit polévku, <hnětnout> těsto, ze spíže snášet sklenice se zavařeninou, které držela proti světlu a otáčela a obracela. V korpusu Araneum je jich několik.*

⁴⁴ Doklad z korpusu Araneum: *Dáme na 48 hodin odležet do ledničky při teplotě 4–6 °C (můžeme během oněch 48 hodin maso jednou „<prohnětnout>“).*

⁴⁵ Doklad z korpusu Araneum: *<Vyhnětnout> do bochánku. Přes noc těsto v ledničce odpočívá.*

⁴⁶ Doklad z korpusu csTenTen: *Adsl.Platím 200,- Kč měsíc mám garantovanou přidělenou rychlost 8kB./s a žádnéj podělanej limit 10GB a jednoduchým počtem si můžete <spočtnout>, kolik jsem schopen za 24h stáhnout, jasně, není to tak rychlé, ale oproti adsl je to zadarmo a občas možná rychlejší.*

⁴⁷ Doklad z korpusu csTenTen: *Bohužel v osobákách už žádá V8ka není, Taurus má jen V6 a Viktorka se už nedělá. Ještě nutno <započtnout> Lincolna a jeho Navigator.*

⁴⁸ Doklad z korpusu csTenTen: *Přidává dalších osm událostí, která otráslý, nebo mohly <otrásnout> světem mezi lety 2000–2008.*

⁴⁹ Doklad z korpusu csTenTen: *Zlověstný meloun si vysedává na vrcholku hory a vaším úkolem bude ho <setřásnout>.*

⁵⁰ Doklad z korpusu csTenTen: *... jediný, co pro ní můžu udělat, je vzít jí za ruku <proplétnout> si s ní prsty a možná se pokusit je trochu zahřát.*

⁵¹ Tvar *vléknout* je vždy překlep (mezera navíc).

⁵² Slovesa *střící/střehnout se a střežit (se)* nejsou ve všech užitích synonymní. V morfologickém slovníku by tudíž neměly být přegenerovány tvary *střežen*. * k lemmatu {*střící, střehnout, ...*} podle {*zmoci, zmohnout, zmožený*}. Tvary *střežen*. * nechť patří k lemmatu {*střežit, střežený*}, k lemmatu {*střící, střehnout, střehnuty*} by měly patřit pouze tvary *střežnut*. *, které ovšem v úzu prezentovaném korpusy doloženy nejsou. Verbální substantivum *střežení (se)* čeho se ovšem vztahuje ke slovesu {*střící, střehnout*}, kdežto *střežení koho* se vztahuje ke slovesu *střežit*. Chybou slovníku *Morfflex* je interpretace imperativu *střež (se)*, *střež[mt]e (se)* lemmatem *střežit*. Tvarem imperativu od lemmatu *střežit* je *střež*. Tvar imperativu *střež* je tvarem slovesa *střící (se)*, viz Internetová jazyková příručka (<http://prirucka.ujc.cas.cz/?slovo=st%C5%99ez>).

⁵³ Doklady jsou pouze ve věnovém korpusu Araneum: *Proto doufám, že se najde dostatek zodpovědných lidí, které případně osoby, jenž <zmohne> alkohol a budou mít potřebu provokovat a dělat bordel, pošlou do patřičných míst a nedopustí, abychom pošpinili důvěru v naše fandění!*

Pravidelné tvary od lemmatu {.*hrýzt, .*hrzynout} jsou .*hryz[[^][aá]].*/*.*hryzn.* (Na, na, co <hryzeš> valacha, ty bagáne? Nesmíte tabak hryzat jak chleba, nebo ho hned <rozhryzete> a bude na nic. ... zase jsem se vracel, <hryzen> výčitkami svědomí ...), a nikoli .*hryže.*/*.*hryz[áa]].*, patřící k lemmatu .*hryzat.

Infinitivy s[í]e]ci a seknout a většina prefigovaných odvozenin nejsou úplnými synonymy. Pokud ovšem zavedeme zásadu přiřazení tvarů l-ových příčestí .*sekl.* lemmatu .*s[í]e]ci a tvarů .*seknul.* lemmatu .*seknout, výsledky lemmatizace nebudou odpovídat intuici rodilého mluvčího. Intuici neodpovídá ani nynější stav desambiguace, kdy je přiřazení dvou různých lemmat tvarům sekl.* dílem statistiky.⁵⁷

Tvary jednoho lemmatu (. *nout) jsou v korpusu SYNv7 doloženy od sloves ozábnout, vyzábnout, dovládnout, ovládnout, povládnout, převládnout, spoluvládnout, zavládnout, zvládnout, nakrást, pokrát, odkrát, dokrát, prokrát, zakrát, překrát, doukrát, prohryznout, přehryznout, rozhryznout, vyhryznout, uhryznout, i dalších sloves 1. třídy vzoru nést neuvedených výše, a dále pak rozvléc[*it*], uvléc[*it*], nařknout, prořecknout se, přeřecknout se, zařecknout se, uřknout, vyřknout, postřechnout, i dalších sloves 1. třídy vzoru péci neuvedených výše, ani od osáhnout,⁵⁸ obsáhnout, přisáhnout, sáhnout, vsáhnout, zasáhnout, odpřisáhnout, zapřisáhnout.

Vícenásobná lemmata necht' mají synonymní slovesa s tvary podle typu *prosit* i *tisknout*:⁵⁹ {chytit, chytnout}, {podchytit, podchytnout}, {přichytit, přichytnout}, {uchytit, uchytnout}, {zachytit, zachytnout}, {odchlípít, odchlípnout}, {rozchlípít, rozchlípnout}, {rozštípnout, rozštípnout}, {oblbit, oblbnout}, {rozglábit, rozglábnout}, {nachladit, nachladnout}, {zatřpytit, zatřpytnout}, {zhasit, zhasnout}, {obstoupit, obstoupnout}, {odstoupit, odstoupnout}.

Teoreticky by bylo možné mít v těchto případech lemmata dvě. Bylo by ovšem třeba důsledně dodržet pravidlo, podle kterého by tvary na .*nut(y) patřily k lemmatu {.*nut, .*nutý} a tvary .*[eě]n(y) k lemmatu {.*it, .*[eě]ný}⁶⁰. Domníváme se ovšem, že jde-li o kolísání mezi konjugačními typy, pak je z hlediska analogie s případy jako *oblečen(y)/obléknut(y)*

⁵⁴ Doloženo pouze v korpusu csTenTen: ... ale v příslovích se všelijaké archaismy udržují rýmem velmi dlouho. V dnešní době <vztěci> se už nadobro vymizelo a máme jen vzteknouti se, ...

⁵⁵ Doloženo pouze v korpusu csTenTen: Dá se silničně <zteknout> třeba Milešovka?

⁵⁶ Infinitiv *poseknout* je doložen pouze v korpusu Araneum: Protože po neděli odjíždíme, tak jsme museli <poseknout> trávu a to bylo něco.

⁵⁷ A my s dědečkem právě sekli /seknout/VpMP--XR-AA--1 trávu. ... sníh uklízeli, sekli /síci/VpMP--XR-AA--trávu, hrabali listí, ...

⁵⁸ Jediný relevantní doklad v korpusu SYNv7 je: Očekáváme, že hráčům přinese zkvalitnění herních činností a také si sami <osáhnout> to, když budou sami v cizím prostředí. Ve všech dalších dokladech se jedná o překlapy (většinou vynechané písmeno, takže ve skutečnosti se jedná o tvary slovesa *dosáhnout*).

⁵⁹ Tento seznam by měl být uzavřený. Akademická gramatika spisovné češtiny (Štícha et al. 2013, s. 486) uvádí pouze sloveso *chytit/chytnout*. Slovníky (především SsJČ) jich uvádějí více. Většinou se v případě příbuzných sloves s tvary podle *prosit* a *tisknout* jedná o dvě různá významově odlišná slovesa. Synonymních sloves je poměrně málo. Synonymie patrně souvisí s tím, že slovesné kořeny označují děj, který probíhá v okamžiku, takže je potlačen rozdíl způsobu slovesného děje (momentánní × nemomentánní) v češtině často vyjadřovaný právě rozdílem mezi 2. a 4. třídou slovesnou. Podobně ojedinelými jsou synonymní dvojice *tančit* a *tancovat*, *soustružit* a *soustruhovat* (viz Šlosar 1981, s. 58).

⁶⁰ Akademická gramatika spisovné češtiny (Štícha et al. 2013, s. 486) uvádí ke slovesu *chytit/chytnout*: „Trpné příčestí se tvoří výhradně podle vzoru *prosit* (viz Tvoření neurčitých slovesných tvarů).“ V korpusu SYNv7 jsou doloženy jen dlouhé/adjektivní tvary *chytnutý* (Myslíte si, že *byly* penalty špatně kopnuté nebo dobře <chytnuté>?).

vícenásobné lemma případnějším řešením. Naopak v okrajových případech synonymních dvojice typu *tančit/tancovat* a *soustružit/soustruhovat* vykazují obě sady tvarů takovou míru autonomie (dané výraznou kmenotvornou příponou), že je u nich možné nadále ponechat lemmata dvě, přestože jde o synonymní slovesa.

5.3 Návrh k typu 3

Existují-li dvě varianty infinitivu, které lze pokládat za synonymní a u jednotlivých tvarů slovesných subparadigmat nelze nikdy určit, ke kterému infinitivu/lemmatu se mají přiřadit (variantní i nevariantní tvary lze na základě analogie morfologického systému češtiny řadit k oběma variantním lemmatům), pak necht' je využít koncept vícenásobného lemmatu.

Slovesa kolísající mezi typy *krýt – prosit/trpět/sázet* s variantním infinitivem na [íe]t:⁶¹ {tlít, tlet}, {zetlít, zetlet}, {chvít, chvět}, {zachvít, zachvět}, {rozechvít, rozechvět}, {prochvít, prochvět}, {pozachvít, pozachvět}, {dochvít, dochvět}, {s(t)kvít,⁶² s(t)květ}, {zas(t)kvít, zas(t)květ}.

Nepravidelná slovesa:

{chtít, chtět}, {jíst, jest⁶³}, {najíst, najest}, {dojíst, dojest⁶⁴}, {pojíst, pojest}, {přejíst, přejest⁶⁵}, {zajíst, zajest}.

Z mluveného úzu jsou známy i další prefigované varianty. Lze je najít na internetu (*projest*,⁶⁶ *ujest*,⁶⁷ *vyjest*⁶⁸). Časté jsou ovšem případy, kdy různé graficky pojaté tvary mají význam související se slovesem *jet*.

5.4 Typ 4

Existují-li dvě varianty infinitivu, které lze pokládat za synonymní a které jsou jedinými variantními tvary, k nimž lze přiřadit pouze nevariantní tvary, pak necht' je využít koncept vícenásobného lemmatu.

⁶¹ Tento seznam by měl být uzavřený. Kromě uvedených sloves by k tomuto typu patřilo vulgární sloveso, které kolísá mezi typem *krýt – trpět – dělat*, a má tudíž čtyři variantní tvary infinitivu: {chčít, ščít, chcát, scát}, {vychčít, vysčít, vychcat, vyscat},

⁶² V korpusech řady SYN se tyto substandardní tvary okrajově vyskytují a jsou lemmatizovány lemmatem *skvět*.

⁶³ Substandardní infinitiv je homonymní s archaickým tvarem 3. osoby indikativu přítomného aktiva slovesa *být*. V korpusech SYNv7 se doklady vyskytují: *Myslela, že budeme <jest> šišky, a zatím jíme zemáky*. Jsou lemmatizovány chybně jako tvary slovesa *být*.

⁶⁴ Tvary jsou doloženy ve webovém korpuse řady Araneum. Doklady svědčí pro substandardní infinitiv slovesa *dojíst* (*K tomu pochopitelně už v rodině zazita hlaska: „To se musí <dojest>“, kterou babička aplikuje vzdy, když jí něco zbyva...*), ale i pro chybně graficky zaznamenané substantivum *dojezd* (... *ale benzín mě vůbec v hadičce neproudí přitom hlavu a vše co k tomu patří jsem o pucoval nevíte co stím díkes a dobrej <dojest> všem ...*).

⁶⁵ Tvary jsou doloženy ve webovém korpuse řady Araneum: *„Za chvíli bude oběd, takže se nesmíš moc <přejest>, abys obědval.“*

⁶⁶ *Má problém závislost na jídle a chlastu. Zkoušela sis s ním promluvit? Nemá zájem se změnit, kam až se chce <projest> na dva metraky nebo pres?*

⁶⁷ ... *mohla jsem se <ujest> rybích salátů ...*

⁶⁸ *tak jsem včera byl <vyjest> tu šelepku za pár peněz a musím říct, že dobrý, dost dobrý, za cenu dvou menu a dvou pivíček pecka.*

Do této skupiny patří:

(a) slovesa typu *trpět/sázet* s variantním infinitivem na [íě]t:⁶⁹

{bdít, bdět}, {bzdít, bzdět}, {čnít, čnět}, {čpít, čpět}, {doznít, doznět}, {hřmít, hřmět}, {lpít, lpět}, {od(e)znít, od(e)znět}, {pomnít, pomnět}, {proznít, proznět}, {rdít, rdět}, {rozeznít, rozeznět}, {setmít, setmět}, {souznít, souznět}, {tkvít, tkvět}, {tmít, tmět}, {ulpít, ulpět}, {utkvít, utkvět}, {vyčpít, vyčpět}, {vyznít, vyznět}, {zardít, zardět}, {zahřmít, zahřmět}, {zatmít, zatmět}, {zaznít, zaznět}, {znít, znět}, popřípadě další prefigované varianty. V korpusu SYNv7 jsou doloženy i další lokální varianty jako například: {odpovědít, odpovědět}, {ubzdít, ubzdět}, {vědít, vědět}.

(b) Slovesa typu *krýt* s variantním infinitivem na [íě]t:⁷⁰

{klít, klet⁷¹}, {plít, ⁷²plet}, {proklít, proklet⁷³}, {odít, odět⁷⁴}, {podít, podět⁷⁵}, {sít, set⁷⁶}, {zaklít, zaklet⁷⁷}, {zít,⁷⁸ zet⁷⁹}.

(c) Slovesa typu *umřít* s variantním infinitivem na [íě]t:⁸⁰

{mlít, mlet}, {semlít, semlet}, {umlít, umlet}, {podemlít, podemlet}, {rozemlít, rozemlet}, {namlít, namlet}, {vymlít, vymlet}, {omlít, omlet}, {domlít, domlet}, {přemlít, přemlet}, {zamlít, zamlet}, {promlít, promlet}, {odemlít, odemlet}, {předemlít, předemlet}, {zmlít, zmlet}, {přimlít, přimlet}, {mlít, mlet}, {mlít, mlet}, {pomlít, pomlet}.

⁶⁹ Tento seznam by měl být uzavřený.

⁷⁰ Tento seznam je uzavřený.

⁷¹ Část dokladů jsou správně interpretované tvary pasivního přičestí, část jsou nerozpoznané variantní infinitivy interpretované mylně stejně a část jsou rozdělená slova. (... budiž <klet>! ... Já teď kleju tak, jak jsem si nikdy nemyslí, že budu někdy <klet>. Vše budeme ladit <klet> nímú mistrovství světa ...).

⁷² Menší část dokladů jsou případy homonymního lemmatu *plít* s významem *plivat*: *Svíčky jak nahrbení skřítci těží tu hebky noční klid rubají tiše do všech stran pod nohy sobě budou <plít> ... Sám Antonín Václavík byl rodkem luhačovickým a připomíná, že v jeho dětství nebylo dovoleno <plít> nebo močit do studánky či řeky. V jistých kruzích patří k dobrému tónu se s dalajlámou vyfotografovat, v jiných zas na něj <plít> a lísat se k Číně.* K této interpretaci samozřejmě variantní infinitiv *plet* vztáhnout nelze.

⁷³ Většina dokladů jsou tvary pasivního přičestí správně rozpoznané automatickou morfolgickou analýzou. Najdou se ovšem i tvary infinitivu: *Na jevišti tak ožije příběh žen, které uměly pomoci a léčit, ale i uřknout a <proklet>.* I ty jsou interpretovány jako tvary pasivního participia. Naopak tvary *zaklet* jsou v korpusu SYN v7 více méně všechny případy pasiva, ačkoliv infinitiv nelze vyloučit.

⁷⁴ Tvar je homonymní s tvarem pasivního participia: „A tak jako blesk z čistého nebe, či spíše z modrého Atlantiku, přišel rozkaz, že mám složit svých pět švestek, <odět> parádní modř a být připraven k transportu do San Franciska!“ ale ... *Gudea je <odět> do černé tuniky ...*

⁷⁵ *Protože nebylo jasné, kam se vlastně mohli <podět>, prohledávali policisté okolí Jáchymova ...*

⁷⁶ Doklady jsou vesměs desambiguovány chybně a je jich poměrně málo (většina dokladů je buď na tvar číslovky *sto*, nebo substantivum *set*). Přesto se variantní infinitivy vyskytují: *Stroj byl ale přepracován, takže je univerzální a může <set> jak do zpracované tak nezpracované půdy. Málo kdy se v tuhle dobu podaří, abychom mohli <set>. Výsevni ústrojí je mechanické (lze <set> mák i kukuřici) distribuce osiva je pneumatická.*

⁷⁷ Až na jediný (*Tak aš pudete do stráně a začne to bejt na zděch těžky, tak musíte pořádně <zaklet>, vono to udělá s rance hop a budete vidět, že pudete jako prázdný.*) jsou doklady v korpusu SYNv7 tvary pasivního přičestí.

⁷⁸ O variantě *zít/zet* svědčí doklady z korpusu SYNv7 z Deníku Bohemia (např. *Celé léto bude zít prázdnou.*). Akademická gramatika spisovné češtiny (Štícha et al. 2013, s. 481), řadí sloveso *zet* ke slovesům „bez alternace samohlásky v tvarotvorném základu“ (spolu se slovesy *pět, spět*). Automatická analýza tvar *zít* řadí k nerozpoznaným tvarům.

⁷⁹ *Řekl bych, že bude akorát <zet> prázdnou.*

⁸⁰ Tento seznam je uzavřený.

6. Závěr

Jedním z kroků automatické morfologické analýzy užívané např. k tagování velkých korpusů je lemmatizace. V dosavadní praxi jsou výsledky automatické morfologické analýzy nekonzistentní v případech, že základní tvar/lemma je variantní (má více podob). Velmi komplikovaná je lemmatizace sloves, jejichž tvarosloví je z principu mnohem komplexnější než tvarosloví jmen. Na základě dosavadních mluvnických popisů, slovníků i jazykových korpusů jsme se zabývali otázkou variantnosti/synonymie infinitivu s ohledem na tvaroslovnou soustavu, kterou infinitiv reprezentuje. Na základě klasifikací a analýz popsanych výše konstatujeme, že konzistentní lemmatizace sloves se synonymními/variantními tvary v infinitivu vícenásobným lemmatem podle návrhu projektu NovaMorrf (Osolsobě et al. 2017; Petkevič et al. 2019) by neměla narazit na výrazné problémy, protože se jedná především o uzavřené skupiny sloves, jejichž výčet opřený o rozsáhlé korpusy máme k dispozici (viz výše).

U uzavřených tříd je vždy možné zvažovat, nakolik je vhodné generovat i další prefigované varianty, které nejsou doloženy ve sledovaných korpusech. Proti hovoří fakt, že se někdy může jednat buď o překlepy,⁸¹ nebo o homonymní tvary.⁸²

Naším cílem nikterak není tvrdit, že se všechna výše uvedená data mají zaznamenávat v morfologickém slovníku. Chtěli jsme ukázat, že pokud zaznamenávána budou, pak by se tak mělo dít konzistentně.

ZDROJE

BENKO, V. (2015): Srovnatelné webové korpusy Aranea. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <<http://www.korpus.cz>>.

CVRČEK, V. – VONDŘIČKA, P. (2013): *Morfio*. Praha: FF UK. Dostupný z WWW: <<http://morfio.korpus.cz>>.

Czech web corpus crawled by SpiderLing in May, October and November 2017, October and November 2016, October and November 2015. Czech Wikipedia downloaded in November 2017. Encoded in UTF-8, cleaned, deduplicated, tagged using majka+desamb pipeline v2. Dostupný z WWW: <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fctesten17_mj2>.

Czech Web 2017 (csTenTen17).

Internetová jazyková příručka. Dostupná z WWW: <<http://prirucka.ujc.cas.cz/>>.

KŘEN, M. – CVRČEK, V. – ČAPKA, T. – ČERMÁKOVÁ, A. – HNÁTKOVÁ, M. – CHLUMSKÁ, L. – JELÍNEK, T. – KOVAŘÍKOVÁ, D. – PETKEVIČ, V. – PROCHÁZKA, P. – SKOUMALOVÁ, H. – ŠKRABAL, M. – TRUNEČEK, P. – VONDŘIČKA, P. – ZASINA, A. (2018): *Korpus SYN, verze 7 z 29. 11. 2018*. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <<http://www.korpus.cz>>.

ŠMERK, P. (2009): *Deriv – webové rozhraní*. Dostupné z: <<http://deb.fi.muni.cz/deriv>>.

LITERATURA

BENKO, V. (2014): Aranea: Yet Another Family of (Comparable) Web Corpora. In: P. Sojka – A. Horák – I. Kopeček – K. Pala (eds.), *TSD 2014*, LNAI 8655, 257–264. Springer International Publishing.

CVRČEK, V. et al. (2010): *Mluvnice současné češtiny*, 1. Praha: Nakladatelství Lidové noviny.

⁸¹ Např. ve webových korpusech se vyskytuje tvar *načtnout*, a to nikoli jako variantní infinitiv k *načíst*, ale jako překlep za *načrtnout*.

⁸² Například tvary pasivních příčestí na t (*odět, podět, zaklet, ...*).

- HAJIČ, J. (2004): *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Vol. 1. Praha: Karolinum Charles University Press.
- HAJIČ, J. – HLAVÁČOVÁ, J. (2013): *MorfFlex CZ*, LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, <<http://hdl.handle.net/11858/00-097C-0000-0015-A780-9>>.
- HLAVÁČOVÁ, J. (2009): *Formalizace systému české morfologie s ohledem na automatické zpracování českých textů*. Disertační práce. Praha: UK. Dostupná z: <<http://utkl.ff.cuni.cz/phpBB3/viewtopic.php?f=11&t=1>>.
- HNÁTKOVÁ, M. – KŘEN, M. – PROCHÁZKA, P. – SKOUMALOVÁ, H. (2014): The SYN-series corpora of written Czech. In: N. Calzolari (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 160–164.
- KOMÁREK, M. et al. (1986): *Mluvnice češtiny*, 2. Praha: Academia. (cit. jako MČ 2)
- OSOLSOBĚ, K. (1996): *Algoritmický popis české formální morfologie a strojový slovník češtiny*. Disertační práce. Brno: Masarykova univerzita.
- OSOLSOBĚ, K. (2011): *Morfologie českého slovesa a tvoření deverbativ jako problém strokové analýzy češtiny*. 1. vyd. Brno: Masarykova univerzita.
- OSOLSOBĚ, K. – HLAVÁČKOVÁ, D. – PALA, K. – ŠMERK, P. (2009): Exploring Derivational Relations in Czech with the Deriv Tool. In: *NLP, Corpus Linguistics, Corpus Based Grammar Research*. Bratislava, Slovakia: Tribun, 152–161.
- OSOLSOBĚ, K. – HLAVÁČOVÁ, J. – PETKEVIČ, V. – ŠIMANDL, J. – SVÁŠEK, M. (2017): Nová automatická morfologická analýza češtiny. *Naše řeč*, 100 (4), 225–234.
- OSOLSOBĚ, K. – ŽIŽKOVÁ, H. (2016): Automatická morfologická analýza z hlediska pokrytí a nepokrytí morfologických variant. In: *Korpusová lingvistika Praha 2016, Od mluvené češtiny k psané, 17. září 2016*, Praha.
- PETKEVIČ, V. – HLAVÁČOVÁ, J. – OSOLSOBĚ, K. – SVÁŠEK, M. – ŠIMANDL, J. (2019): Parts of Speech in NovaMorf, A New Morphological Annotation of Czech. *Journal of Linguistics / Jazykovedný časopis*, 70 (2), 358–369.
- ŠLOSAR, D. (1981): *Slovotvorný vývoj českého slovesa*. Brno: UJEP.
- ŠMILAUER, V. (1972): *Nauka o českém jazyku*. Praha: SPN.
- ŠTÍCHA, F. et al. (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.
- VOLKOVÁ, N. (2017): *Vybraný problém homonymie slovesných tvarů z hlediska automatické morfologické analýzy*. Brno: FF MU. (Nepublikovaná magisterská diplomová práce pod vedením K. Osolsobě.) Dostupná z WWW: <<https://is.muni.cz/auth/th/nrfef/>>.
- ZIKOVÁ, M. (2016): When Prosody Follows Syntax: Verbal Stems in Czech. *Linguistica Brunensia*, 64 (1), 163–185.