



Variace a frekvence variant na příkladu tvrdých neživotných maskulin

NEIL BERMEL

Materiál pro tuto studii vznikl jako součást dvou společných výzkumných projektů, ve kterých jsme používali ČNK pro zjištění stavu české morfologie.¹ Po shromáždění četných dat ohledně nejrozšířenějšího a snad i nejpestřejšího paradigmatu v české morfologii, tedy paradigmatu tvrdých maskulin neživotných, je na místě uvažovat o užitečnosti pojmu „vzor“.

1. K definici pojmu „vzor“

Vzory používáme, abychom zjistili rychle a snadno, který formant použijeme v jistém kontextu. Podle Lyonse (1971, s. 189–190) mají vzory plnit didaktickou funkci spíše než funkci lingvistickou: vzor nevyjadřuje jazykové zákonitosti, nýbrž poskytuje rámec pro spolehlivé vytvoření správných tvarů.

Se vznikem tohoto pojmu ale vzory pronikají do lingvistického myšlení a tím i do pokusů o popis struktury jazyka. Např. Komárek (2006, s. 107) rozlišuje morfémy *paralelní* a *synonymické*. První se liší jen distribucí (Asg *hrad, pána, ženu, duši*) a v rámci téhož gramatického významu jsou komplementární (byť někdy ne plně komplementární). Druhé jsou komplementární jen v poměru k různým morfologickým typům, ale v rámci téhož morfologického typu jsou volitelné (1sg *píší/píšu*, Lsg *v jazyku/v jazyce*). Implikací takového tvrzení je, že vzor je a priori strukturou v jazyce, jelikož morfémy, které se uplatňují důsledně uvnitř něj, mají jednu povahu (jsou s morfémy dalšími „paralelní“), a jiné, které konkurují uvnitř paradigmatu, mají povahu jinou (jsou „synonymické“).

Taková klasifikace je nutná, protože existuje nesoulad mezi morfologickými funkcemi a počtem disponovaných tvarů, které mají danou funkci plnit. Rozsáhlá akademická literatura se zabývá problematikou *nedostatečné specifikace (underspecification)*. V dané situaci jeden prostředek pokrývá více morfologických funkcí, tj. jde o běžný případ *synkretismu*:

Většinu kontinentu pokrývají lesy.

Blížili jsme se k novému kontinentu.

Méně diskutována je *nadměrná specifikace (overspecification/overdifferentiation)*, při které daná morfologická funkce má k dispozici více jazykových prostředků:

Vyrostla na jiném kontinentě/kontinentu.

¹ Možnosti a meze české gramatiky ve světle Českého národního korpusu; Kapitoly z české gramatiky (podporovány GAČR, pod vedením Fr. Stíchy); Comparing corpus data and linguistic intuition in the study of Czech morphology (podporován Britskou akademií).

V článku o jednom případě nadměrné specifikace britský lingvista Brown (2007) dospěl k závěru, že pro adekvátní řešení nadměrně specifikovaných pozic je vhodné operovat s pojmem *neplného pádu* (angl. *sub-case*), který vytyčil podle následujících charakteristik:

- Jedná se o malou skupinu slov (v ruském „druhém lokálu“ 108 podst. jmen), příp. s vysokou frekvencí (přispívá k udržování tvarů).
- Slova se zvláštním tvarem v daném neplném pádě mají formální shody (prozodické, etymologické) a společné sémantické charakteristiky.
- Platí minimální diferenciaci mezi tvary neplného pádu a tvary mimo plný pád (např. zvláštní tvar pro ruský „druhý lokál“ existuje jenom u podst. jmen, nenajdeme vedle něj i zvláštní tvar adjektiv).

Ve vzoru „hrad“ je poměrně bohatě evidována nadměrná specifikace v Gsg (*hradu*, ale *lesa*, někdy *roka/roku*); v Asg (*hrad*, ale *tripla*, někdy *buřt/buřta*); ve Vsg (*hradě*, ale *taxíku*); v Lsg (*hradě/hradu*, ale *trhu*, *ovčíně*); v Lpl (*hradech*, ale *vlacích*, někdy *vlakách*); a v Ipl (*hrady*, někdy *hradama*).

V některých pádech (Vsg, Lpl a Ipl) můžeme motivovat volbu tvaru buď čistě fonologicky, nebo méně čistě podle stylistických intencí mluvčího/pisatele či podle žánru textu (o Lpl viz např. Bermel, 2007). Pro každé slovo se tedy uplatňuje jeden nebo dva tvary v závislosti na komunikační situaci.

Pro zbývající pády (Gsg, Asg a Lsg) je situace složitější. Volba formantů není jasně řízena fonologickými kritérii a není vždy pro každý člen paradigmatu stejná. Ve druhém pádě je tvar běžný *hradu*, nikoli **hrada*, na druhé straně existuje jenom *lesa*, nikoli **lesu*, ale je doloženo *roka* i *roku*.

Z hlediska deskriptivního je tedy na místě se zeptat, zda zařazením minoritních tvarů do „neplných“ pádů získáváme na průhlednosti a adekvátnosti popisu tohoto paradigmatu.

2. Asg – „fakultativní“ životnost

Příznaky „životnosti“ v j. č. u jmen životných vzorů spočívají hlavně v užití zvláštního životného tvaru u podst. jmen v Dsg a Lsg (zavolám tomu *studentovi*, mluvím o *studentovi*) a v synkretismu mezi Asg a Gsg při užití formantu *-a* (kniha toho *studenta* = znám toho *studenta*).²

Termín „fakultativní životnost“ přejatý z americké slavistiky označuje situaci jinou, ve které nejde o životný předmět a rozsah variace je mnohem menší. Dsg, Lsg, Npl se obvykle neliší od jiných jmen neživotných (zvláštní případy tvoří jména s životným významem použítá pro popis věcí: *drak* atd.) a poměr mezi tvary Asg a Gsg není tak jasný:

Gsg ≈ Asg :	vůně turka = pít turka	Gsg -a = Asg -a
	barva buicka ≈ mít buick/buicka	Gsg -a ≈ Asg -Ø/-a
	barva fiata/fiatu ≈ mít fiat/fiata	Gsg -a/-u ≈ Asg -Ø/-a
	barva citroenu ≠ mít citroen/citroena	Gsg -u ≠ Asg -Ø/-a
		(podle SYN2000/2005/2006PUB)

² Není žádný spolehlivý způsob, jak rozpoznat v korpusu slova s fakultativní životností. Tagger s tímto pojmem neoperuje a tvary Asg. na *-a* jsou proto označeny buď jako 2. pád, nebo jako 4. pád živ. rodu, jindy jsou přiřazeny k jinému pádu a deklinaci, či nejsou rozpoznány. Nezbyvalo tedy, než vyhledat každý tvar manuálně podle seznamů z vědecké literatury, a proto nepokládám tento seznam slov za úplný.

Status fakultativní životnosti v korpusech SYN není jednoznačný. Bez variací (tj. s důslednou implementací formantu -a) jsou tato frekventovaná slova (nad 20 výskytů v korpusech): *bach, bych, drak, flok, majzl, panák, pták, slaneček, šlofik, tripl, turek*. Všimneme si, že některá (*drak, pták, Turek* s velkým *t*) jsou v nejčastějším významu slova živ. rodu a neživ. význam je vedlejší. Jiná se užívají prakticky jen v Asg, a to v určitých slovních spojeních (*dát si bacha, honit bycha*), a není vyloučeno, že tu nejde o slova s celým paradigmatem. U dalších slov je proporce dokladů s jedním či druhým formantem rozdílná, uvádíme jen příklady nad 20 výskytů:

- -a nad 90 %: *bavorák (auto), frían, kulich, motýlek (kravata), sólokapr, stoják, šluk*
- -a 61–90 %: *april, buřt, debl, joint, koníček, motýlek (plavání), peugeot, silvestr, špaček, trabant, vajgl, walkman*
- -a 41–60 % : *fiat, ford, kombík, lok, taxík*
- -a 10–40 %: *budík, cadillac, hřib, opel, padák, poštěvák, renault, skútr, volkswagen*
- -a 1–10 %: *budíček, cvik, dudlík, džíp, flek, chevrolet, meloun, mercedes, poštěváček, rum, truhlík, vřed, zlepšovák*
- -a pod 1 %: *doutník, mail/mejl, mobil, moučník, vřed*

Předběžný závěr o Asg podle korpusů SYN tedy je, že se u valné většiny slov fakultativní životnost neuplatňuje důsledně. Jména, která mají k dispozici tento tvar, spadají do sémanticky a tvaroslovně omezených skupin, ale nicméně nemůžeme z členství v té či oné skupině vydedukovat, do jaké míry se používá životný tvar (viz výše o značkách aut).

3. Gsg

V Gsg jde o volbu formantů *-a/-u*. Gsg *-u* má zhruba 13 000 podst. jmen v SYN2005 (tj. 13 tisíc lemmat; výsledky ještě nebyly manuálně procesány, a proto zahrnují i špatně lemmatizovaná slova). Z nich se 112 jmen pojí rovněž s *-a* v různých proporcích a 65 jmen se pojí výhradně s *-a*.

Analýza průměrné frekvence a její mediány ukazují, že Gsg *-a* má vyšší proporce jmen ze základní složky slovní zásoby, než má Gsg *-u*. Pro *-a* je průměrná frekvence tvarů 1455 (medián 10,5), kdežto pro *-u* je průměrná frekvence tvarů 92 (medián 3) a pro jména s variací je průměrná frekvence tvarů 1368 (medián 120,5): slovní tvary na *-u* mají průměrnou frekvenci 817 (medián 22) a slovní tvary na *-a* mají průměrnou frekvenci 551 (medián 17). Vyšší průměrná frekvence a vyšší medián svědčí o častějším užití daných slov a tvarů. I přesto, že data nejsou procesována, si můžeme všimnout shodnosti se závěry Akademické mluvnické, podle které 13 210 slov uplatňuje formant *-u* v Gsg, a 240 uplatňuje formant *-a*.

Gsg na *-a* má proto okrajovou pozici v české morfologii systémově: jde celkem o 177 podst. jmen ze 13tis. v SYN2005, tj. 1,4 %. Nicméně se *-a* používá dost často: 116 364 dokladů ze 1374tis., čili 8,5 % všech dokladů. Pro existenci dvou tvarů v Gsg ne-lze nalézt funkční zdůvodnění (tj. žádné sémantické, prozodické, či fonetické rozdíly).

4. Lsg

V Lsg jde o volbu formantů *-ě (-e)/-u*. V korpusu SYN2005 se 6 801 jmen pojí výhradně s formantem *-u*, 53 jmen se pojí výhradně s formantem *-ě* a 392 jmen se pojí s *-ě* nebo *-u* v různých proporcích.³

³ Jde o ručně vytříbené výsledky provedené Mgr. L. Knittlem v rámci projektu Comparing corpus data and linguistic intuition in the study of Czech morphology.

Frekvenční distribuce tvarů v Lsg v SYN2005 se značně liší od distribuce tvarů v Gsg. Slova s Lsg jenom na *-u* nemají vysokou průměrnou frekvenci (74) ani vysoký medián (3). Slova s Lsg jenom na *-ě* ale mají dokonce nižší průměrnou frekvenci (9) a medián (1). Vysokou průměrnou frekvenci (807) a medián (110) mají slova s kolísáním ve formantech Lsg (*-ě/-u*): Slovní tvary na *-u* mají průměrnou frekvenci 158 (medián 38,5) a slovní tvary na *-ě* mají průměrnou frekvenci 650 (medián 13).

Pro porovnání: Akademická mluvnice uvádí zhruba 12 300 hesel s formantem *-u* v Lsg., přibližně 120 s formantem *-e/-ě* a zhruba 1200 hesel s kolísáním mezi dvěma formanty. Korpus SYN2005 tedy prezentuje asi polovinu možných tvarů v tomto pádě.

Frekvence formantu *-u* je v podstatě podobná v Gsg a v Lsg. U slov, která mají jenom *-e/-ě* v Lsg, je frekvence velmi nízká, jde o málo frekventovaná slova nebo se v mnoha případech jedná o novotvary (nemůžeme vyloučit, že chybí formant *-u* jenom kvůli nízké frekvenci). U slov, která kolísají mezi formanty *-e/-ě* a *-u*, je frekvence značně vyšší.

Lsg na *-e/-ě* tedy nemá tak okrajovou pozici v české morfologii jako Gsg na *-a*: užívá se celkem u 445 podst. jmen ze 6801 v SYN2005, nebo 6,5 %. Formant *-e/-ě* se užívá ve 255 198 dokladech ze 753 231, čili 33,9 % všech dokladů. Počet jmen, která se pojí výhradně s *-e/-ě*, je mizivý a taková jména nejsou často doložena. Téměř každý tvar s formantem *-e/-ě* má alternativní tvar na *-u*, což může svědčit o ústupu tohoto tvaru v rámci systému.

5. Výskyt „alternativních“ formantů ve více pádech

Pokusíme se teď o zjištění možných kombinací formantů (méně používané nazýváme pro průhlednost „alternativními“: Asg *-a*, Gsg *-a*, Lsg *-e/-ě*). První skupina má jenom jeden alternativní formant:⁴

- Gsg *-u*, Asg *-a*, Lsg *-u*: šlofik, spodek, šluk... (N = 55)
- Gsg *-u*, Asg *-a/-Ø*, Lsg *-u*: citroen, čoud, fiat, gól, sólokapr... (N = 45)
- Gsg *-a*, Asg *-Ø*, Lsg *-u*: březan, lesíček, dvanáctiúhelník... (N = 44)
- Gsg *-a/-u*, Asg *-Ø*, Lsg *-u*: betlém, dnešek, kousek, kurník... (N = 61)
- Gsg *-u*, Asg *-Ø*, Lsg *-ě*: úbyt, velehrad, kocourkov, hunt... (N = 42)
- Gsg *-u*, Asg *-Ø*, Lsg *-e/-ě/-u*: autobus, časopis, parlament... (N = 355)

V druhé a ve třetí skupině jde o alternace ve více pádech:

Gsg / Asg

- Gsg *-a*, Asg *-a*, Lsg *-u*: búr, ferbl, frtán, kozáček... (N = 25)
- Gsg *-a*, Asg *-a/-Ø*, Lsg *-u*: apríl, joint, kulich, tuzér... (N = 18)
- Gsg *-u/-a*, Asg *-a*, Lsg *-u*: fiátek, flok, panák... (N = 9)
- Gsg *-u/-a*, Asg *-a/-Ø*, Lsg *-u*: bavorák, bedekr, budík, hřib... (N = 29)

Gsg / Lsg

- Gsg *-a*, Asg *-Ø*, Lsg *-e/-ě*: hřebčín, ovčín, prales... (N = 8)
- Gsg *-a*, Asg *-Ø*, Lsg *-u/-e/-ě*: hřbitov, kravín, les, svět, život... (N = 10)
- Gsg *-u/-a*, Asg *-Ø*, Lsg *-ě*: polosvět, pronárod, vepřín (N = 3)
- Gsg *-u/-a*, Asg *-a/-Ø*, Lsg *-u/-ě*: buřt (N = 1)
- Gsg *-u/-a*, Asg *-Ø*, Lsg *-u/-e/-ě*: dvůr, javor, jazyk, klín, kostel... (N = 25)

⁴ Z tohoto seznamu však nemůžeme automaticky vyvodit, že pro daná slova další možnosti alternace neexistují. Absence dokladů nějakého tvaru v korpusu neznamená, že tento tvar není přípustný, zvláště v případě slov s nízkou frekvencí.

Přehled kombinací naznačuje, že Gsg *-a* a Asg *-a* souvisí jen částečně: Asg na *-a* se uplatňuje mnohem častěji s Gsg na *-u*. Často citovaná souvislost mezi Gsg *-a* a Lsg *-ě* se v korpusu důsledně realizuje jen pro úzkou skupinu slov (N = 8, většinou zemědělské termíny). Mnohem běžnější je částečné překrývání typu „Gsg *-a/-u* = Lsg *-e/-ě/-u*“ (N = 25) a nejfrekventovanější se zdají být slova, která mají jenom jeden „alternativní“ tvar, a to buď v Gsg, Asg, nebo Lsg.

6. Závěry

Jak v gramatice zacházet s nadměrnou specifikací? První způsob je rozložení slov do více vzorů nebo podvzorů (např. vzor „svět“, Lsg *světě*; vzor „kontext“, Lsg *kontextu*; vzor „kontinent“, Lsg *kontinentě* i *kontinentu*). Druhý způsob je, že stanovíme více funkcí, např. rozlišujeme lokální versus prepoziční užití v Lsg a pokusíme se seřadit tvary podle toho. Třetí způsob spočívá ve specifikaci „neplných pádů“ u méně frekventovaných formantů.

Výhodou prvního způsobu je jasnost výkladu pro psané gramatiky. Nevýhodou je ale mnohem více. Znamená to rozmnožení počtu velmi podobných vzorů: z jednoho na nejméně 16, nebo rozšíření kolísání mezi podobnými vzory. Toto řešení nenabízí žádný způsob, jak zacházet s frekvenčním kolísáním mezi tvary a problematika volby mezi tvary zůstává nevyřešena.

Druhý způsob má výhodu, že zachycuje některé faktory ovlivňující volbu toho či onoho tvaru (např. v Lsg *na hradě/o hradu*, v Asg *najít hřib/hřiba*). Za nevýhodu bychom mohli označit implikování funkčních rozdílů tam, kde být nemusí (*kolem rybníku/rybníka, na balkoně/balkonu*). Navíc tento způsob neřeší vůbec situaci tam, kde nejde o žádné funkční rozdíly. Další problém je, že rozšiřuje počet pádů a tím i synkretismus (má každé paradigma dva „lokální“ pády, L1 a L2, podle významu, nebo dva „akuzativní“ pády podle expresivity?).

Co se týče třetího způsobu, má jisté výhody, že zachovává průhlednost „velkých“ paradigm, omezuje rozrůstání se plných „pádů“ a „vzorů“, které by platily jen pro malé skupiny slov. Navíc každý ze zkoumaných případů variace plní alespoň jedno z kritérií pro „neplné“ pády. Problematická zůstává neplná aplikace kritérií stanovených Brownem.

Ze tří zkoumaných jevů se jako neplný pád nejjasněji profiluje Asg. Plní ze všech nejvíce kritérií (vyskytuje se to v malé skupině slov, která mají formální shodnosti jako cizí původ a jisté sémantické shodnosti). Nejproblematictější z tohoto hlediska se zdá být Lsg, u kterého se minoritní tvary vyskytují u početných slov a v podstatné části dokladů Lsg. Naplňuje ale alespoň částečně několik z Brownových kritérií, včetně formálních a sémantických. U Gsg tento popis selhává z jiného důvodu: naprosto chybí formální a sémantické shodnosti, jeho status zdánlivě záleží čistě na tom, že se uplatňuje v nevelké skupině slov.

Vznik neúplných pádů Brown vysvětluje historickými důvody: druhý lokál v ruštině vznikl jako pozůstatek vedlejšího deklinačního typu (tzv. u-kmenů), do kterého v praslověnanštině patřila podstatná jména s určitými společnými rysy a jehož vývoj se v pozdějším vývoji ruštiny zastavil. Fakultativní životnost rovněž představuje „zastavený vývoj“ kategorie, pro Lsg a Gsg je však toto vysvětlení neuspokojivé. Gsg na *-a* a Lsg na *-e/-ě* jsou naopak pozůstatky formantů deklinačního typu, do kterého patřila naprostá většina maskulin (tzv. o-kmenů). Postupující expanze formantů dříve minoritní u-kmenové deklinace v češtině – tzv. *exapce* (viz např. Janda, 1996) – nechává proto v porovnání s ruštinou poměrně velké a různorodé skupiny jmen, u kterých se nadále uplatňují formanty tradiční pro o-kmeny.



Literatura

- BERMEL, N. (2007): What can corpora tell us about centre and periphery in Czech morphology? In: F. Štícha, J. Šimandl (eds.), *Gramatika a korpus / Grammar & Corpora 2005*. Praha: ÚJČ AV ČR, s. 5–16.
- BROWN, D. (2007): Peripheral functions and overdifferentiation: The Russian second locative. *Russian Linguistics*, 31 (1), s. 61–76.
- JANDA, L. (1996): *Back from the Brink: A Study of how Relic Forms in Languages Serve as Source Material for Analogical Extension*. Mnichov: Lincom Europa.
- KOMÁREK, M. (2006): *Příspěvky k české morfologii*. Olomouc: Periplum.
- LYONS, J. (1971): *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- PETR, J. a kol. (1986): *Mluvnice češtiny 2. Tvarosloví*. Praha: Academia.
- ŠULC, M. (2001): Životná koncovka -a v akuzativu singuláru neživotných maskulin. *Jazykovědné aktuality*, 38, s. 117–128.
- Český národní korpus – SYN2000. Ústav Českého národního korpusu FF UK, Praha 2000. Dostupný z WWW: <<http://www.korpus.cz>>.
- Český národní korpus – SYN2005. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <<http://www.korpus.cz>>.
- Český národní korpus – SYN2006PUB. Ústav Českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <<http://www.korpus.cz>>.

Variation and the frequency of variants in hard masculine inanimate nouns

This article examines variation in three case forms of the Czech masculine hard inanimate declension pattern and asks how this variation is best described: by breaking this declension class into smaller classes; through identifying functional differences in the usage of these forms; or by establishing these as sub-cases. Brown's 2007 criteria for sub-cases in Russian are considered and tested against the data. While the sub-case explanation is seen to have certain advantages over the others, the Czech data diverge from Brown's criteria in ways that test the limits of the model.

*Department of Russian and Slavonic Studies,
Faculty of Arts and Humanities, University of Sheffield
n.bermel@sheffield.ac.uk*