

KAREL PALA, PAVEL ŠEVEČEK

VALENCE ČESKÝCH SLOVES

1. Úvod

Jedním z dluhů českého gramatického bádání je skutečnost, že chybí podrobnější materiálové zpracování českých slovesných valencí v podobě rozsáhlejšího a tedy i reprezentativnějšího seznamu českých sloves, který by obsahoval jejich valence.

Ani české slovníky neposkytují příliš spolehlivé a systematické údaje o valencích českých sloves: SSČ (Praha 1994, 2.vyd.) uvádí slovesné valenční vzorce explicitně jen okrajově a v příkladech, SSJČ (1989, 2.vyd.) je sice obsahuje, ale jejich celkové zpracování není v tomto slovníku dostatečně konzistentní a soustavné, kromě toho přibližně 20% údajů v SSJČ je dnes už zastaralých.

V oblasti počítačového zpracování češtiny (a nejen zde) již delší dobu existuje naléhavá potřeba rozsáhlejšího seznamu českých sloves s jejich valencemi, který by zahrnoval pokud možno všechna běžná česká slovesa a měl také formálnější podobu. Situace je podle našeho názoru zralá k pokusu o vytvoření základního valenčního seznamu a pro nás příznivá v tom, že se můžeme opřít o existující počítačové zdroje — strojový slovník českých kmenů, který je jádrem automatického morfologického analyzátoru a současně lemmatizátoru LEMMA (Ševeček 1995, Osolsobě 1996). Tento slovník v současnosti obsahuje celkem téměř 165 000 českých kmenů, z nichž slovesných je cca 36 000 (substantivních kolem 80 000 a adjektivních asi 40 000, zbytek – cca 11 000 připadá na ostatní slovní druhy), takže může sloužit jako dobré východisko k pokusu vybudovat základní valenční slovník v rozsahu minimálně 10 000 českých sloves. Vlastní pokus pak spočívá ve vytvoření výchozího seznamu (slovníku) českých sloves, u něhož počítáme s rozsahem do 12 000 sloves, přičemž jako další zdroj poslouží i Slovník českých synonym (Pala, Všianský, 1995). Dodejme ještě, že tento rozsah je i ve shodě se SSČ (Praha, 1994), který obsahuje něco přes 12 000 slovesných hesel.

Jsme si přirozeně vědomi, že plně reprezentativní seznam českých sloves užívaných v současné češtině budeme moci získat až z právě vznikajícího Českého národního korpusu (ČNK) včetně příslušných frekvenčních charakteristik.

To ale bude vyžadovat ještě určitý čas (nejméně 1-2 roky) a navíc důležitou podmínkou, která musí být splněna, abychom dostali přesnější obraz o distribuci slovních druhů včetně sloves v současné češtině, je gramatické označkování dostatečně velké části ČNK (odhadem nejméně 1 mil. slovních tvarů, ale k získání adekvátního obrazu bude potřeba označkovat podstatně více – nepochybně nejméně 10 mil. slovních tvarů).

Právě tyto časové úvahy nás vedly k rozhodnutí začít budovat Valenční slovník českých sloves (dále VSČS) manuálně již teď a jeho výchozí verze je přirozeně připravována v elektronické podobě: ta dovolí, jak doufáme, dospět k úplnějším a přesnějším popisům českých slovesných valencí doloženým i podrobnějšími údaji statistickými (viz níže). U vybraných sloves bude v konečné podobě obsahovat též základní frazeologická spojení a nejčtenější kolokace. Takto lze navíc získat přirozené východisko pro vytvoření základního valenčního seznamu i pro česká substantiva a adjektiva: takový seznam představuje další chybějící článek gramatického popisu češtiny a je nezbytným předpokladem jejího realistického (nejen) počítačového zpracování.

Celkově se v tomto textu zabýváme teoretickými východisky pro český valenční slovník a notačním systémem pro formální reprezentaci valenčních vzorců u českých sloves. Navržené notační prostředky zachycují jak jednotlivé jednoduché valence, tak i jejich možné kombinace mající pak podobu konkrétních valenčních vzorců. Dále text obsahuje předběžné výchozí údaje o distribuci a četnostech binárních a ternárních valenčních vzorců a pádových příznaků v tabulkové podobě. I když mají předkládané údaje předběžnou povahu, poskytují dostatečně reprezentativní představu o distribuci českých pádových příznaků u pravostranných valencí a nebudou se už pravděpodobně nijak výrazněji měnit. Poznamenejme, že navíc jde i o první statistiku tohoto druhu pro češtinu na tak velkém materiálu. Principy notace pro valenční vzorce uvádíme a objasňujeme níže v odd.4 Typy valencí.

1.1 Zdroje valenčního slovníku

Jak jsme už jsme naznačili, materiál pro VSČS byl celkově získán z několika zdrojů:

1. Slovník českých synonym, NLN, Praha 1995,
2. Počítačový slovník českých kmenů s celkovým rozsahem cca 165 000 jednotek (dodejme jen, že prefigovaná slovesa a pravidelně tvořená deverbativa, adjektiva a adverbia jsou však v tomto slovníku generována automaticky, takže skutečný rozsah tohoto slovníku se může pohybovat kolem 400 000 položek).
3. Slovník spisovné češtiny, Academia, Praha 1994, 2. vyd.
4. Slovník spisovného jazyka českého, Academia, Praha 1989, 2. vyd.

Výchozí seznam získaný z uvedených zdrojů čítal po prvním kroku kolem 10 tisíc českých sloves. Po jeho zpracování a postupném porovnání se SSJČ jsme zatím dospěli k první verzi seznamu obsahujícímu cca 12 000 českých sloves, což je rozsah, který lze z hlediska našich současných potřeb pokládat za plně dostačující.

2. Výchozí pojmy

Jak ukazují Čermák a Holub (1991), základním konceptem, z něhož lze vycházet, je pojem kolokability, což je obecná schopnost slova (a dalších jednotek) spojovat se v textu s jinými.

S tímto termínem by se pravděpodobně dalo vystačit, u sloves je však obvyklé mluvit o valenci zejména proto, že mají schopnost vázat na sebe jistý počet výrazů povrchově realizovaných morfologickými prostředky, tj. pády. Tuto schopnost sloves vázat na sebe gramaticky ostatní slova můžeme formálně reprezentovat v termínech některých slovních druhů — typicky substantiv či zájmen: v našem případě tak činíme pomocí specifických pronominálních výrazů jako *koho, co, čeho, komu, čemu, kom, čem, kým, čím*.

V literatuře se zhusta diskutuje o tom, zda valence je jevem primárně syntaktickým, či sémantickým. Dost často se setkáváme s názorem, že valence je více záležitostí roviny syntaktické, což se opírá především o fakt, že je (v češtině) formálně vyjadřována právě morfologickými prostředky — pády, které se vazebně pojí s jednotlivými slovesy.

Tyto formální prostředky realizace valence by nám však neměly zakrývat podstatu věci, totiž skutečnost, že schopnost slova kombinovat se v textu s jinými slovy je primárně dána sémanticky, tj. významy spojujících se slov. Považujeme proto slovesnou valenci za jev hloubkově sémantický a chápeme ji jako významem determinovanou schopnost slova kombinovat se s jinými slovy. Celkově se snažíme přihlížet co nejuplněji k významové povaze jednotlivých slovesných participantů, i když v dané verzi slovníku zatím pracujeme především s povrchovými pádovými příznaky. Je však jasné, že bez důsledného přihlížení k významu sloves nemůžeme rozumně vysvětlit celou řadu jevů, např. kontextové elipsy typu (v1) *Otec čte dětem před spaním.* (pohádku) nebo významově blízké případy — synonyma jako (v2a) *Matka mluví s otcem o těch penězích.* (v2b) *Matka říká otci o těch penězích.*, i když jejich valenční vlastnosti se povrchově mohou lišit (viz např. Leech, 1981). Podobně i rozdíly jako *vypláznout jazyk [na koho]* proti *vypláznout stovku [na co]* dosvědčují nutnost přihlížet systematicky k významové stránce konkrétních participantů.

Příklady tohoto typu tedy vedou k závěru, že rozumně zpracovaný valenční slovník musí důsledně přihlížet k celkovému lexikálnímu obsazení valenčních vzorců — bez něho nelze dostatečně přesně a adekvátně zachytit vztahy mezi významy a podvýznamy uvnitř jednotlivých hesel a také i mezi jednotlivými hesly navzájem. Jde tu i o postižení víceznačnosti u sloves, kterou naše „prázdné“ valenční vzorce zachycují zatím jen zčásti. Pořízení seznamu valenčních vzorců spolu s jejich úplným typickým lexikálním obsazením je však evidentně velmi pracné a musí se nutně opírat o značkový korpus (ČNK): dospět k němu bude tedy možné až v další fázi. Výsledkem bude ovšem již slovník v plném slova smyslu, který pak může tvořit jádro většího a reprezentativnějšího elektronického slovníku současné češtiny. Dojde tak v plné míře k tomu, co je signalizováno korpusově orientovanými výzkumy odjinud: na jedné straně k lexicalizaci gramatiky a ke gramatikalizaci slovníku na straně druhé (Francis, 1997).

2.1. Obligatornost a fakultativnost

V literatuře se věnuje značná pozornost otázkám obligatornosti a fakultativnosti slovesných participantů: podrobný přehled lze najít zejména u Panevové (Panevová, 1974, 1975). Z její i dalších analýz (např. Grepl, Karlík, 1985) celkem jednoznačně vyplývá, že i když uplatníme řadu různých testů na vypustitelnost, substituovatelnost, opakovatelnost a neopakovatelnost, parafázovatelnost, i když se pokusíme testy navzájem zkombinovat, je téměř jisté, že se nám nepodaří získat dostatečně spolehlivou, konzistentní, jednoduchou a konstruktivní odhalovací proceduru, která by se dala použít jako nástroj při klasifikaci valencí u většího počtu českých sloves. Jistého úspěchu lze dosáhnout, pracujeme-li s malými a vhodně vybranými skupinami sloves. Jakmile se začneme pokoušet o systematickou klasifikaci většího počtu sloves, ukazuje se, že snaha rozlišovat obligatorní a fakultativní participanty nevede k dostatečně konzistentním výsledkům. Naše zkušenost je taková, že jsme se opakovaně dostávali do situací, kdy jsme byli ochotni stejná doplnění jednou považovat za obligatorní a podruhé za fakultativní. Skutečnost, že se nesnáze tohoto druhu pravidelně opakovaly, nás vede k názoru, že kritéria užívaná obvykle pro rozlišování mezi obligatorností a fakultativností slovesných doplnění jsou s největší pravděpodobně definována vágně a zjevně nemají dostatečně vysokou rozlišovací sílu.

Přitom nemůžeme říci, že bychom opomíjeli kritéria významová, naopak, snažili jsme se k nim přihlížet co nejdůsledněji, ale ani tak jsme nedosáhli lepších výsledků: stále se nám nedařilo u velkého počtu sloves (řádově až tisíce) bezpečně a konzistentně klasifikovat slovesné participanty co do obligatornosti a/nebo fakultativnosti. Naše dosavadní zkušenost nás vede k závěru, že má smysl snažit se zachytit všechna doplnění daná významem slovesa a jejich vzájemné vztahy, jejich vzájemnou propojitelnost. Místo snahy rozlišovat za každou cenu mezi obligatorními a fakultativními participanty je podle našeho názoru rozumnější registrovat všechna doplnění typická pro dané sloveso a pokusit se je seřadit do určité škály, v níž na jednom konci budou doplnění nejobyklejší, významově nejprediktabilnější a na konci druhém doplnění významově volnější a také méně běžná. Jsme si vědomi, že takto uvedené charakteristiky nejsou ještě dostatečně přesné: jsme však přesvědčeni, že tyto problémy se podaří spolehlivěji řešit, jakmile budeme mít k dispozici dostatečně rozsáhlá korpusová data a pro dané kolokace budeme moci získat jejich M/I- a T-skóre (Hanks, Church, 1997).

3. Valenční vzorce

Vzhledem k rozsahu našeho seznamu sloves jsme proto zatím do jisté míry rezignovali na striktní rozlišování obligatorních a fakultativních participantů a zaměřili se primárně na jejich co nejúplnější zachycení v souladu s významem toho kterého predikátu. Jednoduše řečeno, jde nám hlavně o to, abychom na prvním místě zachytili, co k danému slovesu patří, a teprve na druhém, jak to k němu patří. Kategorie obligatornosti a fakultativnosti jsme nicméně přes

všechny uvedené výhrady nenechali úplně stranou, ovšem pracujeme s nimi s patřičnou opatrností a dosavadní výsledky zatím pokládáme za předběžné.

V tomto ohledu nejde pochopitelně jen o naši zkušenost, a proto např. ve shodě se Somersem (Somers 1987) navrhuje určitou stupnici slovesných doplňení, v níž rozlišujeme následujících šest stupňů valenční spojitelnosti a popisujeme je pomocí následujících notačních konvencí:

- 1) integrální, lexikálně determinovaná, nevypustitelná doplňení, nepřipouštějící substituce argumentů blízkými synonymy a modifikace adjektivy, nevstupují do syntagmatických substitučních paradigmat (pronominalizace), tj. frazeologická spojení, ev. idiomy, frazémy, např. *držet krok, hubu, mít šanci, mít koholco po ruce, dělat co stůj co stůj* apod. Značíme je samostatným symbolem #, který signalizuje, že ve skutečnosti jde o autonomní a dále nečlenitelnou jednotku. Výše uvedená doplňení *držet hubu* a podobně i *držet na koho* zapíšeme tedy jako:

držet

hubu

<na koho|co>

- 2) obligatorní, nutná doplňení mající pravidelně formu přímých a předložkových pádů a vedlejších vět uvozených např. spojkami *aby, co, jak, že*. Obligatorní přímé i předložkové pády, resp. jejich kombinace, jsou vyznačeny symbolem & a větná doplňení symbolem \$. Můžeme tedy mít:

rozkázat t

= komu & co

= komu \$(aby, co)

= inf

dokázat t

= co

= komu & co

= \$(že)

adresovat t

= co

= komu & co

= co ? na koho|co

- 3) fakultativní, nepovinná doplňení formálně realizovaná podobně jako v předchozím případě přímými a předložkovými pády. Frekvence jejich výskytu u příslušných sloves je pravděpodobně dosti vysoká, ale uplatněním testu na vypustitelnost se lze přesvědčit o jejich fakultativnosti. K označení těchto případů užíváme ?, takže můžeme mít např.:

dopisovat si

= s kým ? o čem ,

což postihuje jak *dopisovat si* bez doplňení, tak i *dopisovat si s kým, dopisovat si o čem* a nakonec i *dopisovat si s kým o čem*.

Podobnou povahu mají i případy jako

zaimponovat i
= komu ? čím

Poněkud jiný typ představují slovesa některá slovesa pohybu, např.
vyprovázet, vyprovodit t

= koho ? (z čeho, do čeho, na co),

u nichž první doplnění je zjevně obligatorní, ale ostatní uvedená v kulatých závorkách jsou fakultativně dána významem příslušného slovesa: mohou a nemusí být přítomna. Lze snad namítnout, že zmíněná doplnění mají adverbialní povahu (označují místo), je však třeba si uvědomit, že v daném případě jde o slovesa pohybu — lokální modifikátory u nich nemají podle našeho názoru povahu naprosto volných doplnění.

- 4) střední doplnění — široce determinovaná sémantickou třídou (významem) slovesa. Nejtypičtějším představitelem tohoto typu doplnění, jak ukazují naše data, je obvykle přímý instrumentál s obecným nástrojovým významem. Dále sem mohou patřit i výrazy způsobové a méně již časové, místní, pokud jsou široce predikovány významem odpovídajících sloves — to platí nejčastěji pro doplnění [jak]. Doplnění tohoto typu jsou notačně reprezentována pomocí hranatých závorek, např.:

dopovat t

= koho [čím]

nebo

dosáhnout t

= čeho, co [čím], [jak]

- 5) volná doplnění — sem řadíme zcela volná doplnění časová, místní a způsobová spojitelná, jak předpokládáme, s každým běžným slovesem. Jsou volná do té míry, že nijak nevyplývají z významu slovesa, proto je u sloves v současné verzi slovníku nijak nevyznačujeme a počítáme v případě potřeby s jejich automatickou doplnitelností. Pokud jsou vyjadřována čistými adverbií jako *doma*, *tady* nebo *ted*, *dnes* a *rychle*, nepředstavuje jejich rozpoznání nebo vygenerování zvláštní potíže, složitější je situace u adverbialních předložkových pádů, u nichž je spolehlivé rozpoznávání jejich sémantické povahy velmi nesnadné. Jedním z možných řešení je zavést ve slovníku sémantické rysy a označkovat jimi co nejvíc substantiv: rysy pak mohou pomoci indikovat, že např. *na ulici* ve spojení *plakala na ulici* má povahu volného lokálního modifikátoru. Po dobudování valenčního slovníku do definitivní podoby počítáme s vyznačením volných doplnění u jednotlivých sloves pomocí speciálních rysů, které mohou předběžně vypadat např. takto:

platit t

= komu & co [čím] [za co] <kdy=dnes> <kde=v obchodě, jak=hotově>.

- 6) periferní doplnění, k nimž nepochybně patří částice různého typu, zejména pak částice mající hlavně pragmatickou povahu. Zatím zvolený způsob značení je v daném okamžiku celkem arbitrární a definitivně bude řešen až v průběhu času. Jako příklad uveďme třeba

poslat

= komu & co |<asi, možná, patrně>|

Je vidět, že použitá škála vede ke klasifikaci, která je širší než klasifikace obvyklé v českých gramatikách. Za její přednost pokládáme právě to, že umožňuje zachytit v jednom rámci jak frazeologická spojení na straně jedné, tak i různé typy partikulí zejména pragmatické povahy na straně druhé. Mezi nimi se pak nacházejí jednotlivá doplnění počínaje obligatorními až po úplně volná.

Celkově tedy zachycujeme valenci českých sloves v připravovaném slovníku tak, že u každého slovesa uvádíme s ohledem na jeho význam jednotlivé přímé nebo předložkové pády, s nimiž se dané sloveso pojí. Nepracujeme zatím samostatně s časovými a místními adverbialními doplněními typu *kam, kde, kudy, kdy, ...* — hlavním důvodem je to, že ve většině případů patří k výše uvedeným volným či automaticky doplnitelným participantům. Dá se namítnout, že i četné adverbialní předložkové pády mají povahu volných doplnění a že tedy není potřeba je u jednotlivých sloves uvádět: důvodem zvoleného řešení je skutečnost, že předložkové i některé přímé pády jsou jasným zdrojem víceznačností ve větných strukturách, proto usilujeme o jejich zachycení s ohledem na další počítačové zpracování. U čistých adverbialních doplnění tomu tak zjevně není, lze je tedy zatím nechat stranou, ale dojde na ně, jakmile do valenčních vzorců začneme systematicky doplňovat jejich lexikální obsazení. V této souvislosti půjde rovněž o ověření běžně vyslovovaného předpokladu o automatické doplnitelnosti časoprostorových určení k většině „běžných“ sloves a o porovnání s korpusovými daty.

Jak je vidět z příkladů uvedených výše, hlavním notačním prostředkem uvozujícím a vyznačujícím jednotlivé valenční vzorce je rovnítko =. Je zřejmé, že takto vyznačené valenční vzorce často odpovídají jednotlivým slovesným „významům“ či „podvýznamům“, přičemž uvozovkami chceme naznačit, že ne vždy musí jít o významy (podvýznamy), které by přesně odpovídaly členění hesel jako třeba v SSJČ, lze však očekávat, že míra shody bude dosti velká. Základní údaje v tomto ohledu poskytuje tab. 4 níže.

V dané verzi slovníku pracujeme tedy, jak jsme uvedli, s pádovými příznaky vyznačenými pomocí pronominálních výrazů jako *koho, co, čeho, komu, čemu, ...*, které jsou vhodné jednak mnemotechnicky a jednak umožňují celkem pohodlně rozlišovat opozici životnost : neživotnost.

Subjektové participanty jsou v dané verzi implicitní a samostatně jsme nepracovali ani aritu sloves (predikátů) např. tak, že bychom každému predikátu přiřazovali aritu pro jeho jednotlivé významy třeba pomocí číslice umístěné před rovnítkem a vyznačující jednotlivé valenční vzorce. Je ostatně patrné, že aritu lze z uvedené notace pohodlně odvodit, aniž bychom ji uváděli samostatně.

Příznaky jako *jak* a *kolik* uvádíme jen u sloves, u nichž mají v závislosti na významu slovesa jednoznačně valenční charakter. Podobně, jak jsme už naznačili, zacházíme i s adverbialními pády jako *na čem, v čem, do čeho, z čeho, ...*, které uvádíme jen tam, kde jsou viditelně podmíněny významem slovesa — tak je tomu zjevně u sloves pohybu.

4. Předběžná statistika valencí

Seznam čítající téměř 12 tisíc českých sloves poskytuje bohaté výchozí podklady pro získání řady slovesných tříd sdružujících slovesa se stejnými valenčními vzorci (pádovými příznaky a jejich kombinacemi). Díky celkové složitosti notace a velkému počtu různých valenčních vzorců představuje třídění sloves s jejich valenčními vzorci dosti komplikovaný úkol, pro jehož úplné vyřešení je třeba nejprve v postupných krocích testovat konzistenci zápisu valencí v aktuální verzi valenčního seznamu. O to se nyní pokoušíme u sloves pohybu, která se vyčleňují jako poměrně dobře uchopitelná a samostatná skupina. Učinili jsme zatím jen první pokus a pomocí valencí jsme se pokusili vytřídit z našeho seznamu slovesa pohybu. Použili jsme k tomu valencí do čeho a z čeho, které lze považovat za spolehlivé signály místního doplnění. Takto získaný podseznam jsme ještě prošli manuálně a vyřadili některá slovesa, jež se sice vyskytují s valencí do čeho, ovšem označují velmi specifickou variantu pohybu jako např. *bít, bouchat do čeho*: výsledkem je seznam sloves pohybu, který zatím čítá cca 1700 sloves (z cca 12 tis. sloves). Tento seznam se ještě zjevně rozpadne na menší a sémanticky kompaktnější skupiny podle jednotlivých typů pohybu — k tomuto jemnějšímu třídění použijeme dalších valencí vyskytujících se u sloves v seznamu jako např. na co a v čem a dalších. Detailní popis této slovesné třídy však zde ponecháváme stranou a hodláme se jím zabývat v samostatném článku.

V tomto textu můžeme zatím nabídnout jen svého druhu první sondu poskytující údaje o základních typech valencí a valenčních vzorců. Jsou to především tabulky udávající četnosti jednotlivých valencí (pádových příznaků) a také četnosti jejich základních kombinací. Jinak řečeno, jde o valenční vzorce binární (s jednou levou a jednou pravou valencí) a ternární (s jednou levou a dvěma pravými). Vzorcům s aritou vyšší než 3 se budeme věnovat v samostatném textu.

První soubor údajů se týká tranzitivity a intranzitivity: chápeme je formálně tak, že za tranzitiva pokládáme všechna slovesa, která mají ve svém valenčním vzorci akuzativ — i v kombinaci s jinými pády, zatímco mezi intranzitiva řadíme ta slovesa, u nichž se akuzativ nevyskytuje. Počítáme tu i s případy, kdy je sloveso víceznačné.

K rozlišení možných variant slouží příznaky *t* a *i* uvedené u jednotlivých heslových slov.

Podobně je zachycena i reflexivita, a to tak, že u heslového slova je podle potřeby uvedeno *se* nebo *si*, které pak slouží jako rozlišující příznak.

Náš materiál potvrzuje, že existuje pravidelný vztah mezi tranzitivou a reflexivou, jak to dokládají následující příklady, např.

barvit *t*

= co [čím], [jak]

barvit *se*

= čím, jak.

První stručná tabulka tedy dává představu o vztazích mezi tranzitivou a intranzitivou a o četnostech sloves majících u sebe *si* a *se* v našem téměř 12 tisícovém seznamu.

Tabulka 1 — tranzitiva, intranzitiva, reflexiva

příznak	četnost	procenta
intranaktivní	1700	cca 15 %
tranzitivní	6471	cca 55 %
se	2780	cca 25 %
si	572	cca 5 %
celkem	11523	100 %

4.1 Předběžná statistika valencí (a pádových příznaků)

V následujícím textu uvádíme v tab. 2a předběžné údaje o četnostech jednotlivých přímých i předložkových pádů, jak se vyskytují u sloves v našem současném valenčním seznamu. Ve všech tabulkách jsou zatím uvedeny jen absolutní četnosti (z celkových téměř 12 tisíc), detailnější statistiky s ohledem na celkovou různorodost a komplikovanost valenčních vzorců budeme moci nabídnout až v dalším. Zatím počet různých valenčních vzorců v poměru k celkovému počtu cca 12 000 sloves vychází na 4000 — z nich 2849 se vyskytuje s četností 1. To je nečekaně vysoký počet, který je podle našeho názoru způsoben tím, že v současné verzi slovníku jsme se jednak pokusili zachytit u jednotlivých sloves co nejvíc valenčních vzorců a jednak jsme se snažili popsat i složené valenční vzorce. Ty jsou ovšem dosti komplikované, což zřejmě plyne z méně striktního rozlišování obligatorních a fakultativních doplnění a také ze soustavnějšího uvádění adverbialních pádů tam, kde by jiní autoři pravděpodobně použili nerozčleněných příznaků typu *kam*, *kde*, *odkud*, *kudy*, *kdy*, *jak dlouho*, *dokdy*,

Statistiku složitějších vzorců zde zatím necháváme stranou, k jejich podrobnému třídění bude možno systematicky přistoupit až po určitých úpravách. Podotýkáme, že pro všechny tabulky uvedené níže je potřeba přihlídnout k tomu, že vztah mezi celkovým seznamem a součty v tabulkách není přímočarý: řada sloves se pojí, jak ukazuje tab. 4, s více valenčními vzorci a poměrně velký počet vzorců je složených.

Tabulka 2a — souhrnné četnosti jednotlivých pádů

	valence	četnost
genitiv	přímý	354
"	předložkový	540
dativ	přímý	295
"	předložkový	331
akuzativ	přímý	2376
"	předložkový	645

valence		četnost
lokál		2154
instrumentál	přímý	862
"	předložkový	1098
větná doplnění		246
celkem		8901

Tab.2a poskytuje souhrnný přehled o distribuci přímých a předložkových pádů, které se vyskytují s jednotlivými slovesy samy o sobě, tj. je to základní přehled dvoumístných valencí tvořených na levé straně nominativem, který se implicitně objevuje téměř u všech sloves (počet sloves bez nominativu se řádově pohybuje nejvýše ve stovkách), a na pravé straně příslušným pádem z tabulky. Tabulka potvrzuje očekávanou převahu akuzativu, následuje podle očekávání lokál a teprve za ním instrumentál. Na třetím místě se objevuje genitiv a jako poslední vychází dativ, u něhož podle očekávání převažuje dativ přímý. Tab. 2a ovšem nijak nepostihuje skutečnost, že jednotlivé pády se samostatně vyskytují u jednoho slovesa i vícekrát v závislosti na jeho různých významech.

Tabulka 2b — Přehled četností jednotlivých pádů

genitiv	četnost	dativ	četnost	akuzativ	četnost
čeho	64	komu	295	co	1471
kohočeho	290	čemu	54	kohočo	880
do čeho	44	komučemu	84	na koho	57
do kohočeho	222	k čemu	104	na kohočo	211
z čeho	20	ke komučemu	56	na co	247
z kohočeho	13	proti komučemu	33	o co	33
od čeho	24			o kohočo	24
od kohočeho	161			pro kohočo	28
u koho	56			za kohočo	19
				přes co	10
				v co	16
celkem	894		626		2996

lokál	četnost	instr.	četnost	větná dopl.	četnost
o čem	595	čím	862	\$(co, jak, že)	98
o komičem	47	kýmčím	46	\$(že)	83
v čem	1089	s kým	192	\$(aby)	28
v komičem	30	s čím	246	koho \$(aby)	13
na čem	265	s kýmčím	281	komu \$(že)	15
na komičem	16	nad čím	88	komu \$(aby)	9
po čem	23	nad kýmčím	59		
po komičem	55	před kýmčím	76		
o čem	13	před čím	85		
o komičem	21	za kýmčím	25		
				infinitiv	77
celkem	2154		1960		246

Tab.2b nabízí v porovnání s tab.2a navíc základní přehled četností konkrétních přímých a předložkových pádů získaných vyříděním z našeho seznamu. Lze z ní tedy vyčíst základní tendence u variant jednotlivých předložkových pádů podle konkrétních předložek a také rozdíly u přímých pádů. Např. u akuzativu (ale i u ostatních pádů) je vidět, že *co* je preferováno proti *koho/co*, což ukazuje na rozdílnou distribuci vzhledem k opozici životnost : neživotnost. Navíc jsou v tabulce uvedeny i základní údaje o infinitivní valenci a dále o vět-
ných doplněních a spojkách, které je uvozují.

Tabulka 3a — základní přehled dvoumístných (pravých) valencí

valence	četnost
genitiv – gen.předl.	24
dativ – gen.předl.	64
dativ – akuzativ	810
dativ – lokál	92
dativ – instr.	33
akuzativ – genitiv př.	579
akuzativ – lokál	490
akuz. – instr.	1375
akuz. – ak.předl.	243
akuz. – dat.předl.	69
akuz. – jak	159
instr. – lokál	52
větná dopl.	145
celkem	4811

Tabulka 3b — nejčetnější trojmístné valence

akuz. – instr.		akuz. – gen.předl.		akuzativ – dativ předl.	
co & čím	16	co ? z čeho	132	co ? k čemu	30
co ? čím	250	co [z čeho]	63	co ? ke komu/čemu	3
co [čím]	246	co & z čeho	23	co [k čemu]	15
koho & čím	11	co & z koho/čeho	16	koho/co ? ke komu/čemu	6
koho ? čím	153	co ? do čeho	148	koho/co ? k čemu	15
koho [čím]	86	co [do čeho]	39		
koho/co ? čím	264	co & do čeho	24		
koho/co [čím]	284	koho/co ? do čeho	52		
co ? s kým	34	koho ? do čeho	24		
co [s kým]	18	koho/co ? z čeho	22		
co ? s čím	13	koho/co [do čeho]	20		
		koho/co [z čeho]	16		
celkem	1375		579		69

akuzativ – lokál		akuzativ – ak.předl.		dativ – akuzativ	
co ? na čem	22	co ? na co	82	komu & co	460
co [v čem]	126	co [na co]	24	komu & koho/co	22
co [na čem]	69	co ? na koho	32	komu & do čeho	20
co & na kom	13	co ? na koho/co	22	komu & v čem	18
koho/co [v čem]	95	co & na koho/co	15	komu ? koho/co	18

akuzativ – lokál		akuzativ – ak.předl.		dativ – akuzativ	
co ? v čem	73	co & na koho	25	komu ? čím	33
koholco [v čem]	53	co & na co	15	komu ? co	20
co ? na čem	22	co & za co	14	komulčemu ? co	10
koholco ? na čem	17	koholco [na co]	14	komu & na co	8
				komu ? na co	19
				komu [čím]	16
				[komu] co	166
celkem	490		243		810

instrumentál – lokál		dativ – lokál		dativ – gen.předl.	
s kým ? o čem	26	komu & v čem	18	komu & z čeho	15
s kýmliším [v čem]	9	komu ? v čem	30	komu & do čeho	20
čím ? v čem	5	komu [v čem]	12	komu ? z čeho	14
čím [v čem]	7	komulčemu [v čem]	9	komu ? do čeho	15
čím ? po čem	5	komulčemu ? v čem	16		
		[komu] v čem	7		
celkem	52		92		64

akuzat. – jak

co [jak]	70
koholco [jak]	46
na koholco [jak]	15
co ? jak	28
celkem	159

Tab. 3a a 3b poskytují základní představu o nejčetnějších trojmístných valencích včetně údajů o konkrétních kombinacích přímých i předložkových pádů. K tomu zásadně poznamenejme, že pořadí, v němž jsou jednotlivé pády v tabulkách uváděny, je dáno zvoleným způsobem notace, takže si lze představit, že s ohledem na volný slovosled v češtině by tabulky 3a i 3b mohly vypadat jinak, ale je zjevné, že podstatu věci by to neovlivnilo. Pracujeme tedy se zápisem věnovat komu & co, ale stejně tak bychom díky volnému slovosledu mohli mít věnovat co & komu, informace o valenci a v daném případě též o její obligatornosti by tím nebyla nijak dotčena. Tab. 3a ukazuje jasnou převahu dvojice akuzativ-instrumentál vyplývající pravděpodobně z vyšší četnosti sloves s obecným významem *dělat něco něčím*. Druhé místo kombinace dativ-akuzativ rovněž není nijak překvapující a je nepochybně dáno nemalým počtem sloves s významy typu dávání či poskytování v širokém slova smyslu.

Z notace použité v tab. 3b lze také vyčíst rozdíly v distribuci obligatorních a fakultativních doplňení, např. zápis valence komu & co vyjadřuje, že jde o valenci obligatorní patřící pod zmíněný již typ 2, zatímco zápis komu ? co v naší notaci znamená, že obě valence jsou v dané kombinaci fakultativní — mohou se spolu vyskytovat obě nebo jen jedna z nich — viz typ 2 výše. Zápis tvaru [komu] co pak chápeme tak, že hranaté závorky vyznačují volné doplňení typu 3, které se u příslušného slovesa může a nemusí objevit — v případě valence dativ – akuzativ jde s velkou pravděpodobností o volné dativy.

Tabulka 4 — počet valenčních vzorců (podvýznamů) u sloves

slovesa bez valenčního vzorce (rovnítka)	266	cca 2,0%
slovesa s jedním valenčním vzorcem (=)	8429	70,9%
se dvěma valenčními vzorci (=)	2196	18,5%
se třemi " "	647	5,4%
se čtyřmi " "	224	1,8%
s pěti " "	73	0,6%
s šesti " "	33	0,3%
se sedmi " "	21	0,1%
s osmi " "	6	0,05%
s devíti " "	6	0,05%
s desíti " "	3	0,025
s dvanácti " "	1	0,008%
s patnácti " "	1	0,008%
se sedmnácti: být	1	0,008%
s padesáti osmi: jít (včetně frazémů)	1	0,008%
celkem	11908	100,0%

Tab.4 poskytuje zajímavé údaje o distribuci významů (podvýznamů) u sloves v seznamu: ukazuje se, že počet víceznačných sloves je celkově poměrně nízký, sloves se třemi a více valenčními vzorci je v našem seznamu celkem 1017, tj. 8,5% z celkového počtu. Tento výsledek dává naději, že celkové zpracování valenčního slovníku včetně lexikálního obsazení valenčních vzorců je zvládnutelné v přijatelném čase a s rozumnými náklady.

5. Závěry

Získané výsledky mají ještě v řadě ohledů předběžnou povahu, ale na druhé straně poskytují celkem zřetelnou představu o hlavních frekvenčních charakteristikách vybraných pádových a valenčních příznaků na materiálu čítajícím téměř 12 000 českých sloves. Dosud získaný valenční seznam představuje rovněž dobré východisko pro získání sémantické klasifikace českých sloves na základě valenčních vzorců, které jsou slovesům již přiřazeny.

Výchozí valenční seznam je i nesmírně potřebným nástrojem pro další pokrok v oblasti počítačového zpracování češtiny:

- a) najde uplatnění v oblasti korpusu, zejména při jeho značkování,
- b) představuje jádro realistické formální gramatiky češtiny,
- c) stane se východiskem pro počítačové lexikografické aplikace — systematický počítačový slovník češtiny a z něho vycházející vícejazyčné překladové slovníky použitelné např. v budoucích systémech strojového překladu.

Za zajímavý výsledek je třeba pokládat i to, že jsme narazili na opakované komplikace při pokusech systematicky rozlišit obligatorní a fakultativní doplnění ve valenčních vzorcích: zdá se, že tato teoretická distinkce se zatím opírá o vágní kritéria. Dosavadní data, jejichž analýze se hodláme věnovat v dalším textu, bude potřeba konfrontovat s korpusovými daty: výsledek této konfrontace, doufáme, naznačí spolehlivě směr dalšího postupu.

LITERATURA

1. ČERMAK, F., HOLUB, J., Syntagmatika a paradigmatica českého slova I (Valence a kolokabilita), skriptum LŠSS, UK Karolinum, Praha 1991.
2. FRANCIS, G., Korpusový přístup ke gramatice (Principy, metody a příklady), čes. překlad, Sb. článků z korpusové lingvistiky, v tisku, Praha 1997.
3. GREPL, M., KARLÍK, P., Skladba spisovné češtiny, SPN Praha, 1985, s.120-21.
4. CHURCH, K. W., HANKS, P., Normy slovních asociací, vzájemná informace a lexikografie, čes. překlad, Sb.článků z korpusové lingvistiky, v tisku, Praha 1997.
5. OSOLSOBĚ, K., Algoritmický popis české formální morfologie, disertační práce, Brno 1996.
6. PANEVOVA, J., On Verbal Frames in Functional Generative Description, Part I, The Prague Bulletin of Mathematical Linguistics 22, Prague 1974, pp.3-39; Part II, Prague 1975, pp.17-71
7. PALA, K., VSIANSKÝ J., Slovník českých synonym, NLN Praha, 1995
8. PETR J., a kolektiv autorů, Akademická mluvnice češtiny I, II (Academic Grammar of Czech I, II), Academia, Praha, 1986.
9. Slovník spisovné češtiny, Academia, Praha, 2.vyd. 1995
10. Slovník spisovného jazyka českého, Academia Praha, 1.vyd.1960, 2.vyd. 1989.
11. SOMERS, H., L., Valency and Case in Computational Linguistics, eds. S. Michaelson and Y.Wilks, Edinburgh Information Technology Series, Edinburgh University Press, 1987, pp. 4-29
12. SVOZILOVA N. et al., Valenční seznam českých sloves, rukopis, Praha, ÚJČ, 1995.

RÉSUMÉ

In the article we present the preliminary results of the research devoted to the compilation of the Czech verb frames list. The list size is almost 12 000 items, for each verb in the list we have found its respective verb frame(s) and represented it in a formal notation suitable for computer processing. The notation has been developed for future computer applications, particularly in computer lexicography and automatic syntactic analysis of Czech.

We offer the first statistical findings: the tables 2a and 2b contain frequencies of the verb participants occurring in simple verb frames with two participants consisting of nominative and other possible direct and prepositional cases — i.e. genitive, dative, accusative, locative and instrumental. The tables 3a and 3b present the similar information for the verb frames with three participants: nominative on the left side and the particular case pairs on the right.

The obtained results will serve as a basis for a detailed semantic classification of Czech verbs: we are going to deal with the semantic verb classes in a close future.

Karel Pala & Pavel Ševeček
 Fakulta informatiky MU
 Botanická 68a
 602 00 Brno