

Zamyslenie nad výučbou zámen a čísloviek pri príprave morfolologickej anotácie Slovenského národného korpusu

Lucia Gianitsová

Oddelenie Slovenského národného korpusu, Jazykovedný ústav L. Štúra SAV, Bratislava

*Príspevok bol prednesený na medzinárodnej konferencii
„Tradiční a netradiční metody a formy práce ve výuce českého jazyka na základní škole“
na PdF UP v Olomouci (19. 3. 2004) a bude publikovaný v zborníku.*

1. Zámená a číslovky možno považovať za špecifické slovné druhy v súvislosti s ich pôvodom, charakterom i funkciami. Zámená sú špecifické ako slová s vysokým stupňom zovšeobecnenia a abstrakcie, číslovky zas kvantitatívnym významom i ich morfológickou komplexnosťou, resp. formálnou i funkčnou rôznorodosťou jednotlivých prvkov skupiny. Aj rôzne prístupy a pokusy o klasifikáciu týchto slov, opisy ich deklinácie, na ktoré upozorníme, poukazujú na to, že ide o problematický súbor prvkov, v ktorom sa systematika hľadá viac či menej úspešne.

S problémom zachytenia relevantných formálnych vlastností (aj sémantických príznakov) zámen a čísloviek sme sa stretli najmä pri príprave *morfológického tagsetu* (súboru značiek) pre ručné značkovanie textov *Slovenského národného korpusu* (SNK). Komplexné riešenia a zásady morfológickej anotácie sme predstavili v kolektívnej štúdii (Forróová – Garabík – Gianitsová – Horák – Šimková, 2003) a aktuálna verzia morfológického tagsetu je prístupná na webovej stránke SNK¹ (Garabík – Gianitsová – Horák – Šimková, 2004). Samostatne sme sa zaoberali aj osobitnými problémami pri morfológickej anotácii substantív a adjektív (Gianitsová, 2003).

V tomto príspevku sa zameriame na problematiku zámen a čísloviek, najmä na ich spracovanie pre potreby počítačovej analýzy textov v korpuse. Sme toho názoru, že tento formálny (korpusový) prístup môže v mnohom pomôcť aj deskriptívnemu opisu jazyka a školskej praxi.

2. Slovenský národný korpus (SNK) a jeho využitie v lingvo-didaktike

V roku 2002 bol vládou schválený vedecko-výskumný projekt budovania elektronického korpusu v *Jazykovednom ústave L. Štúra SAV* (Šimková, 2003, s. 17). V pláne samostatného oddelenia, ktoré vzniklo, je vybudovať všeobecný jednojazyčný korpus, ktorý bude vyvážený, tzn. bude obsahovať v primeranej miere texty z rôznych oblastí používania súčasného slovenského jazyka (z rokov 1955 – 2005). Výber textov v rozsahu 200 miliónov slov sa obmedzí na písané texty a postupne sa bude rozširovať o texty z ďalších období a sfér používania slovenského jazyka (Šimková, 2003, s. 18).

V súčasnosti je používateľom, odbornej verejnosti i ďalším záujemcom, prístupná verzia *Prim0.2* obsahujúca približne 170 miliónov textových jednotiek (*tokenov*)² a texty sú prevažne z oblastí publicistiky. Intenzívne sa pracuje na ďalšom spracovaní korpusových dát, ktoré môžu poskytnúť bohaté informácie o jazyku a jeho konkrétnom použití. Ide najmä o externú (bibliografickú, štýlovo-žánrovú) a internú (morfológickú, syntaktickú, sémantickú) anotáciu textov (Forróová – Horák, 2004, s. 175).

¹ Pozri <http://korpus.juls.savba.sk/korpus/biblioteka/publikacie/Tagset-aktualny.pdf>.

² Pozri http://korpus.juls.savba.sk/vyhľadavanie/korpusovy_manazer/verzia/.

O význame a využití korpusov sme už formou príspevku informovali (Gianitsová, 2004) a v tejto súvislosti sme uvažovali najmä o využití korpusu v pedagogickej praxi. Zdôraznili sme, že korpus môže byť vhodným zdrojom fráz a viet potrebných pri výučbe cudzích jazykov. Zároveň ako zdroj textov na vytváranie pravopisných, gramatických a štylistických cvičení otvára možnosti problémového vyučovania jazyka v počítačovej učebni (bližšie pozri Gianitsová, 2004). Uplatnenie práce s korpusom je teda možné vo viacerých činnostiach učiteľa i žiaka:

- pri individuálnej príprave učiteľa na hodinu – príprava cvičení;
- ako motivácia na vysvetlenie nového učiva – učiteľ zadá problémovú úlohu, na základe výsledkov hľadania žiaci dedukujú s pomocou učiteľa teoretické princípy;
- pri výklade učiteľa, oboznamovaní sa s novým gramatickým javom – žiaci si overujú teóriu v praxi, ako sa gramatické prostriedky reálne vyskytujú v texte;
- pri precvičovaní, upevňovaní nového učiva – vyhľadávanie príkladov priamo na hodine, učiteľ zadá úlohu a žiaci pracujú s programom sami;
- pri opakovaní – problémové úlohy.

Korpusy majú naozaj široké uplatnenie v rôznych vedeckých výskumoch a aplikáciách (pozri aj Šimková, 2003, s. 17). Pri využívaní korpusových dát SNK (aj v pedagogickej praxi) je možné použiť *www.rozhranie* priamo na webovej stránke SNK, možnosť rozšíreného vyhľadávania ponúka *korpusový manažér Manatee s klientom Bonito*³ pre zaregistrovaných používateľov. Ten poskytuje aj informácie o širšom kontexte, frekvencii výskytu slova, *kolokáciách* (spoluvýskyt slova), štýlových a žánrových charakteristikách textu a tiež údaje o pôvode textov a pod. Textom SNK teda boli pridané bibliografické a štýlovo-žánrové charakteristiky.

3. Morfológická analýza SNK

Korpusový manažér má samozrejme aj ďalšie možnosti. V spolupráci s českými partnermi z *Ústavu formálnej a aplikovanej lingvistiky MFF UK v Prahe*⁴ je v týchto dňoch sprístupňované vyhľadávanie podľa *lemy* (základného tvaru) v textoch SNK. Spolupráca sa týka aj automatickej morfológickej analýzy slovných tvarov na základe skúseností a nástrojov vyvinutých českými partnermi. Morfológicky anotovaný korpus by poskytol ďalšie možnosti využitia SNK. Zároveň prešiel úpravami aj samostatne navrhnutý súbor morfológických značiek (Garabík, R. – Gianitsová, L. – Horák, A. – Šimková, M., 2004) a v súčasnosti prebieha na jeho základe ručná lematizácia a morfológické značkovanie.

Proces ručnej morfológickej anotácie pozostáva z dvoch krokov: 1. *Lematizácia*. *Tokenu* (textovej jednotke) určíme základný (slovníkový) tvar. 2. *Morfológické značkovanie*. K tvarom slov priradíme značku (*tag*), ktorá je súborom znakov (kombináciou písmen a číslíc), symbolickým vyjadrením slovnodruhovej príslušnosti tokenu a formálnych kategórií, ktoré považujeme pri danom tvare za relevantné. Takáto analýza nepredstavuje väčšinou pre človeka, ktorý je jazykovo kompetentný, problém, pre počítač je to však náročná úloha. Pritom máme na mysli aj problém tvarovej homonymie slov, ktorá vzniká pri automatickej morfológickej anotácii a vyžaduje si následnú *dezambiguáciu* (zjednotenie výstupu morfológickej analýzy, proces

³ Program je možné nainštalovať po stiahnutí z *www* stránky SNK. Každý, kto sa zaviaže k nekomerčnému používaniu sprístupneného korpusu, môže pracovať s testovacou verziou Slovenského národného korpusu *Prim0.2* cez vlastné konto (heslo). Heslo bude záujemcovi pridelené na základe poštu zaslaného vyplneného a podpísaného formulára, ktorý sa tak isto nachádza na webovej stránke SNK.

⁴ V rámci grantovej úlohy *Využitie spoločných vlastností češtiny a slovenčiny na budovanie anotovaných národných jazykových korpusov* sa rozvíja spolupráca najmä s J. Hajičom.

selekcie správnej značky).

Na začiatku práce na súbore morfológických značiek sme si uvedomovali, že takýto súbor má zachytávať lingvistické vlastnosti textu, teda istým spôsobom text interpretovať. Stáli sme pred rozhodnutím, či navrhnutý súbor značiek bude výsledkom nového formálneho opisu jazyka alebo budeme vychádzať z lingvistických opisov, ktoré existujú a len sa ich pokúsime formalizovať.

V teoretickej rovine sa príprave morfológického tagsetu pre ručné značkovanie SNK venovali M. Forróová a A. Horák (2004, s. 182), ktorí vymedzili pevné východiská. Návrh tagsetu je teda založený na predpoklade, že používateľ korpusu:

- má isté znalosti o korpuse a možnostiach jeho použitia;
- je aspoň minimálne lingvisticky vzdelaný;
- má schopnosť lingvistickej abstrakcie.

Do popredia by sme chceli postaviť najmä zásadu prístupnosti (koncovému používateľovi, ktorým býva aj nelingvista) a zásadu konsenzu vedeckých teórií, teoretickej „neutrality“, ktoré sme v prvom rade mali na zreteli pri príprave súboru značiek.

3.1. Tradičné gramatické opisy ako východiská morfológickej analýzy

Pri príprave morfológického tagsetu pre analýzu textov SNK bola teda v centre záujmu i otázka, do akej miery môžeme vychádzať z tradičných gramatických opisov slovenskej morfológie. Za relevantné sme považovali prihliadať najmä na systematický opis podaný v akademickej *Morfológii slovenského jazyka* (ďalej MSJ, 1966), prípadne na ďalšie morfológické práce (Oravec – Bajzíkova – Furdík, 1984; Dvonč, 1984). Konflikt medzi reprezentáciou zaužívaných gramatických kategórií a možnosťou automatického spracovania jazyka sa odráža aj v koncepcii morfológického tagsetu pre SNK. Už M. Forróová a A. Horák (2004) si uvedomili nejasnosť kritérií morfológickej klasifikácie. Upozorňujú, že ako klasifikačné kritérium pri zaraďovaní slov do slovných druhov sa uplatňuje komplex morfológických, syntaktických, lexikálnych a sémantických vlastností slov (MSJ, 1966; Oravec – Bajzíkova – Furdík, 1984) a v rámci morfológie hovoríme tradične i o lexikálno-gramatických kategóriách (intencia, vid, stupeň). Kritériá sú teda často vágne, či nejednoznačné. Nás zaujal najmä problém klasifikácie a opisu zámen a čísloviek.

3.1.1. Zámená v lingvistickej tradícii a ich obraz v didaktike. Základné morfológické opisy slovenského jazyka pracujú predovšetkým s lexikálno-sémantickou klasifikáciou. V týchto intenciách sa potom pohybujú aj opisy a výklady v učebniciach. Analyzovali sme najmä učebnice pre druhý stupeň základných škôl, keďže gramatika a morfológia sa sprostredkúva žiakom v celosti a úplnosti predovšetkým na tomto stupni.⁵ V prípade zámen ide o 6. ročník ZŠ (Tibenská a kol., 2002a), kde sa žiaci prvý raz detailne oboznámia s problematikou klasifikácie zámen. V školskej praxi je v popredí najmä rozlišovanie zámen „podľa vecného významu (funkcie)“ (Navrátil, 1998, s. 49) alebo podľa „slovného významu“ (Tibenská a kol., 2001, s. 59) na zámená *osobné (základné a privlastňovacie)*, *zvrtné (základné a privlastňovacie)*, *opytovacie*, *ukazovacie*, *vymedzovacie* a *neurčité* (pozri riadky v tabuľke č. 1). V ôsmom ročníku sa opäť k problematike klasifikácie zámen autori učebníc vracajú a tentoraz ide o sémanticko-gramatický pohľad (Navrátil, 1998, s. 48; Tibenská a kol., 2001, s. 53), čiže zámená sú delené na základe ich všeobecného gramatického významu *predmetnosti*, *kvalitatívnosti*, *kvantitatívnosti* alebo *okolnosti* (Ondrus, 1978, s. 84; pozri aj stĺpce v tabuľke č. 1). Tieto poznatky možno zhrnúť do

⁵ Stredná škola s týmito poznatkami len pracuje pri výklade štylistickej problematiky, teda ráta so základnými znalosťami z oblasti morfológie, na ktorých ďalej stavia.

tabuľky, ktorú sme vypracovali na základe cvičení obsiahnutých v učebnici pre 8. ročník (Tibenská a kol., 2001, s. 53):⁶

Tabuľka č. 1: Klasifikácia zámen podľa tradičných gramatík (MSJ, 1966; Oravec – Bajzíkova – Furdík, 1984)

	SUBSTANTÍVNE	ADJEKTÍVNE	NUMERÁLNE	ADVERBIÁLNE
OSOBNÉ ZÁKL.	<i>ja, ty, on, ona, ono my, vy, oni, ony</i>			
OSOBNÉ PRIVL.		<i>môj, tvoj, jeho jej, ich, náš, váš</i>		
ZVRATNÉ ZÁKL.	<i>seba/sa</i>			
ZVRATNÉ PRIVL.		<i>svoj, svoja, svoje</i>		
OPYTOVACIE (VZŤAŽNÉ)	<i>Kto? Čo?</i>	<i>Aký? Aká? Aké? Ktorý? Ktorá? Či? Čia? Čie?</i>	<i>Koľký? Koľko? Koľkonásobný? Koľkoraký?</i>	<i>Kde? Kam? Kadiaľ? Kedy? Odkedy? Ako? Prečo?</i>
UKAZOVACIE	<i>ten, tá, to tento, táto, toto</i>	<i>taký, taká, také takýto, onen</i>	<i>toľký, toľko toľkonásobný toľkoraký, toľkoto</i>	<i>tam, tadiaľ, tamto vtedy, dovedy tak, preto</i>
VYMEDZOVACIE	<i>nikto, nič všetci, všetko</i>	<i>iný, taký istý nijaký, každý samý, žiaden</i>	<i>všetok všetka všetko</i>	<i>nikde, všade, inde nikdy, vždy nijako, inak</i>
NEURČITÉ	<i>niekto, niečo dakto, dačo ktosi, čosi</i>	<i>nejaký, nejaká čís, ktorýkoľvek dajaký, akýsi</i>	<i>niekoľko dakkoľko koľkosi</i>	<i>niekde, dakde kedysi, niekedy nejako, akosi</i>

Všeobecné gramatické významy predmetnosti, kvalitatívnosti, kvantitatívnosti či okolnosti, ktoré chápeme ako vnútorné vlastnosti zámen, určujú ich vonkajšie vlastnosti – lexikálny význam a najmä gramatické vlastnosti (Ondrus, 1978, s. 84 – 85). Na základe týchto vonkajších vlastností sú teda zámená jednoducho klasifikovateľné a aj pedagogická prax potvrdzuje, že takúto klasifikáciu žiaci intuitívne využívajú.⁷ Aj z komunikačného hľadiska by bolo vhodné, keby sa výučba intenzívne sústredila na zisťovanie gramatickej povahy zámen, keby sa viac precvičovalo zisťovanie ich funkcie vo vete a vyzdvihovala sa zástupná funkcia v texte (otázkami typu: Aký slovný druh zastupuje toto zámeno?).

No aj spomenutá klasifikácia prináša často rozpory a odlišné hodnotenia v morfológických prácach a slovníkových heslách. Na tomto mieste spomenieme najmä najvýraznejšie problémy, ktoré podľa nášho názoru sú zásadným ústupkom logike systému a v mnohom komplikujú aj výklad v triede.

V *Krátkom slovníku slovenského jazyka* (KSSJ, vo všetkých vydaniach, aj v najnovšom 4. vydaní z roku 2003) nachádzame pri každej lexéme jej slovnodruhovou príslušnosť vyjadrenú pomocou kvalifikátora. Pri zámenách ide o kvalifikátor *zám.*, ktorý býva doplnený aj bližším

⁶ Aj tu si uvedomujeme, že platné učebnice pre základné školy skrývajú v sebe množstvo podnetov, ktoré by mohol vnímavý učiteľ využiť na skvalitnenie výučby.

⁷ Poukazuje na to častá situácia, ktorá je hodnotená väčšinou negatívne, keď žiak považuje zámená *tam, tu, niekde, nikde, vždy, inde* za príslovky. No práve táto tendencia uprednostňovať sémanticko-gramatické vlastnosti môže byť v komunikačne zameranej výučbe vhodne využitá.

určením, či ide o zámeno osobné základné (*osob.*), osobné privlastňovacie (*privl.*), zvrtné základné (*zvrat. zákl.*), zvrtné privlastňovacie (*zvrat. privl.*), opytovacie (*opyt.*), ukazovacie (*ukaz.*), vymedzovacie (*vymedz.*) alebo neurčité (*neurč.*).⁸ Pri niektorých nájdeme aj kvalifikátory označujúce vnútorné vlastnosti – všeobecnogramatické významy. Zachytené sú zámená numerálne (*číslov.*), adverbialne (*príslov.*), prípadne označené ako nesklonné (*neskl.*). Pri analýze inventára zámen sme však narazili na logický rozpor. V prípade ukazovacích a opytovacích číslovkových zámen máme zachytené zámenné koreláty takmer pre všetky druhy čísloviek, pri neurčitých zámenách však nachádzame len jedného zástupcu číslovkových zámen:

UKAZOVACIE ZÁM.	OPYTOVACIE ZÁM.	NEURČITÉ ZÁM.
<i>tol'ko</i> (neskl. zám. číslov.)	<i>kol'ko</i> (zám. číslov.)	<i>niekol'ko</i> (zám. číslov.)
<i>tol'koto</i> (zám.)		
<i>tol'ký</i> (zám.)	<i>kol'ký</i> (zám. číslov.)	
<i>tol'kýto</i> (zám.)		
<i>tol'ký raz</i> (zám. číslov.)		
<i>tol'kokrát</i> (neskl. zám. číslov.)	<i>kol'kokrát</i> (neskl. zám. číslov.)	
<i>tol'ko rás</i> (neskl. zám. číslov.)	<i>kol'ko rás</i> (neskl. zám. číslov.)	
	<i>kol'konásobný</i> (zám. číslov.)	
	<i>kol'konásobne</i> (zám. príslov.)	
<i>tol'koraký</i> (zám. číslov.)	<i>kol'koraký</i> (zám. číslov.)	

Ďalšie očakávané lexémy, ktoré vznikajú analogicky z opytovacích zámen predponami *nie-*, *da-*, *voľa-*, *hoci-* a pod. v slovníku síce nájdeme, ale sú označené ako neurčité číslovky: *niekol'kýkrát*, *niekol'ký raz* (čísl. neurč.), *niekol'kokrát*, *niekol'ko rás* (neskl. čísl. neurč.), *niekol'konásobný* (čísl. neurč.), *niekol'konásobne* (čísl. príslov.), *niekol'koraký* (čísl. neurč.), *niekol'korako* (čísl. príslov.). Toto riešenie, ktoré je pravdepodobne dôsledkom niektorých konštatovaní MSJ (1966) o číslovkách a číslovkovej platnosti niektorých zámen, nepovažujeme za prijateľné. Vychádzame z formulácie, ktorá sa často opakuje v gramatikách, a to, že neurčité zámená vznikajú z opytovacích zámen (MSJ, 1966, s. 235, 293). Ak teda medzi opytovacie zámená autori MSJ (1966) i KSSJ (2003) zaradili nielen lexému *kol'ko*, ale aj lexémy *kol'ký*, *kol'kokrát*, *kol'ko rás*, *kol'konásobný*, *kol'konásobne*, *kol'koraký*, od nich odvodené lexémy s predponou *nie-*, *da-*, *voľa-*, *všeli-*, *hoci-*, *bárs-* alebo s príponami *-si*, *-kol'vek* musíme považovať za zámená. Druhou možnosťou je nazerať na lexémy *niekol'kokrát*, *niekol'konásobný*, *niekol'ko rás*, *niekol'koraký* a pod. ako na deriváty zámena *niekol'ko*, podobne ako zámená *kol'kokrát*, *kol'konásobný*, *kol'ko rás*, *kol'koraký* sú derivátmi zámena *kol'ko*. V oboch prípadoch zo slovotvorného hľadiska ide len o modifikáciu, ktorá sa deje v rámci toho istého slovného druhu. V žiadnom prípade takýmto spôsobom nemôže dôjsť k vzniku lexémy s odlišnou slovnodruhovou príslušnosťou.

Ako sme už spomenuli, tieto rozpory sa odrazili aj vo výučbe. Lexémy *niekol'ko*, *niekol'kokrát* a pod. sú síce v učebniciach zaraďované k zámenám (Tibenská a kol., 2001, s. 59)⁹,

⁸ V KSSJ (2003) nájdeme aj nesprávne priradený kvalifikátor. Zámeno *zvoľakadiaľ* je hodnotené ako vymedzovacie, aj keď nasledujúce doklady dosvedčujú, že ide o neurčité zámeno (*odkiaľsi, odniekiaľ: zvolakadiaľ zaznela hudba*, KSSJ, 2003, s. 972).

⁹ V učebnici pre 7. ročník (Tibenská a kol., 2002b, s. 40) sa upozorňuje len na zámeno *niekol'ko*, ktoré sa takto interpretuje aj v KSSJ (2003). Práve problematické prípady ako *niekol'kokrát*, *niekol'ko rás*, *niekol'koraký* a podobne nie sú v učebniciach zachytené. Len v učebnici pre 8. ročník sa ako zámeno zastupujúce neohybné číslovky uvádza lexéma *niekol'kokrát* (Tibenská a kol., 2001, s. 59).

no to odporuje kodifikačným príručkám (MSJ, 1966; KSSJ, 2003), ktoré majú slúžiť a často aj slúžia ako pomoc v prípade neistoty. Navyše výučba v takýchto prípadoch nemôže využiť ani základné logické postupy, akými sú napríklad analógia a paralela a žiak sa nemôže oprieť ani o základnú analýzu. Pri analýze týchto tvarov totiž jasne vyplynie, že základom je zámenný korelát, doplnený sémantickým príznakom neurčitosti.

Opakovanie uvedených zaradení (podľa MSJ, 1966) udivuje najmä v súvislosti s tým, že už vysokoškolská učebnica morfológie koriguje niektoré názory MSJ (1966) a uvádza, že medzi neurčité číslovky nepatria číselné zámená *koľko, koľkokrát, koľkonásobný, koľkoraký, tolko, toľkokrát, toľkonásobný, toľkoraký*, rovnako aj ich deriváty s predponou *nie-, da-, hoci-, bárs-*, a príponou *-si, -koľvek* (Oravec – Bajzíkova – Furdík, 1984, s. 97 – 98). V spomínanej učebnici sa aj lexémy *niekoľkokrát, niekoľko ráz, niekoľkonásobný* hodnotia ako číslovkové zámená (op. cit., s. 104, 110). Pri navrhovaní morfológického tagsetu pre ručnú anotáciu SNK sme teda pracovali s takto upraveným systémom zámen.

3.1.2. Číslovky v lingvistickej tradícii a ich obraz v didaktike. S problematikou čísloviek ako slovného druhu sa žiaci oboznamujú komplexne v 7. ročníku základnej školy (Tibenská a kol., 2002b). Základné vedomosti o používaní týchto slov a ich komunikačnej platnosti sú však prítomné v učebniciach od 3. ročníka. Teoreticky pritom klasifikácia čísloviek vychádza opäť z klasických gramatických opisov (MSJ, 1966; Ondrus, 1969, 1978; Oravec – Bajzíkova – Furdík, 1984; KSSJ, 2003). Je len prirodzené, že tento komplexný slovný druh, do ktorého sa zaraďujú slová s rôznymi gramatickými kvalitami, nie je v jednotlivých prácach opísaný a hierarchizovaný rovnako. Aj tu sme vypracovali tabuľku, ktorá zachytáva klasifikáciu na základe morfológických, syntaktických, ale najmä sémantických vlastností (pozri riadky tabuľky č. 2). Formálny pohľad na číslovky ponúka ešte jednu možnosť klasifikácie: podľa flexie, resp. deklinačnej paradigmy (pozri stĺpce v tabuľke č. 2). Táto klasifikácia je tiež implicitne prítomná v komplexnom prístupe, ktorý zvolili autori učebníc v 8. ročníku základnej školy (Tibenská a kol., 2001):

Tabuľka č. 2: Klasifikácia čísloviek podľa tradičných gramatík (MSJ, 1966; Oravec – Bajzíkova – Furdík, 1984)

OHYBANIE DRUH	samostatné: <i>jeden, dva, tri, štyri, päť</i>	ako príd. mená: <i>pekný, cudzí</i>	ako podst. m.: <i>mesto, stroj, žena...</i>	neohybné
ZÁKLADNÉ	<i>jeden, dva, tri štyri, päť</i>		<i>sto, tisíc milión, miliarda</i>	<i>veľa, málo, dvadsaťjeden</i>
RADOVÉ		<i>prvý, druhý posledný, mnohý</i>		
NÁSOBNÉ		<i>dvojnásobný mnohonásobný</i>	<i>raz, (dva) razy (päť) ráz</i>	<i>päťkrát, päťnásobne päť ráz, mnohokrát mnoho ráz, mnohonásobne</i>
DRUHOVÉ		<i>dvojaký, pätoraký mnohoraký</i>		<i>dvojako, pätorako mnohorako</i>
SKUPINOVÉ				<i>dvoje, pätoro, viacero</i>

Na rozporuplnosť slovnodruhového hodnotenia takých lexém ako *koľko, niekoľko, tolko* a pod. sme už poukázali. No aj v rámci čísloviek nájdeme niekoľko ďalších problémov. Autori učebnice pre 7. ročník iste vychádzali aj z KSSJ (2003), kde pri číslovkách *druhykrát, druhý raz; ôsmykrát, ôsmy raz; piatykrát, piaty raz* a pod. nájdeme kvalifikátor *rad. nás. príslov. neskl.* (Ide

teda o radové násobné príslovkové nesklonné číslovky.) V slovenskej gramatickej tradícii (MSJ, 1966) dokonca nájdeme aj hodnotenie, že uvedené číslovky sú len radové. Takto sa interpretujú aj v školskej praxi (Tibenská a kol., 2002b, s. 43), pričom sa neprihliada na formálne kritérium i morfológickú štruktúru slov. Podľa nášho názoru sa výlučným zaradením týchto slov k radovým číslovkám stráca dôležitá informácia o sémantickom príznaku, teda o tom, že kategoriálny význam číselného poradia je tu obohatený o príznak „násobnosť“, podobne ako sa pri lexémach *dvakrát*, *dva razy*, *osemkrát*, *osem ráz* obohacuje základný kategoriálny význam. Schematicky by sme mohli proces vzniku a vzťah týchto lexém naznačiť takto:

päť → *päťkrát*, *päť ráz*
 → *piaty* → *piatykrát*, *piaty raz*

Z naznačeného vyplýva, že číslovky *druhykrát*, *druhý raz*; *ôsmkrát*, *ôsmý raz*; *piatykrát*, *piaty raz* zaraďujeme k násobným číslovkám utvoreným od radových čísloviek, zatiaľ čo násobné číslovky *dvakrát*, *dva razy*; *osemkrát*, *osem ráz*; *päťkrát*, *päť ráz* sú utvorené od základných čísloviek. Z tohto rozdielu potom plynú aj ďalšie sémantické odlišnosti, no v rámci skupiny násobných čísloviek. Tento formálny prístup sa nám javí vhodnejší aj v školskej praxi, napokon, slovotvorné kritérium a stavba slova sú prvými signalizátormi slovnodruhovej príslušnosti a žiaci sa na ne pri jazykovej intuícii spoliehajú.

V opise systému čísloviek v akademickej MSJ (1966) nájdeme aj niektoré slová majúce aj kvantitatívne odtienky, ktoré je možné na základe ich všeobecnogramatických významov a lexikálno-gramatických vlastností zaradiť k iným slovným druhom, napr. k substantívam (*tretina*, *štvrtina*), k adjektívam (*dvojitý*, *dvojný*, *podvojný*, *jednorazový*), k zámenám (*tolko*, *niekoľkokrát*), prípadne k príslovkám (*dvojmo*, *trojito*). Bohužiaľ, aj tento problém sa prenáša do oblasti školskej praxe. V učebniciach sú medzi násobnými číslovkami zaradené aj lexémy *dvojitý*, *dvojito*, *dvojmo* (Tibenská a kol., 2002b, s. 43), podobne aj v popularizačných príručkách (Navrátil, 1998, s. 73). Okrem týchto slov, ktoré podľa nášho názoru nepatria k číslovkám, by sa dalo uvažovať aj nad slovnodruhovou príslušnosťou radových čísloviek (*druhý*, *tretí*), adjektívnych násobných čísloviek (*dvojnásobný*), adverbiálnych násobných čísloviek (*dvojnásobne*) a pod., ako o tom uvažujú P. Ondrus (1969, 1978) aj F. Sabol (1993). Svoje výhody má i delenie čísloviek na primárne – *základné*, *skupinové*, *zlomkové*, *násobné adverbiálne* – a sekundárne – *radové*, *druhovú*, *adjektívne násobné číslovky* (Sokolová, 1995, s. 140).

Situácia v systéme čísloviek a jeho opise však ostáva naďalej problémová. Preto sme boli nútení pri návrhu morfológických značiek pre tento slovný druh pristúpiť k niektorým riešeniam, ktoré nebudú v súlade s klasickým, prípadne školským opisom.

3.2. Morfológická anotácia korpusu a jej prínos

Klasický lingvistický prístup založený na komplexe morfológických, syntaktických, lexikálnych a sémantických kritérií môže teda, ako sme ukázali, v rovine automatického spracovania predstavovať množstvo problémov. Pri príprave tagsetu pre morfológickú anotáciu korpusových textov sa ponúkal aj iný zdroj inšpirácie. V oblasti korpusovej lingvistiky je možné sledovať a analyzovať viaceré prístupy k morfológickej anotácii, ktoré už konkrétne fungujú pri anotácii textov iných národných korpusov. Uvažovanie o výhodách i nevýhodách morfológických tagsetov *Českého národného korpusu* (ČNK), projektu *Multext-East* i korpusu vznikajúceho v *Ústave základov informatiky Poľskej akadémie vied* (IPI PAN) (Forróová – Horák, 2004) nás priviedlo k záverom, že sa pri návrhu morfológického tagsetu pre SNK nechceme vybrať len výlučne formálnou cestou. Tento prístup viedol aj autorov tagsetu ČNK k tomu, že na úrovni

bližšieho určenia slovného druhu (SUBPOS) vymedzujú až 74 možných hodnôt, z toho len pre zámená je uvedených 19 hodnôt.¹⁰ Podobne v tagsete IPI PAN sa rozoznáva až 29 gramatických tried slov (bližšie pozri Forróová – Horák, 2004).

Po porovnaní doterajších prístupov k obsahu tagsetu, ktoré Forróová a Horák (2004, s. 178) vymedzujú ako *lingvisticky optimistický*¹¹ a *lingvisticky pesimistický*¹², bol pre potreby značkovania vybraný formálno-gramatický princíp, ktorý však s ohľadom na domácu lingvistickú, resp. gramatickú tradíciu má isté špecifiká.

Konvenčný lingvistický prístup (zapojenie celého komplexu sémantických, syntaktických a gramatických kritérií v morfológii) sa naďalej uplatňuje pri určení slovného druhu, formálno-gramatický prístup pri zámenách a číslovkách sa uplatňuje najmä pri určovaní hodnôt novej kategórie *paradigmy*, ktorá má vyjadrovať formálne vlastnosti tvaru (*tokenu*).¹³ Toto riešenie teda vyplynulo najmä z poznania opisu morfológie v tradičných gramatikách (MSJ, 1966; Oravec – Bajzíkova – Furdík, 1984; Dvonč, 1984, Sokolová, 1995), z poznania vývinových zákonitostí slovenskej deklinácie (Pauliny, 1990) i z postrehov a skúseností s morfológickou anotáciou inojazyčných korpusov (Forróová – Horák, 2004).

Kategóriu *paradigmy* chápeme ako vymedzenie špecifickej formy konkrétneho člena slovného druhu a vymedzuje sa pri menných slovných druhoch na druhom mieste v značke (*tagu*). Slovný tvar (*token*) môže teda mať *substantívnu*, *adjektívnu*, *zámennú* (*pronominálnu*), *číslovkovú* (*numerálnu*), *príslovkovú* (*adverbiálnu*), *zmiešanú* alebo *neúplnú paradigmu*.

Substantívnu, *adjektívnu*, *pronominálnu*, *numerálnu* a *adverbiálnu paradigmu* priradíme slovám, ktorých deklinácia má formálne znaky typické pre daný slovný druh a tieto slová nadobúdajú gramatické, najmä syntaktické vlastnosti, ktoré sú vlastné danému slovnému druhu. Substantívna paradigma je priradená tvarom, ktoré sa skloňujú ako typické substantíva podľa vzorov *chlap*, *hrdina*, *dub*, *stroj*, *žena*, *ulica*, *dlaň*, *kosť*, *mesto*, *srdce*, *vysvedčenie*, *dievča*. (Okrem mnohých substantív ide napr. o číslovky *nula*, *milión*, *bilión*, *miliarda*, *raz*.) Adjektívnu paradigmu majú tokeny, ktoré patria do deklinačných typov *pekný*, *cudzí*, *pávi* (napr. *veľký*, *taký*, *aký*, *tretí*, *štvrtý*), pronominálna paradigma sa priraduje len zámenám so samostatným skloňovaním *ja*, *ty*, *my*, *vy*, *seba*, numerálna zas len číslovkám *dva*, *tri*, *štyri*. Adverbiálnu – príslovkovú paradigmu majú také zámená a číslovky, ktoré sa neskloňujú a v KSSJ (2003) bývajú označené kvalifikátorom *neskl.* alebo *príslov.*, resp. majú adverbiálnu funkciu (*kol'konásobne*, *kol'korako*, *tam*, *tu*, *vtedy*, *vždy*, *viacnásobne*, *dvojako*). V týchto prípadoch sa tag určením *paradigmy* končí. Ďalšie kategórie ako rod, číslo alebo pád sa neuvádzajú.

Zmiešaná paradigma platí pre slová, ktoré majú len sčasti kongruentnú paradigmu. Vývoj ich deklinácie prešiel zložitými procesmi a nie sú jednoznačne zaraditeľné k čistým deklinačným typom. Patria sem slová typu *kuli*, *gazdiná* a všetky rovnako sa skloňujúce substantíva, *otcov*, *matkin* (všetky individuálne privlastňovacie prídavné mená), a zo zámen a čísloviek najmä *on*, *ona*, *ono*, *kto*, *čo*, *nikto*, *nič*, *môj*, *tvoj*, *náš*, *váš*, *ten*, *tá*, *to*, *sám*, *onen*, *žiadan*,

¹⁰ Ide napr. o samostatné hodnoty pre: 1. zámeno *on* v tvaroch po predložke (*něj*, *něho*), 2. privlastňovacie zámeno *svůj*, 3. reflexívne zámeno *se* v dlhých tvaroch (*sebe*, *sobě*, *sebou*), 4. reflexívne zámeno *se*, *si* iba v týchto tvaroch a tvaroch *ses*, *sis*, 5. opytovacie alebo vzťahné zámeno *kdo* vrátane tvarov *s -ž* a *-s*, 6. ukazovacie zámeno *ten*, *onen*, 7. vzťahné privlastňovacie zámeno *jehož*, *jejiž*, 8. vzťahné zámeno s adjektívnym skloňovaním (*jaký*, *kerý*), 9. vzťahné zámeno *jenž*, *již* po predložke (*něhož*, *níž*), 10. vzťahné zámeno *což*, 11. vzťahné zámeno *jenž*, *již* bez predložky, atď. (Hajič, 2000; Hana – Hanová, 2002).

¹¹ Lingvisticky optimistický prístup znamená snahu o uplatnenie čo najväčšieho počtu gramatických kategórií v tagsete, následnú dezambiguáciu pomocou syntaktických pravidiel. Predstaviteľmi sú V. Petkevič a K. Oliva.

¹² Ide o kompromis medzi lingvistickým a „inžinierskym“ prístupom, resp. obsah tagsetu sa prispôsobuje matematickému modelu tagera ako u J. Hajiča a v projekte Multext-East.

¹³ Napr. zámeno *taký* má formu adjektíva.

všetok, jeden.

Neúplná paradigma bude priradená tým substantívam, adjektívam, zámenám a číslovkám, ktoré sa v tradičnej gramatike označujú ako nesklonné (*kupé, super, jeho*), s tendenciou k sklonnosti (*kanoe – G pl. od kanoí*) alebo ako obyčajne nesklonné (*pani, koľko, tolko, viacero, päť, sto, tisíc*). Tieto tvary chápeme ako také podoby slov, ktorých morfológické vlastnosti nie sú z formy slova „viditeľné“, teda tieto slová nie sú dostatočne formovo transparentné. Ich gramatické vlastnosti však môžeme určiť z kontextu, ktorý je pre potreby ručného značkovania neobmedzený. Prihliadame najmä na kongruenciu v rámci syntagmy alebo na valenčnú väzbu.

Pri niektorých slovách s neúplnou paradigmou nachádzame aj sklonné, aj nesklonné podoby. Ich distribúcia je závislá od mnohých činiteľov. Za substantívum s neúplnou paradigmou môže byť považovaný aj tvar takého substantíva cudzieho pôvodu, ktoré je v zásade sklonné (*Phillips, Tesco, Kia*), no pisateľ uprednostnil v danom páde nesklonnú podobu (*od Phillips*), hoci sa vyskytuje zväčša sklonná podoba (*od Phillipsu*). Podobne je tomu pri číslovke typu *päť*. V textoch nachádzame sklonné (*piatim chlapom*) i nesklonné podoby (*päť chlapom*). Všetky takéto nesklonné tvary potom majú priradenú neúplnú paradigmu.

Pri zámenách a číslovkách teda nachádzame takéto kombinácie znakov pre slovný druh a paradigmu:

- PS – zámeno so substantívnou paradigmou (*koľkátka*)
- PA – zámeno s adjektívnou paradigmou (*taký*)
- PP – zámeno so samostatnou zámenou paradigmou (*ja*)
- PF – zámeno so zmiešanou paradigmou (*on, sám, žiaden*)
- PU – zámeno s neúplnou paradigmou (*jeho, jej, ich, koľko, tolko*)
- PD – zámeno s adverbálnou paradigmou (*tam, niekedy*)
- NS – číslovka so substantívnou paradigmou (*milión, raz*)
- NA – číslovka s adjektívnou paradigmou (*štvrtý*)
- NN – číslovka so samostatnou číslovkovou paradigmou (*tri*)
- NF – číslovka so zmiešanou paradigmou (*jeden*)
- NU – číslovka s neúplnou paradigmou (*päť, sto, tisíc, päťoro, veľa*)
- ND – číslovka s adverbálnou paradigmou (*dvakrát, mnohonásobne*)

3.2.1. Spracovanie zámen v tagsete SNK sa vyznačuje niekoľkými špecifikami:

1. V ručnej lematizácii i pri značkovaní nerozlišujeme medzi zvratným zámenom *sa, si* a *sa, si* ako (zvratnou) súčasťou slovesa. Tvary *sa, si* majú priradenú osobitnú značku R a ich spracovanie ponechávame na osobitné štúdie vypracované na základe korpusu, prípadne na syntaktickú anotáciu. Napr.:

<i>rozprávali sa</i>	lema (l) = <i>sa</i>	tag (t) = R
<i>povedali si</i>	l = <i>si</i>	t = R
<i>umývali sa</i>	l = <i>sa</i>	t = R

Toto riešenie vyplynulo z nejednoznačného hodnotenia zvratných elementov v odborných prácach,¹⁴ z poznania problémov s ich určovaním v školskej praxi¹⁵ i z niektorých

¹⁴ V MSJ (1966, s. 251) sa zvratný element *sa* v spojeniach *brat sa zabil, utopil sa, popálil sa* hodnotí ako gramatická morféma, hoci ide o derivačnú morfému (Sokolová, 1995).

¹⁵ O problémoch s odlišovaním zvratných sloviess (*smiať sa, prechádzať sa*), zvratných tvarov nezvratných sloviess (*ide sa*) a spojení zvratného zámena s nezvratným slovesom (*umývať sa*) v školskej praxi sme sa zmienili v príspevku *Pokus o nové vymedzenie syntetických a analytických tvarov v slovenčine s ohľadom na školskú prax* (Gianitsová, 2003/2004, s. 74 – 75).

nových relevantných príspevkov do diskusie o ich statuse (Jarošová, 2003, s. 204 – 208).

2. Zrastené tvary (*oňho, preňho, naňho, oň, preň, zaň*) označujeme ako zámená, predložková časť sa zachytáva ako súčasť lemy (Garabík – Gianitsová – Horák – Šimková, 2004, pozri 2.2.3.3), ich aglutinovanosť je vyjadrená aj pomocou formálnej kategórie a jej hodnoty aglutinované (g) v rámci tagu. Napr.: *starám sa oňho* (l = o_on; t = PFms4g).¹⁶

3. Pri lematizácii dôsledne rozlišujeme homonymiu tvarov. Tvary *jeho, jej, ich* lematizujeme N sg. príslušného rodu ako osobné základné zámeno *on, ona, ono* (*jeho* – G, A sg. tokenu s lemov *on; jej* – G, D sg. tokenu s lemov *ona; ich* – G, A pl. tokenu s lemov *on, ona, ono*) alebo lemov *jeho, jej, ich*, ak ide o osobné privlastňovacie zámeno. Rozlíšime ich na základe postavenia a kontextu. Napr.: *Jeho to mrzí.* (lema – *on*), *Dám to jeho priateľom.* (lema – *jeho*).

Tvary *samého, samému, samého, o samom, so samým, samých, samým, samých, o samých, so samými* sú homonymné nepriame pády zámen *sám* a *samý*. Aj tu pri určovaní lemy vychádzame z kontextu. Napr.: *Mrzí to jeho samého.* (lema – *sám*), *Ušili šaty zo samého zlata.* (lema – *samý*).

Tvary zámena *koľko* (*koľkí, koľké...*) a *tol'ko* (*tol'kí, tol'ké*), ktoré odkazujú na základnú číslovku, sa lematizujú N sg. spoločným pre všetky rody (*koľko, tol'ko*). Tvary *koľkí, tol'kí, koľkých, tol'kých* môžu byť popri tom aj zámená odkazujúce na radové číslovky, lema bude N sg. mask. (*koľký, tol'ký*). Napr.: *Koľkí žiaci prišli?* (lema – *koľko*), *Koľkí v poradí boli naši pretekári na súťaži?* (lema – *koľký*).

4. Zámená môžu nadobúdať uvedené hodnoty kategórie paradigma:

Tabuľka č. 3: Paradigma a jej hodnoty pri zámenách.

ZNAK	HODNOTA	PRÍKLAD
S	SUBSTANTÍVNA	<i>koľkátka, všetučko</i>
A	ADJEKTÍVNA	<i>oný, iný, aký, -aký-, nijaký, ktorý, -ktorý-, cí, -čí-, ničí, taký koľký, -koľký-, koľkoraký, -koľkoraký-, koľkonásobný, -koľkonásobný- tol'ký, tol'koraký, tol'konásobný, každý, samý</i>
P	ZÁMENNÁ	<i>ja, ty, my, vy, seba/sebe</i>
F	ZMIEŠANÁ	<i>on, ona, ono, onô, kto, -kto-, nikto, čo, -čo-, nič, ten, tá, to, -ten-, -tá-, -to môj, tvoj, náš, váš, svoj, onen, všetok, všetka, všetko sám, sama, samo, žiaden, žiadny, žiadna, žiadne</i>
U	NEÚPLNÁ	<i>koľko, -koľko-, tol'ko, jeho, jej, ich</i>
D	PRÍSLOVKOVÁ	<i>koľkonásobne, -koľkonásobne-, koľkokrát, -koľkokrát-, koľkýkrát -koľkýkrát-, tol'konásobne, tol'kokrát, tol'kýkrát ako, -ako-, nijako, tak, onak, ináč, inak, inakšie, takisto, všakovak kde, -kde-, nikde, inde, všade, odvšade, zovšade, kade, -kade-, tade, -tade-kam, nikam, inam, tam, tam-, sem, -kam-, kadiaľ, -kadiaľ-, tadiaľ, -tadiaľ- všadiaľ, odvšadiaľ, zovšadiaľ, dokiaľ, odkiaľ, -odkiaľ-, odniekiaľ, pokiaľ odnikiaľ, odinikiaľ, dokiaľkoľvek, dotiaľ, zatiaľ, odtiaľ, potiaľ kedy, -kedy-, inokedy, vtedy, -vtedy, nikdy, vždy, dosiaľ, posiaľ prečo, -prečo-, nakoľko, natoľko, načo, začo, nato, preto, zato</i>

3.2.2. Spracovanie čísloviek v tagsete SNK je sčasti aj kompromisným riešením, keďže v písaných textoch SNK sa číslovky vyskytujú v číselnej i slovnej podobe. Pritom číselná podoba čísloviek môže ponúkať rôzne slovné interpretácie (*videl 5 chlapov*), často ide o poštové smerovacie čísla, telefónne čísla, súčasti značiek, nekontextové použitie. Preto číslice (rímske i

¹⁶ Znak v tagu čítame: P – zámeno, F – zmiešaná paradigma, m – mužský životný rod, s – singulár, 4 – akuzatív, g – aglutinovaný tvar.

arabské), resp. číselné znaky a kombinácie číselných znakov vyčleňujeme ako samostatnú triedu *číslice* (0) a ďalej ich neanalyzujeme. Slovná podoba číseloviek potom (a) v samostatnom použití (napr. *v roku dvetisíc, región číslo jeden, vhod číslo šesť* - NX) nevyžaduje ďalšie určenie gramatických kategórií, tie priradujeme len (b) číselkovým tokenom, ktoré sa úzko zapájajú do syntaktických štruktúr a valenčných väzieb (*milión hviezd, s dvojnásobným víťazom, od jednej hodiny, veľa ľudí, šesťkrát zvíťazil*). Aj pri týchto číselkoch uvádzame na druhom mieste, hneď za určením slovného druhu, typ paradigmy. Príklady na jednotlivé typy paradigmy, ktoré môžu nadobúdať číselky uvádzame v tabuľke č. 4:

Tabuľka č. 4: Paradigma a jej hodnoty pri číselkoch.

ZNAK	HODNOTA	PRÍKLADY
S	SUBSTANTÍVNA	<i>nula, milión, bilión, miliarda, stámilióny, státisíce, raz</i>
A	ADJEKTÍVNA	<i>jediný, mnohý, prvý, druhý, tisíci, milióny, miliardy, nultý, posledný dvojnásobný, trojnásobný, mnohonásobný, viacnásobný jednaký, dvojaký, trojaký, obojaký, mnohoraký</i>
N	ČÍSELKOVÁ	<i>dva, dvaja, dve, tri/traja, štyri/štyria</i>
F	ZMIEŠANÁ	<i>jeden, jedna, jedno</i>
U	NEÚPLNÁ	<i>päť/piati, šesť/šiesti, sedem/siedmi, sto, tisíc, dvesto, tristo, dvetisíc, tritisíc dvoje, troje, štvoro, päťoro, tisícoro, obidvoje, oboje, viacero veľa, viac, mnoho, moc, plno, nadostač, málo, trocha, pár, zopár, dosť</i>
D	PRÍSLOVKOVÁ	<i>prvýkrát, druhýkrát, tretíkrát, šesťkrát, sedemkrát, viackrát dvoj-, troj-, štvor-, sto-, tisíc-, mnoho-, viac- -násobne jedn-, dvoj-, troj-, štvor-, päťor-, stor-, tisícor-, mnohor- -ako</i>

Zložené číselky sú rozdelené na samostatné tokeny oddelené medzerou, preto ich značíme zvlášť. Pri týchto a podobných slovách rozdelených na viac tokenov (napriek tomu, že tvoria jednu jazykovú jednotku – *dvadsiateho tretieho, dva razy*) by preto osamostatnenou morfológickou interpretáciou mohlo dôjsť k ich dezinterpretácii. Súčasné riešenie však viac vyhovuje počítačovému spracovaniu a nevylučuje možnosť zapojenia logického modulu do spracovania textu, ktorý by sa neskôr uplatnil ako vhodnejší základ morfológickej lematizácie a značkovania.

4. Čo ponúka morfológicky anotovaný korpus lingvistiky a výučbe (zámen a číseloviek)?

Morfológická anotácia textového materiálu v korpuse so sebou prináša množstvo nových možností. Lematizácia pomáha zefektívniť vyhľadávanie cez korpusový manažér, a tak získať z korpusu špecifickejšie informácie využiteľné nielen vo výskume, ale aj v pedagogickej praxi pri vytváraní cvičení či analýze vybraných gramatických javov, ktoré sa tu navyše vyskytujú v prirodzenom kontexte. Formálny prístup k morfológickej anotácii môže poskytnúť vhodnú pomôcku pri príprave učiteľa i pri samotnej výučbe vo všetkých fázach vyučovacieho procesu, ako sme ich načrtli na začiatku. Kategória paradigmy i anotácia tzv. neohybných, resp. formovo netransparentných slov (*jeho priateľka, o sto ľuďoch, do kupé, s fajn dievčaťom*) posilňuje uvedomovanie si vzájomných väzieb morfológie a syntaxe, orientuje žiaka na morfosyntaktické vzťahy a predstavuje jazyk ako komplex, systém, v ktorom jednotlivé prvky spolupôsobia a komplementárne sa dopĺňujú pri vyjadrovaní významu. Práca so živým jazykovým materiálom zas posilňuje analytické myslenie, vytvára možnosti pre osvojenie si a ďalšie uplatnenie základných myšlienkových operácií (dedukcia, analýza, syntéza a analógia). Hádám ani netreba

pripomínať to, že práca s počítačom žiaka motivuje, zabezpečuje kladný postoj k výučbe materinského jazyka i gramatiky cudzích jazykov.

Predpokladom však musí byť, že na korpus sa pozeráme ako na konkrétny obraz používania jazykových prvkov v textoch a anotáciu vnímame ako pokus o zachytenie formálnych systematických štruktúr. Potom analýza korpusových dát môže posilniť prirodzený pohľad na jazyk ako na živý nástroj.

Sme toho názoru, že takýto nový prirodzený pohľad na skutočnú realizáciu jazykového systému je nevyhnutný aj v oblasti gramatických opisov. O potrebe novej slovenskej gramatiky sa uvažuje intenzívne najmä v posledných rokoch.¹⁷ Pri tomto úsilí bude nevyhnutné opierať sa o niekoľko pilierov, za ktoré M. Sokolová (2004) považuje nielen a) systémovo-lingvistický prístup uplatnený v MSJ (1966), b) slovakistické práce po r. 1966, c) explanačné princípy a teóriu prirodzenej gramatiky (J. Dolník a P. Žigo), d) komunikatívno-pragmatické výskumy (D. Slančová, J. Kesselová, J. Kupcová, M. Kášová), ale aj e) nové počítačové a korpusové výskumy. Na záver svojej prednášky totiž uvádza: „Korpusový výskum má mimoriadny potenciál. Počítač ako charakteristický nástroj súčasnej epochy globalizácie môže pomôcť zmierniť handicap malých kultúr ..., navyše pomáha oslobodiť sa „od sugescie špeciálneho príkladu“ (Páleš) a znamená aj „korekciu teórie“ (Štícha), aj veľkú inšpiráciu.“ (Sokolová, 2004). Získané nové zistenia na tomto základe by potom mohli priniesť aj želané a očakávané zmeny vo výučbe morfológie v školách.

Literatúra

- Dvonč, L.: Dynamika slovenskej morfológie. Bratislava, Veda 1984. 124 s.
- Forróová, M. – Garabík, R. – Gianitsová, L. – Horák, A. – Šimková, M.: Návrh morfológického tagsetu SNK. Medzinárodná vedecká konferencia Slovánske jazyky v počítačovom spracovaní. Bratislava 24. – 25. 10. 2003, v tlači.
- Forróová, M. – Horák, A.: Morfológická anotácia korpusu. In: Slovenčina na začiatku 21. storočia. Red. M. Imrichová. Prešov: Fakulta humanitných a prírodných vied PU 2004, s. 174 – 183.
- Garabík, R. – Gianitsová, L. – Horák, A. – Šimková, M.: Tokenizácia, lematizácia a morfológická anotácia Slovenského národného korpusu. Bratislava: SNK JÚLŠ 2004, <http://korpus.juls.savba.sk/korpus/biblioteka/publikacie/Tagset-aktualny.pdf>.
- Gianitsová, Lucia: Pokus o nové vymedzenie syntetických a analytických tvarov v slovenčine s ohľadom na školskú prax. Slovenský jazyk a literatúra v škole, 50, 2003/2004, 3 – 4, s. 65 – 77.
- Gianitsová, L.: Problémy prípravy morfológického tagsetu pre SNK II. Lema a tag. XIII. kolokvium mladých jazykovedcov. Modra-Piesok 3. – 5. 12. 2003, v tlači.
- Gianitsová, L.: Možnosti využitia SNK v pedagogickej praxi. Vedecký seminár Učebné materiály vo vyučovaní materinského jazyka, slohu a literatúry. Banská Bystrica 3. – 4. 2. 2004, v tlači.
- Hajič, J.: Popis morfológických značiek – poziční systém. Praha: ÚČNK – ÚFAL MFF UK 2000, <http://ucnk.ff.cuni.cz/manual/znacky.html>.
- Hana, J. – Hanová, H.: Manual for morphological annotation. Praha: ÚFAL MFF UK 2002.
- Jarošová, A.: *Umývať sa*: predmetová syntagma alebo zvrtné sloveso? In: Tradícia a perspektívy

¹⁷ V tomto duchu sa niesla aj vedecká konferencia s gramatickou tematikou nazvaná *Tradícia a perspektívy gramatického výskumu na Slovensku* (Smolenice 8. – 9. októbra 2001). Zborník z tejto konferencie vyšiel pod rovnakým názvom v roku 2003, editorka M. Šimková (porov. Ološtiak – Gianitsová, 2004).

- gramatického výskumu na Slovensku. Ed. M. Šimková. Bratislava: Veda 2003, s. 204 – 208.
- Krátky slovník slovenského jazyka. 4. vyd. Red. J. Kačala, M. Pisárčiková, M. Považaj. Bratislava: Veda 2003. 985 s.
- Morfológia slovenského jazyka. Red. J. Ružička. Bratislava: Veda 1966.
- Navrátil, L.: Prídavné mená – zámená – číslovky. Nitra: Enigma 1998. 83 s.
- Ološtiak, M. – Gianitsová, L.: O tradícii a perspektívach akademickej gramatiky slovenčiny. [Tradícia a perspektívy gramatického výskumu na Slovensku.] Jazykovedný časopis, 55, 2004, č. 1, s. 43 – 53.
- Ondrus, P.: Číslovky v súčasnej spisovnej slovenčine. Bratislava: Univerzita Komenského 1969. 147 s.
- Ondrus, P.: Kapitoly zo slovenskej morfológie. Bratislava: SPN 1978. 187 s.
- Oravec, J. – Bajzíkova, E. – Furdík, J.: Súčasný slovenský spisovný jazyk. Morfológia. Bratislava: SPN 1984.
- Pauliny, E.: Vývin slovenskej deklinácie. Bratislava, Veda 1990. 266 s.
- Sabol, F.: Nové vymedzenie čísloviek. Slovenská reč, 58, 1993, č. 5, s. 302 – 305.
- Sokolová, M.: Kapitoly zo slovenskej morfológie. Prešov, Slovacontact 1995. 178 s.
- Sokolová, M.: Možnosti vytvorenia slovenskej morfológie na korpusovom základe. Prednáška pre SJS pri SAV v Bratislave (28. 6. 2004), nepublikované.
- Šimková, M.: Počítačové spracovanie prirodzeného jazyka a Slovenský národný korpus. In: Počítačová podpora prekladu. Zborník prednášok. Budmerice 22. – 23. 5. 2003, s. 15 – 19.
- Tibenská, E. a kol.: Slovenský jazyk pre 8. ročník základných škôl. 1. diel. 1. vyd. Bratislava: Orbis Pictus Istropolitana 2001. 64 s.
- Tibenská, E. a kol.: Slovenský jazyk pre 6. ročník základných škôl. 2. diel. 2. vyd. Bratislava: Orbis Pictus Istropolitana 2002a. 64 s.
- Tibenská, E. a kol.: Slovenský jazyk pre 7. ročník základných škôl. 1. diel. 2. vyd. Bratislava: Orbis Pictus Istropolitana 2002b. 64 s.
- URL: <http://korpus.juls.savba.sk>