

**Lingvistická analýza výrazu *to*
v pracovnej verzii hovoreného korpusu slovenčiny**

DIPLOMOVÁ PRÁCA

Katarína Gajdošová

**UNIVERZITA KOMENSKÉHO V BRATISLAVE
FILOZOFICKÁ FAKULTA
KATEDRA SLOVENSKEHO JAZYKA**

Slovenský jazyk a literatúra

Juraj Dolník, Prof. PhDr., DrSc.

BRATISLAVA 2008

Čestne vyhlasujem, že som diplomovú prácu vypracovala samostatne
s použitím uvedenej literatúry.

Bratislava 30. júla 2008

.....

Katarína Gajdošová

Za odbornú pomoc pri písaní diplomovej práce, za povzbudivé slová, trpezlivosť a usmernenia i za patronát nad vedením práce ďakujem môjmu školiteľovi prof. PhDr. Jurajovi Dolníkovi, DrSc.

Takisto ďakujem PhDr. Márii Šimkovej, vedúcej oddelenia Slovenského národného korpusu JÚLŠ SAV v Bratislave, za cenné pripomienky a postrehy pri písaní mojej práce, ako aj za prijatie a zaradenie mojej pracovnej verzie hovoreného korpusu, vytvorenej na výskumné účely tejto práce, do väčšieho celku – Slovenského hovoreného korpusu, ktorý sa pripravuje na pôde Slovenského národného korpusu JÚLŠ SAV v Bratislave.

Osobitne ďakujem PhDr. Mire Nábělkovej, CSc., za podnetné myšlienky do diskusie o analyzovanej problematike, za konštruktívne poznámky a usmernenia.

Moje poďakovanie patrí aj RNDr. Radovanovi Garabíkovi, ktorý pre môj výskum vytvoril lematizovanú a morfológicky anotovanú pracovnú verziu hovoreného korpusu zo zozbieraného a prepísaného materiálu.

Ďakujem všetkým, ktorí svojou pomocou prispeli k prepísaniu zozbieraných zvukových nahrávok.

A na záver, ale nie na poslednom mieste, ďakujem všetkým, ktorí ma podporovali svojimi modlitbami, a hoci dnes ani neviem, koľkí to boli, raz budú za to iste bohato odmenení.

Vybudovanie pracovnej verzie hovoreného korpusu slovenčiny bolo možné vďaka finančnej podpore spoločnosti Konto Orange, n. f., v rámci grantového projektu *Je ti tvoja škola malá?*

Obsah

Úvod	6
1. Stav a možnosti výskumu hovorenej reči	9
1.1. Výskumy hovorenej reči na Slovensku	9
1.2. Výskumy hovorenej reči vo svete	11
1.2.1. Hovorené korpusy	11
1.2.2. Rečové archívy	14
2. Hovorený korpus slovenčiny	16
2.1. Štruktúra metadát	16
2.2. Spôsob prepisu	18
2.3 Utvorenie pracovnej verzie hovoreného korpusu slovenčiny, jej charakteristika	19
3. Prehľad prístupov k výrazu <i>to</i>	22
4. Vlastný výskum výrazu <i>to</i> na materiáli hovoreného korpusu slovenčiny	28
4.1. Funkcie výrazu <i>to</i>	28
4.2. Výraz <i>to</i> z formálneho hľadiska	36
Záver	44
Literatúra	47
Zdroje	55
Prílohy	
1. Značky použité v prepise textov	57
2. Metadáta o respondentovi v zvukovej nahrávke	59
3. Metadáta o nahrávke	60
4. Frekvenčná distribúcia výrazu <i>to</i> (lemma=1<KWIC)	61
5. Frekvenčná distribúcia výrazu <i>to</i> (lemma=-1<KWIC)	62
6. Frekvenčná distribúcia výrazu <i>to</i> (word=1<KWIC)	63
7. Frekvenčná distribúcia spojenia <i>to</i> + tvar slovesa <i>byť</i>	

(word=1<KWIC).....	63
6. Frekvenčná distribúcia výrazu <i>to</i> (word=1<KWIC)	63
7. Frekvenčná distribúcia spojenia <i>to</i> + tvar slovesa <i>byť</i> (word=1<KWIC)	63

Úvod

Skúmanie hovorenej reči a jej osobitostí, ako aj porovnávanie komunikácie v spontánných prejavoch a v písaných prejavoch je v súčasnosti veľmi aktuálne. Na Slovensku sa však dosiaľ nenachádza systematicky spracovaný materiál hovorenej podoby spisovnej slovenčiny, na ktorom by sa takýto výskum dal realizovať. Pritom „z hľadiska rozvoja sociolingvistiky a sociolingvistického myslenia na Slovensku má osobitnú, v našich podmienkach priam strategickú dôležitosť práve výskum hovorenej podoby spisovnej slovenčiny, zorganizovaný začiatkom 60. rokov 20. storočia“ a pretrváva „nádej, že sa aspoň v istej podobe podarí zopakovať po 40 rokoch túto výnimočnú akciu“ (Ondrejovič, 2007, s. 9 a s. 16). Táto výzva a trend budovania hovorených korpusov vo svete boli inšpiráciou aj na tvorbu hovoreného korpusu slovenčiny.

S finančnou podporou spoločnosti Konto Orange, n. f., v rámci grantového projektu pre študentov vysokých škôl *Je ti tvoja škola malá?* získaného v akademickom roku 2007/2008 sme mohli na účely nášho výskumu vytvoriť pracovnú verziu hovoreného korpusu slovenčiny, ktorá sa stane súčasťou Slovenského hovoreného korpusu budovaného v rámci projektov Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV v Bratislave. Po utvorení pracovnej verzie hovoreného korpusu sme sa zamerali na skúmanie výrazu *to* (výlučne v tomto tvare) v materiáli zozbieranom z rôznych lokalít Slovenska. Výber bol ovplyvnený viacerými dôvodmi: hľadali sme jazykové prostriedky, ktoré by boli typické pre hovorený prejav, boli by v ňom dostatočne frekventované, ich zápis by nebol ovplyvnený spôsobom realizácie (výslovnosťou) a súviseli by s aktuálnym členením, ktorému sa plánujeme venovať. Všetkým týmto kritériám najlepšie vyhovovali niektoré zámená, z nich sme si vybrali výraz *to*. Vzhľadom na to, že systematické skúmanie jazykových prostriedkov slovenčiny sa dosiaľ realizovalo takmer výlučne na materiáli z písaných textov, je náš materiálový výskum na báze elektronicke spracovaného hovoreného korpusu do istej miery jedinečný.

Metodologicky sme postupovali od sondážneho terénneho zberu a prepisu materiálu k jeho analýze (manuálne aj pomocou počítačových nástrojov) so zameraním na funkciu a spôsoby použitia výrazu *to* v hovorenej komunikácii vrátane jeho spájateľnosti. Zistenia z tejto analýzy a komparácie uplatnenia výrazu *to* v hovorenom korpuse a v korpuse písaných textov nás priviedli aj k čiastkovej syntéze v závere predkladanej práce.

V prvej kapitole našej práce sme sa pokúsili podať komplexnejší pohľad na výskum hovorenej komunikácie nielen na Slovensku, ale stručne uvádzame aj projekty, ktoré uskutočňujú výskum hovorenej reči vo svete. Problematika tvorby hovoreného korpusu a najmä spôsobov prepisu zvukového záznamu nie je dosiaľ vyriešená vo viacerých krajinách sveta napriek tomu, že mnohé jazyky disponujú hovorenými korpusmi v rozsahu niekoľkých stoviek tisícov či niekoľkých miliónov textových jednotiek. Snažíme sa podať stručné informácie o existencii jednotlivých hovorených korpusov vo svete, budovaných na základe rôznych teórií a prístupov a o jestvovaní rečových archívov rôzneho druhu.

V druhej kapitole sa osobitne venujeme tvorbe vlastnej pracovnej verzie hovoreného korpusu, analýze zdrojov, postupom vytvárania hovoreného korpusu slovenčiny, spôsobu prepisu získaného materiálu. Samostatne charakterizujeme nami pripravený pracovný hovorený korpus slovenčiny.

Tretia kapitola predstavuje sumár vedecko-výskumných prác, ktoré sa s naším výskumným cieľom bezprostredne spájajú. Ide o slovenské i české gramatiky a lexikografickú príručku slovenského jazyka, ale aj o čiastkové štúdie lingvistov, ktorí sa výrazu *to* venujú z rôznych hľadísk. Osobitný pohľad predstavuje analýza uplatnenia výrazu *to* v koreferenčných vzťahoch.

Štvrtú časť našej práce tvorí vlastný výskum výrazu *to* v zozbieranom materiáli. Niektoré charakteristiky výrazu *to* uvedené v slovenskej gramatike, najmä v Morfológii slovenského jazyka, podrobujeme čiastkovej analýze. Pokúsili sme sa odpútať sa od tradičných prístupov k výrazu *to* prevažne chápaného ako zámeno a odhaliť aj jeho ďalšie použitia. Snažíme sa o dokladovanie textovej funkcie výrazu *to* z hľadiska koreferenčných vzťahov a

osobitnú podkapitolu našej práce venujeme formálnemu pohľadu na výraz *to* a jeho spájateľnosť v hovorenom korpuse. Pri tejto časti analýzy uvádzame aj presne formulované príkazy na vyhľadávanie jednotlivých použití výrazu *to* v elektronickej databáze.

Záver našej práce obsahuje zhrnutie doterajších výsledkov materiálového výskumu a vymedzenie možností perspektívneho pokračovania výskumu výrazu *to*.

1. Stav a možnosti výskumu hovorenej reči

1.1. Výskumy hovorenej reči na Slovensku

Otázka skúmania hovorenej reči nie je na Slovensku novinkou, hoci sa súčasne neraz konštatuje, že hovorená podoba spisovnej slovenčiny nie je dodnes dostatočne preskúmaná. Dôvodom je na jednej strane úzky okruh lingvistov, ktorí sa špecializujú na tento druh lingvistického bádania, na druhej strane i problém zhromaždiť dostatok materiálu, na ktorom by bolo možné realizovať výskum. Napriek neľahkej situácii sa v slovenskej lingvistike realizovalo niekoľko špecializovaných výskumov z tejto oblasti. Sondy do hovorenej podoby spisovnej slovenčiny predstavujú buď dotazníkové výskumy, alebo výskumy na získanom audio či audiovizuálnom materiáli, ktorý najzreteľnejšie zachytáva skutočný stav hovorenej podoby spisovnej slovenčiny. Najobťažnejšie a zo strany sociolingvistov súčasne najpožadovanejšie je získanie spontánneho audio či audiovizuálneho materiálu, ktorý poskytuje širokú bázu na skúmanie hovorenej podoby jazyka, jeho variet či špecifikácií. Terénny zber materiálu sa však dosiaľ uskutočňoval len pre potreby jednotlivých bádateľov, veľmi výberovo a útržkovito. V súčasnosti nie je k dispozícii žiaden komplexnejší súbor (korpus) zvukových záznamov, ktoré by mohli slúžiť všetkým lingvistom na ich konkrétne výskumné ciele. Vzhľadom na časovú a personálnu náročnosť sa doterajšie nahrávky získavali značne nekonzistentne, nepokrývajú celé územie Slovenska, plnia funkciu sondy, ktorá vznikla pre potreby naplnenia čiastkového cieľa. V minulosti boli zvukové záznamy a ich prepisy uchovávané len čiastočne, po opísaní sondáže bol materiál neraz zničený alebo nahradený iným bez akéhokoľvek archivovania.

V širšom rozsahu uchovávané zvukové záznamy a ich prepisy z dialektologických výskumov (zhromaždené v Jazykovednom ústave Ľ. Štúra SAV v Bratislave) sú orientované primárne na výskum fungovania slovenských nárečí. V našej práci sa však sústreďujeme na výskum používania hovorenej

slovenčiny, bežne hovoreného jazyka, ktorého nositelia v súčasnosti nárečie často ani nepoužívajú.

Skúmanie hovorenej podoby spisovnej slovenčiny predstavuje širokú oblasť a mnohé jej miesta sú stále biele. Východiská celistvejšieho sociolingvistického bádania na Slovensku, metodologické postupy či podnety na skúmanie konkrétnych javov sú koncipované dostatočne široko (Ondrejovič, 1991; Ondrejovič, 1994; Ondrejovič, 1995; Ondrejovič, 1996), no komplexné výskumy jednotlivých javov pre materiálový problém absentujú.

Pri výskume hovorenej podoby slovenčiny venoval sústredenú pozornosť jazyku miest J. Muránsky. Snažil sa o podanie celistvejšieho pohľadu na jazyk východoslovenských miest (Muránsky, 1978; 1979; 1980; 1981; 1982; 1983; 1984; 1986; 1988; 1990). Jeho bádania však boli len ojedinelou sondou do tejto problematiky. Variety slovenčiny v hovorenej komunikácii na Slovensku so zameraním na východné Slovensko skúmali M. Sokolová a D. Slančová opierajúc sa o stratifikáciu národného jazyka J. Horeckého (Sokolová – Slančová, 1995; 1996). Špecifickej vrstve jazykových prostriedkov jazyka mesta – vulgarizmom – sa komplexnejšie venovala D. Slančová (2000). V súčasnosti D. Slančová v nadväznosti na svoje dávnejšie výskumy osobitne analyzuje jazyk mládeže, jazyk detskej komunikácie nielen na východnom Slovensku a v okolí Prešova (1993; 1997; 1999; 2001; 2002; 2003; 2004; 2005). Výskumami mestského jazyka najmä v oblasti stredného Slovenska sa zaoberali V. Patráš (1991; 1992a; 1992b; 1994; 1995a; 1995b; 1996; 1999; 2000) a L. Urbancová (1999; 2000a; 2000b; 2004; 2005). Niekoľko výskumov realizoval na južnom Slovensku S. Ondrejovič (1996), komparatívne skúmal západné a stredné Slovensko (Ondrejovič, 1991).

V 60. rokoch minulého storočia bola širšia slovenská sociolingvistická obec zapojená do celoslovenského projektu s názvom *Hovorená podoba spisovnej slovenčiny*, ktorého cieľom bolo zmapovať podobu spisovnej slovenčiny v 100 slovenských mestách prostredníctvom 1 000 respondentov.

Spracované materiály boli donedávna k dispozícii len v JÚLŠ SAV v Bratislave v podobe dvoch zväzkov interných materiálov (Hovorená podoba spisovnej slovenčiny I., II., 1965). V roku 2007 došlo k prvému oficiálnemu vydaniu textov z rovnomennej konferencie, na ktorej sa hodnotili výsledky projektu, v rámci série zborníkov Sociolinguistica Slovaca (2007), aby k nim mala prístup aj celá lingvistická obec. Zvukové nahrávky, ktoré boli základným východiskom výskumu, sú archivované v JÚLŠ SAV, no dosiaľ sa nedigitalizovali, a tak záujemcovia nemajú prístup k zvukovému prameňu. Predpokladaná digitalizácia a záchrana týchto zvukových záznamov a ich zaradenie do vznikajúceho Slovenského hovoreného korpusu (SHK) sa zatiaľ neplánujú vzhľadom na vysokú finančnú náročnosť takéhoto procesu.

1.2. Výskumy hovorenej reči vo svete

Výskumami hovorenej reči vo svete sa vzhľadom na počty odborníkov v oblasti sociolingvistiky a iných vedných odborov zaoberá omnoho viac vedcov ako na Slovensku. V súvislosti s témou našej diplomovej práce sa budeme prehľadovo venovať niektorým výskumom, a to najmä hovoreným korpusom. Osobitné skupiny tvoria hovorené korpusy, ktoré sú súčasťou zväčša veľkých národných korpusov. Ich zastúpenie v celku národných korpusov je malé, percentuálny podiel sa pohybuje v rozmedzí 10 – 20 % z celého korpusu textov.

1.2.1. Hovorené korpusy

V roku 1988 vznikla celosvetová organizácia ESCA, ktorá sa od svojho vzniku venuje podpore výskumu hovorenej reči. V roku 1999 bola premenovaná na ISCA (International Speech Communication Association; <http://www.isca-speech.org/index.html>). Táto organizácia má za úlohu podporovať výskumy hovorenej reči po celom svete, organizovať konferencie a semináre rôzneho zamerania, ponúka možnosti podania grantov a pracovných/študijných pobytov

pre záujemcov. Spolupracuje na formálnej úrovni s množstvom pracovísk na výskum hovorenej reči.

London-Lund Corpus (<http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM>) vznikol pôvodne ako súbor prepisov nahrávok z projektu Survey of English Usage z roku 1959 v rozsahu 0,5 milióna slov. Neskôr bol doplnený o ďalšie texty. Tie zachytávajú už aj paralingvistické informácie: tónové jednotky, začiatky tónového nasadenia, intonáciu, dva druhy páuz, typy prízvukov, aktuálneho hovoriaceho, sprievodné informácie typu smiech, plač, kašeľ a pod.

British Academic Spoken English Corpus (BASE) (<http://www2.warwick.ac.uk/fac/soc/celte/research/base/>) vznikol z prepisov nahrávok prednášok a seminárov z rôznych vedných oblastí. Obsahuje podrobnú žánrovú anotáciu prepisov, časové vymedzenie páuz, označenie častí prehovoru, striedania partnerov v komunikácii a spôsobu preberania iniciatívy počas rozhovoru.

Longman Spoken American Corpus (<http://www.pearsonlongman.com/dictionaries/pdfs/Spoken-American.pdf>) tvorí 5-miliónová databáza prepísaných textov. Tie vznikli z nahrávok, ktoré sa zozbierali z 12 regiónov po celej Amerike. Nahrávky boli následne transkribované podľa zvolených zásad.

British National Corpus (<http://www.natcorp.ox.ac.uk/>) tvorí 90 % písaných textov a 10 % textov vzniklo prepisom zvukových nahrávok. Obsah hovorenej časti Britského národného korpusu tvoria nahrávky formálnych obchodných a vládnych stretnutí, ktoré sa realizovali priamo v rádiách alebo boli prezentované v rádiách prostredníctvom telefonických rozhovorov účastníkov diskusií. Respondenti boli delení vyvážené na základe veku, bydliska, sociálneho postavenia.

American National Corpus (<http://www.americannationalcorpus.org/>) sa takisto skladá z písanej a hovorenej časti, pričom hovorená časť obsahuje v súčasnosti 3,2 milióna slov. Na prepis boli použité nahrávky z osobných rozhovorov (asi 198 tisíc slov) a telefonických rozhovorov (asi 3 milióny slov).

Spoken Language Corpus of Swedish (<http://www.qualitative->

research.net/fqs-texte/3-00/3-00allwoodetal-e.htm) vznikol na Univerzite v Göteborgu. Obsahuje asi 1,26 milióna slov, ktoré vznikli prepisom nahrávok z komunikácií počas rôznych aktivít respondentov. Prvá polovica prepisov bola urobená z audio nahrávok, druhá z video nahrávok.

C-ORAL (<http://lablita.dit.unifi.it/coralrom/>) je projekt hovoreného korpusu štyroch hlavných románskych jazykov – španielčiny, portugalčiny, francúzštiny a taliančiny. Každý jazyk je zastúpený počtom 300 000 slov prepísaných z formálnych aj neformálnych rečových prejavov. K transkripcii sú doplnené informácie o metadátach, ako aj označkovanie celého korpusu. Korpus bol distribuovaný na 8 DVD nosičoch.

Corpus of Spoken Israeli Hebrew (<http://www.tau.ac.il/humanities/semitic/cosih.html>) je projekt hovoreného korpusu modernej hebrejčiny. Má bohato rozpracovanú metodológiu i štruktúru dát a metadát, ktorá slúži často ako inšpirácia pre tvorcov iných národných korpusov.

Spoken Dutch Corpus (<http://lands.let.kun.nl/cgn/ehome.htm>) vznikol v roku 2004 ako databáza prepisov zo zvukových nahrávok dospelých obyvateľov Flámska a Holandska v rozsahu 1 000 hodín (10 miliónov slov). Dve tretiny materiálu pochádzajú od respondentov z Holandska, jedna tretina z Flámska. Tento hovorený korpus je jedným z mála korpusov, ktoré poskytujú súčasne zvukovú nahrávku aj ortografický prepis. Celok asi 1 milión slov bol prepísaný foneticky, časť textu bola lematizovaná. Následne prebehla štandardizácia textu s cieľom realizácie ďalšej úrovne – syntaktickej anotácie. Asi 250 tisíc slov bolo anotovaných aj prozodicky.

Vo svete vznikajú korpusy hovorenej reči aj na komerčné využitie. Ich tvorcami sú zväčša veľké počítačové či telekomunikačné spoločnosti, ktoré na zhromaždených dátach testujú nástroje na analýzu a syntézu reči. Výskum hovorenej reči podporuje napríklad počítačová spoločnosť IBM, ktorá má osobitné oddelenie na rozpoznávanie reči.

Veľkým komerčným projektom bol medzinárodný projekt Verbmobil zameraný na rozpoznávanie reči, jej analýzu a preklad do iného cudzieho jazyka,

konkrétne medzi jazykom nemeckým, anglickým a japonským. Cieľom projektu bolo vyvinúť nástroj, ktorý by umožňoval používateľom mobilných telefónov uskutočňovať rozhovory s inou osobou hovoriacou odlišným jazykom bez toho, aby museli účastníci rozhovoru používať rovnaký jazyk. Vyvinutý automatický nástroj mal zabezpečiť rozhovor oboch účastníkov vo svojom materinskom jazyku.

1.2.2. Rečové archívy

Na rozdiel od hovorených korpusov, ktoré majú zvyčajne menší rozsah a presnejšie špecifikovaný obsah vrátane rôznych anotácií, rečové archívy majú za cieľ zhromaždenie audio a audiovizuálneho materiálu vo veľkom rozsahu tak, aby slúžil predovšetkým na účely spracovania prirodzenej reči, t. j. napríklad ako trénovací materiál pre potreby analýzy a syntézy reči. Rečové archívy rôznych jazykov sa vyznačujú rozdielnym stupňom komerčného využitia buď v oblasti telekomunikácií, alebo v oblasti automatických hovorených nástrojov v počítačovom priemysle. Projekty rečových archívov zahŕňajú v mnohých, nielen európskych krajinách aj existenciu hovoreného korpusu. Rečový archív je teda databázou, ktorá obsahuje niekoľko druhov špecificky označovaných alebo aj neoznačovaných subdatabáz. Každé označovanie má svoje priority, či už sú to v podstate gramatické a širšie sociolingvistické metadáta v prípade hovorených korpusov, fonetické v prípade špecifického fonetického výskumu¹ či suprasegmentálne anotácie.

Bavorský rečový archív rečových signálov (<http://www.phonetik.uni-muenchen.de/Bas/BasKorporaeng.html>.) sa skladá z niekoľkých druhov zvukových nahrávok – spontánnych rečových prejavov, čítaných rečových prejavov, nárečových prejavov, telefónnych nahrávok aj nahrávok s vysokou kvalitou zvukového záznamu (tieto nahrávky môžeme považovať do istej miery

¹ Sem môžeme zaradiť napr. trvanie jednotlivých hlások, ktoré sú značkové pomocou nástrojov na zaznamenávanie dĺžky hlások, alebo časový interval geminovaných hlások tam, kde by ich ľudské ucho nebolo schopné zachytiť.

za laboratórne).

Speech Accent Archive (<http://accent.gmu.edu/>) zhromažďuje nahrávky anglicky hovoriacich respondentov (rodených aj nerodených). Respondenti čítajú vybrané texty do zvukového zariadenia a následne sa zvukové nahrávky ručne prepisujú. Archív slúži ako zdrojový materiál na komparatívne výskumy. Okrem zvukovej nahrávky a jej prepisu sú k dispozícii aj sociologické informácie o respondentovi.

CNN archive (<http://www.cnn.com/SPECIALS/2004/reagan/stories/-speech.archive/index.html>) obsahuje napríklad súbor verejných prejavov bývalého prezidenta Spojených štátov amerických Ronalda Reagana.

Ambíciu a potrebu budovať rečový archív slovenčiny by mali aj niektoré pracoviská na Slovensku, napríklad Ústav informatiky SAV v Bratislave (oddelenie Spracovania reči), ale ide o finančne veľmi náročný projekt.

2. Hovorený korpus slovenčiny

Ako sme už vyššie uviedli, hovorený korpus môže byť súčasťou rečového archívu, avšak to len v prípade jeho existencie. Budovanie rečového archívu je omnoho náročnejšie ako budovanie užšie vymedzeného hovoreného korpusu, hoci aj to, ako ukážeme neskôr, má svoje špecifiká. Mohli by sme konštatovať, že pre potreby lingvistických výskumov je potrebnější, ba dokonca nevyhnutný práve hovorený korpus, pričom existencia rečového archívu ním nie je podmienená.

Ak by sme však chceli hovoriť o komplexnosti zachytenia akejkoľvek hovorenej reči, s istotou môžeme potvrdiť neoceniteľný prínos existencie rečových archívov v tých jazykoch, ktoré nimi momentálne disponujú. Keďže sa však v našej práci chceme primárne venovať vybudovaniu pracovnej verzie hovoreného korpusu, nebudeme sa ďalej zaoberať existenciou a procesmi utvárania rečových archívov.

Vybudovanie hovoreného korpusu slovenčiny je v súčasnosti v oddelení Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV v Bratislave (ďalej SNK) jednou z primárnych úloh (Šimková – Garabík – Karčová – Gajdošová, 2007). Keďže však používateľom nie je zatiaľ sprístupnená žiadna verzia Slovenského hovoreného korpusu (ďalej SHK), vybudovali sme si vlastnú pracovnú verziu hovoreného korpusu slovenčiny, ktorá bude neskôr zaradená do databázy SHK. Stručne predstavíme relevantné informácie týkajúce sa štruktúry metadát a samotného prepisu nahrávok.

2.1. Štruktúra metadát

Štruktúrovanosť a vyváženosť hovoreného korpusu nielen vo vlastnom výskume, ale aj pri utváraní reprezentatívneho niekoľkomiliónového Slovenského hovoreného korpusu podmieňuje podrobná špecifikácia

jednotlivých vybraných sociolingvistických, sociologických i geograficko-demografických parametrov.

Za základný parameter považujeme spolu s pohlavím respondenta jeho vek, ktorý zisťujeme a uvádzame v rozmedzí desiatich rokov, v prípade, že je respondent ochotný uviesť svoj vek presne, bude táto informácia rovnako k dispozícii pre ďalší výskum.

Vzdelanie je štandardne členené do štyroch skupín – žiadne, nižšie (základné), stredoškolské a vysokoškolské. V našom hovorenom korpuse je rozšírené o položku odborné vzdelanie, do ktorej zaraďujeme respondentov – absolventov odborného učilišťa bez maturity.

Z geografického hľadiska sú pre náš výskum dôležité informácie o mieste narodenia respondenta, mieste jeho najdlhšieho pobytu, ktoré často výrazne ovplyvňuje respondentov prehovor, a mieste jeho súčasného pobytu.

K metadátam o respondentovi a zvukovej nahrávke zaraďujeme aj informáciu o materinskom jazyku respondenta, o jeho jazykovej kompetencii vo vzťahu k iným jazykom. Nezanedbateľnou je aj položka ne/používania nárečia, v prípade kladnej odpovede jeho špecifikácia.

Na podrobnejšie vyhľadávanie v utváranej databáze bude slúžiť aj uvádzanie témy rozhovoru v prípade, ak sa bude nejaký konkrétny výskum orientovať na špecifické tematické okruhy rozhovorov.

Informácie o priebehu nahrávky uvádzame v dvojiciach riadený – spontánný, formálny – neformálny rozhovor. Za riadený rozhovor považujeme taký, v ktorom sú respondentovi kladené otázky a on na ne odpovedá. Spontánný je taký druh rozhovoru, v ktorom prebieha rozhovor voľne, bez kladenia otázok. Takýto druh rozhovoru vzniká najmä v situáciách, keď respondent nie je dopredu informovaný o prítomnosti nahrávacieho zariadenia, teda nahrávanie sa uskutočňuje tajne, v skrytosti. Až následne je respondent oboznámený s nahrávaním a dodatočne sa od neho získa súhlas so začlenením nahrávky do databázy.

Nezanedbateľnou je informácia o ne/známosti účastníkov rozhovoru a o

ich ne/rovnocennosti (spoločenskej alebo profesionálnej).

Ak je respondentov prejav niečím charakteristický, v metadátach o respondentovi nesmie táto informácia chýbať. Ide najmä o informácie, ktoré by mohli zdôvodniť, prípadne spresniť či odhaliť spôsob prejavu alebo artikulácie respondenta (operovaný nos, fajčiar, piercing na jazyku, ale aj zhoršený zdravotný stav respondenta a pod.).

K nahrávke a jej spracovaniu neodmysliteľne patria položky počet účastníkov rozhovoru, dátum, keď bola nahrávka urobená, a miesto nahrávania – nielen začlenenie do geografickej oblasti, ale aj konkrétna špecifikácia, napríklad školská trieda počas prestávky, námestie, kaviareň.²

2.2. Spôsob prepisu

Transkripčné systémy na prepis zvukových záznamov sa vo svetových korpusoch rôznia. Po dôkladnom štúdiu a oboznámení sa s jednotlivými prístupmi doma i vo svete, s ich kladmi a zápormi sme spolu s pracovným kolektívom oddelenia SNK a Ústavu informatiky SAV v Bratislave vypracovali zásady prepisu zvukových záznamov, ktoré budú súčasťou nášho pilotného hovoreného korpusu slovenčiny a neskôr súčasťou Slovenského hovoreného korpusu.

V rámci nášho výskumného projektu sme využili ortografickú zložku prepisu používaného v SHK. Okrem štandardného prepisu zvukovej nahrávky do ortografickej podoby sa v SHK pracuje aj so značkami, ktoré zachytávajú šumy hovoriaceho, príp. šumy iných aktívnych účastníkov komunikácie. Okrem týchto ruchov sa zachytávajú aj výslovnostné odlišnosti a ďalšie súčasti rečových prejavov, ako sú lapsusy, opakovania, prekrývania a pod. Tie tvoria druhú rovinu anotácie. Zoznam značiek pre ruchy a realizované výslovnosti uvádzame v prílohe.

² Viac informácií o štruktúre metadát v hovorenom korpuse a dôvodoch jej zvolenia porov. Gajdošová (2007).

2.3. Utvorenie pracovnej verzie hovoreného korpusu slovenčiny, jej charakteristika

Naša pracovná verzia hovoreného korpusu vznikla na základe finančnej podpory grantového projektu spoločnosti Konto Orange, n. f., s názvom *Je ti tvoja škola malá?* v akademickom roku 2007/2008 a s podporou Slovenského národného korpusu v rámci spolupráce pri utváraní koncepcie budovania SHK.

V pracovnej verzii hovoreného korpusu sa opierame o nahrávky hovorených prejavov, ktoré môžeme štruktúrovať do niekoľkých skupín:

- a) spontánne rozhovory,
- b) poloriadené rozhovory,
- c) homílie,
- d) diskusné televízne relácie.

Nahrávky sme získali prevažne vlastným nahrávaním v rôznych lokalitách Slovenska. Na výskum v predkladanej práci sme použili všetky nahrávky, ktoré sme spracovali rovnakou metódou, teda prepisom do ortografickej podoby. Databáza nami utvoreného korpusu obsahuje 164 764 tokenov (textových jednotiek) a zahŕňa spolu 64 textových prepisov 26 zvukových nahrávok:

a) 13 nahrávok patrí do skupiny spontánnych prehovorov. Nahrávanie bolo uskutočnené tajne, teda respondent nebol vopred informovaný o nahrávaní, nahrávacie zariadenie bolo skryté. Až po ukončení rozhovoru bol respondent dodatočne oboznámený s nahrávaním a požiadany o vyjadrenie súhlasu s použitím nahrávky.

b) 8 nahrávok má charakter poloriadených rozhovorov. Po položení otázky respondent formuluje odpoveď, avšak otázky sú orientované tak, aby respondentu primäli k širšej výpovedi, príp. k vybočeniu z rámca otázky. Tak sme zabezpečili istú spontánnosť rozhovoru, ktorý sa z poloriadeného rozhovoru miestami presunul do polohy spontánneho dialógu.

c) 1 nahrávka je prepisom homílie.

d) 4 nahrávky sú prepisom diskusných televíznych relácií. Do nášho korpusu sme ich zaradili napriek tomu, že na ich použitie, na rozdiel od ostatných prípadov, nemáme súhlas. Považovali sme však za vhodné, aby časť našich dát tvoril tento druh nahrávok. V prípade diskusných relácií ide v podstate o poloriadené prejavy účinkujúcich. Scenár relácií je síce vopred daný a diskutéri sú s ním vopred oboznámení, avšak v prípade relácií, ako je Pod lampou, O 5 minút dvanásť, Na streche či 7 s. r. o., ktoré sme do nášho korpusu zaradili, môžeme hovoriť o značnej voľnosti pri vyjadrovaní, prejavy nie sú formálne, ale uvoľnené, od hlavnej témy sa často rôznym spôsobom odbočuje.

Pracovný hovorený korpus slovenčiny je na rozdiel od SHK zložený len z ortografických prepisov zvukových nahrávok. SHK obsahuje okrem ortografického prepisu aj druhú rovinu prepisu – tzv. ortoepickú, ktorá je osobitne špecifikovaná na stránkach projektu SHK.³ Obe roviny prepisu sú v SHK navzájom prelinkované, aby mal používateľ k dispozícii aj zvukový záznam prepisu.⁴

Pracovný korpus obsahuje štyri prepisy televíznych relácií (Na streche špeciál, Pod lampou, 7 s. r. o., O 5 minút dvanásť). Tieto prepisy ani nahrávky k nim SHK neobsahuje.

Databáza SHK naopak obsahuje niekoľko desiatok ďalších nahrávok, ktoré sme do nášho pracovného korpusu nezaradili, pretože neboli uskutočnené v rámci grantového projektu spoločnosti Konto Orange, n. f., *Je ti tvoja škola malá?* v akademickom roku 2007/2008.

Pracovnú verziu nášho hovoreného korpusu vytvorili programátori oddelenia Slovenského národného korpusu JÚLŠ SAV v Bratislave štandardnými metódami vypracovanými a overenými pri tvorbe písaných korpusov. S

³ <http://data.juls.savba.sk/oral>

⁴ Pre potreby nášho výskumu nebolo potrebné pracovať s prepisom zlinkovaným s jemu prislúchajúcou nahrávkou.

materiálom v korpuse sme pracovali prostredníctvom korpusového manažéra Manatee s klientom Bonito. Pracovná verzia hovoreného korpusu bola lematizovaná a morfológicky anotovaná automatizovanými nástrojmi používanými pri značkovaní textov písaného korpusu, čo nám umožnilo vyhľadávať nielen podľa položky word (konkrétny slovný tvar), ale aj prostredníctvom položky lemma (základný tvar, pomocou ktorého získame vo výstupe všetky tvary slova v daných kontextoch) a tag (príslušná morfológická značka). Špecifické príkazy na vyhľadávanie jednotlivých funkcií výrazu *to* v korpusovom manažéri Manatee s klientom Bonito uvádzame pri jednotlivých analýzach.

3. Prehľad prístupov k výrazu *to*

V slovenskej lingvistike sa výraz *to* hodnotí zväčša ako zámeno. Vo všeobecnosti je zámeno *to* chápané ako demonštratívum, pre ktoré je typické ukazovanie či odkazovanie na čosi (vec, vlastnosť, osobu) v blízkosti zasahovaného objektu (javu), oproti zámenám s prvkom *on-* (*onen*, *onaký*, *onehdy*) ukazujúcimi na čosi vzdialené (Morfológia slovenského jazyka, 1966, s. 256; ďalej MSJ).

V MSJ (1966, s. 267 – 269) sa vyčleňuje niekoľko základných charakteristík zámena *to* v písaných textoch:

- a) odkazovanie na vytýčený vetný člen bez ohľadu na rodovú príslušnosť a gramatické číslo mena vo vytýčenom člene;
- b) odkazovanie na viac vecí súčasne v spojení s vymedzovacím zámenom *všetok*;
- c) odkazovanie na infinitívnu väzbu a jej zdôraznenie vo vete;
- d) pripájanie prístavku v citovo nadnesených prejavoch;
- e) odkazovanie na obsah reči vôbec;
- f) konkurencia rodovým odkazovacím zámenám *ten*, *tá*, *tí*, *tie* v hovorovom štýle;
- g) odkazovanie na práve vnímaný jav, uvádzanie vysvetľujúcej vety;
- h) odkazovanie na udalosť a vyjadrenie citového záujmu hovoriaceho o ňu;
- i) odkazovanie na časovú okolnosť s platnosťou príslovkového zámena *vtedy* a jeho citové podfarbenie;
- j) odkazovanie na nasledujúci dej v súvetiach – význam zámena *to* vysvetľuje druhá veta súvetia;
- k) substantivizovanie adjektíva, príslovky, predložkových pádov i celých viet;
- l) spájanie pri radení striedavých dejov a okolností za sebou s významom *raz – raz*, *i – i*, *alebo – alebo*;
- m) usúvzťažňovanie dvoch za sebou nasledujúcich dejov do príčinného súvisu, vysvetľovanie príčinného charakteru prvého deja a druhého deja ako dôsledku.

V Mluvnici češtiny. 2 (1986, s. 81 – 95) sa zdôrazňuje okrem kontextovej deixy v prípade zámena *to* jeho využitie v situačnej deixe. Toto zámeno sa

považuje za najneutrálnejšie ukazovacie zámeno, ktoré sa v mnohých ďalších funkciách – najčastejšími bývajú funkcie zdôrazňovacia, zosilňujúca, vytyčovacia a expresívna – používa pred substantívom, ktoré chce autor výpovede podľa svojho zámeru osobitne vyčleniť.

Autori Krátkeho slovníka slovenského jazyka (2003; ďalej KSSJ) radia výraz *to* 1. k zámenám, zachytávajú ho aj 2. v spojení *to jest*, ktoré ako celok označujú ako priradovaciu spojku a osobitne ho vyčleňujú aj ako 3. hovorovú časticu (*to vietor tak zavýja? to on len žartoval; to musíš tak kričať?*). Výraz *to* v časticovej funkcii má podľa KSSJ funkciu nadväznosti na kontext a uvedenia novej vety v nadväzovacej funkcii. Ak sa však pozrieme na vyššie uvedenú charakteristiku funkcií zámena *to* v MSJ, javí sa nám lexikografický výklad ako nedôsledný. V delení funkcií podľa MSJ sa totiž stretávame s nadväzovacou funkciou niekoľkokrát.

E. Bajžíková (Oravec – Bajžíková, 1982) vo svojej klasifikácii jazykových konektorov zaraďuje zámená všeobecne k opakovacím konektorom, môžu to byť anaforické či kataforické konektory. Ukazovacie a privlastňovacie zámená spolu s inými slovnými druhmi (zámenné príslovky, adjektíva) plnia aj funkciu exoforických konektorov.

Vo svojom príspevku Ukazovacie zámená v textovej funkcii sa E. Bajžíková (1979) už skôr zamýšľala aj nad výrazom *to* a synonymným výrazom *toto*. Konštatovala, že tieto ukazovacie zámená v písaných prejavoch nepomenúvajú len entity stredného rodu, ale častejšie zahŕňajú význam celej predchádzajúcej konštrukcie, resp. textovej jednotky. Túto vlastnosť, ako ukážeme ďalej, sme už skôr zistili aj v spontánných prejavoch pri využívaní výrazu *to*.

Funkcie výrazu *to* sa spomínajú v prácach venovaných zámenám všeobecnejšie, ale aj v osobitných výskumoch, napr. z hľadiska jeho slovnodruhovej príslušnosti a jednotlivých funkcií.

M. Dudok (2007) predkladá komplexný pohľad na kategóriu zámen, ako je v slovenskej lingvistickej teórii tradíciou, a podrobuje ju skúmaniu a triedeniu

z funkčného hľadiska. Výraz *to*, ako aj iné zámená zaradil z hľadiska vecnej príznakovosti a diskurzívnej funkcie do 14 relatívne samostatných tried podobne, ako sú zámená triedené v Mluvnici češtiny. 2 (1986).

Možnosti podrobnej analýzy výrazu *to* a jeho identifikácie v českom korpuse písaných textov SYN2000 predstavila I. Kolářová (2004). Vzhľadom na rozsiahlosť materiálu, ktorý bolo potenciálne možné nájsť, zvolila autorka pragmaticky premyslený postup tvorby príkazu podľa postavenia výrazu *to* v nadväzovacej funkcii vo vete tak, aby sa čo najviac eliminovala možnosť vyhľadania dokladov vo funkcii zámena. Preto formulovala vyhľadávacie príkazy tak, aby obsahovali slovesá a bezpodmetové prísudkové výrazy pomenúvajúce prírodné deje a stavy, časové deje a časové úseky, časti dňa, vek a slovesá a prísudkové výrazy označujúce pohyb, lokalizáciu, pozíciu, polohu a stav.

Vychádzala zo skúseností s prácou s Českým národným korpusom, ale aj zo svojich poznatkov o vetnom postavení výrazu *to*. Použitím jednotlivých príkazov získala materiál, v ktorom selektovala výraz *to* primárne vždy ako konektor, ale pokúsila sa súčasne aj o identifikáciu jeho ďalšieho použitia, a to vo význame vysvetľovacom, nadväzovacom, časovom. Vo svojej štúdií nazerá na výraz *to* a jeho kontextovú funkciu z hľadiska substitúcie, pričom jeho jednotlivé významy v syntaktických štruktúrach vždy nahrádza konkrétnou či potenciálnou lexémou, prípadne vetnou konštrukciou.

F. Štícha (2004) analyzuje korelatívum *to* v spojeniach so slovesami. Na základe materiálu z českého korpusu písaných textov SYN2000 vyčleňuje 5 typov sloves, ktoré sa ne/viažu s korelatívom *to* v istých kontextoch. Ide o slovesá s pevnou normou absencie korelatíva *to* (*tvrdit', zistit'*); slovesá s príležitostne a riedko používaným korelatívom *to* (*riect', dokázat', kritizovat', ocenit', fascinovat'*); slovesá s nevýrazným rozdielom v používaní a absencii korelatíva *to* (*brzdit'*); slovesá s prevládajúcim používaním korelatíva *to* (*hovorit'*) a slovesá používané len s korelatívom *to* (*charakterizovat', kompenzovat', obnášat', odbúrat', odcudzovat', miznúť, obhájit', uviaznuť, zahojit', vykúpiť,*

vynucovat). Osobitne vyčleňuje konkurenčné slovesá s výrazom *to* vo funkcii kataforika a bez neho: podmetové vedľajšie vety (*bolieť, brzdiť, hovoriť*) a predmetové vedľajšie vety (*ctiť, dokázať, kritizovať*) bez kataforika *to*; podmetové vety len s kataforikom *to* (*budiť, fungovať, charakterizovať, odbúrať, odcudzovať, obhájiť, uviaznuť*) a predmetové vety výhradne s kataforikom *to* (*kompenzovať, obnášať, zahojiť, vykúpiť, vynucovať si*). Osobitne vyčleňuje korelatívne *to* v špeciálnych súvetných štruktúrach. V závere svojho príspevku formuluje dva dôvody používania kataforika *to*. Významový dôvod napĺňa idea, že sa kataforikum *to* stáva rémou obsahu vedľajšej vety. V prípade použitia korelatíva *to* ide o spojenie slovesa – predikátu vety – so závislým aktantom, ktorým je vedľajšia veta.

Ďalším príspevkom I. Kolářovej (2006) do tejto problematiky je analýza konkurenčného postavenia výrazov *to* a *takový*, s ktorými sa čitateľ, resp. prekladateľ stretáva v prípade prekladu istých vetných konštrukcií do češtiny. Konkurenčné postavenie týchto výrazov na začiatku vety dokladá I. Kolářová textami z českého korpusu SYN2000. Anaforické nadväzovanie skúma najmä v prípadoch, keď ide o odkazovanie na jazykové výrazy charakterizujúce vlastnosti živých a neživých objektov. V závere zhrňa svoje poznatky do konštatovania, že výraz *takový* sa v češtine používa viac ako prostriedok odkazovania na pomenovanie vlastností, resp. charakteristiky. Výraz *to* sa zas používa frekventovanejšie v prípade odkazovania na vlastnosti pomenované prísudkovým adjektívom, substantívom či frazémami, ale aj v prípade odkazovania na stav či prechodnú vlastnosť. S anaforikom *to* sa spája jeho nevyhnutné postavenie na začiatku vety, pred ním zvyčajne stojí len častica, prípadne podrad'ovacia spojka *že*.

V nasledujúcom výskume sa I. Kolářová (2007) venovala rozlíšeniu nadväzovacej a expletívnej funkcie výrazu *to* v písaných textoch českého korpusu SYN2000. Overuje informácie o týchto funkciách známe z českých gramatík a dopĺňa ich vlastnou analýzou. Primárne sa zameriava na výraz *to* v spojení so slovesami *mrznúť, mrholiť, snežiť* a pod., sponovými výrazmi *je mi*

zima, je pekne a pod. v bezpodmetových spojeniach. Týmto výskumom dospela I. Kollárová k záveru, že pri rozlišovaní funkcie výrazu *to* v bezpodmetových konštrukciách – platnosť nadväzovacej alebo emocionálnej častice, koreferenčného výrazu alebo expletívneho výrazu – sú dôležité tri faktory: význam a vlastnosti slovesa alebo predikátového výrazu, pozícia výrazu *to* vo výpovedi a možnosť substitúcie výrazu *to* inými výrazmi. Ako autorka uviedla v závere svojho príspevku, je nevyhnutné podrobnejšie skúmanie tohto javu nielen v korpuse písaných textov, ale aj v hovorenom korpuse, prípadne porovnanie výsledkov výskumu v jednotlivých druhoch komunikácie.

V súvislosti s analýzou komunikačných funkcií výrazu *to* je dôležitá kategória koreferencie. S pojmom koreferencie sa v slovenskej lingvistickej teórii nepracuje v takom rozsahu, ako je to v českej lingvistickej teórii. Do istej miery sa tejto oblasti terminologicky dotýka v článkoch, ktoré sme aj vyššie spomenuli, E. Bajžíková.

Termín koreferencia pomenúva referenčný vzťah zhody medzi výrazmi v texte (Encyklopedický slovník češtiny, 2002, s. 233). V teórii rozoznávame koreferenciu gramatickú a textovú. Textovú koreferenciu tvorí anafora a katafora. Prostriedky na vyjadrenie anafory či katafory sú rozmanité, jednou z možností je ukazovacie zámeno *to*. Pri vymedzovaní realizácie referenčných vzťahov uvádza F. Daneš (1979) pronominalizáciu vo všeobecnosti – vrátane nadväzovania ukazovacími zámenami – hneď na treťom mieste svojej klasifikácie. Štandardne považuje anaforu a kataforu za dve zložky pronominalizácie. Ako uvádza, jav pronominalizácie je nepreskúmaným a neopísaným javom – môžeme doplniť, že nielen v hovorených, ale aj v písaných textoch.

Antecedentom v prípade odkazovania môže byť pri referende *to* akékoľvek podstatné meno, zámeno, číslovka, ale aj príslovka. V ďalšej kapitole uvádzame niektoré obmedzenia tohto tvrdenia, ktoré je zrejme všeobecnejšie platné pre písané texty. Z hľadiska konexie sa výraz *to* najčastejšie používa ako univerzálny

anaforický konektor (Kollářová, 2006, s. 249). Ako antecedent či postcedent v spontánných prejavoch často vystupuje celá vypoved', resp. rôzne druhy syntagiem, ako to vo svojej štúdií, založenej na materiáli písaných textov, rovnako interpretuje E. Bajzíkova (1980).

Z hľadiska teórie textovej koreferencie je zrejmé, že foriká – akéhokoľvek druhu vrátane výrazu *to* – sa vzťahujú len na endoforické vyjadrenie koreferencie. V hovorených prejavoch, najmä v spontánných, sa však zatiaľ ukazuje istý posun ku koreferencii v prípade antecedentu či postcedentu exofory. Zmienku o tomto jave v písaných textoch uvádza aj Encyklopedický slovník češtiny (2002, s. 41), spontánne prejavy túto tézu potvrdzujú aspoň do tej miery, do akej nám to obmedzený rozsah hovoreného korpusu umožnil.

4. Vlastný výskum výrazu *to* na materiáli hovoreného korpusu slovenčiny

Pracovná verzia hovoreného korpusu obsahuje 164 764 textových jednotiek (tokenov), výraz *to* sa v ňom nachádza 3 696-krát. Už tento materiál priniesol zaujímavé výsledky, ktoré jednak potvrdzujú niektoré vyššie uvedené charakteristiky zámena *to*, jednak dokladujú aj náznaky ďalšieho využitia výrazu *to* v hovorených prejavoch. Pokúsime sa ich bližšie opísať.

4.1. Funkcie výrazu *to*

Z trinástich funkcií zámena *to* vymedzených v MSJ (s. 267 – 269) podrobnejšie analyzujeme prvé štyri. Najskôr uvádzame funkciu podľa MSJ, potom príkaz alebo príkazy, ktorými sme vyhľadávali doklady na jednotlivé charakteristiky výrazu *to* v pracovnej verzii hovoreného korpusu.⁵

a) Odkazovanie na vytýčený vetný člen bez ohľadu na rodovú príslušnosť a gramatické číslo mena vo vytýčenom člene

```
[(tag="S.*") | (tag="A.*") | (tag="P.*") | (tag="N.*")] [(lemma="a") | (lemma="ani") | (lemma="i")] [(tag="S.*") | (tag="A.*") | (tag="P.*") | (tag="N.*")] [word=","] [] {0,3} [word="to"] – Vyhľadávali sme akékoľvek substantívum, adjektívum, zámeno alebo číslovku v koordinácii s akýmkoľvek substantívom, adjektívom, zámenom alebo číslovkou, ak za nimi nasleduje čiarka a v pozícii nula až tri akékoľvek slová vrátane výrazu to.
```

Uvedenej charakteristike zodpovedá v našom korpuse niekoľko výskytov, ide o odkazovanie na substantívum v ženskom a strednom rode:

⁵ Dokladový materiál je v podobe ortografického, resp. textového prepisu podľa zásad spoločne vypracovaných v oddelení SNK, neuvádzame však neverbálne súčasti komunikácie, príp. rôzne druhy ruchov, lapsusov a pod.

*On bol a. Ta tento, pod nami v matematickom bol, ale v inom oddelení. Ja som bola v inom oddelení, takže zase tam také **analytické a syntéza**, to mi pomohlo. [Braidy⁶]*

*No a potom, už keď som učil, tak tie **zemiakové brigády**, to bola nekonečná legenda, neviem kto ich dneska zbiera. [Prwixa]*

***Riečnica**, to je malá dedinka, ktorá toho času je už za zatopená, je tam vybudovaná priehrada, takže prakticky už teraz neexistuje moje rodisko alebo nebude existovať. [Brgirg]*

*Ale asi najviac mám najradšej mám **taliančinu**, to je taká moja láska. [Brgirg]*

b) Odkazovanie na viac vecí súčasne v spojení s vymedzovacím zámenom *všetok* a jeho tvarmi, v našom prípade *všetko*

```
[(tag="S.*") | (tag="A.*") | (tag="P.*") | (tag="N.*")] [(lemma="a") | (lemma="ani") | (lemma="i")] [(tag="S.*") | (tag="A.*") | (tag="P.*") | (tag="N.*")] [word="," ] [] {0,3} [word="všetko"]
```

Pokúsili sme sa zistiť, či sa v hovorenom korpuse nachádza vymedzovacie zámeno *všetko* pred výrazom *to*, keďže sme si vedomí pestrosti spontánnych prejavov, ale takéto použitie výrazu *to* v korpuse spontánnych prejavov sa nám nepodarilo dokázať.

c) Odkazovanie na infinitívnu väzbu a jej zdôraznenie vo vete

⁶ Nahrávky získané pri terénnom zbere materiálu sú kódované podľa kľúča, ktorý predstavuje: dátum nahrávania, prvé dve písmená názvu mesta, v ktorom bola nahrávka urobená, a počítačovým systémom vygenerovaný refazec znakov. Tým získava nahrávka jedinečný kód, s ktorým vystupuje v databáze. Dátum získavania nahrávok v našej práci neuvádzame.

[tag="VI.*"] []{0,2} [word="to"]

Po selekcii 145 nájdených výskytov zámena *to*, v ktorého blízkosti (2 slová zľava alebo sprava) sa nachádza infinitív, sme len 3 výskyty mohli zaradiť k definícii uvedenej v MSJ, pričom dva z nich neobsahujú plnohodnotne zdôraznenú infinitívnu väzbu, t. j. infinitívnu väzbu v hypotaxe, kde významovo ide o zhrnutie celého súvetia.

*A môžem vám povedať, že **mala by sa viacej pestovať** nielen slovensko č česká vzájomnosť, ale aj slovensko slovenská vzájomnosť medzi západným stredným a východným Slovenskom **ľudia by mali viac spoznávať**, to patrí takisto k tomu by som povedala k k tomu zámernému pestovaniu si toho vlasteneckého cítenia. [Brtupt]*

*Myslím si, že človek by by **mal k tomu siahat**, keď mu m kto poradiť, a **mal by sa o to usilovať**, to je môj názor aj keď s by si mohol zachovať aj keď by si mohol zachovať to svoje nárečie. [Brtupt]*

*Na **jednej strane sme počuli**, že vraj víťaz architektonickej súťaže nemá zaručené právo, **že to bude kresliť**, to odznelo v úvode, a prednávanom sme počuli, že ako by sa cítil niekto v koži víťaza po prvom a druhom kole, keby to dostal kresliť niekto iný. [Podlampou⁷]*

d) Pripájanie prístavku v citovo nadnesených prejavoch

[tag="S.*"] [word="to"]

V MSJ uvedené príklady (*Pažiť rozrývali dlhé trhliny, stopy **to** granátov*

⁷ Prepisy televíznych relácií boli označené názvom jednotlivých relácií, nie sú osobitne kódované.

[Letz]; *Na rukách nemal tvrdých mozol'ov, známky to vytrvalej roboty. [Kukučín]*) sú skôr štylistickou figúrou charakteristickou pre písané texty. V spontánných prejavoch sme nenašli takéto uplatnenie, keďže v plynulom prúde reči hovoriaci zvyčajne nedbá na používanie ozdobných figúr. Spontánne prejavy sú charakteristické skôr lapsusmi, ktoré vznikajú v procese komunikácie pri striedaní myšlienok a tém. Použitie uvedené v MSJ by sme teda mohli nájsť v spontánných prejavoch len v prípade, žeby v prúde reči došlo k neplánovanej prešmyčke, resp. lapsusu, ktorý by sa zhodou okolností vyskytol práve na takom mieste, ako je pozične uvedené v tomto prípade v MSJ. Hovorený korpus síce obsahuje 128 výskytov zámena *to* pozične umiestneného podľa príkazu uvedeného vyššie, ale pri ručnej selekcii sme zistili, že ani jeden výskyt nie je príkladom použitia výrazu *to* podľa charakteristiky d) v MSJ.

Ďalšie funkcie výrazu *to* uvedené v MSJ osobitne nedokladujeme, keďže niektoré z nich sú podobne ako v predchádzajúcom prípade typické len pre písané texty. K iným uvádzame doklady v časti o kontextovej nadväznosti (anafora, katafora, exofora).

Vlastnou analýzou takmer 4 000 výskytov výrazu *to* v hovorenom korpuse sme zistili, že výraz *to* sa v spontánných prejavoch používa ako atribút na zdôraznenie nadradeného vetného člena. Do tejto skupiny zaraďujeme spojenia s neutrom, ktoré je očakávané po výraze *to* vystupujúcom vo funkcii určitého člena neutra. V týchto postaveniach plní výraz *to* podobnú funkciu ako určitý člen *der, die, das* v nemčine či určitý člen *the* v angličtine. V jazykových prejavoch máme veľa konštrukcií, ktoré sú zdôrazňované týmto spôsobom, ako dokazuje aj 32 732 výskytov spojenia výrazu *to* s neutrom v 319-miliónovom korpuse písaných textov prim-3.0-public-sane⁸. Pracovná verzia hovoreného korpusu obsahuje takýchto bezprostredne vedľa seba stojacich spojení 89.

⁸ Všetky ďalšie porovnávacie analýzy na materiáli písaného korpusu realizujeme v tejto verzii.

*To bolo tak, že zastavil ho, zastavil ten poľský policajt a on stiahol **to okienko** a teraz ten poľský policajt mu povedal: „Pšprasam, pan nema pneumatik na samochudže. [7s.r.o.]*

*A keď ten mŕtvy nepríde a vlastnoručne nepodpíše rozviazanie kontraktu s operátorom ani nezavolá na **to** spomínané číslo do dvoch týždňov, čo nasleduje? [7s.r.o.]*

*A to, čo sa deje, je teraz, že vlastne verejnosť, my všetci, čakáme, aká odpoveď tam vznikne na **to** kultúrne barbarstvo, ktoré tam vzniklo za totality. [Podlampou]*

*No a my sme si takúto masu historického jadra zbúrali úplne dobrovoľne, sčasti v rámci výstavby mosta, ktorý rozčesol **to** historické **jadro** až po Suché mýto, čiže celé Podhradie je Suché mýto a sčasti teda iba z takého, z takej samopaše, lebo samozrejme, že Zuckerman del je tak ďaleko od mosta, že ten sa búrať nemusel. [Podlampou]*

*Ae myslím že vo veľa prípadoch je to dosť ťažké takže prípadne aj prijať **to** cudzie slovo. [Brgirg]*

*..., pretože je jasné, že ide o vec, aj keď je teda **to slovo** zaužívané ako osoba teda alebo tuto dvoch Oskarov, pretože je to osoba a nie dva Oskare, áno čiže vychádza sa z toho významu z toho základu slova. [Brgirg]*

*No nie, v ateliéri nie, ale vieš, keď si priamo na mieste a teraz naozaj tam máš **to bábätko**, ktoré má kilo osemdesiat, môžem ti povedať, máš dosť. [Nastreche]*

*Aj keď by si mohol zachovať **to** svoje nárečie. [Brtupt]*

*..., pretože ten jazyk vtáka poznáš po perí, poznáte **to** porekadlo, rozhodne. [Brtupt]*

V našej sondážnej analýze výrazu *to* v korpuse hovoreného jazyka (Gajdošová, 2008) sme sa zmienili o osobitnom výskyte výrazu *to* pred substantívami mužského a ženského rodu. Pracovný korpus diplomového výskumu je štyrikrát väčší ako sondážny korpus, no výskyty spojení výrazu *to* v postavení určitého člena a maskulína či feminína sa vo väčšom materiáli dosiaľ nepotvrdili. V predchádzajúcom korpuse išlo o výskyty týchto spojení len v nahrávkach a prepisoch z východného Slovenska, kde je príznačné používanie nielen výrazu *to*, ale aj *ta* či *tie*. Ide teda o lokálny nárečový prvok.

Ako osobitný druh zdôrazňovania by sme mohli vyčleniť anaforické nadväzovanie, pri ktorom výraz *to* odkazuje na skutočnosť známu z predchádzajúceho textu, pričom zastupuje nominalizáciu predikátu predchádzajúcej vety:

*Ak sú výsledky volieb známe ešte pred voľbami, tak **to** zaručuje istý stupeň kultivovanosti. [7s.r.o] – [známosť výsledkov volieb]*

*No ja neviem, či ste si všimli, ale pani Belousovová si dala **upraviť aj vlasy**. Si, si dala teda tak trošku ako prestrihať odborníkom a ja si myslím, že **to** isté by mala urobiť aj so svojimi vyhláseniami. [7s.r.o.] - [upravenie vlasov]*

Anaforické nadväzovanie na konkrétne slovo alebo na celú výpoveď je v spontánnych prejavoch značne frekventované.

Dozorná rada sociálnej poisťovne, ktorej predsedníčkou je

*ministerka Tomanová, pôvodne navrhla, aby generálny riaditeľ poisťovne Ivan Bernátek **dostal** ročnú **odmenu** vo výške 788 000 korún. Lenže keď **to** médiá zverejnili, Viera Tomanová oznámila, že ona ako ministerka návrh na takú odmenu nepredloží. [7s.r.o.]*

*Ale, keď ja si spomínam. My **sme sa** tiež **učili samé nezmysly**. Čo **sme sa** my všetko naučili? Akurát na lúštenie križoviek **sa to** hodí. [7s.r.o.]*

*V Bratislave by som povedala, že rôzne vrstvy rozlične hovoria a myslím si, že **to bratislavské nárečie** si skôr **zachovali** starší ľudia, ktorí tu ktorí sú Bratislavčanmi a ktorí tu dlhšie žijú, napríklad naša pani upratovačka, ktorá nám tu upratovala, tá má asi sedemdesiat rokov nie šesťdesiatpäť rokov, tak na nej som si **to** všimla, pretože ja som nemala až takú veľkú možnosť s s stýkať sa s rodenými Bratislavčanmi. [Brtuť]*

***Moderátor v televízii alebo konferenciér, ten by sa nemal dopustiť takýchto chýb**, nemal by sa toho dopustiť, bolo **to** kritizované napríklad keď bola voľba Miss Slovensko, tam bol nejaký konfar konferenciér a my sme boli radi, že **to** niekto kritizoval, v Smene bol na to nejaký ohlas túto jazykovú stránku a podľa mňa to bolo veľmi správne. [Brtuť]*

Funkciu anaforika môžeme priradiť výrazu *to* aj v prípadoch, keď neodkazuje na žiadne konkrétne realizované slovo z predchádzajúceho prehovoru, avšak v konečnom dôsledku a pri dôraze na kontext, v ktorom je použité, je zrejmé, ktorá konštrukcia je antecedentom. Nejde len o prísne kontextové nadväzovanie, ale o endoforu celých výpovedných konštrukcií do jediného výrazu:

*Však keď je to ACER alebo ten ASUS, že jednoducho, my vám dáme preinštalovať systém celý, navštívime stránku a ono si stiahne pro potrebné na inštaláciu všetkých. Uložíme mu **to** a keď bude chcieť si preinštalovať. [Prtweo]*

Tradične je v spontánných prejavoch prítomné exoforické nadväzovanie, keď je zámenom *to* označovaný práve kontextovo zasahovaný predmet alebo vykonávaná činnosť, prípadne ide o naznačenie gestom, jazykovo vyjadrené výrazom *to*.

*Inak pred týždňom sme tu spomínali, čo sa dialo na železničnej stanici Poprad – Tatry, ak si ešte pamätáte. M: Áno, tomu som rozumel. V: No a teraz vidíš sa **to** isté zopakovalo aj na strednej škole v Novom Meste n/Váhom. [7s.r.o.]*

*Dupli na **to**. (smiech) Dupli na **to** a tí policajti v jednej zákrute im vystrelili z cesty. [7s.r.o.]*

Kataforické nadväzovanie je v spontánnej reči zriedkavejšie v prípade jedného slova a častejšie v prípadoch katafory celých výpovedných celkov.

*Som **to** nedávno čítal takú (smiech) viem ktorá (smiech) takú **propagáciu** v jednom časopise. Že chcete vychádzať dobre so svojim šéfom? Chcete byť v kolektíve obľúbená? Chcete postupovať rýchlo v zamestnaní? A iné tajomstvá orálneho sexu. [7s.r.o.]*

*A **to** nie je pekné od pani ministerky, že tomu **dotyčnému závidí odmenu**, ktorú mu sama chcela udeliť. [7s.r.o.]*

Vlastne aj ja som sa to ešte učila to používanie predpony s a z alebo so a zo zostúpiť. [Brgirg]

4.2. Výraz *to* z formálneho hľadiska

Pomocou príkazu *[word="to"]; frekvenčná distribúcia lemma=1<KWIC* sme v pracovnej verzii hovoreného korpusu z hľadiska frekvenčnej distribúcie zistili, ktoré slová nasledujú bezprostredne za výrazom *to*. Môžeme konštatovať, že z pravej strany sa so skúmaným výrazom najčastejšie spája sloveso *byť* (768 výskytov), ďalej sloveso *mať* (84 výskytov), výrazy *tak* (62 výskytov), *už* (55 výskytov), *čo* (50 výskytov) a *sa* (45 výskytov).⁹ Z uvedeného je zrejmé, že najčastejšou je práve konštrukcia, kde je výraz *to* v postavení subjektu vety.

Ľavostranné kolokáty (*[word="to"]; frekvenčná distribúcia lemma=-1<KWIC*) predstavujú v najvyšších frekvenciách opäť sloveso *byť* (674 výskytov), ďalej výrazy *a* (166 výskytov), *že* (153 výskytov), predložka *na* (151 výskytov) a *pod*.¹⁰

Z oboch strán výrazu *to* sú najfrekventovanejšie spojenia s tvarmi slovesa *byť*, často je *to* členom koordinácie (spojka *a*), prípadne časťou zloženej spojky (*a to*), rovnako stojí ako vetný člen, ktorý býva rozvítený atributívnou hypotaxou so spojkou *že*.

Ak sa výraz *to* nachádza na začiatku vety (*[word=". "] [word="To "]; frekvenčná distribúcia lemma=1<KWIC*; 219 výskytov) najčastejšie sa spája s neplnovýznamovým slovesom *byť* (92 výskytov), potom so slovesom *znamenat'* (17 výskytov). Podobná situácia je aj v písaných textoch, kde sa výraz *to* v tomto postavení spája s rovnakými slovesami ako v našom pracovnom hovorenom

⁹ Výskyty po hranicu 10 uvádzame v prílohe č. 4.

¹⁰ Výskyty po hranicu 10 uvádzame v prílohe č. 5.

korpusu.

Zo všetkých možných začiatkov výpovedí (`[word="\."] [tag=".*" & lemma="([ABCDEFGHJKLMNOPRSTUVZ]).*"]`, `N filter:[lemma=":"]` $l < KWIC$) sa výraz *to* nachádza na treťom mieste frekvenčnej distribúcie s frekvenciou 223, predchádzajú ho len výraz *a* a výraz *no*¹¹. V písaných textoch sa výraz *to* ako začiatok výpovede nachádza až na 17. mieste.

Ak je na začiatku vety spojka, výraz *to* za ňou nasleduje v našom pracovnom korpusu 38-krát, pričom najfrekventovanejšou spojkou je prirodzene spojka *a* (16 výskytov), ďalšou spojkou *takže* (8 výskytov), potom *ale* (4 výskyty) a výrazy *no* a *lebo* (2 výskyty).

A to je pre nás to najdôležitejšie, že historizujúcu architektúru ja som zatiaľ nevidel. [Podlampou]

A to sú tie magické miesta. [Podlampou]

A to je chyba. [Hatwie]

A to je problém veľký. [Hatwie]

A to je jedno. [Prcany]

Pomocou príkazu `[word="to"]` `[word="\."]` `[word="([AÁBCČDĎEÉFGHIÍJKLLMNŇOÓPRSŠTŤUÚVZŽ]).*"]` sme hľadali výraz *to* v postavení na konci vety. Za príznakové pokladáme postavenie výrazu *to* na konci uzavretých viet. Vo všetkých 34 prípadoch ide o krátke vety, napríklad:

¹¹ Vzhľadom na to, že v automatizovanej lematizácii sa nerozlišujú zaradenie homonymných tvarov do konkrétnych slovných druhov, používame všeobecné pomenovanie výraz.

Stálo to za to. [7s.r.o.]

A zhaslo to. [7s.r.o.]

Dupli na to. [7s.r.o.]

Takto je to. [7s.r.o.]

Nemyslím si to. [Podlampou]

No ved' to. [Nastreche]

Spojenie *to* + tvar slovesa *byť* ([*word="to"*] [*lemma="byť"*] & [*tag="V.*"*]); resp. ([*word="to"*] [*lemma="nebyť"*] & [*tag="V.*"*]) je frekventované aj s časticami, ktoré v spontánných prejavoch stoja bezprostredne vedľa nich.¹²

Osobitne nás zaujalo spojenie s výrazom *tak*, ktoré je bohato zastúpené v korpuse písaných textov. V spontánných prejavoch však nie je natoľko frekventované (7 výskytov), ako by sme predpokladali. Pôsobí skôr ako nejasné, zahmlievajúce vyjadrenie myšlienky.

*Mať také základy. Tiež neviem, či bola cena. Tam asi nebola. Dvestopäťdesiat. Nestálo toto sedemsto korún? Tam vraj **to** je tak. Teraz ty sa schovaj. Mňa nič nestálo. Určite to nestálo sedemsto korún, stoja ja neviem. [Prfrid]*

Tak to je tak to je. [Prwixa]

V rámci súvetných konštrukcií sa výraz *to* najčastejšie nachádza v postavení nadradeného vetného člena, ktorý je rozvíjaný vedľajšou vetou s podradovacou spojku. Takto hodnotíme uvedené postavenie výrazu *to* na

¹² Najfrekventovanejšie výrazy uvádzame v prílohe.

základe funkčno-generatívneho prístupu uplatneného pri syntaktickej anotácii Pražského závislostného korpusu a Slovenského závislostného korpusu. V slovenskej syntaxi sa takéto postavenie výrazu *to* tradične hodnotí ako *t*-ové odkazovacie zámeno. Najčastejšie za ním nasleduje spojka *že* (77 výskytov), korefenčne sa viaže na antecedent *to* aj spojka *čo* v podrad'ovacích súvetiach (43 výskytov). Nasleduje spojka *aby* (24 výskytov) a lemma *to* (10 výskytov).

Najfrekventovanejším substantívom v spojení *to* + tvar slovesa *byť* + substantívum je substantívum *otázka* (5 výskytov), s odchýlkou jedného ľubovoľného miesta je to všeobecnejšie substantívum *vec* (6 výskytov), *názor* (5 výskytov), *problém* (4 výskytov).

V prípade opačného poradia konštrukcie, t. j. tvar slovesa *byť* + výraz *to* + substantívum, ide o substantívum *pravda* (8 výskytov), ostatné sú iba v jednotlivých výskytach. S odchýlkou jednej pozície je to opäť substantívum *problém* (4 výskytov). Uvedené substantíva sú vhl'adom do všeobecnosti ľudskej komunikácie: riešenie *problémov* či hodnota *pravdy*, kladenie *otázok*, zaujatie stanoviska k nejakej *veci*.

Spojenie *to* + *všetko* (osobitne vyčlenené aj v MSJ) sa v hovorenom korpuse vyskytuje 12-krát (*[word="to"] [word="všetko"]*). Je veľmi frekventované v písaných textoch (15 330 výskytov). Nazdávame sa, že nejde výsostne len o jav prítomný v písaných textoch, ale aj o prvok spontánnych prejavov.

Áno, je to všetko kvôli tebe. [Nastreche]

Tak, veď nie je dobré keď to všetko poslúcha? [7s.r.o.]

To bolo pár študentov, ktorí mali tuná počítač a proste tam to všetko zapisovali. [Peyasp]

Spojenie *všetko* + *to* ([word="všetko"] [word="to"]) sa v korpuse písaných textov nenachádza ani raz, pričom v našej malej pracovnej verzii hovoreného korpusu sa táto konštrukcia objavila hneď 5-krát.

*Je tu rozpísaný každý jednotlivý dom, je tu urobené kompletne **všetko to**, o čom hovoríme, včítane jeho, nazvime to, historizujúcej hodnoty. [Podlampou]*

*To organizoval v tomto čo sa týka tejto registrácie a **všetko to** tak to organizoval študentský parlament. [Peyasp]*

Výraz *to* považujeme v prvom prípade za zdôrazňujúci, v druhom prípade za nadradenejšiu časť spojenia, na ktorej je zväčša aj vetný dôraz.

Podobným prípadom je aj konštrukcia *ono* + *to* ([word="ono"] [word="to"]); 3 výskyty), v ktorej ide rovnako o zdôrazňovaciu funkciu viac v prospech výrazu *to*, pričom výraz *ono* slúži ako podporná časť.

*Ale **ono to** sedí aj krížom. [7s.r.o.]*

*Viete, **ono to** nie je o jednotlivých budovách, pán Huba povedal jednu vec. [Podlampou]*

*A keď sme pri tomto prechyľovaní, aj včera sme hovorili o väzení a **ono to** väzenie tam svojím spôsobom v pokuse o recenziu je. Čo vy na na ženský tvar od mužského väžňa? [Brgirg]*

Spojenie *ono to* sa v písanom korpuse vyskytuje 251-krát, v našej pracovnej verzii hovoreného korpusu sa vyskytuje len trikrát. Uvedené spojenie má funkciu neoddeliteľnej zložky, ktorá vo vete stojí pevne spätá zväčša s

predikátom. Ani jednu zložku z tohto spojenia nemôžeme nazvať atribútom druhej. V písaných textoch má toto spojenie tendenciu spájať sa nielen s neplnovýznamovým slovesom *byť*, ale aj s plnovýznamovými slovesami *vypáliť*, *fungovať*, *dopadnúť*, *znieť*, *vychádzať*, *prejsť*, *vyjsť*, *prísť* atď.

V hovorenom korpuse je takýto výskyt zatiaľ minimálny, spojenie sa vyskytuje v kolokáciách so slovesom *byť* a *sediť*.

Ale ono to sedí aj krížom. [7s.r.o.]

*Ono to väzenie tam svojím spôsobom v pokuse o recenziu je.
[Brgirg]*

Viete, ono to nie je o jednotlivých budovách. [Podlampou]

[word="to"] [word="ono"]

Konštrukcia *to + ono* je značne frekventovaná v korpuse písaných textov (206 výskytov), najmä v spojení s neplnovýznamovým slovesom *byť*, keďže ide o vety typu *je to ono* a spojenia so slovesom *byť* v iných tvaroch. V hovorenom korpuse sme takúto konštrukciu nenašli.

V spontánnych prejavoch nechýba ani použitie výrazu *to* v časticovej funkcii. Zväčša ide tak, ako to opisuje MSJ a sčasti KSSJ, o postavenie na začiatku vety:

To o akých problémoch to hovoríte? [7s.r.o.]

*To proste budú fasády, za ktorými budú nové domy, hej?
[Podlampou]*

To večne nejaká skupinka pri dverách, aj keď zvonilo už, keď idem po chodbe. [Braidy]

Výraz *to* v časticovom použití vystupuje v niektorých prípadoch ako synonymum častice *teda*, avšak v niektorých spojeniach môže mať viacero interpretácií. V uvedenom príklade možno uvažovať napríklad o anaforickom nadväzovaní či elipse.

... *to* by bolo dostať nejaké základy doma a potom sa venovať
tomu jazyku *to* už proste budovať na tých základoch ktoré
dostane doma. [Brgirg]

Keďže v spontánných prejavoch je naozaj veľká variabilita, skúsili sme sa pozrieť na podrad'ovacie súvetia so spojku *že* v takomto postavení: sloveso v akomkoľvek tvare + výraz *to* + , + podrad'ovacia spojka *že* ([tag="V.*"] [word="to"] [word="\,"] [lemma="že"]). Tento druh konštrukcie sme konfrontovali s významovo zväčša totožnou konštrukciou sloveso v akomkoľvek tvare + , + podrad'ovacia spojka *že* ([tag="V.*"] [word="\,"] [lemma="že"]). Inšpirovali sme sa čiastočne výskumom F. Štíchu na písaných textoch korpusu SYN2000, ktorý sme spomínali vyššie.

Náš pracovný korpus je značne malý na to, aby sa v ňom mohol ukázať kompletný materiál. Po porovnaní jednotlivých výskytov sloviess s výrazom *to* a bez neho sme zistili, že hovorený korpus ponúka len niekoľko rovnako frekventovaných sloviess (*byť*, *myslieť*, *povedať*, *znamenáť*, *ubezpečiť*). Rovnakým spôsobom sme porovnali tento výskyt v korpuse písaných textov a zistili sme, že slovesá, ktoré sú zatiaľ frekvenčne rovnako umiestnené v našom hovorenom korpuse¹³, majú v korpuse písaných textov výskyt: *byť* (2), *myslieť* (23), *povedať* (29), *znamenáť* (1), *ubezpečiť* (v písanom korpuse sa nenachádza). Keďže spontánne prejavy sú charakteristické vysokou mierou uvoľnenosti, je prirodzené, že sa neplnovýznamové sloveso *byť* nachádza v hovorenom korpuse na najvyššej frekvenčnej úrovni s hodnotou 11 výskytov. Písané texty, najmä publicistické texty, obsahujú konštrukciu *znamená to*, teda frekvenčné prvenstvo

¹³ 72 totožných výskytov po frekvenčnú hranicu 10.

slovesa *znamenať* je primerané. Ostatné slovesá je v tejto chvíli ťažšie analyzovať, keďže materiál hovoreného korpusu je zatiaľ malého rozsahu. Komplexnejšie závery môže priniesť väčší materiál z nového korpusu.

Záver

V diplomovej práci sme sa pokúsili o analýzu výrazu *to* (iba v tomto tvare) na materiáli pracovnej verzie hovoreného korpusu, ktorý vznikol s finančnou podporou spoločnosti Konto Orange, n. f., v grantovom programe *Je ti tvoja škola malá?* v akademickom roku 2007/2008 a s podporou oddelenia Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV v Bratislave.

Jadrom skúmaného materiálu sa stali nahrávky spontánnych prehovorov respondentov z rôznych častí Slovenska, z rôznych vekových a vzdelanostných kategórií, na rozličné témy. Zo zhromaždeného materiálu sme v spolupráci s programátormi Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV v Bratislave vytvorili pracovnú verziu hovoreného korpusu slovenčiny. Materiál v rozsahu približne 164 000 tokenov (textových jednotiek) obsahoval výraz *to* temer 4 000-krát. Všetok materiál sme analyzovali ručne a pomocou príslušných počítačových nástrojov.

Zámeno *to* patrí do skupiny odkazovacích aj ukazovacích zámen. Radí sa k zámenám, ktoré nie sú parciálne viazané len na naplnenie gramatickej kongruencie realizovanej v morfolologickej zhode v rode, čísle a páde, ale je disponované aj na deiktickú funkciu spätosti s adverbiálnymi prvkami. Do tejto druhej skupiny zaraďujeme vlastnosti výrazu *to* podľa bodov i), l) a m) (okolnostné, resp. adverbiálne príznaky) z charakteristiky v Morfológii slovenského jazyka (1966). Materiálový výskum potvrdil, že medzi funkcie výrazu *to* patrí funkcia vymedzená v Morfológii slovenského jazyka (1966) ako prvá – odkazovanie na vytýčený vetný člen bez ohľadu na jeho rodovú príslušnosť a gramatické číslo. Hoci by sme mohli predpokladať, že dokladov na túto funkciu bude v spontánnych prejavoch mnoho, nie je to tak. Niekoľko prípadov tohto druhu odkazovania sa však zistilo. Ďalšia charakteristika – odkazovanie na viac vecí súčasne v spojení s vymedzovacím zámenom *všetok* (v tvare *všetko*) sa však v spontánnych prejavoch nepotvrdila vôbec, dokonca ani v

inverznom poradí. Môžeme tu konštatovať istý rozdiel v písaných a hovorených prejavoch, ktorý zrejme vyplýva z podstaty spontánnych prejavov. V nich sa jednotliví hovoriaci nekontrolujú tak, ako je to pri formulovaní (a prípadnom spätnom čítaní) písaných textov a súčasne im nejde primárne o upravenosť, hladkú syntaktickú štruktúrovanosť prejavu. V súvislosti s touto charakteristikou sa vynára aj ďalší dôvod – hovoriaci pri vymenúvaní nezaznamenávajú vo svojej pamäti všetky prvky konkrétnej koordinácie, a preto v podstate ani nie je možné, aby na ne v prúde reči odkazovali týmto spôsobom. Odkazovanie na niekoľko prvkov v koordinácii sa častejšie deje pomocou ďalšieho opakovania konkrétneho výrazu zväčša doslova.

V Morfológii slovenského jazyka (1966) sa vyčleňuje aj funkcia odkazovania výrazu *to* na infinitívnu väzbu, pričom výraz *to* plní funkciu zdôraznenia. V pracovnom hovorenom korpuse sa takéto použitie zistilo len v troch prípadoch, hoci nie vždy išlo o naplnenie danej funkcie stopercentne. Ide skôr o odkazovanie na infinitívne väzby v rámci podrad'ovacej vety, teda o zhrnutie celého súvetia.

Za najfrekvencovanejšiu funkciu výrazu *to* považujeme jeho funkciu zdôrazňovania pred substantívami stredného rodu. Ide o prípad analogický s určitými členmi v nemeckom či anglickom jazyku (*der/die/das* či *the*).

Ako nosná sa aj v hovorených prejavoch potvrdila funkcia výrazu *to* ako anaforika, ktoré nadväzuje či odkazuje nielen na antecedent vyjadrený jedným slovom, ale aj na antecedent vyjadrený celou výpoveďou či syntagmou. Podobné závery na sondážnom materiáli hovorených prejavov konštatovala aj E. Bajžíková vo svojom článku z roku 1973, ktorý sme dostali k dispozícii v záverečnom štádiu písania práce a potvrdili sme ním naše zistenia.

Okrem týchto vlastností sa výraz *to* uplatňuje aj pri kataforickom odkazovaní na širší kontext, ide napr. o vyrozprávanie nejakej udalosti zachytenej vo viacerých výpovediach, na ktoré sa v ďalšom texte nadväzuje len použitím zámena *to* bez akéhokoľvek ďalšieho vysvetľovania. V niektorých prípadoch – najmä pri nadväznosti na citové výpovede – slúži zámeno *to* ako náhrada

expresívne vyjadrenej udalosti.

Materiál hovorených prejavov sme sa pokúsili podrobiť aj formálnej analýze, najmä z hľadiska distribúcie, pričom sme využili znalosť regulárnych výrazov a postupov pri vyhľadávaní v korpuse písaných textov. Za zaujímavé považujeme napríklad zistenie, že medzi najfrekvencovanejšie substantíva, ktoré sa spájajú s výrazom *to* po slovese *byť* patria feminína *otázka*, *vec* a maskulína *názor* a *problém*. Aj z toho sa dá usudzovať, akým smerom sa uberá naša komunikácia, a konfrontovať túto distribúciu s budúcim väčším a inak štruktúrovaným hovoreným korpusom.

Na výraz *to* môžeme nazerať z viacerých strán – z hľadiska formálneho, textového, kontextového, gramatického, sémantického, pragmatického a pod. V predkladanej práci sme sa pokúsili zhrnúť prvé tri stránky, aj keď iste nie úplne vyčerpávajúco. Rozšírenie poznatkov môže priniesť jednak širší materiál Slovenského hovoreného korpusu, ktorý sa intenzívne buduje v oddelení Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV v Bratislave, jednak analýza ďalších aspektov fungovania výrazu *to* v komunikácii, osobitne napríklad v aktuálnom členení výpovede. To je námiet na ďalšiu samostatnú prácu, ktorá však presahuje rozsah a možnosti diplomovej práce. Výskum výrazu *to* by sme chceli perspektívne zamerať práve na jeho uplatnenie v tematicko-rématickej štruktúre spontánnych prehovorov.

Literatúra

BAJZÍKOVÁ, Eugénia: O daktorých konektoroch v hovorových textoch. In: Slovenská reč, 1973, roč. 38, č. 3, s. 154 – 162.

BAJZÍKOVÁ, Eugénia: Ukazovacie zámená v textovej funkcii. In: Zborník Filozofickej fakulty Univerzity Komenského. Bratislava: Slovenské pedagogické nakladateľstvo 1979, s. 19 – 24.

BAJZÍKOVÁ, Eugénia: Sémantika ukazovacích zámen v textovej syntaxi. In: Jazykovedné štúdie XVI. Materiály z vedeckej konferencie o sémantike, (Smolenice, 14. – 17. januára 1980). Bratislava: Veda 1980, s. 123 – 126.

DANEŠ, František: O identifikaci známé (kontextově zapojené) informace v textu. In: Slovo a slovesnost, roč. 15, č. 49, 1979, s. 257 – 270.

DUDOK, Miroslav: Deiktické identifikátory v jazyku a komunikácii. In: Jazyk a komunikácia v súvislostiach II. Ed. O. Orgoňová. Bratislava: Univerzita Komenského 2007, s. 341 – 353.

Encyklopedický slovník češtiny. Praha: Nakladatelství Lidové noviny 2002, s. 233 – 234, s. 41 – 42.

FEČKOVÁ, Svetlana – SLANČOVÁ, Daniela: K niektorým východiskám výskumu a modelovania rečového vývinu dieťaťa. In: Logopaedica. 4. Zborník Slovenskej asociácie logopédov. Bratislava: Liečreh Gúth 2001, s. 38 – 45.

GAJDOŠOVÁ, Katarína: Metadáta v slovenskom hovorenom korpuse. In: Varia XVII – Materiály z kolokvia mladých jazykovedcov, 2007. (V tlači.)

GAJDOŠOVÁ, Katarína: Analýza výrazu *to* v pracovnej verzii hovoreného korpusu slovenčiny. (Príspevok bol súčasťou Študentskej vedeckej konferencie organizovanej v roku 2008 Katedrou slovenského jazyka FiF UK v Bratislave.) Rkp. 12 s.

Hovorená podoba spisovnej slovenčiny I., II. Referáty a diskusné príspevky z konferencie dňa 6. – 9. októbra 1965.

Hovorená podoba spisovnej slovenčiny. Sociolinguistica Slovaca 6. Ed. S. Ondrejovič. Bratislava: Veda 2007, s. 9 – 17.

CHABAĽOVÁ, J. – SLANČOVÁ, D.: Reč prešovských vysokoškolákov a jej kontexty. In: Text a kontext. Zborník z medzinárodnej vedeckej konferencie. Text v priestore jazykovej komunikácie. Text v priestore literárnej komunikácie. Text v priestore didaktickej komunikácie. Prešov 18. – 19. novembra 1993, s. 105 – 116.

KOLÁŘOVÁ, Ivana: Významy slova *to* v navazovací platnosti v textech Českého národního korpusu. In: Korpus jako zdroj dat o češtině. Ed. P. Karlík. Brno: Masarykova univerzita v Brně 2004, s. 165 – 174.

KOLÁŘOVÁ, Ivana: Možnosti odkazování k jazykovému vyjádření vlastnosti v češtině zájmeny *TO* a *TAKOVÝ*. In: Možnosti a meze české gramatiky. Ed. F. Štícha. Praha: Academia 2006, s. 249 – 267.

KOLÁŘOVÁ, Ivana: Kritéria a možnosti rozlišení navazovací a expletivní platnosti slova *to* (*V hodinách to hrklo – To hrklo v hodinách*). In: Gramatika a korpus. Ed. F. Štícha – J. Šimandl. Praha: Ústav pro jazyk český 2007, s. 143 – 151.

Krátky slovník slovenského jazyka. 4., doplnené a upravené vydanie. Red.: J. Kačala – M. Pisárčiková – M. Považaj. Bratislava: Veda 2003. 985 s.

Mluvnice češtiny (2). Tvarosloví. Praha: Academie 1986. 536 s.

Morfológia slovenského jazyka. Bratislava: Vydavateľstvo Slovenskej akadémie vied 1966. 896 s.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Bardejove. In: Nové obzory. Spoločenskovedný zborník východného Slovenska. 20. Red. I. Michnovič. Košice: Východoslovenské vydavateľstvo 1978a, s. 319 – 336.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Prešove. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 13., zv. 3. Slavistika. Red. M. Novák et al. Bratislava: Slovenské pedagogické nakladateľstvo 1978b, s. 37 – 55.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Michalovciach. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J.

Šafárika v Košiciach, roč. 14., zv. 3. Slavistika. Red. M. Novák et al. Bratislava: Slovenské pedagogické nakladateľstvo 1979, s. 111 – 129.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Humennom. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach. Red. M. Novák et al. Bratislava: Slovenské pedagogické nakladateľstvo 1980, s. 167 – 187.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Trebišove. In: Nové obzory. Spoločenskovedný zborník východného Slovenska. 24. Red. I. Michnovič. Košice: Východoslovenské vydavateľstvo 1982, s. 217 – 235.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Košiciach. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 17., zv. 3. Slavistika. Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1983, s. 83 – 105.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Levoči. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 19., zv. 3. Slavistika. Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1984a, s. 131 – 160.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Poprade. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 18., zv. 3. Slavistika. Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1984b, s. 180 – 202.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Krompachoch. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 20., zv. 3. Slavistika. (Jazykoveda.) Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1986, s. 106 – 146.

MURÁNSKY, Jozef: Osobitosti hovorenej podoby spisovnej slovenčiny v Lipanoch. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 22., zv. 3. Slavistika. Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1988, s. 157 – 181.

MURÁNSKY, Jozef: Osobitosti hovorevej podoby spisovnej slovenčiny v Stropkove. In: Zborník Pedagogickej fakulty v Prešove Univerzity P. J. Šafárika v Košiciach, roč. 24., zv. 3. Slavistika (jazykoveda). Red. J. Muránsky et al. Bratislava: Slovenské pedagogické nakladateľstvo 1990, s. 117 – 142.

ONDREJOVIČ, Slavomír: Bratislava versus stredné Slovensko I (sociolingvistický pohľad). In: Všeobecné a špecifické otázky jazykovej komunikácie. Komunikáty z vedeckej konferencie konanej v Banskej Bystrici 3. – 5. septembra 1991. 2. diel. Red. P. Odaloš – V. Patráš. Banská Bystrica: Pedagogická fakulta 1991a, s. 226 – 232.

ONDREJOVIČ, Slavomír: K sociolingvistike na Slovensku. In: Studia Academica Slovaca. 20. Prednášky XXVII. letného seminára slovenského jazyka a kultúry. Red. J. Mistrík. Bratislava: Alfa 1991b, s. 177 – 191.

ONDREJOVIČ, Slavomír: Sociolingvistické aspirácie slovenskej jazykovedy. In: Slovenské pohľady, 1994, roč. 114, č. 4, s. 127 – 130.

ONDREJOVIČ, Slavomír: Sociolingvistické aspekty jazykovedného výskumu na Slovensku. In: Sociolingvistické aspekty výskumu súčasnej slovenčiny. Sociolinguistica Slovaca. 1. Red. S. Ondrejovič – M. Šimková. Bratislava: Veda 1995, s. 9 – 13.

ONDREJOVIČ, Slavomír: Z výskumu jazykovej situácie na južnom Slovensku. In: Sociolingvistika a areálová lingvistika. Sociolinguistica Slovaca. 2. Red. S. Ondrejovič. Bratislava: Veda 1996a, s. 141 – 147.

ONDREJOVIČ, Slavomír: Metamorfózy sociolingvistického výskumu na Slovensku. In: Sociolingvistika a areálová lingvistika. Sociolinguistica Slovaca. 2. Red. S. Ondrejovič. Bratislava: Veda 1996b, s. 11 – 17.

ONDREJOVIČ, Slavomír: Výskum hovorenej podoby spisovnej slovenčiny po 40 rokoch. In: Hovorená podoba spisovnej slovenčiny. Sociolinguistica Slovaca 6. Ed. S. Ondrejovič. Bratislava: Veda 2007. s. 9 – 17.

ORAVEC, Ján – BAJZÍKOVÁ, Eugénia: Súčasný slovenský spisovný jazyk. Syntax. Bratislava: Slovenské pedagogické nakladateľstvo 1982. 261 s.

PATRÁŠ, Vladimír: Exponovanie komunikačných noriem v hovorenej

podobe slovenčiny v meste. In: Všeobecné a špecifické otázky jazykovej komunikácie. Komunikáty z vedeckej konferencie konanej v Banskej Bystrici 3. – 5. septembra 1991. 2. diel. Red. P. Odaloš – V. Patráš. Banská Bystrica: Pedagogická fakulta 1991, s. 198 – 205.

PATRÁŠ, Vladimír: Anketové sondy do bežnej komunikácie v meste. In: Slovenská reč, 1992(a), roč. 57, č. 1, s. 10 – 19.

PATRÁŠ, Vladimír: Normatívny aspekt hovorenej podoby slovenčiny v meste. In: Varia. 1. Materiály z Kolokvia mladých jazykovedcov (Modra – Piesok 21. – 22. októbra 1991). Red. M. Nábělková – M. Šimková. Bratislava: Slovenská jazykovedná spoločnosť pri SAV 1992b, s. 85 – 89.

PATRÁŠ, Vladimír: Germanismens Spuren in der gesprochenen Urbanomform des Slowakischen. In: Satz – Text – Diskurs. Akten des 27. linguistischen Kolloquiums, Münster 1992. Zv. 1. Red. S. Bechmann – S. Frilling. Tübingen: Max Niemeyer Verlag 1994, s. 283 – 289.

PATRÁŠ, Vladimír: Sociálne vektory a dynamika variet v komunikácii obyvateľov miest. In: Wielkie miasto. Czynniki integrujące i dezintegrujące. Red. D. Bieńkowska. Łódź Wydawnictwo Uniwersytetu Łódzkiego 1995a, s. 179 – 185.

PATRÁŠ, Vladimír: Sociolingvistický profil bežnej hovorenej komunikácie v Banskej Bystrici (90. roky). In: Sociolingvistické aspekty výskumu súčasnej slovenčiny. Sociolinguistica Slovaca. 1. Red. S. Ondrejovič – M. Šimková. Bratislava: Veda 1995b, s. 119 – 131.

PATRÁŠ, Vladimír: Generačno-sociálne súvislosti a bežná hovorená komunikácia v Banskej Bystrici. In: Banská Bystrica. Sociokultúrne obrazy a portréty. Acta Universitatis Matthiae Belii. Vedecko-výskumný ústav regionalistiky. Zborník. 2. Red. J. Darulcová. Banská Bystrica: Univerzita Mateja Bela 1996, s. 125 – 135.

PATRÁŠ, Vladimír: Metodologické špecifiká pri výskumoch živej komunikácie v (slovenských) mestách. In: Miesto – teren koegzistenciji pokoleń. Red. Z. Staszewska. Łódź: Wydawnictwo Uniwersytetu Łódzkiego 1999, s. 162 –

169.

PATRÁŠ, Vladimír: „Jazyk mesta“ na domovských komerčných stránkach v slovenskej sieti internetu? In: Mesto a jeho jazyk. Sociolinguistica Slovaca 5. Red. S. Ondrejovič. Bratislava: Veda 2000, s. 242 – 250.

SLANČOVÁ, Daniela: Oslovenie a kontakt v dialógu dospelého a dieťaťa. In: Studia Academica Slovaca. 26. Prednášky XXXIII. letného seminára slovenského jazyka a kultúry. Red. J. Mlacek. Bratislava: Stimul – Centrum informatiky a vzdelávania FF UK 1997, s. 148 – 158.

SLANČOVÁ, Daniela: Reč autority a lásky. Reč učiteľky materskej školy orientovaná na dieťa – opis registra. 1. vyd. Prešov: Filozofická fakulta Prešovskej univerzity 1999. 224 + 52 s.

SLANČOVÁ, D – MARIČOVÁ, D.: Vulgarizmy v súčasnej (mestskej) komunikácii. In: Mesto a jeho jazyk. Sociolinguistica Slovaca. 5. Red. S. Ondrejovič. Bratislava: Veda 2000, s. 194 – 202.

SLANČOVÁ, Daniela: Reč orientovaná na dieťa (o niektorých aspektoch a výsledkoch jej výskumu). In: Logopaedica. 5. Zborník Slovenskej asociácie logopédov. Bratislava: Liečreh Gúth 2002, s. 91 – 102.

SLANČOVÁ, Daniela: Jazyk a komunikácia v dialógoch dospelých a detí. In: Slovo o slove. Zborník Katedry slovenského jazyka a literatúry Pedagogickej fakulty Prešovskej univerzity. 9. Red. Ľ. Sičáková. Prešov: Pedagogická fakulta Prešovskej univerzity v Prešove 2003, s. 42 – 56.

SLANČOVÁ, Daniela: Mesto a jeho jazyk. In: Moje mesto Prešov. Magazín o jednom meste, 2004, roč. 1, č. 1, s. 63 – 67.

SLANČOVÁ, Daniela: Vývin detskej reči v slovenčine (od tušenia k deskripcii). In: Jazyk a komunikácia v súvislostiach. Zborník príspevkov z medzinárodnej vedeckej konferencie konanej 24. a 25. júna 2004 na Filozofickej fakulte Univerzity Komenského v Bratislave. Zost. J. Dolník. Bratislava: Univerzita Komenského 2005, s. 59 – 70.

SOKOLOVÁ, Miloslava – SLANČOVÁ, Daniela: Variety hovorenej podoby slovenčiny. In: Studia Academica Slovaca. 23. Prednášky XXX. letného

seminára slovenského jazyka a kultúry. Red. J. Mlacek. Bratislava: Stimul – Centrum informatiky a vzdelávania FF UK 1994, s. 225 – 240.

SOKOLOVÁ, Miloslava – SLANČOVÁ, Daniela: Výskum podoby hovorenej komunikácie na východnom Slovensku. In: Sociolingvistické aspekty výskumu súčasnej slovenčiny. In: Sociolinguistica Slovaca. 1. Red. S. Ondrejovič – M. Šimková. Bratislava: Veda 1995, s. 132 – 143.

SOKOLOVÁ, Miloslava – SLANČOVÁ, Daniela: Výsledky a perspektívy sociolingvistického výskumu prešovských vysokoškolákov. (K jazykovej situácii na východnom Slovensku.) In: Slovenčina na konci 20. storočia, jej normy a perspektívy. Sociolinguistica Slovaca. 3. Red. S. Ondrejovič. Bratislava, Veda 1997, s. 213 – 224.

ŠIMKOVÁ, Mária – GARABÍK, Radovan – KARČOVÁ, Agáta – GAJDOŠOVÁ, Katarína: Hovorený korpus slovenčiny. Príspevok prednesený na konferencii Čeština v mluveném korpusu v októbri 2007 v Prahe. (v tlači)

ŠTÍCHA, František: Kataforické korelatívum *to* v souvětí s větami podmětovými a předmětovými. In: Korpus jako zdroj dat o češtině. Ed. P. Karlík. Brno: Masarykova univerzita v Brně 2004. s. 145 – 153.

URBANCOVÁ, Lujza: Výskum hovorenej podoby slovenčiny v meste. In: Zborník mladých filológov Univerzity M. Bela I. Banská Bystrica: Fakulta humanitných vied Univerzity Mateja Bela 1999.

URBANCOVÁ, Lujza: Projekt výskumu hovorenej podoby slovenčiny v meste Zvolen. In: Mesto a jeho jazyk. Ed. S. Ondrejovič. Bratislava: VEDA, vydavateľstvo Slovenskej akadémie vied 2000(a), s. 213 – 216.

URBANCOVÁ, Lujza: Niekoľko poznámok k reči najstaršej generácie v meste Zvolen. In: Jazyková komunikácia v 21. storočí. Ed. J. Klincková. Banská Bystrica: Fakulta humanitných vied Univerzity Mateja Bela 2000(b), s. 278 – 281.

URBANCOVÁ, Lujza: Staršia generácia vo výskumoch jazyka mesta. In: Súčasná jazyková komunikácia v interdisciplinárnych súvislostiach. Ed. Vladimír Patráš. Banská Bystrica: Univerzita Mateja Bela 2004, s. 179 – 186.

URBANCOVÁ, Lujza: Faktor času vo výskume jazyka obyvateľov mesta.
In: Čas v jazyce a literatúre. Ústí nad Labem: Univerzita J. E. Purkyně 2005, s.
173 – 178.

Zdroje

Bavarian Archive for Speech Signals Corpora. Dostupný z WWW: <http://www.phonetik.uni-muenchen.de/Bas/BasKorporaeng.html>

Longman Spoken American Corpus. Dostupný z WWW: <http://www.pearsonlongman.com/dictionaries/pdfs/Spoken-American.pdf>

Pracovná verzia hovoreného korpusu slovenčiny. (Jún 2008).

Slovenský národný korpus. Verzia prim-3.0-public-sane. Dostupný z WWW: <http://korpus.juls.savba.sk>

The American National Corpus. Dostupný z WWW: <http://www.americannationalcorpus.org/>

The British Academic Spoken Corpus. Dostupný z WWW: <http://www2.warwick.ac.uk/fac/soc/celte/research/base/>

The British National Corpus. Dostupný z WWW: <http://www.natcorp.ox.ac.uk/>

The C – ORAL. Dostupný z WWW: <http://lablita.dit.unifi.it/coralrom/>

The CNN Archive. Dostupný z WWW: <http://www.cnn.com/SPECIALS/2004/reagan/stories/speech.archive/index.html>

The Corpus of Spoken Israeli Hebrew. Dostupný z WWW: <http://www.tau.ac.il/humanities/semitic/cosih.html>

The London – Lund Corpus. Dostupný z WWW: <http://khnt.hit.uib.no/icame/manuals/LONDLUND/INDEX.HTM>

The Speech Accent Archive. Dostupný z WWW: <http://accent.gmu.edu/>

The Spoken Dutch Corpus. Dostupný z WWW: <http://lands.let.kun.nl/cgn/ehome.htm>

The Spoken Language Corpus of Swedish. Dostupný z WWW: <http://www.qualitative-research.net/fqs-texte/3-00/3-00allwoodetal-e.htm>

Prílohy

1. Značky použité v prepise textov Slovenského hovoreného korpusu

ŠUMY HOVORIACI

[rh]	dýchanie (hlasné) hovoriaceho
[nh]	smrkanie/fučanie hovoriaceho
[txh]	kašeľ/odfhanie hovoriaceho
[tx2h]	kýchanie hovoriaceho
[rih]	smiech hovoriaceho
[sifh]	šepot hovoriaceho
[zivh]	zívanie hovoriaceho
[stikh]	štikútanie hovoriaceho

ŠUMY POČÚVAJÚCI

[rp]	dýchanie (hlasné) počúvajúceho
[np]	smrkanie/fučanie počúvajúceho
[txp]	kašeľ/odfhanie počúvajúceho
[tx2p]	kýchanie počúvajúceho
[rip]	smiech počúvajúceho
[sifp]	šepot počúvajúceho
[zivp]	zívanie počúvajúceho
[stikp]	štikútanie počúvajúceho

ŠUMY POZADIE

[poz]	zvuky v pozadí
[mm]	zvuky hovoriaceho (<i>ee, hm</i>)
[hh]	zvuky počúvajúceho (<i>ee, hm</i>), zvuky, slová udržiajúce konverzáciu

VÝSLOVNOSŤ

[lapsus]	prerieknutie
[lapsus2]	prerieknutie vzťahujúce sa na 2 slová dozadu
[lapsus3]	prerieknutie vzťahujúce sa na 3 slová dozadu

[lapsus4]	prerieknutie vzťahujúce sa na 4 slová dozadu
[lapsus5]	prerieknutie vzťahujúce sa na 5 slov dozadu
[lapsus6]	prerieknutie vzťahujúce sa na 6 slov dozadu
[lapsus7]	prerieknutie vzťahujúce sa na 7 slov dozadu
[lapsus8]	prerieknutie vzťahujúce sa na 8 slov dozadu
[lapsus9]	prerieknutie vzťahujúce sa na 9 slov dozadu
[lapsus10]	prerieknutie vzťahujúce sa na 10 slov dozadu
[skom]	skomolené slovo
[pi]	nezrozumiteľne vyslovené
[ch]	šuškание
[mum]	mumlavý, zachrípnutý štýl prehovoru
[ned]	nedokončené slovo
[lu]	prečítaná skratka
[luc]	prečítaná skratka v pôvodnom jazyku
[ep]	vyhláskovaná skratka
[ps-]	krátka pauza
[ps--]	stredne dlhá pauza
[ps---]	dlhá pauza
[***]	anotátor nedokáže identifikovať vypovedané slovo

VLASTNÉ PRVKY

[pers]	označenie mena osoby
[loc]	označenie miesta alebo iného zatretého zvuku okrem mena osoby

2. Metadáta o respondentovi k zvukovej nahrávke č.

Pohlavie: muž žena

Vek: vek<20

20≤vek<30

30≤vek<40

40≤vek<50

50≤vek<60

60≤vek<70

70≤vek<80

80≤vek

Vzdelanie: vysokoškolské

stredoškolské

odborné

nižšie

žiadne

Miesto narodenia:

Miesto najdlhšieho pobytu:

Miesto súčasného pobytu:

Materinský jazyk

Ďalšie jazyky

Nárečie: používa nepoužíva

Ak používa, aké:

3. Metadáta o nahrávke č.

Kódy respondentov zúčastnených v nahrávke:
.....
.....
.....
.....
.....
.....
.....

Téma/y rozhovoru:

.....

.....

Rozhovor:	riadený	spontánny
Rozhovor:	formálny	neformálny
Účastníci rozhovoru sa:	poznajú	nepoznajú
Účastníci rozhovoru sú:	rovnocenní	nerovnocenní partneri

Komentáre:

Dátum rozhovoru:

Miesto rozhovoru:

Počet účastníkov rozhovoru:

Zdroj: (napr. rozhlas, televízia, špecifikovaná inštitúcia...)

Nahrávacie zariadenie:

Meno nahrávajúceho:

Aktívna účasť nahrávajúceho v rozhovore: áno nie

Informácia o nahrávaní poskytnutá respondentovi/om vopred: áno nie

Formát zvukového záznamu:

Veľkosť zvukového záznamu:

4. Frekvenčná distribúcia výrazu *to*
(lemma=1<KWIC) – kolokáty lem sprava

byť	768	nebyť	28	že	16	vyzerat'	
,	234	aj	26	tu	15		11
mat'	84	taký	25	dat'	15	a	11
.	64	ja	24	len	13	chciet'	10
tak	62	si	21	s	13	pre	10
už	55	veľmi	21	ani	12	ísť	10
čo	50	znamenat'	21	istý	12	dobry'	10
sa	45	robiť	20	on	12	urobit'	10
/	45	na	20	za	12	potom	10
:	45	môcť	20	všetok	12	teda	10
nie	44	ešte	20	nevediet'	12	z	10
by	33	v	20	?	12		
tam	29	povedat'	17	vidiet'	12		
to	28	ako	17	musiet'	11		

5. Frekvenčná distribúcia výrazu *to*
(lemma=-1<KWIC) – kolokáty lem zľava

byť	674	ako	54	no	30	čiže	16
,	310	/	43	my	30	mať	16
a	166	by	42	takže	30	ty	15
že	153	čo	39	aby	29	.	14
na	151	aj	37	či	25	vy	12
sa	144	to	34	však	25	kto	11
ja	127	za	33	:	24)	10
tak	88	lebo	32	o	23	môcť	10
ale	69	ved'	32	už	23	ešte	10
on	62	keď	31	ona	19		

6. Frekvenčná distribúcia výrazu *to*
(word=-1<KWIC) – kolokáty tvarov slov zľava

je	376	ale	43	to	25	.	14
,	310	by	42	či	24	ste	14
na	147	sme	39	:	24	jej	13
že	147	aj	33	o	23	im	12
sa	143	bolo	32	ved'	23	no	11
a	102	za	31	No	19	nám	11
si	82	sú	30	už	19	mne	11
som	76	ja	30	však	18	ma	10
tak	71	čo	29	oni	17	nás	10
A	64	keď	28	Tak	1		
mi	56	aby	28	takže	16		
Je	50	lebo	27	Takže	14		
ako	49	Ale	26	mu	14		

7. Frekvenčná distribúcia spojenia *to* + tvar slovesa *byť*
(word=2<KWIC)

aj	8	tiež	6	iba	3	naozaj	2
ešte	8	asi	5	skôr	3	proste	2
už	8	teda	5	zase	3	hlavne	2
len	6	vlastne	4	tak	2	práve	2