SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

Korpus DEAF:
psané texty českých neslyšících

Klára Richterová
Ústav českého jazyka a teorie komunikace, FF UK v Praze
klara.richterova@ff.cuni.cz

Abstract:
This article is an introduction to a special corpus of Czech texts written by the Deaf – a unique project resulting from cooperation of two institutes of the Faculty of Arts at Charles University, Prague, namely the Institute of the Czech Language and Theory of Communication and the Czech National Corpus Institute. The two principal aims of the corpus DEAF are (1) to contribute to the improvement of teaching of the Czech language at schools for students with impaired hearing and (2) to provide a repository for records of the culture and history of the Czech Deaf. The corpus, with a number of specific features reflecting the specificity of the language data, is also designed to provide the first firm basis for scientific analysis of written Czech used by the Deaf.

Key words:
Czech Deaf, Deaf history, Deaf culture, written language, special corpora, annotation.

Jeden ze směrů, kterým se v posledních desetiletích ubírá využití jazykových korpusů, připomenul v minulém čísle SALi K. Šebesta (2010) příspěvkem věnovaným problematice akvizičních korpusů. Představil v něm české projekty orientované jednak na osvojování prvního jazyka (korpus SCHOLA a korpus SKRIPT), jednak na osvojování druhého nebo cizího jazyka (learner corpus / žákovský korpus CZESL). Náš příspěvek informující o projektu DEAF na problematiku stati K. Šebesty do jisté míry navazuje: jde rovněž o nově zpracovávaný projekt, korpus DEAF lze rovněž přiřadit k akvizičním korpusům a podobně jako korpus CZESL se i korpus DEAF zaměřuje na budování speciálního korpusu jazyka nerodilých mluvčích. Heslovitě lze tento projekt charakterizovat následujícími údaji: Jde o korpus, který zahrnuje česky psané texty českých neslyšících a který je součástí plánu velké národní infrastruktury Český národní korpus (projekt MŠMT bude realizován od 1. ledna 2012). Na budování korpusu DEAF spolupracují členové Ústavu Českého národního korpusu (ÚČNK) a pracovníci oboru čeština v komunikaci neslyšících (ČNES) na Filozofické fakultě UK v Praze; k hlavním jeho tvůrcům zde patří M. Křen s K. Kučerou (ÚČNK), A. Macurová a K. Richterová (ČNES), na sběru a přípravě textů se však ve velké míře podílejí neslyšící i slyšící studenti oboru ČNES a zčásti i studenti bohemistiky. Přípravné práce na struktuře korpusu DEAF byly – souběžně se sběrem prvních textů – zahájeny v roce 2010. Odhadované tempo růstu korpusu DEAF je 40 000 textových slov ročně, se zpřístupněním jeho první části se počítá nejpozději v roce 2014.

V uvedeném relativně pomalém tempu růstu korpusu DEAF se odráží vliv řady specifických problémů, které s sebou přináší jak získávání textů českých neslyšících (patří k nim zejména značně omezený počet autorů a obtížná dostupnost textů), tak jejich zpracování (mimořádně časově náročná manuální lemmatizace a anotace). Těmto a dalším rysům projektu DEAF se

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

budeme podrobněji věnovat níže, hned úvodem je však třeba zdůraznit, že přes naznačená specifika je při budování korpusu zachováván obdobný obecný princip jako u většiny jazykových korpusů, totiž že složení dat i způsob anotování odráží konkrétní oblasti předpokládaného využití korpusu a snaží se jim podle možností přizpůsobit. V případě projektu DEAF jde o dvě prioritní oblasti (lingvisticko-didaktickou a kulturně-historickou), v jejichž rámci si korpus klade dva konkrétní cíle:

(a) vytvořit předpoklady ke zkvalitnění výuky češtiny na školách pro sluchově postižené (elektronický korpus by měl nejen sloužit lingvistickým a lingvodidaktickým výzkumům, ale jako výuková pomůcka by mohl být využíván učiteli přímo v hodinách českého jazyka);

(b) poskytnout platformu pro uchovávání záznamů o kultuře a historii českých neslyšících ve formě osobních vzpomínek, názorů a zkušeností.

Mimo to lze výhledově počítat s využitím korpusu DEAF i v dalších směrech: po dosažení dostatečného rozsahu může korpus fungovat jako spolehlivá základna například pro výzkum kontaktových jevů mezi češtinou a českým znakovým jazykem nebo pro hlubší poznání málo uvědomovaných jazykových zákonitostí v češtině „většinové", „standardní" – zákonitostí, které si rodilí mluvčí češtiny uvědomují často teprve tehdy, když se setkají s jejich porušováním.

Jak je zřejmé z bodu b) výše, akcentováním osobních výpovědí neslyšících se projekt v některých ohledech přibližuje zaměření orální historie, která se sice v mezinárodním měřítku začala rozvíjet již koncem 40. let, avšak v Česku nachází výraznější uplatnění až v posledních desetiletích.[1] Kulturně-historická hodnota vzpomínkových aj. autentických výpovědí, která stojí v popředí zájmu orální historie a na kterou se do značné míry zaměřuje i korpus DEAF, ostatně rozhodně není opomíjena ani v již existujících zahraničních korpusech znakových jazyků; setkáváme se s ní např. v korpusu holandského znakového jazyka (*NGT Corpus*) ve formě elicitovaných dat tematického okruhu *Deaf Issues* a *Event*.[2] Budování korpusu ja-

1 První centrum orální historie bylo založeno na *Columbia University* z iniciativy Allana Nevinse a Louise Starra už v roce 1948 (*Columbia Oral History Office*); v roce 1966 pak vzniká celonárodní zastřešující organizace *Oral History Association*, k níž se na počátku 70. let připojila britská *Oral History Society* (1973). V České republice bylo pro systematický rozvoj metody orální historie a její aplikace v historiografii zřízeno k 1. lednu 2000 v Ústavu pro soudobé dějiny AV ČR *Centrum orální historie*. Na styčné body orální historie s jinými typy shromažďování autentických pramenů v této souvislosti upozornil M. Vaněk: „K metodě orální historie se blíží, či se jí okrajově dotýkají, další způsoby a cesty vedoucí k zachycení minulosti, které za autentický historický pramen považují sdělení jednotlivce. Takovým pramenem mohou být *paměti*, *deníky*, *osobní dopisy*, *privátní filmy*, *školní slohové práce* atd., které sice (s větší či menší mírou autocenzury) zachycují individuální prožitky, zkušenosti a názory, ale postrádají nezbytné kritérium orální historie: formu rozhovoru vedeného z očí do očí mezi tazatelem a narátorem" (Vaněk, 2007, s. 75; srov. také Vaněk, 2004).

2 Jazyková data obsažená ve vznikajících (všeobecných) korpusech znakových jazyků obvykle bývají částečně jiné povahy než data v korpusech jazyků mluvených. Zásadní rozdíl spočívá v tom, že v případě jazyků znakových – kromě spontánního či polospontánního projevu (např. vyprávění o tom, co má respondent rád) – jde o řízenou elicitaci zahrnující různé jazykové úkoly a zaměřující se na vybraná témata. Například v případě korpusu NGT jde o oblasti *Introduction*, *Narrative (fable)*, *Narrative (comic)*, *Narrative (TV-sketch)*, *Narrative (picture books)*, *Spot the difference*, *Event* a *Deaf Issues* (blíže k vymezení těchto oblastí viz <http://www.ru.nl/corpusngtuk/methodology/data_collection/>; k obdobné problematice v korpusu německého znakového jazyka srov. <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/corpus.html>).

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

kéhokoli znakového jazyka je však ve srovnání s vytvářením korpusů audio-orálních jazyků po všech stránkách mnohem náročnější,[3] a v českých podmínkách proto v dohledné době nelze s existencí plně reprezentativního korpusu českého znakového jazyka počítat. Zaznamenávání a uchovávání vzpomínek a názorů českých neslyšících bude tedy v současné době suplovat korpus DEAF, pochopitelně ovšem jen do jisté míry, neboť na rozdíl od českého znakového jazyka je psaná čeština pro neslyšící jazykem cizím, a proto nevyhnutelně snižuje autenticitu a spontánnost výpovědí. Že tato funkce, do značné míry kulturní a historická, není v projektu DEAF pokládána za pouze přidruženou či nepodstatnou, ale přinejmenším za plně srovnatelnou s druhou výše uvedenou funkcí (uplatnění korpusu v didaktické praxi), se snaží alespoň zčásti naznačit již samotný název korpusu. Na vysvětlenou: anglické pojmenování *Deaf* lze považovat za synonymum českého výrazu *Neslyšící*, které označuje příslušníka jazykové a kulturní menšiny a které se běžně – standardně s výrazně pozitivními konotacemi – vyskytuje v různých pravopisných obměnách (především *Deaf* a *deaf*, méně často *Def*, *def* aj.) v textech českých neslyšících – srov. např. *Defové zvládnete všechno sami – Deaf jednoduchy rozumet co je obsah*. Proč je v názvu korpusu upřednostněno anglické pojmenování před českým, napovídají doklady, v nichž se výraz *deaf* užívá jako adjektivum, resp. první část složeného slova, jako např. *náš deafsvět / deaf svět – naše deafkultura / deaf kultura: „*Skoda, vy [slyšící ovládající český znakový jazyk] nejste uplne deafs, proto nic nevite jak to vypada deafs kultura i svet.; Protoze ma dva svet, a slysici furt mysli ze jeden svet slysici, ale bohuzel slysici furt debil. jasne ma dva svet deafs i slysici. “[4] Jak je zřejmé z uvedených dokladů i z toho, že výrazu *Deaf* se podobně užívá i v řadě jiných jazyků,[5] být Deaf znamená (nejen) pro české neslyšící být začleněn do mezinárodního společenství Neslyšících/Deafs, které sdílí podobné vzorce chování, podobné životní zkušenosti, problémy aj., tj. společnou kulturu odlišnou od různých národních kultur slyšících.

Psaná čeština neslyšících

Vzhledem k tomu, že jedním z hlavních důvodů budování korpusu DEAF je – jak jsme naznačili v úvodu – vytvořit předpoklady ke zkvalitnění výuky češtiny neslyšících, je na místě alespoň krátce shrnout, co o psané češtině neslyšících a její odlišnosti od češtiny „většinové“ v současné době vlastně víme. Předně: je zřejmé, že tato odlišnost – resp. značná specifičnost

3 Jde především o pořízení, obsluhu a údržbu velmi nákladného technického vybavení, k němuž patří kvalitní kamery, servery schopné spravovat velké objemy videodat aj., a o mimořádně časově náročnou manuální anotaci, která je v případě víceúčelových korpusů znakového jazyka nezbytným požadavkem. Vytváření těchto korpusů je navíc spojeno se specifickými problémy při sběru materiálu, k nimž patří zejména nemožnost zachovat anonymitu respondentů vzhledem k tomu, že na videozáznamu musí být vždy dobře rozeznatelný i obličej, neboť jeho mimika vyjadřuje mj. některé gramatické významy. Samotnému nahrávání proto většinou předchází pečlivě připravená informační kampaň.

4 Čeští neslyšící obvykle užívají slovo *Deaf* s koncovkami podle typu *pán* (srov. např. *Deafové; pro Deafy; znáš nějakého deafa?* atp.). Výjimečné však není ani užívání výrazu *Deaf* s anglickým plurálem (*Deafs*) nebo v nesklonné podobě substantivní (např. *podle deaf*, *kvuli deaf*, *DRBY O DEAF* atp.) nebo adjektivní (viz už výše uvedené *deaf svět*, *deaf kultura*). Uvedené doklady pocházejí z internetu, především ze stránek <www.neslysici.cz>, a zde i na jiných místech tohoto článku jsou citovány vždy v autentické podobě.

5 Srov. např. následující internetové doklady ze slovenštiny (*Urobil si niečo dobré pre deaf?*), z němčiny (*Babyphone mit Vibration und Lichtsignal für Deaf*), francouzštiny (*Le protocole de communication standard pour le deaf ou malentendants est appelé téléphone texte*), polštiny (*klub federacja dla deaf*; *tam nie ma udogodnień dla Deaf*) aj.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

psaných projevů neslyšících vůbec – vyplývá z několika vnějších faktorů, mezi nimiž je jako nejdůležitější obvykle uváděna smyslová nedostupnost audio-orálního jazyka, která neslyšícím neumožňuje, aby si mluvený jazyk (na rozdíl od vizuálně-motorického znakového jazyka) osvojili ve smyslu přirozeného nabývání (srov. např. Macurová, 1998). Odezírání jako způsob zprostředkování mluvené podoby jazyka lze považovat v této souvislosti za dysfunkční jednak vzhledem k tomu, že úspěšnost vizuální interpretace hlásek dosahuje maximálně 40 %, jednak (a to hlavně) vzhledem k tomu, že neznámá slova nelze odezírat (srov. např. Strnadová, 1998).[6] Zásadním faktorem je i skutečnost, že 90–97 % neslyšících dětí se rodí slyšícím rodičům. Může se proto stát (a není to zcela výjimečné), „že ani mluvený, ani znakový jazyk se k malému neslyšícímu dítěti nedostává v takové kvantitě či kvalitě, která dostačuje pro snadné osvojení prvního jazyka" (Gallaway, 2001, s. 55). Stav takového raného „bezjazyčí" následně zamezuje funkčnímu osvojení jakéhokoli jazyka v dospělosti (srov. např. Mayberry, 1998).

V českém prostředí se problematice psané češtiny v širším kontextu komunikace neslyšících věnuje zhruba od poloviny 90. let A. Macurová, která také – především na materiálu soukromé korespondence českých neslyšících ve věku 15–19 let (Macurová, 1995) a textů školních prací (Macurová, 1998) – jako první popsala její typické rysy.[7] Jako na obecný signifikantní znak psané češtiny neslyšících poukázala na její „jednorodost", tj. na skutečnost, že se v ní neprojevuje „ani typová rozdílnost adresátů (interkulturní vs. intrakulturní komunikace), ani rozměr důvěrnosti vs. oficiálnosti, ani různá míra připravenosti, ani různost funkcí, jež se textům přiřazují. Nic podstatného neřekne tato čeština ani o svých původcích (různých co do věku, profese, původu, místa bydliště, pohlaví atd.)" (Macurová, 1998, s. 180–181).

Ke konkrétním odchylkám, které lze většinou přičítat vlivu interferencí z českého znakového jazyka, patří podle výzkumů A. Macurové (tamtéž, s. 184) např. záměna sloves označujících existenci a vlastnictví (*Sestra je krásné rty*; *do školy nechodím, protože mám nemocný*), vypouštění sponových sloves (*Doktor pitomec*), lexikální vyjadřování času a modality (*bude možná koupit motorku*; *ptá se bude domů*), absolutní zápor (*tatínek je líný, vůbec pracovat*) a nerozlišené užívání „proto" a „protože", případně vyjádření důvodově-příčinných vztahů slovem *důvod* (*zítra pojedeme na vlaku lyžovat, proto zítra tu budou málo lidí, proto oni chodí do práce*; *Jsem trochu závislý na tlumočníka, protože jsem velice rád, že tlumočník mi dal plné informace.*; *[Neslyšící] Rád chodí na sport, ale důvod nemocné občany, olympiáda problém*). Některé ze zmiňovaných odchylek jsou typické i pro psané texty anglických neslyšících, jako např. záměna sloves *to have* a *to be* (*The boy have a sick*. *The boy is a sweater*; srov. Quigley – Paul, 1984, s. 124) či vynechávání pomocných sloves (např. *it very important*; srov. Anderson, 1993, s. 96–97). S anglickými texty sdílí psaná čeština neslyšících i některé další obecné rysy, které charakterizovali např. Quigley a Paul (1984): psaný projev neslyšících se podle těchto autorů vyznačuje především kratšími a jednoduššími větami (např. v textech sedmnáctiletých neslyšících dosahují věty stejné průměrné délky jako věty v textech osmiletých slyšících), dále častým výskytem konvencionalizovaných a stereotypních konstrukcí (jako *There is...*) či jinou frekvenční distribucí slovních druhů – neslyšící děti užívají více determinátorů (členů

6 Jazyk většinový (v našem případě psaná čeština) by proto zákonitě měl být vyučován jako jazyk cizí, jako jazyk druhý. Je však třeba říci, že přes řadu pozitivních změn, k nimž u nás v posledních letech došlo ve výuce na speciálních školách pro sluchově postižené, nelze současnou situaci považovat za ideální (srov. disertační práci – zatím nepublikovanou – A. Hudákové [2008] *Čeština ve vzdělávání dětí s vadou sluchu.*)

7 Srov. také Macurová, 2001.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

a ukazovacích zájmen) a více substantiv a sloves, naproti tomu se v jejich textech uplatňuje méně příslovcí, pomocných sloves a spojek.

Větší překážku srozumitelnosti textů než tyto poměrně snadno odhalitelné (tzv. mechanické) chyby – k nimž patří také např. četná vynechávání slov, přidávání slov, nahrazování slov a chyby ve slovosledu – představuje podle Andersonové (1993, s. 19–20) častý výskyt odboček a celková tematická nesoudržnost, tj. vesměs méně nápadné chyby na úrovni textové, kterým lingvisté doposud nevěnovali příliš pozornosti. I tyto textotvorné odchylky, identifikované v textech psaných anglickými neslyšícími, bychom mohli zaznamenat v psaných textech neslyšících u nás.

Při celkovém pohledu je třeba konstatovat, že jazyk psaných projevů českých neslyšících je zatím prozkoumán jen okrajově – a totéž lze s jistými modifikacemi říci i o psaných textech neslyšících jiných národů. Pokud jde o neslyšící české, vedle statí A. Macurové byly jejich psané projevy analyzovány vlastně jen v několika diplomových pracích studentů oboru čeština v komunikaci neslyšících, případně studentů bohemistiky, které se zaměřovaly na dílčí témata z roviny morfologické (sloveso v psané češtině neslyšících zkoumala M. Racková [1996],[8] jmenný rod J. Wágnerová [2008] a vyjadřování času K. Richterová [2009]), na některé aspekty lexikální (problematikou *být* a *mít* se zabývala W. Ziebikerová [2009]) nebo na popis odchylek v rámci jednoho textového typu (analýza internetové komunikace od H. Šebkové [2008]). Celostní systematický popis psané češtiny neslyšících však dosud chybí.[9]

Reprezentativnost korpusu DEAF

Jedním z hlavních důvodů, proč zmíněný celostní popis psané češtiny neslyšících nebyl dosud realizován, je skutečnost, že k němu v současné době neexistuje dostatečně reprezentativní základna autentických textů. Aby jako taková základna mohl sloužit korpus DEAF, musí splňovat kritérium reprezentativnosti, v níž se ve spojitosti s novodobými jazykovými korpusy obvykle spatřuje ukazatel jejich funkčnosti a která bývá tradičně poměřována jak jejich rozsahem, tak jejich strukturou (odpovídající zastoupení jednotlivých typů textů) a kvalitou (autentičnost textů).

V případě korpusu DEAF se naplnění těchto kritérií setkává se specifickými problémy. Rozsah tohoto korpusu je nevyhnutelně determinován jednak už výše naznačenými dvěma faktory, tj. malou základnou autorů, jejichž texty mohou být do korpusu zahrnuty,[10] jednak – a to zvlášť – značně obtížným přístupem k autentickým textům. (I slyšící pisatelé váhají

8 Viz též Zýková, 1997.

9 K problematice kompetencí českých neslyšících v oblasti recepce srov. Poláková 2000; 2001; 2002.

10 Uvést přesný údaj o počtu neslyšících v ČR je poměrně obtížné, a to i kvůli zohlednění různých faktorů; např. J. Hrubý (1999, s. 53) uvádí, že mezi osobami, jejichž vada sluchu trvá od narození nebo vznikla před započetím školní docházky, je asi 3 700 osob s úplnou hluchotou a 3 900 osob s praktickou hluchotou (tzn. ztráta sluchu je větší než 70 dB); podle výzkumu Českého statistického úřadu z roku 2007 (dostupné z <http://www.czso.cz/csu/2008edicniplan.nsf/t/4100269DD7/$File/330908j3.pdf>) je v ČR 7 746 osob s velmi těžkým sluchovým postižením (bez ohledu na dobu jeho vzniku) a 18 061 osob s těžkou vadou sluchu (rovněž bez ohledu na to, kdy postižení vzniklo). Srov. také značné rozpětí údajů v zahraničí: počet uživatelů amerického znakového jazyka v USA i v Kanadě je odhadován na 250 000–500 000 (Baker-Schenk – Cokely, 1980, s. 47); podle T. Johnstona a A. Schembriho (2007, s. 27–28) se sice počet neslyšících v Austrálii tradičně odhaduje poměrně přesně na 0,1 % populace, tj. 20 000 neslyšících, nicméně odhad počtu uživatelů australského znakového jazyka kolísá mezi 7 000–30 000.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

s poskytnutím svých textů k zařazení do korpusu, jsou-li si vědomi jejich odchylek od textů standardních, tj. především chyb, které v nich jsou nebo mohou být, a je tedy pochopitelné, že se s velkým váháním setkáváme i u neslyšících, pro něž je čeština cizím jazykem.)[11]

Souvýskyt těchto dvou vnějších faktorů zásadním způsobem ovlivňuje i otázku samotné skladby textů v korpusu DEAF. V situaci, v níž každý text neslyšících, který se podaří získat, představuje pro korpus znatelný přínos, lze hledisko vyváženosti dat (ať již podle žánrových, typových aj. charakteristik textů, nebo podle charakteristik autorů) uplatňovat jen v omezeném rozsahu. Omezenou možností zohledňovat složení textů se tak korpus DEAF velmi blíží akvizičním korpusům, u nichž se také – na rozdíl od dominantní percepce v případě velkých moderních korpusů (srov. Čermák – Králík – Kučera, 1997) – spíše přihlíží k recepci (srov. Šebesta, 2010). Jediným přístupem reálně aplikovatelným v této situaci je zjednodušená zásada *more data = better data* (srov. např. Merce – Church, 1993; Hanks, 1990), jež korpusu DEAF zajistí alespoň ten druh reprezentativnosti, který vychází z možného využití korpusu a který se do jisté míry uplatnil i v obecných úvahách o struktuře textových zdrojů pro reprezentativní korpus SYN2000.[12]

Omezený ohled na vyváženost dat při budování korpusu však s sebou nepřináší jen nevýhody. Ve skutečnosti právě v důsledku tohoto omezení může být korpus DEAF (podobně jako například kdysi Pražský mluvený korpus, srov. http://ucnk.ff.cuni.cz/pmk.php) významně rozšířen i o texty řízené, tj. o texty, které vznikají většinou na předem zadané téma a jejichž rozsáhlejší zařazení do všestranně vyváženého korpusu by jinak bylo těžko obhajitelné vzhledem k porušení skladby textů. V případě korpusu DEAF jde primárně o řízené texty zaměřené na kulturu a historii neslyšících a o texty, v nichž se neslyšící vyjadřují k vybraným aktuálním otázkám, souhrnně tedy o texty, které jsou mimořádně závažné z hlediska komunity českých neslyšících a jejichž význam přesahuje hranice lingvistického výzkumu. Pilotní verze navržených témat, která byla konzultována s neslyšícími studenty a která je v současné době předkládána k připomínkám širšímu okruhu neslyšících, má podobu tematických okruhů s připojeným výčtem možných konkrétních témat. K tématům/otázkám, k nimž se v těchto textech neslyšící mají vyjádřit, patří např. Co si myslíte o zrušení deaflympiády na Slovensku? Liší se nějak humor neslyšících a slyšících? Jak si představujete štěstí? atp.

Texty psané na zadané téma nicméně přes svůj zmíněný kulturní (a v širší časové perspektivě i historický) význam nepředstavují jádro korpusu DEAF. To je tvořeno texty, jež nevznikly za účelem výzkumu, respektive nevznikly přímo za účelem zařazení do korpusu. I při výše zmíněných omezeních je rozpětí korpusových textů v mnoha ohledech velmi široké: jde o texty formálně realizované různými způsoby (rukopis vs. elektronický záznam), o texty pohybující se na škále mezi řízeností a spontánností, o texty náležející k různým žánrovým typům, o texty, v nichž se vedle různých subjektivních stylotvorných faktorů uplatňují i různé faktory objektivní, jejichž vlivem se odlišují texty soukromé vs. veřejné, připravené vs. nepři-

11 Stejně jako v případě korpusů znakových jazyků by tyto obtíže snad mohlo pomoci překlenout detailní seznámení neslyšících se smyslem korpusu DEAF. Velká pozornost se proto soustředí jak na popularizační přednášky pro neslyšící, tak na individuální přístup k neslyšícím autorům.

12 Srov. „Vzhledem k tomu, že neexistuje univerzální, všeobecně uznávaná představa o využití korpusů (neboť neexistuje ani představa o všech možných budoucích požadavcích na korpusy jako takové), staly se východiskem pro ČNK odhady možných očekávání ze strany dnes uvažovaných potenciálních zájemců o práci s korpusem“ (Kocek – Kopřivová – Kučera, 2000, s. 16–17).

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

pravené, monologické vs. dialogické, texty s přítomným vs. nepřítomným adresátem, texty intrakulturní (psané neslyšícími autory neslyšícím adresátům) vs. interkulturní (adresované slyšícím). Konkrétně se jedná o školní slohové práce, dopisy, internetovou komunikaci (chaty, profily na Facebooku a zvlášť vyčleněné e-maily), dále zprávy SMS, faxy a vzácně i o odborné a umělecké texty. Ke skladbě textů zařazovaných do korpusu DEAF je navíc třeba dodat, že kromě textů od neslyšících autorů zahrnuje korpus DEAF v omezené míře i texty autorů nedoslýchavých, ohluchlých či osob s kochleárním implantátem. Tyto texty se obvykle liší od „většinové" češtiny v mnohem menší míře, případně v jiném ohledu než texty autorů neslyšících, a tudíž budou dobře využitelné ke komparativní analýze s texty neslyšících, která dosud nebyla realizována ani u nás, ani v zahraničí.

Anotování

Správní a strukturní anotování

Údaje o textu a o autorovi, které by mohly být relevantní pro lingvistickou analýzu a které stejně jako u ostatních korpusů ČNK soustřeďuje hlavička každého textu, se u korpusu DEAF skládají ze 13 položek.

údaje o textu a o typu komunikace	**údaje o autorovi**
(1) identifikační kód textu, ID	(6) pohlaví
(2) typ textu	(7) věk
(3) doba vzniku textu	(8) vzdělání
(4) typ komunikace	(9) původ
	(10) škola
	(11) stav sluchu
	(12) mateřský jazyk
	(13) preferovaný jazyk
(5) poznámka	(5) poznámka

Hodnoty jednotlivých položek s vysvětlivkami: 1) ID se skládá ze zkratky typu textu + čísla, např. sms910; 2) anotátor má možnost výběru z hodnot dopis/email/fax/sms/internetová diskuse/odborný text/školní text/text na zadané téma/umělecký text; 3) rok/časové rozpětí v letech, např. 1995–2000/nelze určit atp.; 4) interkulturní/intrakulturní/nelze určit: hodnota *interkulturní* – jak už bylo zčásti naznačeno – je určena k označení komunikace mezi neslyšícím, resp. sluchově postiženým komunikantem a slyšícím/slyšícími, příp. skupinou komunikantů, v které – jako v případě textů veřejně přístupných na internetu – není vyloučen slyšící adresát; 5) pole *poznámka* je určeno pro jakékoli další relevantní informace, např. že jde o školní práci na téma zadané a závazně zformulované učitelem, že jde o špatně čitelný rukopis; zařazeny sem bývají i relevantní poznámky týkající se mimojazykové reality aj.; 6) muž/žena/nelze určit; 7) konkrétní věk autora v době vzniku textu, popř. věkové rozpětí, např.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

16–20/nelze určit; 8) nedokončené základní/dokončené základní/nedokončené střední/dokončené střední/nedokončené VŠ/dokončené VŠ/nelze určit; 9) místo (město/kraj/oblast), kde autor žije, popř. místo pobytu v dětství; 10) místo (město, v případě Prahy i bližší určení/kraj/oblast) školy/škol; 11) nedoslýchavý/neslyšící/osoba s kochleárním implantátem/ohluchlý/nelze určit; 12) jazyk/jazyky, případně kód, kterým/kterými respondent komunikoval v úzké rodině: znakový jazyk/znakovaná čeština/čeština/nelze určit; 13) jazyk nebo kód, který respondent upřednostňuje: znakový jazyk/znakovaná čeština/čeština/nelze určit. Výčet těchto hodnot je závazný, není však do budoucna uzavřený: v odůvodněných případech, k jakým by patřilo např. získání textů nového typu, ho lze snadno rozšířit.

Jak je patrné z tabulky i z vysvětlujících poznámek, zahrnují údaje o autorovi vedle standardních sociolingvistických kategorií (např. věk, pohlaví atp.), tj. položek běžně zaznamenávaných i u mluvených korpusů, také položky zaměřené na konkrétní skupinu respondentů/autorů textů, v našem případě neslyšících, resp. sluchově postižených (položky jako např. stav sluchu; jazyk, kterým autor komunikoval s rodiči, apod.). V souvislosti se sledovanými faktory je třeba uvést, že spodní věková hranice autorů, jejichž texty jsou do korpusu DEAF zařazovány, sice není explicitně stanovena, nicméně nejmladšími autory, od nichž se podařilo získat autentické psané projevy (tj. projevy, které nebyly korigovány učitelem), jsou – alespoň v současné době – žáci druhého stupně základních škol.

Ve výše uvedené tabulce stojí za zmínku také položka *škola*, v níž se uvádí geografické místo školy nebo škol navštěvovaných autorem textu. Zařazení této informace mezi faktory ovlivňující texty sluchově postižených je motivováno dvěma hlavními důvody – jednak tím, že některé jazykové konstrukce by mohly být typické pouze pro žáky/studenty některých konkrétních škol,[13] a jednak tím, že díky této položce by také mohli učitelé konkrétních škol pro sluchově postižené využívat korpus DEAF jako zpětnou vazbu reflektující efektivitu výukových metod, které sami užívají.

Repertoár zjišťovaných údajů by mohl být bezpochyby rozšířen ještě o další položky, např. věk, kdy začala výuka češtiny (a jaký byl způsob této výuky); znalost dalších jazyků atp., ale z dosavadních zkušeností vyplývá, že ani při současném rozsahu hlavičky textu není zdaleka vždy možné zjistit všechny požadované informace (to se – jak naznačují poznámky k uvedené tabulce – projevuje mimo jiné tím, že v rámci hodnot jednotlivých položek se standardně nabízí možnost *nelze určit*). Navíc je třeba být si obecně vědom toho, že i některé z údajů, které se podaří získat, je nutno přijímat s jistou rezervou: jde o informace sdělené samotnými autory – a ty mohou být z různých důvodů do jisté míry zkreslující.

Na rozdíl od poměrně propracovaného značkování správního, jemuž jsme se doposud věnovali, je značkování strukturní v současné době jen velmi omezené – spočívá v pouhém označování odstavců kódem *<p>*. První fáze zpracování textu obnáší kromě vkládání hlavičky a značek pro odstavce ovšem ještě i jiné operace, především anonymizaci, v rámci níž jsou kódovány části textu, které by mohly pisatele nebo jinou osobu, instituci atp. poškodit nebo které – pokud autor nedal výslovný souhlas k jejich zachování – by mohly vést k určení identity autora. Konkrétní způsob anonymizace vychází z praxe uplatňované v korpusech řady ORAL,

13 Např. konstrukce obsahující sloveso *dělat/udělat* a substantivum nebo verbum, příp. verba, jako *děláš diskriminace menšina*, *jsi děláš vymlouvat* apod., tedy vyjádření, z nichž některá jsou blízká tzv. rozloženým vazbám slovesným (srov. Šmilauer, 1952, s. 253), nebyla zatím zaznamenána u jiných autorů než u studentů jedné pražské školy pro sluchově postižené.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

která spočívá v nahrazování anonymizovaných výrazů standardizovanými zkratkami (např. příjmení jsou nahrazena zkratkou *NJ*, místní názvy *NM* atp.). Systém užívaných značek v korpusech ORAL byl pro potřeby korpusu DEAF rozšířen ještě o některé další: např. telefonní číslo je zastoupeno zkratkou *NT*, datum (zejm. datum narození) nahrazuje zkratka *ND* apod.[14]

Lingvistické anotování

S jak specifickými problémy se setkává lingvistická anotace v korpusu DEAF, je možno naznačit několika krátkými ukázkami textů neslyšících:

> „[...] Máme na dovolenou prýč nikam někdý tu mě už vždý co dělat šla. [...]“ (soukromá korespondence – neslyšící adresát, autor: žena, 15–19 let, učňovský obor)

> „NESLYŠÍM, VŮBEC NEREAGUJE, POUZE ZRAK, PROTO CO JE. TAK SÁMOTNĚ PRACUJE, MÁLO VZTAHUJE, MOC PRÁCE, NÍZKÉ MZDY. TELEVIZNÍ PROGRAM, ASPOŇ KAŽDÝ DEN PRO NESLYŠÍCÍ POUZE ZA PŮL HODIN. [...] ZTRÁTY SLUCHU, NĚKDO POMALU POCHOPÍ DOMA VAŘÍ, VŮBEC NESLYŠÍ, JEN NA POZOR. LUPIČ CHODÍ TICHO, PES ŠTĚKÁ, VŮBEC NESLYŠÍ MOBIL VYBRAČÍ, VŮBEC NESLYŠÍ UŽ POZDĚ...“ (text na zadané téma *Jak se žije neslyšícím?*; autor: muž, 54 let, dokončené střední vzdělání)

> „[...] Sobota 9. ledna 2010: Denesky ráno já byla vsdavala 7:05 potom oblecení zima. Autobus já byla s deti legraní, potom já jíslo maslo a s chled. Já byla deti hrát sníh a vrát chatu Marteka řikala pokoj s detiv aku a Brachym, Petra a já. Vereče vjá byla jislo špagety. [...]“ (školní slohová práce; autorka: žena, 16–19 let, nedokončené střední vzdělání)

> „[...] prosim pišt my ja nesem total hluchy aspatne mluv ušni lekař mie napiš proto nesm řidim auto. Od 19 56 řidim bez problem v CZ slovensko a zahraniče hlvnk Neměck vicod a zapad. auto mam nutne po cdraf a spolecni. [...]” (fax adresovaný organizaci pro sluchově postižené; autor: muž)

Už z uvedených čtyř ukázek je nepochybně zřejmé, že texty neslyšících vyžadují ruční anotaci a že tato anotace nemůže mít standardní podobu, s níž se setkáváme např. v korpusech řady SYN v Českém národním korpusu. Nelze zde například vůbec uvažovat o systematickém

14 Že je anonymizace nezbytným postupem v případě malé komunity, jakou jsou neslyšící, není třeba rozvádět. Stejně tak je evidentní, že při tomto způsobu anonymizace se ztrácí některé lingvisticky důležité informace, zejména dochází ke ztížení, resp. znemožnění identifikace chyb, které se v anonymizovaných výrazech vyskytly (srov. *Jsem z Liberec*; *bydlím v Prahe* > *Jsem z NM*; *bydlím v NM*; *jmenují se Libor Dvořák* > *jmenují se NJ NP*.). Systém vlastních jmen je ale natolik specifický a rozmanitý (srov. už jenom různorodost víceslovných názvů jako *Ústí nad Labem*, *Hradec Králové nad Labem*, *Nové Město pod Smrkem, Moravská Třebová* atd.), že není možné postupovat tak, aby např. byl zvolen jeden reprezentativní lexém, který by nahrazoval jiné lexémy náležející ke stejnému typu (slovotvornému, deklinačnímu atp.). Neřešitelným problémem je v této situaci také zaznamenávání hláskových odchylek typu *Prahe*, *Vysoké Tetry* aj.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

morfologickém značkování, neboť v textech neslyšících se nezřídka vyskytují věty, jimž nejsme s to přiřadit smysl a které nejsme s to jednoznačně gramaticky interpretovat, a jednotlivé slovní tvary v nich tedy není možné jednoznačně anotovat (srov. např. věty jako *Máme na dovolenou prýč nikam někdý tu mě už vždý co dělat šla.*; *Petr Schwarz říká mně od já mi napsat k tobě jeho.*; *Nemocný (á) protože přivěd pro auta, může a vozů.*). V této situaci zůstává jedinou morfologickou značkou, s jejímž zavedením se v korpusu DEAF počítá, označení slovního druhu, které lze bez větších problémů spojit s lemmatizací (srov. dále).

Ze stejného důvodu nelze v korpusu DEAF využít ani chybovou anotaci,[15] která vychází z navržené opravy v podobě gramaticky správné věty / větného celku. Tento druh chybové anotace není na korpus DEAF aplikovatelný také proto, že v textech neslyšících se velmi často setkáváme s konstrukcemi, jejichž obecný smysl sice chápeme (nebo se tak domníváme), ale přitom nejsme s to z množiny možných noremních, gramaticky správných větných celků vybrat jeden, který by dané formulaci odpovídal přesněji než ostatní; v řadě textů takové konstrukce dokonce převažují. Tak například závěr věty *Denesky ráno já byla vsdavala 7:05 potom oblecení zima* můžeme interpretovat jako svého druhu ekvivalent formulace *... potom jsem si vzala oblečení do velké zimy (zimní oblečení)* nebo *... potom jsem se teple oblékla* nebo *... potom jsem si vzala teplé oblečení, protože venku byla zima* nebo *... rychle jsem se oblékla, protože mi byla zima* atd.

V současnosti se tedy v korpusu DEAF užívá systematická chybová anotace jen na nejnižší úrovni, konkrétně na úrovni formy jednotlivých lexémů (k tomu srov. u Šebesty [2010, s. 29–30] první dvě roviny vícerovinné anotace CZESL). Na této úrovni pak například kód *<#>* označuje řetězec písmen, který neodpovídá žádnému běžně rozšířenému českému slovu ani slovnímu tvaru, resp. podobě. Užití tohoto kódu, vkládaného za příslušná textová slova, je možno demonstrovat na anotaci první věty z výše uvedených ukázek: *Máme na dovolenou prýč <#> nikam někdý <#> tu mě už vždý <#> co dělat šla.* V dohledné době by však tuto elementární anotaci měla doplnit anotace propracovanější, která se zaměří i na některé další vybrané jevy z vyšších rovin, např. na vyjadřování shody (zejména u přísudku s podmětem, v konstrukcích s genitivem počítaného předmětu a u shodných přívlastků) a na záměnu, případně vynechávání *být* a *mít*, které je charakteristické pro texty téměř všech neslyšících. Po aplikaci tohoto rozšířeného pojetí bude například věta *Já jeli na hory* označena kódy signalizujícími porušení shody přísudku s podmětem. Ani v takových relativně jednoduchých případech však nebude uživatelům korpusu předkládán noremní ekvivalent daného vyjádření, neboť i ten by mohl být zavádějící. I takové větě jako *Já jeli na hory* je totiž jen zdánlivě možno jednoznačně přiřadit noremní ekvivalent *Já (jsem) jel na hory*: podle některých neslyšících mohl autor chtít ve skutečnosti napsat *My jsme jeli na hory*, přičemž užití zájmena *já* by mohlo být v dané větě výsledkem interference s českým znakovým jazykem, v němž zájmeno *my* může být vyjádřeno i inkluzivním plurálem JÁ + TY. Protože v předešlé větě autor psal o rodičích, nelze také vyloučit, že mělo jít o větu typu *Rodiče a já jsme jeli na hory*, *Já s rodiči jsme jeli na hory* atp.

Uvedené případy jsou jen ukázkami problémů, s nimiž se anotace textů pro korpus DEAF setkává. Úplný výčet anotačních zásad a kódů (vedle zmiňované značky *<#>* se využívá např. kódů pro značení cizojazyčných částí textu, pro části textu, které psal slyšící, atp.) bude zveřejněn na stránkách Ústavu Českého národního korpusu. Na tomto místě už jen stručně

15 K chybové anotaci blíže Šebesta, 2010, zvl. s. 26–27.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

upozorněme na to, že četné problémy, pro které je třeba najít řešení, se objevují už na zcela elementární úrovni. Například pro nastíněné – zdánlivě banální – značkování neexistujících slov a tvarů je třeba formulovat zvláštní zásady při anotování textů internetové komunikace, které jsou často psány bez diakritiky, nebo při anotování slov s nenoremní distribucí velkých a malých písmen.

Lemmatizace

Lemma je pro potřeby korpusu DEAF chápáno – z čistě praktických důvodů – poněkud šířeji, než bývá tradičně zvykem: kromě své obvyklé funkce (reprezentativní tvar zastupující množinu existujících tvarů lexému) zastupuje lemma v korpusu DEAF i nestandardní a nenoremní tvary či podoby, které v daném konkrétním kontextu byly užity na místě tvarů standardních a noremních (např. ve větě *Napsal byl jsem, ze prijedui* je k zápisu *ze* přiřazeno lemma *<že>* a k tvaru *prijedui* lemma *<přijet>*; podobně srov. výše řetězec *vereče* lemmatizovaný jako *<večeře>, vsdavala* jako *<vstávat>* atp.; vedle lemmatu je k těmto tvarům ovšem samozřejmě přiřazen i kód *<#>*). Tento přístup, při němž lemma slouží k vyhledávání všech (tedy i nenáležitých, nenoremních) podob a tvarů, jež neslyšící v textu užil, se snaží vyjít vstříc uživatelům korpusu a co nejvíce jim usnadnit vyhledávání v textech. Ze stejného důvodu byla modifikována i lemmatizace významově nejednoznačných výpovědí, s nimiž se lze v textech neslyšících poměrně často setkat. V korpusu DEAF bylo například zavedeno tzv. alternativní (dvojí) lemma pro případy, k jakým patří např. výše uvedené *jíslo* (vzhledem ke kontextu interpretovatelné, a tedy i lemmatizované jednak jako *<jíst>*, jednak jako *<jídlo>*). Možnosti přiřadit jednomu řetězci písmen dvě lemmata se využívá také v případě nenoremního spojitého psaní dvou slov (srov. např. zápis *podstolem* s dvojicí lemmat *<pod>* a *<stůl>*).

Za zmínku stojí i zavedení formálního lemmatu *<xxx>* pro neidentifikovatelné, resp. neinterpretovatelné zápisy, jako je výše zmíněné *cdraf* (*auto mam nutne po cdraf a spolecni*) nebo *aku* (*Marteka řikala pokoj s detiv aku a Brachym, Petra a já*). Stejné formální lemma *<xxx>* se uplatňuje rovněž v případech, kdy by počet možných interpretací zápisu byl vyšší než dvě – srov. např. formu *zpozdeji* užitou v kontextu *kdyby letadlo zpozdeji* interpretovatelnou jako defektní zápis tvarů patřících přinejmenším ke třem různým lemmatům: *zpoždění*, *později* a *zpozdit*.

Kromě specifických rysů komunikace neslyšících se lemmatizace textů pro korpus DEAF snaží reflektovat i některé zvláštnosti internetové komunikace (viz již výše), vyznačující se značnou variabilitou podob v rámci některých slovních druhů, konkrétně částic a citoslovcí (srov. např. varianty pozdravu na rozloučenou typu *pá*, *páá*, *papapa*, *papa*, *papá* atp.). Vzhledem k tomu, že podobné varianty částic a citoslovcí nelze dost dobře, resp. nenásilně soustředit k jednomu lemmatu, ponechávají se jim samostatná lemmata shodná s formou užitou v textu (a *pá* je tedy lemmatizováno jako *<pá>*, *páá* jako *<páá>*, *papapa* jako *<papapa>* atd.). Odlišný postup se naproti tomu volí při lemmatizaci obdobných variant v rámci jiných slovních druhů než částic nebo citoslovcí, tj. variant v rámci slovních druhů, kde je naznačená variabilita (např. *mmmmóóóócccc, chciii*) celkem výjimečná: s ohledem na nemožnost přenést v takových případech formální odlišnost užitého tvaru do podoby lemmatu a s ohledem na snazší vyhledávání dostávají podobné příležitostné varianty standardní lemmata, konkrétně v uvedených dvou příkladech *<moc>* a *<chtít>*.

Výše naznačené specifické lemmatizační zásady přijaté v počáteční fázi budování korpusu DEAF neuvádíme na tomto místě jako definitivní, dále zcela neměnné principy, ale primárně

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

jako ilustraci problémů, na něž může v praxi narážet lemmatizace méně běžných textů, nikoli nutně jen textů neslyšících. Vzhledem k tomu, že lemmatizace – stejně jako anotování – korpusu DEAF se v současné době teprve zahajuje a ověřuje, i vzhledem k neobyčejné různorodosti textů neslyšících lze naopak očekávat, že jednotlivá naznačená řešení bude třeba v budoucnu podle potřeb částečně modifikovat a doplňovat.

Základní informace o korpusu DEAF, které jsme se pokusili stručně shrnout v tomto příspěvku, zahrnují formulaci cílů projektu a jeho principů, obecnou charakteristiku jeho výchozího stavu i nástin vybraných řešení a zčásti otevřených problémů. Jak je zřejmé, DEAF – projekt, který zatím i v mezinárodním měřítku zůstává prvním ojedinělým krokem v naznačeném směru – stojí dosud na počátku své existence. Jisté předznamenání jeho přínosnosti lze dnes spatřovat v tom, že už začátky jeho budování upozorňují na některé okrajově prozkoumané jevy a jsou spojeny s řešením dosud jen málo uvědomovaných problémů nejen z oblasti psané češtiny neslyšících, ale i z oblastí jiných, zejména korpusové lingvistiky (lemmatizace nenoremních tvarů, defektních nebo neinterpretovatelných zápisů aj.). Jeho přínos bude ovšem možné zhodnotit až s delším odstupem, podle toho, jak bude plnit svá dvě hlavní praktická poslání (jako nástroj ke zkvalitnění výuky češtiny na školách pro sluchově postižené a jako platforma pro uchovávání záznamů o kultuře a historii neslyšících), a také podle toho, jak spolehlivou základnu bude s to při naznačených obtížích poskytnout výzkumu psané češtiny neslyšících. Jako o plně funkčním nástroji bude však možno o korpusu DEAF mluvit nepochybně teprve tehdy, až začne sloužit především českým neslyšícím, bez jejichž vstřícnosti a porozumění by mohl zůstat jen planým nerealizovaným plánem.

Literatura:

ANDERSON, Jacqueline (1993): *Deaf Student Mis-Writing, Teacher Mis-Reading (English Education and the Deaf College Student)*. Burtonsville, MD: Linstok Press.

BAKER-SCHENK, Charlotte – COKELY, Dennis (1980): *American Sign Language (A Teacher's Resource Text on Grammar and Culture)*. Washington, DC: Gallaudet University Press.

ČERMÁK, František – KRÁLÍK, Jan – KUČERA, Karel (1997): Recepce současné češtiny a reprezentativnost korpusu. *Slovo a slovesnost*, 58, s. 117–124.

GALLAWAY, Clare (2001): Raná interakce. In: Susan Gregory – Pamela Knight – Wendy McCracken – Stephen Powers – Linda Watson (eds.), *Problémy vzdělávání sluchově postižených*. Praha: FF UK, s. 54–61.

HANKS, Patrick W. (1990): Evidence and Intuition. In: Jerzy Tomaszczyk – Barbara Lewandowska-Tomaszczyk (eds.), *Meaning and Lexicography*. Amsterdam: John Benjamins, s. 31–41.

HRUBÝ, Jaroslav (1999): *Velký ilustrovaný průvodce neslyšících a nedoslýchavých po jejich vlastním osudu. 1. díl*. Praha: Federace rodičů a přátel sluchově postižených.

HUDÁKOVÁ, Andrea (2008): *Čeština ve vzdělávání dětí s vadou sluchu*. Praha: FF UK v Praze. [Disertační práce].

CHURCH, Kenneth W. – MERCER, Robert L. (1993): Introduction to the Special Issue on Computational Linguistics Using Large Corpora. *Computational Linguistics*, 19, s. 1–24.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

JOHNSTON, Trevor – SCHEMBRI, Adam (2007): *Australian Sign Language – Auslan (An Introduction to Sign Language Linguistics)*. Cambridge: Cambridge University Press.

KOCEK, Jan – KOPŘIVOVÁ, Marie – KUČERA, Karel (eds.) (2000): *Český národní korpus (Úvod a příručka pro uživatele)*. Praha: FF UK v Praze – Ústav Českého národního korpusu.

MACUROVÁ, Alena (1995): ... protože já bavím spolu vypravovat. (Komunikace v dopisech českých neslyšících). *Slovo a slovesnost*, 56, s. 23–33.

MACUROVÁ, Alena (1998): Naše řeč? *Naše řeč*, 81, s. 179–188.

MACUROVÁ, Alena (2001): Jazyk a hluchota. *Slovo a slovesnost*, 62, s. 92–104.

MAYBERRY, Rachel I. (1998): The critical period for language acquisition and Deaf child's language comprehension : a psycholinguistic approach. In: *Bulletin d'Audiophonologie: Annales Scientifiques de L'Université de Franche-Comté*, 15, s. 349–358. Dostupné z WWW: <http://www.acfos.org/publication/ourarticles/pdf/acfos1/intro_mayberry.pdf>.

POLÁKOVÁ, Marie (2000): *Čtení s porozuměním? Čeští neslyšící a české texty*. Praha: FF UK v Praze. [Diplomová práce].

POLÁKOVÁ, Marie (2001): Čtení s porozuměním a čeští neslyšící. *Speciální pedagogika*, 11, s. 207–219.

POLÁKOVÁ, Marie (2002): Specifické strategie využívané neslyšícími respondenty v testech čtenářských dovedností. *Čeština doma a ve světě*, 10, s. 168–173.

QUIGLEY, Stephen P. – PAUL, Peter V. (1984): *Language and Deafness*. San Diego, CA: College-Hill Press.

RACKOVÁ, Markéta (1996): *České sloveso v psaných projevech českých neslyšících*. Praha: FF UK v Praze. [Diplomová práce].

RICHTEROVÁ, Klára (2009): *Vyjadřování času v psané češtině českých neslyšících*. Praha: FF UK v Praze. [Diplomová práce].

STRNADOVÁ, Věra (1998): *Hádej, co říkám aneb odezírání je nejisté umění*. Praha: Ministerstvo zdravotnictví ČR.

ŠEBESTA, Karel (2010): Korpusy češtiny a osvojování jazyka. *Studie z aplikované lingvistiky / Studies in applied linguistics*, 2010/2, s. 11–33.

ŠEBKOVÁ, Helena (2008): *Psaná čeština českých neslyšících v internetové komunikaci*. Praha: FF UK v Praze. [Diplomová práce].

ŠMILAUER, Vladimír (1972): *Nauka o českém jazyku*. Praha: SPN.

VANĚK, Miroslav (2004): *Orální historie ve výzkumu soudobých dějin*. Praha: Centrum orální historie Ústavu pro soudobé dějiny AV ČR.

VANĚK, Miroslav – MÜCKE, Pavel – PELIKÁNOVÁ, Hana (2007). *Naslouchat hlasům paměti (Teoretické a praktické aspekty orální historie)*. Praha: Ústav pro soudobé dějiny AV ČR.

SALI 2011/1 /6.9.2011 – SALI_2011_01_04.indd

WÁGNEROVÁ, Jana (2008): *Kategorie rodu v češtině českých neslyšících*. Praha: FF UK v Praze. [Diplomová práce].

ZIEBIKEROVÁ, Wanda (2009): *Interference z českého znakového jazyka do psané češtiny českých neslyšících*. Praha: FF UK v Praze. [Diplomová práce].

ZÝKOVÁ, Markéta (1997): Sloveso v psané češtině neslyšící mládeže. *Speciální pedagogika*, 7, s. 16–25.