

PŘÍPRAVA KORPUSU AFATICKÉ ČEŠTINY¹

V průběhu roku 2015 začal na FF UK vznikat projekt, jehož cílem je vybudování korpusu mluvené češtiny osob s afázií. Korpus se stane součástí rodiny korpusů mluvené češtiny a technické zázemí bude zajištěno Ústavem Českého národního korpusu. Zamýšlenými uživateli jsou jak lingvisté pracující s češtinou či slovanskými jazyky obecně, tak kliničtí pracovníci. V následujícím textu představíme základní rysy korpusu a informujeme o průběhu končící první fáze, jejímž cílem bylo především připravit postupy pro sběr a zpracování dat.

Afázie je získané neurogenní narušení jazykových a komunikačních schopností, které je způsobeno cerebrálními lézemi v oblastech spojovaných s jazykovými schopnostmi. Jedná se o závažnou poruchu, která významně zasahuje do každodenního fungování člověka. Při výzkumu tohoto fenoménu se setkává neurologie, logopedie, lingvistika a mnohé další disciplíny. V českém prostředí se lingvistická práce zaměřená na jazykové poruchy omezuje na poměrně malý rozsah studií (např. Lehečková, 2001; Flanderková et al., 2014). V zahraničí se naproti tomu objevují jak datově orientované analýzy konkrétních řečových profilů, tak především obecněji orientované práce, jež při formulaci teoretických východisek zohledňují patologická data či z nich přímo vycházejí.² Nutno poznamenat, že většina výzkumů probíhá zejména na angličtině, případně na dalších, strukturně jí poměrně blízkých jazycích s nízkou mírou morfologické komplexnosti. V oblasti klinické logopedie je situace v českém kontextu lepší, především co se vyšetřovacích nástrojů týče. Vedle standardizované adaptace skříninkového testu MASTcz (Košťálová et al., 2008) existuje např. původní Vyšetření fatických funkcí (Cséfalvay — Košťálová — Klimešová, 2002).

Bez větších pochybností lze říci, že jedním z důvodů, proč jsou tato specifická data nejen v českém lingvistickém prostředí málo využívána, je jejich obtížná dostupnost. Cílová klinická populace je poměrně malá a velmi heterogenní, a je proto obtížné získat v uspokojivé míře vnitřně srovnatelnou skupinu. Obtížný přístup k populaci pak v českém kontextu může v některých případech plynout z komunikace s ošetřujícími terapeuty, kteří logicky nemají s tímto druhem výzkumu příliš mnoho zkušeností. Práce s cílovou populací je eticky mimořádně citlivá. V rovině lidské pak přináší interakce s osobami s afázií pochopitelně též výrazná specifika. Abychom zmírnili tyto nároky a podpořili rozvoj této výzkumné oblasti, je cílem představovaného projektu vytvoření korpusu afatické řeči, který bude dostupný nejenom výzkumníkům zabývajícím se touto tematikou, ale i lingvistům pracujícím s češtinou či slovanskými jazyky, kteří by chtěli své předpoklady aplikovat i na tato speciální data. Korpus je

1 Tento výstup vznikl v rámci projektu Mezilidská komunikace v interdisciplinární perspektivě, podprojektu Příprava korpusu afatické řeči řešeného na FF UK z prostředků specifického vysokoškolského výzkumu na rok 2015. Autor děkuje V. Diatkovi, T. Papayové a K. Vaníčkové, kteří na projektu spolupracují.

2 To platí především pro formálně a generativně orientované autory (např. Friedmann, 2001), v poslední době se však objevují též studie vycházející z funkčně a povrchově orientovaného prostředí (Gahl, 2002).



tedy budován pro poměrně široké publikum, proto je snahou poskytnout maximální možné množství informací.

Naše práce zároveň navazuje na podobné projekty v zahraničí. Nutno poznamenat, že i ve světovém měřítku se jedná o oblast nedlouho se rozvíjející a omezenou opět především na anglofonní prostředí. Z existujících korpusů lze jmenovat především AphasiaBank (MacWhinney et al., 2011), který je postaven na známém systému CHILDES. Tato databáze je složena z nahrávek sezení, během nichž byl administrován rigidně formulovaný protokol zaměřený na různé oblasti jazykového systému. Korpus obsahuje přibližně hodinové nahrávky 258 participantů s afázií a 180 neurologicky zdravých participantů. V současné době je rozšiřován o další jazyky, např. španělštinu či mandarínskou čínštinu. Druhým příkladem je pak Cambridge Cookie Theft Corpus (Williams et al., 2010), kde jsou shromážděny narativy elicítované s pomocí stejnojmenného stimulového obrázku z testu Western Aphasia Battery (Kertesz, 2007) a také krátké spontánní projevy. Pro lepší srovnatelnost korpus opět obsahuje i texty získané od kontrolní skupiny zdravých mluvčích.

Cílem první, přípravné fáze projektu bylo shromáždit projevy 10 osob s afázií v přibližné délce 10 hodin a na tomto materiálu především formulovat pravidla pro přepis nahrávek a celkové zpracování dat.

Po navržení hrubé struktury sezení byl připraven formulář informovaného souhlasu. Jak bylo již zmíněno výše, etické hledisko má ve výzkumu afázie, podobně jako ve výzkumu dětské řeči, zvláštní význam. Snahou výzkumníka by zde ideálně mělo být, aby byl potenciální participant v rozhodovacím procesu maximálně zapojen, čemuž by měl odpovídat styl a jazyk informovaného souhlasu. Samotný formulář jsme připravili na základě doporučení v literatuře (Palmer — Paterson, 2011). Informace byly prezentovány v bodech, větším písmem a s využitím neterminologického jazyka. Klinický profil jednotlivých participantů navíc dovoľoval předpoklad, že by s takovým textem neměl být výrazný problém. Při samotném procesu vždy administrátor (ML) seděl s participanty a procházel jednotlivé body a případně dovysvětlil nejasná místa.³

Z možných pojetí, která přicházela v úvahu při přípravě osnovy sezení, jsme na rozdíl od zmíněné AphasiaBank zvolili přístup zaměřený výhradně na elicítaci různých typů diskurzu s důrazem na pokud možno přirozený konverzační diskurz. Pro jednotlivá sezení tak byly připraveny podněty, jejichž cílem bylo elicítovat deskriptivní, narativní a procedurální diskurz. V prvním případě se jednalo o několik obrázků Josefa Lady. Narativní diskurz jsme získali jednak za pomoci dvou krátkých obrázkových příběhů, dle nichž participanti tvořili jednoduché příběhy, dále bylo úkolem participantů převyprávět krátký výstup Charlese Chaplina. Procedurální diskurz spočíval v popisu jednoduchého receptu, případně popisu cesty, při němž participanty pracovali s kresleným plánkem.

Konverzační diskurz se držel okruhu předem připravených témat, která zahrnovala například studijní a pracovní zkušenosti, participantovy zájmy a volnočasové aktivity či průběh logopedické terapie. Sezení ovšem neměla pevně stanovenou

³ Než byl vůbec umožněn přístup k potenciálním participantům, byl samozřejmě formulář zkontrolován a schválen příslušnou institucí, která kontakt zprostředkovala.



strukturu, jednotlivá témata, úkoly a jejich pořadí byly přizpůsobovány situačnímu kontextu. Sezení pak byla nahrávána videokamerou, v samotném korpusu je ovšem použita pouze zvuková stopa, zatímco obraz měl jen podpůrnou funkci při přepisu. Rozhodnutí vyloučit z korpusu obraz bylo vedeno především tím, že by tento faktor mohl odradit velké množství potenciálních účastníků.

Z oslovených klinických pracovišť se podařilo navázat kontakt s FN Brno, Nemocnicí Písek a občanskou společností Klub afasie, která organizuje terapeutické skupiny. Ve spolupráci s těmito pracovišti jsme pořídili nahrávky s 11 účastníky v délce 35–70 minut. Pro výběr účastníků nebyla předem stanovena žádná kritéria — zařadili jsme všechny oslovené, kteří měli zájem se projektu zúčastnit. Zároveň jsme však v prvotní fázi projektu z důvodů především praktických cílili primárně na osoby v chronické fázi postižení s lehkými anamnestickými profily a určitou mírou fluence. Takové účastníky je jednak poměrně snazší oslovit, zároveň jsme předpokládali, že získané projevy budou spíše méně problematické z hlediska transkripce, což se zdálo výhodné ve fázi, kdy byla pravidla přepisu teprve formulována. Samozřejmostí jsou v tomto případě kromě standardních demografických údajů o účastnících též anamnestická data, která umožní identifikaci a porovnatelnost projevů jednotlivých účastníků. Klinický profil účastníka zaznamenává dominantní ruku, rok vzniku a počet měsíců od vzniku poruchy, etiologii, lokalizaci léze, typ poruchy po vzniku, případně výkony v některém z užívaných diagnostických testů.

Zvolená transkripční pravidla kombinují přepisovací pravidla užívaná v korpusch mluvené češtiny řady ORAL s některými zvyklostmi konverzačněanalytickými a některými dalšími úpravami. Změny se oproti stylu korpusu ORAL2013 (Benešová — Křen — Waclawíčová, 2015) týkají především heziticí, pauz a dysfluencí obecně. Vedle samotného transkriptu, který je propojen se zvukovou stopou, bude korpus obsahovat morfologické značky, lemmatizaci a chybovou anotaci. Zatímco pro lemmatizaci a morfologické značkování užíváme software používaný Ústavem Českého národního korpusu (Straková — Straka — Hajič, 2014), v chybové anotaci jsou zachyceny jevy typické pro projev osob s afázií. Taková značka zaznamenává typ chyby (např. parafázie, agramatismus), cílový segment a následnou opravu včetně toho, zda byla úspěšná a zda do ní zasáhl komunikační partner. V relevantních místech jsou v transkriptu připojeny též drobné komentáře, které např. dovyšetlují situační kontext či zaznamenávají pomocná a kompenzační gesta provázející problematické úseky.

V současné době probíhá zpracování sebraných dat, korpus přitom bude postupně zveřejňován v první polovině roku 2016. Po zveřejnění bude korpus přístupný registrovaným uživatelům z řad akademických a klinických pracovníků. Věříme, že korpus najde využití jak v čistě lingvistickém výzkumu, tak v klinické logopedii, ať už při přípravě budoucích terapeutů, nebo ve výzkumu. Naším cílem je v další fázi databázi otevřít všem výzkumníkům, kteří by chtěli sdílet relevantní data. Zároveň bude korpus postupně rozšiřován tak, aby pokryl problematiku češtiny v afázii v celé její šíři a aby v něm byly zastoupeny všechny dominantní diagnostické typy.

LITERATURA:

- BENEŠOVÁ, Lucie — KŘEN, Michal — WACLAWIČOVÁ, Martina (2015): Korpus spontánní mluvené češtiny ORAL2013. *Časopis pro moderní filologii*, 97(1), s. 42–50.
- CSÉFALVAY, Zsolt — KOŠTÁLOVÁ, Milena — KLIMEŠOVÁ, Marta (2002): *Diagnostika a terapie afázie, alexie, agrafie: (Manuál)*. Praha: Asociace klinických logopedů ČR.
- FLANDERKOVÁ Eva — MERTINS, Barbara — BEZDÍČEK, Ondřej — BABOROVÁ, Eva — ČERNÁ, Marie (2014): Posuzování gramatičnosti v Brocově afázii — příklad dvou pacientů. *Česká a slovenská neurologie a neurochirurgie*, 77/110(2), s. 202–209.
- FRIEDMANN, Na'ama (2001): Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research*, 30(1), s. 71–90.
- GAHL, Susanne (2002): Lexical biases in aphasic sentence comprehension: an experimental and corpus linguistic study. *Aphasiology*, 16(12), s. 1173–1198.
- KERTESZ, Andrew (2007): *Western Aphasia Battery Revised*. San Antonio, TX: PsychCorp.
- KOŠTÁLOVÁ, Milena — BÁRTKOVÁ, Eva — ŠAJGALÍKOVÁ, Kateřina — DOLENSKÁ, Alice — DUŠEK, Ladislav — BEDNAŘÍK, Josef (2008): A standardization study of the Czech version of the Mississippi Aphasia Screening Test (MASTcz) in stroke patients and control subjects. *Brain Injury*, 22(10), s. 793–801.
- LEHEČKOVÁ, Helena (2001): Manifestation of aphasic symptoms in Czech. *Journal of Neurolinguistics*, 14(2), s. 179–208.
- MACWHINNEY, Brian — FROMM, Davida — FORBES, Margaret — HOLLAND, Audrey (2011): AphasiaBank: methods for studying discourse. *Aphasiology*, 25(11), s. 1286–1307.
- PALMER, Rebecca — PATERSON, Gail (2011): One size does not fit all: obtaining informed consent from people with aphasia. *Advances in Clinical Neuroscience and Rehabilitation*, 11(2), s. 30–31.
- STRAKOVÁ, Jana — STRAKA, Milan — HAJIČ, Jan (2014): Open-source tools for morphology, lemmatization, POS tagging and named entity recognition. In: Kalina Bontcheva — Zhu Jingbo (eds.), *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, MD: Association for Computational Linguistics, s. 13–18.
- WILLIAMS, Caroline — THWAITES, Andrew — BUTTERY, Paula — GEERTZEN, Jeroen — RANDALL, Billi — SHAFTO, Meredith — DEVEREUX, Barry — TYLER, Lorraine (2010): The Cambridge Cookie-Theft Corpus: a corpus of directed and spontaneous speech of brain-damaged patients and healthy individuals. In: Nicoletta Calzolari — Khalid Choukri — Bente Maegaard — Joseph Mariani — Jan Odijk — Stelios Piperidis — Mike Rosner — Daniel Tapias (eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta: European Language Resources Association (ELRA), s. 2824–2830.