

# Mluvená čeština v televizních debatách: korpus DIALOG\*

SVĚTLA ČMEJRKOVÁ – LUCIE JÍLKOVÁ – PETR KADERKA

## Spoken Czech in television debates: The DIALOG corpus

ABSTRACT: The DIALOG corpus is one of two collections of spoken language gathered in the audio-visual studio at the Czech Language Institute of the Czech Academy of Sciences. The article begins by recalling the establishment of the corpus in 1997 as part of the project “Dialogue in a World of People and Machines”, defines the aim motivating the collection of data for this corpus, formulates distinctive criteria for this corpus as a specifically “spoken” one in terms of time, interaction and genre and partially even as topic-specific, and attempts to define the types of spoken dialogues which the corpus can aid in analysing. It characterizes speech in the media, which makes up a focal point here, and details the procedures for storing audio and video recordings of this speech and the resulting transcriptions. The second part provides an overview of the fundamentals of transcription systems and offers theoretical support for transcription method selection as determined by the aim of capturing segmental, supra-segmental, sequential, para-linguistic and extra-linguistic phenomena, including several examples of practical solutions. The third part reports on how this corpus has been thus far utilized in linguistic research, both in the creation of a contemporary Czech theory of dialogue and in the analysis of specific features of spoken Czech. The article concludes by detailing the prospects for further use of this corpus.

## 1. Lingvistická analýza mluveného jazyka

### 1.1. *Specifičnost lingvistické analýzy mluveného vs. psaného jazyka*

Lingvistická analýza mluveného jazyka má k dispozici kritéria, nástroje a mechanismy, jež byly vypracovány při analýze jazyka psaného; o ty se však může opřít jen v omezené míře a navíc, opře-li se o ně, může se mluvený projev – tento svébytný existenční modus jazyka mající výraz v souhře slov a gest nejen zvukových, ale i motorických, ve výrazu tváře a pohybech rukou provázejících jednak proud vlastní řeči, jednak recepci řeči cizí – jevit jako „nedostatečný“, zaostávající za vyšší mírou obsahové úplnosti a vnější uspořádanosti textu psaného. Lingvistická analýza mluveného jazyka se takovému neadekvátnímu a zkreslujícímu hodnocení mluvenosti (dialogičnosti) samozřejmě brání a snaží se proniknout k tomu, co je pro mluvenost a dialogičnost konstitutivní, čím je dána její svébytnost a specifičnost.<sup>1</sup> Setkává se ovšem v tomto svém úsilí s poměrně značnou variabilitou mluvených projevů. Mluvené projevy mohou být totiž připravené, obsahově promyšlené, syn-

\* Tento článek vznikl s podporou grantu GA AV ČR č. KJB9061304.

<sup>1</sup> Lingvistická literatura zabývající se jak charakteristikou psaného a mluveného modu jazyka, tak teorií monologu a dialogu je dnes již velmi bohatá, a máme-li si připomenout alespoň základní přístupy, jež při studiu tohoto tématu nelze zejména v českém prostředí opomenout, pak je třeba zmínit tradici Pražské školy, v níž vznikla Vachkova teorie psaného a mluveného jazyka i Mukařovského průkopnické práce o monologu a dialogu. Zájem o myšlenky českých, ale i světových lingvistů týkající se psanosti a mluvenosti provázal českou lingvistiku i v dalších desetiletích, výrazně ožil v 70. a 80. letech v pracích Hausenblasových a Chloupkových i v pracích jejich žáků, zejména J. Hoffmannové a O. Müllerové, stejně tak jako v pracích českých rusistů, V. Barneta, D. Brčákové, K. Koževnikovové, O. Kafkové, I. Camutaliové a dalších. V 90. letech se tomuto tématu věnovala pražská mezinárodní konference Writing vs Speaking pořádaná na půdě Ústavu pro jazyk český, jejímž příspěvkům bylo kromě stejnojmenného sborníku (viz Čmejrková – Daneš – Havlová, 1994) věnováno i speciální číslo Slova a slovesnosti, roč. 54, 1993, č. 3 (případní zájemci o danou oblast lingvistického zkoumání tu naleznou bibliografické odkazy, jejichž výčet přesahuje možnosti této stati).

takticky koherentní a stylisticky vybroušené, takže vcelku splňují i nároky kladené na texty psané; projevy mluvcích mohou být ale i zcela spontánní a v důsledku kontextového zapojení a situační podmíněnosti – poměrujeme-li je kritérii psaného modu jazyka – obsahově neúplné a syntakticky nekoherentní. Takové projevy mohou být pro vnějšího pozorovatele (a tedy i lingvistu), který nebyl účasten situace vzniku projevu (nebyl jejím participantem), vzhledem ke své útržkovitosti a implicitnosti dokonce nesrozumitelné nebo těžko interpretovatelné (srov. např. Weigand, 2003).

Lingvistická charakteristika mluveného jazyka závisí tedy nepochybně do značné míry na tom, zda učiní předmětem analýzy projevy veřejné, určené širšímu publiku, včetně mediálního (srov. Čmejrková – Hoffmannová, 2003), řeč vedenou v zaměstnání či na úřadech (srov. Hoffmannová – Müllerová, 2000), anebo řeč, jakou se mluví v prostředí neveřejném, např. domácím, rodinném (srov. Hoffmannová – Müllerová – Zeman, 1999).

Proměna sociálních a kulturních hodnot v průběhu posledních desetiletí způsobila, že ostrou hranici mezi lingvistickými rysy mluvených projevů veřejných a neveřejných dnes vést nelze, vzhledem ke klesající prestiži řeči charakterizované rysy oficiálnosti a formálnosti a rostoucí prestiži řeči záměrně prezentující neoficiálnost, neformálnost, spontánnost a bezprostřednost, ať už přátelskou, srdečnou a vlídnou, anebo kritickou, konfliktní až útočnou. Zřetel k prostředí, v němž se řeč odehrává, a k jejímu žánrovému zařazení však může přesto přinést důležité poznatky o tom, jaké lingvistické rysy současné mluvené češtiny lze chápat jako neutrální, nepříznačové a polyfunkční a jaké rysy se naopak jeví jako žánrově vázané a příznačové.

## 1.2. *Materiálové zdroje popisu mluveného jazyka*

Lingvistická analýza současné mluvené češtiny si klade za cíl (srov. např. výše citované práce českých autorů a autorek) analyzovat mluvené projevy všeho druhu, monologické i dialogické, neveřejné i veřejné. Česká lingvistika postoupila v této své snaze poměrně daleko, díky tomu, že na řadě bohemistických pracovišt se systematicky shromažďují autentické nahrávky mluvené řeči (dříve její audiozáznamy a dnes už i videozáznamy), které jsou přepisovány podle různých transkripčních pravidel a posléze analyzovány z rozmanitých hledisek (fónického, morfologického, syntaktického, stylistického, interakčního apod.). Vznikající korpusy mluvené řeči se od sebe odlišují svým rozsahem, časovou hloubkou, zaměřením tematickým, žánrovým, teritoriálním atd. Poslední hledisko od sebe jasně odlišuje tzv. Pražský mluvený korpus (PMK), budovaný v Ústavu českého národního korpusu na FF UK v Praze, a podle paralelních zásad vystavěný Brněnský mluvený korpus (BMK), budovaný na FF MU v Brně.<sup>2</sup> V jiných ohledech jsou oba zmíněné korpusy srovnatelné, včetně toho, že využívají v zásadě stejné transkripční principy.

Ústav pro jazyk český disponuje v současné době (kromě bohatých materiálových sbírek pořizovaných po desetiletí pro účely dialektologického výzkumu, jež jsou v poslední době rovněž digitalizovány, srov. Bachmannová, 1999) dvěma sbírkami běžně mluveného jazyka, z nichž starší, vznikající iniciativou a péčí O. Müllerové od 70. let a průběžně doplňovaná,

<sup>2</sup> Český národní korpus – PMK, BMK, Ústav Českého národního korpusu FF UK; dostupné online na adrese: <<http://ucnk.ff.cuni.cz>>.

obsahuje nahrávky a přepisy autentické řeči žánrově rozmanité, druhá a mladší sbírka soustřeďuje od poloviny 90. let nahrávky řeči mediální, tj. především nahrávky mluvených projevů určených televizním divákům; okrajově jsou v ní zastoupeny též nahrávky pořadů rozhlasových, ale těch je přepsáno méně (13) a užívají se např. při srovnání mluvené dialogické řeči televizní a rozhlasové (charakterizované např. vyšším stupněm explicitnosti odkazů k řečové situaci, častější identifikací mluvčích pojmenováním nebo oslovením, zmínkami o neverbálním chování hostů a dění ve studiu apod.). Vzhledem k tomu, že se jedná o nahrávky dialogických žánrů, jako je interview, diskuse, debata, polemika, talkshow a další typy rozhovorů, nazvali jsme korpus vznikající na základě televizní části této sbírky mluvených textů korpus DIALOG.

Korpus DIALOG odpovídá nárokům na analýzu mluvené řeči, zejména v její primární, dialogické podobě. Jde o tzv. specializovaný korpus (o specializovaných korpusech viz např. Čermák 1995, s. 123), jehož pomocí lze ověřovat výsledky, k nimž korpusová lingvistika dospěla analýzou jiných zdrojů, zejména psaného jazyka, soustředěného, pokud jde o češtinu, v korpusu SYN2000.<sup>3</sup> Stejně důležité a pozoruhodné jako porovnání výsledků získaných z dat psaného a mluveného jazyka je však i porovnání výsledků získaných z jednotlivých korpusů mluveného jazyka; v některých případech se data získaná z korpusu DIALOG mohou lišit i od dat obsažených v PMK a BMK (srov. Čmejrková, v tisku): Zda tomu tak je jen v případech ojedinělých, či častěji, je věcí dalšího zkoumání, časově poměrně dosti náročného. Porovnávání obrazu češtiny v korpusech obecných, resp. nespecifických na jedné straně a specializovaných na straně druhé však rozhodně stojí za to, neboť přispěje jak k pochopení možností a mezí korpusové lingvistiky, tak k vytvoření lingvisticky propracovaného obrazu současného českého jazykového a komunikačního úzu. Představa, že korpusová lingvistika, založená na analýzách kvantitativních a frekvenčních, může uspět jen tehdy, bude-li si vědoma vázanosti svých výsledků na materiálové zdroje, z nichž tyto výsledky pocházejí, je samozřejmou součástí současného uvažování o reprezentativnosti lingvistických korpusů. Maxima korpusové lingvistiky radí: Nespoléhej se na jeden velký korpus, ale porovnávej výsledky získané z různých korpusů (srov. Stubbs, 2000: „It is unwise to rely on a single corpus, however large or well designed it might be: all corpora have in-built biases, and findings should therefore be checked in different independent corpora“).

### 1.3. *Vznik a současný stav korpusu DIALOG: žánrová charakteristika*

V roce 1997 bylo na půdě Ústavu pro jazyk český AV ČR založeno audiovizuální studio, ve kterém se systematicky pořizují a archivují audionahrávky a videonahrávky mluvených projevů dialogického typu, vysílaných jednak rozhlasem, jednak televizí. Archiv obsahuje vedle nahrávek kratších, několikaminutových (krátká studiová interview v rámci zpravodajsko-publicistického pořadu „21“), nahrávky pořadů většinou čtyřiceti- až šedesátiminutových (debaty typu Sedmička, Nedělní partie a jiné publicistické pořady, v nichž jsou ve studiu spolu s moderátorem či moderátory přítomni jeden, dva i více diskutujících hostů, jako je tomu např. v pořadu Netopýr, Na hraně apod.), a také nahrávky zábavných pořadů dva-

<sup>3</sup> Český národní korpus – SYN2000, Ústav Českého národního korpusu FF UK; dostupné online na adrese: <<http://ucnk.ff.cuni.cz>>.

ceti- až třicetiminutových, v nichž si moderátor zve do studia jednoho hosta (typ Z očí do očí, Na plovárně s...) i více hostů k přátelskému popovídání (typ Krásný ztráty aj.). Popovídání může být i velmi neformální (Katovna), pohybovat se svou provokativností na hranici konstituovaného televizního žánru (Síto), případně jej zpochybňovat (Ženský hlas) nebo parodovat (Další, prosím aj.).

Korpus DIALOG se vyznačuje žánrovou uceleností a určitou komplementárností dialogických žánrů, pořady exponující mluvenou dialogickou řeč však v žádném případě nejsou stylově homogenní. Jak je patrné z výše uvedených příkladů zpravodajsko-publicistických a zábavných pořadů, korpus obsahuje diskuse na závažná témata společenská, sahající k stylovému rejstříku neutrálnímu až vyššímu, i hovor o záležitostech privátních, svádějící přirozeně k čerpání ze stylového rejstříku nižšího. Zatímco v některých debatách jsou promluvy hosta omezeny pečlivě střeženými pravidly strohého interview, redukcími hostovy repliky na informační minimum (typ Naostro, Otázky Václava Moravce aj.), v jiných debatách je hostům dána možnost, aby se volně rozhovořili o tématech, jež si sami vyberou, jež jsou jim blízká a jimž je moderátor ochoten pozorně a bez vměšování naslouchat. To se promítá do stylu rozhovoru: na jednom konci stylové škály nacházíme rychlý sled otázek a odpovědí charakteristický pro žánr interview, v němž redaktor hosta neustále usměřňuje, aby jeho řeč nevybočila ze směru udaného otázkou, na druhém konci nacházíme mluvené projevy s prvky vyprávění, úvah, vzpomínek, sebereflexí, zpovědí, humoristických vstupů a postupů ironizujících i sebeironizujících. Klademe-li si otázku, co tyto útvary mluvené řeči spojuje, pak je to právě jejich zasazení do žánrového rámce dialogu, a ještě přesněji do žánru mediálního dialogu, který se odehrává mezi moderátorem a hostem/hosty a který je zároveň obrácen k televizním divákům (srov. Čmejrková, 1999). Tato okolnost je pro analýzu a interpretaci mediálního dialogu důležitá, neboť zasahuje do výstavby dialogu nejen způsobem explicitním (diváci jsou opakovaně oslovováni a zmiňováni), ale i implicitním (divák je úběžníkem strategií užívaných moderátorem i interviewovanými hosty):

PŠ: =ale já mám strach že toto, (.) v tudle chvíli naše diváky moc nezajímá. klíč je, jsou ekonomické problémy...

(DIALOG: Nedělní partie, Prima, 12. 9. 1999; PŠ – Petr Šimůnek, moderátor)

MZ: [ale já uznávám že (.) vaše (.) diváky, ani tak nezajímá, moje chaloupka, tedy fiktivní chaloupka, (.) ale: e zajímá (.) mimo jiné akce olovo.

(DIALOG: Naostro, ČT 1, 3. 9. 2000; MZ – Miloš Zeman, předseda vlády a předseda ČSSD)

JL: jasný. já jenom aby tady (.) n:bylo (.) tvrzení proti tvrzení že abysme: abysme diváky nevodili za nos, ...

JV: [...jen abysme to] přeložili pro diváky. čili chcete říct...

(DIALOG: 7 čili Sedm dní, Nova, 4. 1. 1998; JL – Josef Lux, předseda KDU-ČSL; JV – Jan Vávra, moderátor)

MM: můj sta- m- m- můj vzkaz divákům je jediný. pokud to tak poslouchají, je zcela jasné. myslíte na zadní kolečka, nespolehejte na stát ale na sebe a spojte na stáří.

(DIALOG: Sedmička, Nova, 22. 10. 2000; MM – Miroslav Macek, místopředseda ODS)

#### 1.4. *Technické údaje*

V současné době (údaje uvádíme k 1. 8. 2004) se v archivu nachází 350 videokazet většinou o délce 240 minut; celkem je nahráno 1971 pořadů, což představuje přibližně 80 000 minut dialogické řeči. Část nahrávek je přepsána: existuje 349 prepisů představujících tex-

tový korpus čítající podle odhadu asi jeden a půl až dva miliony slov. Nahrávky jsou přepsány podle jednotných transkripčních zásad, jejichž specifickým znakem je soubor značek pro zachycení prozodických rysů mluveného projevu (viz níže). Tím se transkripční zásady užitě v korpusu DIALOG odlišují od zásad užitých v PMK a BMK. Důvodem je využívání korpusu nejen pro analýzu morfologickou, syntaktickou a lexikální, ale i pro analýzu mechanismů komunikačních, interakčních a pragmatických. Pro vyhledávání v korpusu využíváme korpusový software WordSmith Tools.<sup>4</sup>

Transkripty televizních pořadů byly pořizovány pomocí textového editoru MS Word a ukládány do paměti počítače ve formátu DOC. Pro další počítačové zpracování bylo výhodné převést veškeré transkripty do standardizovaného formátu XML (dle zásad Text Encoding Initiative, viz <<http://www.tei-c.org/Guidelines2/index.html>>). V tomto formátu se neztratí žádná z informací vložených do transkriptu a formát navíc umožňuje využít informace kódované pomocí transkripčních značek pro vyhledávání. Uživatelé to např. poskytuje možnost vyhledávat v korpusu podle jednotlivých mluvčích nebo se zaměřit na výrazy či pasáže, které byly jednotlivými mluvčími zdůrazněny, apod. Převedení do formátu XML je také předpokladem pro obohacení transkriptu o další informace (kódování/anotace). V lingvistické praxi přichází v úvahu především kódování lingvistických informací, zejména gramatických kategorií (k různých úrovním lingvistické anotace viz Leech, 2000).

V první fázi konverze do formátu XML byly zpracovány 203 přepisy.<sup>5</sup> Tyto přepisy obsahují kolem 1,3 miliónu slovních tvarů. Konvertovaná data byla posléze automaticky lemmatizována (každému slovnímu tvaru bylo přiřazeno lemma, tj. základní tvar slova) a obohacena o morfologické značky (podle systému J. Hajiče, viz Hajič, 2000).

Pro další využívání korpusu DIALOG je velmi podstatné, že původní nahrávky na videokazetách VHS postupně digitalizujeme.<sup>6</sup> V budoucnu tak bude možné pohodlně pracovat nejen s transkripty, které považujeme za dobrou, ale omezenou pomůcku, ale především s původními jazykovými daty, která jsou nejlépe reprezentována audiovizuálním záznamem. Při současných technických možnostech je dobře představitelné synchronní propojení transkriptu a zvuku,<sup>7</sup> ideální by ovšem bylo synchronní propojení transkriptu nejen se zvukem, ale také s obrazem.<sup>8</sup>

---

<sup>4</sup> WordSmith Tools, version 3, (viz <<http://www.lexically.net/wordsmith/index.html>>) obsahuje několik korpusových nástrojů: *Concord* vyhledává slova nebo slovní spojení ve vybraných textech a vypočítává kolokace; *WordList* vytváří seznamy slov a slovních spojení z vybraných textů a vypočítává frekvenci, umožňuje srovnávat mezi sebou texty, poskytuje statistické údaje o celkovém počtu slov, délce slov, počtu vět atd., vypočítává vzájemnou informaci (*MI-score*); *KeyWords* identifikuje klíčová slova ve vybraných textech (klíčová slova jsou zde chápána jako výrazy, jejichž frekvence je nezvykle vysoká, porovnáme-li ji s frekvencí v jiném, referenčním souboru textů), umožňuje charakterizovat text nebo žánr.

<sup>5</sup> Jako převodní nástroje byl použit program AntiWord a textové formátovací skripty v jazyce Perl vyvinuté v Ústavu formální a aplikované lingvistiky MFF UK.

<sup>6</sup> Videonahrávky jsou digitalizovány do komprimovaného formátu MPEG-2 (video: 2 Mbit/s, 480x576; zvuk: MP3 220 kBit/s) zvukovou kartou ATI-8500. Jedna hodina nahrávky vyžaduje kolem 1 GB diskového místa. Ukládán je i nekomprimovaný zvuk ve formátu WAV (22 kHz, 16 bit, mono). Jedna hodina těchto zvukových nahrávek má velikost asi 160 MB. Dosud bylo digitalizováno 28 VHS kazet o celkové délce 108 hodin.

<sup>7</sup> Takovou možnost nabízí např. software Transcriber, který lze zdarma získat na internetu (viz recenzi T. Bitriča a Z. Konopáskova, 2001).

<sup>8</sup> Za pomoc s konverzí, lemmatizací a značkováním transkriptů a rovněž s digitalizací videonahrávek bychom rádi poděkovali N. Peterkovi a M. Čmejkovi z Ústavu formální a aplikované lingvistiky MFF UK.

## 2. Otázky transkripce

Výzkum mluveného jazyka si bez zvukových či audiovizuálních nahrávek autentických projevů dnes neumíme představit. S rozvojem nahrávací techniky ovšem nemizí význam přepisování nahraného materiálu. Naopak. Dá se říct, že jeho význam roste právě tím, jak možnost opakovaného přehrávání záznamu dovoluje všimnout si i na první pohled nevýznamných jevů, které však mohou hrát v komunikaci důležitou roli; tyto jevy je možno fixovat transkripcí pro další analýzu. Nárůst množství jevů, které lze v transkriptu zachytit, klade ovšem nároky na podobu transkripčního systému a vzbuzuje také otázku, které jevy jsou pro popis mluvené komunikace relevantní, a má tudíž dopad i na teorii mluveného jazyka a jazykové interakce; tato obecná teorie, na jejímž pozadí výzkumník vybírá jevy pro transkripci, pak spolu s formulací výzkumného cíle zpětně působí na transkripční praxi.

Z toho, co jsme řekli, již vyplývají první, předběžná zobecnění. Za prvé je to poznatek, že transkripce je závislá na teorii jazykové interakce a že soudobý stav poznání a především výzkumný záměr (cíl) zásadním způsobem ovlivňují podobu transkriptu. Za druhé je možné formulovat obecný požadavek, že transkripční systém by měl umožňovat další prohlubování a zjemňování přepisu.

Vzájemnou souhrou transkripce a obecné teorie komunikace se však nevyčerpává otázka teoretické zatíženosti transkriptu. Nesmíme zapomínat na to, že převedení akusticko-časového jevu na jev graficko-prostorový má sémiotické důsledky. V transkripčním systému by měl být tento intersémiotický překlad uspokojivě vyřešen. Kromě toho je třeba mít na paměti jednak úlohu přepisovatele, který v roli intersémiotického překladatele nejprve interpretuje zvuk jako jazykový projev a příp. i další zvukové či vizuální jevy jako signifikantní, jednak úlohu uživatele transkriptu, který transkript čte a interpretuje.

### 2.1. *Obecné principy výstavby transkripčního systému*

V bohaté teoretické literatuře o transkripci mluvených projevů (srov. Ehlich – Rehbein, 1976, 1979; Du Bois, 1991; Hoffmannová, 1992; Edwards – Lampert, 1993; O'Connell – Kowal, 1995; Selting et al., 1998; Lapadat – Lindsay, 1999; Kowal – O'Connell, 2000; Selting, 2001) uvažují jednotliví autoři o různých kritériích, která by měl transkripční systém splňovat. O jednom z požadavků jsme se zmínili již výše: byla to možnost prohlubování a zjemňování transkriptu. Principiální nehotovost transkriptu a možnost jeho neustálého zjemňování si uvědomovali jak tvůrci starších transkripčních systémů (srov. Ehlich – Rehbein, 1976, 1979), tak tvůrci systémů nových (srov. Selting et al., 1998), když zdůrazňovali pracovní charakter přepisu a snažili se definovat jak jeho základní, tak rozšířenou podobu. Možnost prohlubování a zjemňování transkriptu s sebou ovšem nese požadavek neproblematického přechodu od základní podoby k detailní podobě přepisu, tj. přechodu bez nutnosti revize základní verze (Selting et al., 1998).

Ze zaměření na uživatele transkriptu vyplývá další, všeobecně přijímaný požadavek, a to snadné a rychlé čitelnosti transkriptu (srov. Ehlich – Rehbein, 1976; Hoffmannová, 1992; Selting et al., 1998; Kowal – O'Connell, 2000). Z tohoto důvodu se pro přepis dialogu většinou nedoporučuje přesný fonetický přepis, tj. přepis s využitím mezinárodní fonetické abecedy. Někteří lingvisté dokonce výslovně uvádějí, že transkript by měl být čitelný i pro nelingvisty (např. Selting et al., 1998). Tento požadavek je

pochopitelný, když si uvědomíme, že transkripce mluvených projevů se dnes stále častěji využívá i v nelingvistických oborech, jako je psychologie, antropologie, sociologie (srov. např. Edwards – Lampert, 1993; Lapadat – Lindsay, 1999; Kowal – O’Connell, 2000).

Na požadavek čitelnosti je však možné se podívat i z pozice zhotovitele transkriptu a definovat jej jako požadavek *jednoduchosti a snadného zvládnutí transkripčního systému* (tento rys zdůrazňují např. Ehlich – Rehbein, 1976). Spolehlivost transkriptu je jistě závislá i na stupni požadované expertní znalosti. Čím méně speciálních znalostí bude využívání transkripčního systému vyžadovat, tím méně chyb při jeho používání vznikne a tím spolehlivější bude výsledek – transkript.

S požadavkem čitelnosti transkriptu a jednoduchosti a snadného zvládnutí transkripčního systému úzce souvisí celá řada dalších kritérií. Především je to požadavek *jednoznačnosti*: každá značka by měla odpovídat jen jedné kategorii a měla by být jasně definovaná (Selting et al., 1998; srov. též Hoffmannová, 1992). Požadavek jednoznačnosti musíme ovšem vztáhnout na značky jako sémiotické celky, nikoli na jejich části, ze kterých mohou být složeny; při vytváření komplexních (složených) značek je samozřejmě důležité, aby vždy obsahovaly výrazný, snadno vnímatelný distinktivní rys (kontrast) a byly od sebe lehce odlišitelné (např. tečka může v transkriptu označovat buď klesnutí hlasem, nebo krátkou pauzu, podle toho, stojí-li samostatně, nebo je-li z obou stran uzavřena do kulatých závorek). Ke snadnému a rychlému čtení transkriptu přispívá také *ikoničnost* použitých notačních značek čili preferování ikonické notace před arbitrární značkou (Edwards, 1993; Selting et al., 1998); příkladem ikonické notace může být užívání různého počtu teček pro vyjádření délky pauzy apod.

Do oblasti technického řešení patří kritérium *robustnosti transkripčního systému* (Du Bois, 1991; Selting et al., 1998) čili odolnosti vůči nejrůznějším vlivům prostředí při aplikaci systému a dalším zpracovávání transkriptů. V praxi to znamená, že transkript by měl (přínejmenším ve své základní, nerozšířené podobě) co nejvíce využívat běžné tiskařské typy a obejít se bez zvláštních značek (Du Bois, 1991; Selting et al., 1998). Důležitá je jak všeobecná použitelnost a jednoduchost transkripčního systému, tak také jeho nezávislost na technologických změnách, především v oblasti počítačového zpracování.

K vědecké komunitě, především k tvůrcům nových transkripčních systémů, ale i k uživatelům, kteří vybírají ze systémů stávajících, se pak obrací požadavek *kompaktivity ostatními transkripčními systémy* (Selting et al., 1998). V široce chápaném výzkumu dialogu se tento požadavek často řeší tak, že badatelé přebírají alespoň část notačních značek z etnometodologické konverzační analýzy, v níž byl základní repertoár značek vytvořen již v sedmdesátých letech (srov. Sacks – Schegloff – Jefferson, 1974).

## 2.2. *Transkripční systém používaný v korpusu DIALOG*

Transkripční systém, který aplikujeme při přípravě textů pro korpus DIALOG, se primárně zaměřuje na zachycení těch jevů, které se v dosavadním výzkumu mluvených dialogických projevů ukázaly být relevantní (srov. souhrnné práce o konverzační analýze, např. Psathas, 1995; Hutchby – Wooffitt, 1998; ten Have, 1999; z prací o češtině srov. Müllerová – Hoffmannová – Schneiderová, 1992). Zachycujeme jednak sekvenční průběh rozhovoru

(tj. střídání účastníků rozhovoru v řeči vč. simultánního mluvení více mluvčích, okamžitého, bezpauzového navázání na předcházející repliku apod.), jednak segmentální složku řečového projevu (vč. odchylek od noremního vyjadřování) a vybrané jevy suprasegmentální složky (především kadence, pauzy a zdůraznění slova či slabiky). Své místo má v transkripčním systému i záznam parajazykových jevů (hezitačních a responzních zvuků, smíchu, hlasitých nádechů/výdechů apod.), zaznamenat lze i mimojazykové jevy (např. stříh).

Transkripční systém používaný v korpusu DIALOG byl vybrán tak, aby splňoval všechna kritéria výstavby transkripčního systému, která jsme uvedli v kap. 2.1. Zastavme se ještě krátce u jednoho z těchto kritérií, které se stává postupem let stále důležitější, u kritéria kompatibility. V korpusu DIALOG používáme transkripční systém, který je do značné míry kompatibilní se systémy běžně užívanými v anglosaské konverzační analýze (srov. použitou notaci v Psathas, 1995; Hutchby – Wooffitt, 1998; ten Have, 1999; srov. k tomu též Nekvapil, 1999/2000, 2002) a který se víceméně shoduje i se základní podobou transkripčního systému GAT (Gesprächsanalytisches Transkriptionssystem, viz Selting et al., 1998), jenž patří k nejpoužívanějším transkripčním systémům v německojazyčných zemích. Výběr transkripčního systému, který je z velké části kompatibilní s důležitými mezinárodně používanými systémy, se jistě pozitivně odrazí i v oblasti snadné a rychlé srozumitelnosti transkriptu.

### 2.2.1. Segmentální složka

Vyhne-li se z důvodů obtížné čitelnosti, náročného osvojování a nutnosti používat speciální písmena fonetické transkripci, zbývají v zásadě ještě dvě možnosti přepisu segmentální složky mluveného projevu: ortografický přepis a tzv. literární přepis.

Využití ortografického přepisu mluvených projevů, tj. přepisu založeného na pravopisných pravidlech včetně pravidel o interpunkci a psaní velkých písmen, může být za jistých okolností smysluplné. Ve vědě se jej využívá především tam, kde se badatel zaměřuje více na obsah sdělení a méně na formu, např. v orální historii, etnografii nebo sociologii (nikoli však ve výzkumech založených etnometodologicky, kdy badatel většinou sáhne po aparátu sesterské konverzační analýzy).<sup>9</sup> Pro jazykovědná zkoumání, v nichž dominuje otázka *jak?* nad otázkou *co?*, se však ortografický přepis nehodí. Příliš mnoho rysů charakteristických pro mluvený projev by v takto zhotoveném přepisu zůstalo nezachyceno, ortografický přepis by vydal jen neúplný obraz řečové události a zakryl by řadu jevů, které by se při analýze mohly ukázat důležité.

Pro účely lingvistické analýzy dialogu se proto zdá být vhodnější hledat kompromisní řešení mezi ortografickým a fonetickým způsobem přepisu. Takovému kompromisnímu způsobu transkripce se v zahraniční literatuře někdy říká *literární přepis* (*literarische Umschrift / literary transcription*; srov. Ehlich, 1993; O'Connell – Kowal, 1995; Selting et al., 1998; Kowal – O'Connell, 2000). Je to přepis založený na pravopisných pravidlech, který se ovšem snaží zaznamenat veškeré významné odchylky od standardní variety jazy-

<sup>9</sup> Mimo vědu se přepisu založeného na pravopisných pravidlech využívá např. v žurnalistické či nakladatelské praxi. Příkladem ortografické transkripce mohou být přepisy Interview BBC, zveřejňované spolu se zvukovým záznamem na internetových stránkách stanice (viz <<http://www.bbc.co.uk/czech/interview>>). Ortografické přepisy televizních diskusních pořadů zhotovují komerční agentury zabývající se tzv. monitoringem médií.



ka, a to pomocí běžné sady písmen.<sup>10</sup> K takovému kompromisnímu řešení jsme se uchýlili i v korpusu DIALOG:

a přečíst to úplně dětskejma vočima, a zjistit co je v tý hře napsáno,  
(DIALOG: Na plovárně s Tomášem Töpfrem, ČT 1, 24. 4. 2002)

Ke zvláštnostem našeho způsobu transkripce patří to, že pro přepis mluveného projevu používáme výlučně malá písmena (a to i pro přepis vlastních jmen) a nedodržujeme zásady standardní interpunkce (interpunkční znaménka mají v transkriptu svou speciální funkci).

Od přepisu segmentální složky vyžadujeme, aby odlišná grafická podoba slova reprezentovala vždy jen jeho odlišnou zvukovou realizaci, aby tedy např. slovo zasažené (ortoepickou) znělostní asimilací nebylo zapsáno odlišným pravopisem. Na základě tohoto požadavku pak můžeme formulovat základní instrukci pro přepis: co je vysloveno ortoepicky, je zapsáno ortograficky (bez ohledu na pravidla upravující psaní velkých písmen). K posuzování ortoepické kvality projevu ovšem přistupujeme s dostatečnou mírou tolerance, aby se výsledný přepis neprohřešoval proti základnímu požadavku snadné a rychlé srozumitelnosti přepisu. Nepřepisujeme tedy např. odchylky v kvalitě vokálů, tj. případy otevřené nebo zavřené výslovnosti samohlásek, přestože se tyto jevy – přísně vzato – v televizních pořadech vyskytují; přepisovatel má ale možnost tento rys, pokud je nápadný, popsat v komentáři k přepisu, ať už formou shrnující poznámky v záhlaví přepisu nebo přímo v textu přepisu za dotyčnou pasáží (text poznámky se pro odlišení od ostatního textu uzavírá do dvojité kulaté závorky). Odchylky v kvantitě vokálů přepisovatel zaznamenává systematicky, a to jak případy protahování, tak zkracování vokálů: *a to prostě tady zatím není* (DIALOG: Sněží, ČT 2, 4. 1. 2000); pro přepis výrazně protažené hlásky slouží přepisovateli dvojtečka umístěná za hlásku: *včetně třeba: líných soudců, jó, nebo: nebo: lidí který zdržujou vyšetřování* (DIALOG: Sněží, ČT 2, 4. 1. 2000). Míru prodloužení hlásky může přepisovatel ikonicky znázornit různým počtem dvojteček. Dvojtečka znázorňuje i protažení souhlásky: *vidím: jistá jistá zlepšení* (DIALOG: Krásný ztráty, ČT 1, 3. 8. 2000). Z dalších odchylek od ortoepické výslovnosti si všímáme elize hlásek: *dobře eště dvě kila a dete z pořadu* (DIALOG: Další, prosím, Prima, 23. 11. 2000); redukované výslovnosti: *člověk dycky musí jít nák dopředu* (DIALOG: Sněží, ČT 2, 5. 12. 2000); protetických hlá-

<sup>10</sup> Variantou literárního přepisu je tzv. *eye dialect*. Jako jednu z možností přepisu segmentální složky jej pro angličtinu vedle ortografického a literárního přepisu uvádějí S. Kowalová a D. O'Connell (O'Connell – Kowal, 1995; Kowal – O'Connell, 2000) a M. Seltinová (2001). Na rozdíl od literárního přepisu je zde kladen ještě větší důraz na impresionistické znázornění výslovnosti, aniž by se ovšem užilo fonetického přepisu. Dnes se od tohoto extrémně „věrného“ způsobu přepisu většinou upouští. Jeho problematičnost spočívá jednak ve špatné srozumitelnosti (až nesrozumitelnosti) pro nerodilého mluvčího, jednak v negativních konotacích, které takový přepis může ve čtenářích vzbudit o mluvčích (o jejich nízkém vzdělání apod.; srov. Gumperz – Berenz, 1993, s. 96–97; to si uvědomovali již H. Sacks, E. Schegloff a G. Jeffersonová, když připouštějí, že modifikované využití ortografie může evokovat „derogatory connotations“, čtenáře však ujišťují: „we certainly mean no disrespect to the parties cited“; Sacks – Schegloff – Jefferson, 1974, s. 734). Z našeho hlediska je problematičké především to, že se tzv. *eye dialect* někdy využívá i pro zápis standardní výslovnosti a že tedy jiná grafická podoba nereprezentuje odchýlnou zvukovou realizaci; srov. heslo z Encyklopædia Britannica: „Eye dialect = the use of misspellings that are based on standard pronunciations (such as *sez* for *says* or *kow* for *cow*) but are usually intended to suggest a speaker's illiteracy or his use of generally nonstandard pronunciations. It is sometimes used in literature for comic effect“ (Eye dialect, 2004). V češtině bychom mohli případy, kdy nestandardní pravopis reprezentuje standardní výslovnost, ilustrovat např. zápisem *takový malý kluci* nebo *kluci blbý*. Je zřejmé, že takový zápis má především symbolickou funkci (charakterizovat mluvčího, jeho akcent apod.), jeho popisná hodnota je ovšem sporná.

sek: *že to podstatný drama se v odehrává vopravdu v tý hudbě* (DIALOG: Katovna, ČT 2, 27. 6. 2000).

Jak již bylo řečeno v úvodu této kapitoly, v transkriptu důsledně zaznamenáváme také všechny nespisovné slovní tvary. Přepisovatel věrně zaznamenává rovněž případy opakování slov, které je v mluvených projevech velmi časté: *ale nabralo to takovej divnej divnej divnej vobrat* (DIALOG: Další, prosím, Prima, 23. 11. 2000). Pomlčkou se v transkriptech vyjadřuje případ nedokončeného slova; také tento jev je při mluvení dost častý, ať už se jedná o prosté přechnutí nebo o výraz formulačního úsilí; po nedokončeném slově následuje oprava: *že nebude na své pos- na své senátory vykonávat tlak* (DIALOG: V pravé poledne, ČT 1, 4. 6. 2000). Pokud je vyslovený výraz nedobře srozumitelný a přepisovatel si není jist jeho podobou, uzavře tento výraz do jednoduchých kulatých závorek; zcela nerosrozumitelný výraz (nebo pasáž) je v transkriptu naznačen „prázdnými“ kulatými závorkami, počet mezer mezi levou a pravou závorkou by měl zhruba odpovídat délce vysloveného slova nebo pasáže (princip ikoničnosti).

Aplikace transkripčního pravidla „co je vysloveno ortoepicky, je zapsáno ortograficky“ naráží ovšem také na dílčí problémy. Především tu je existence dublet, a to jak v ortografii, tak v ortoepii, které komplikují „čisté“ uplatnění pravidla. Např. existence ortoepické dublety u počátečního *sh-* ([sch] i [zh]) při aplikaci pravidla způsobí, že signifikantní krajová rozdílnost ve výslovnosti bude v přepisu nivelizována. Někteří z našich přepisovatelů si to zřejmě uvědomovali, a tak v korpusu najdeme vedle převažující ortografické formy *shoda* také podoby *schoda* a *zhoda*. Podobná situace nastává u výslovnosti číslovek 7 a 8, které mají také dvě rovnocenné výslovnostní podoby ([sedm]/[sedum] a [osm]/[osum]). Přepisovatelé se v tomto případě již samozřejmě mohou řídit skutečnou výslovností a zapsat buď *sedum*, nebo *sedm*, nelze však vyloučit, že někteří z našich přepisovatelů zapisovali obě zvukové realizace uniformní pravopisnou formou. U ortografických dublet je situace snazší, přepisovatel se pro jednu z dublet rozhodne právě na základě výslovnosti, přestože při čistě formálním uplatnění pravidla nic nebrání tomu zapsat *diskusní*, i když bylo vysloveno [diskuzní]. Dubletní tvary typu *zeširoka / ze široka* nepředstavují pro transkripci vážnější problém, ač pro případné statistické vyhodnocení materiálu by bylo snad dobré i zde distribuci dubletních tvarů regulovat. Dalším, závažnějším problémem je zapisování slovních tvarů *jsem, jsi, jsme, jste, jsou*, které se mohou korektně vyslovovat bez souhlásky *j-* (srov. Palková, 1994; Hůrková, 1995), měly by být tedy – při aplikaci pravidla – zapisovány ve spisovné podobě *s j-*. Stejně jako v předchozích případech i zde dáváme přednost zápisu skutečně vyslovené podoby, tedy většinou bez *j-*, pohledem do našeho korpusu však zjistíme, že zapisování tvarů slovesa *být* je silně zautomatizované, a to ve prospěch spisovné formy. Podobně jako v dalších případech i zde je pak důsledkem častá nivelizace rozdílu mezi dvěma zvukovými realizacemi slova, z nichž jedna je signifikantní (vzácná až „knižní“ výslovnost *s j-*).

Jiným problémem je přepis proprií a apelativ cizího původu. U zdomácnělých cizích slov se držíme našeho pravidla a zapisujeme např. *diskutovat* (a ne *dyskutovat*), *vietnamský* (a ne *vjetnamský*), u slov relativně nových a ještě nezdomácnělých, příp. u slov s původní pravopisnou podobou značně odlišnou od zvukové podoby, přepisujeme podle výslovnosti, např. *otázky rejtinku/rejtingu*, *lisabonský samit*, *šengenský prostor* (toto řešení uplatnily již dříve Müllerová – Hoffmannová – Schneiderová, 1992; srov. s. 7–8). Také u zkratk zapisujeme

malými písmeny výslovnost, nikoli psanou podobu, např. *čé es es dé, dabljú tý ou*. Přepisovatel má vždy možnost uvést ve dvojitých kulatých závorkách (značka pro komentář přepisovatele) ortografickou podobu slova, což může usnadnit např. identifikaci osoby, o které se mluví: *pan šrédr ((míněn Gerhard Schröder))*.

Při aplikaci českého pravopisu na přepis substandardních jazykových projevů se ukazuje i jiné meze českého pravopisu. Jak upozornil Z. Starý (1992), některé obecněčeské věty nelze v souladu s platným českým pravopisem zapsat. Z. Starý ve své práci rozebírá především případy selhání principu analogie při psaní *-i/-y* v kongruentech plurálního subjektu a *l*-ovém participiu, tj. ve větách typu [kluci kteri: sta:li před domem už odešli], kde výrazy *kluci* a *který* dávají různé instrukce pro zapsání koncového */i/* v *l*-ovém participiu (Starý, 1992, s. 75). Podívejme se, jak tento rozpor řeší ve svých textech spisovatelé a překladatelé (zdůraznili S. Č. – L. J. – P. K.):

Ale nejdůležitější dokument má ten chlápek od horský služby, protože se mu tam podepsali všichni lumpové, *který* tam byli.

(Eda Kriseová: Václav Havel. Brno: Atlantis, 1991.)

...a já, ač jsem byl ještě pikolo, jsem se divil, a proto mi byli odporní naši štamgastí, *který* na ulici byli tak nějak *vznešený*, ale u stolu, když se rozdováděli, tak byli *učiněný* kořata, někdy až jako opice...

(Bohumil Hrabal: Obsluhoval jsem anglického krále. Praha: Pražská imaginace, 1993.)

A proto dělali všechno podle knížek, protože tak je to *ty* jejich *nanicovatý, vypatlaný, idiotský, prdelatý* oficií, *který* ve válce taky nikdy nebyli, učili.

(Richard Marcinko: Profesionální válečník, 2. Praha: Ivo Železný, 1997. Z angličtiny přeložil Roman Marhold.)<sup>11</sup>

Ve všech těchto příkladech jsou nestandardní tvary zachyceny na základě téhož pravidla, které jsme zavedli v našem transkripčním systému: To, co lze zapsat ortograficky (zde na základě pravopisných pravidel o mluvnické shodě), je tak zapsáno, odlišně se zapisují slova s odlišnou zvukovou realizací (viz výrazy vyznačené v příkladech kurzivou): např. v Hrabalově textu „naši štamgastí, který na ulici byli tak nějak vznešený“ se v tvaru slovesa *být* píše *-i* na základě mluvnické shody s antecedentem (*naši štamgastí*), ve slově *vznešený* je tvar s *-ý*, protože tvar s *-í* by vedl ke zvukové představě s [ň], tedy k tvaru spisovnému. Kromě poučky o psaní [ďi], [ťi], [ňi] jako *di, ti, ni* zde ke slovu přichází navíc ještě pravidlo o tzv. (pravopisně) tvrdých souhláskách, a tak se ve výše uvedených příkladech nesetkáme ani s tvarem *kterí* ani s tvarem *idiotská*.

Snaha o co nejpřesnější zachycení mluveného projevu však s sebou nese také problém vzdělání a tréninku přepisovatele. Jaké znalosti musí člověk mít, aby mohl přepisovat? Ideální je samozřejmě situace, kdy si analytik přepíše záznam mluveného projevu sám. Při budování rozsáhlejšího korpusu textů však musíme spoléhat také na služby druhých. Základní podmínkou, kterou musí přepisovatel splňovat, je samozřejmě to, aby přepisovatel českých projevů byl rodilý mluvčí češtiny, aby nejen dobře rozuměl tomu, co přepisuje a uměl např. rozhodnout, jak určité slovo zapsat, ale aby rozuměl i významům zvukových nuancí mluveného projevu v češtině a tyto nuance uměl v transkriptu přepsat, příp. popsat. To samo o sobě ale ještě nestačí. Přepisovatel by měl mít také základní vzdělání bohemistické. Aby

<sup>11</sup> Všechny uvedené příklady jsou vzaty z korpusu SYN2000. Další doklady lze vyhledat pomocí dotazu: [tag="N.MP1....."] [word=","] [word="který"].

mohl zapisovat odchylky od standardní normy, musí znát českou ortografii a ortoepii (k základům ortoepie viz např. Palková, 1994, s. 320–345; Hůrková, 1995). Jako důležitá se ukázala především znalost zákonitostí asimilace a neutralizace znělosti. Stávalo se, že přepisovatelé ve snaze zachytit přesně výslovnost přepisovali odlišně právě místa zasažená neutralizací nebo asimilací znělosti, např. z *velkým pocitem křivdy, zbírali každého, žádným způsobem nespochybňuju nutnost tvrdé jasně účinné odvety, abych moch té důvěře dostat*, což bylo v rozporu s naší zásadou zapisovat odlišně pouze odlišnou výslovnost.

### 2.2.2. Suprasegmentální (prozodická) složka

Výzkum prozodické složky autentických rozhovorů patří patrně k nejdůležitějším úkolům současné lingvistiky dialogu (srov. Couper-Kuhlen – Selting, 1996). O tom, jak funguje prozodie v přirozené interakci dvou či více lidí, toho víme zatím sice málo, je ale zřejmé, že lidé považují suprasegmentální složku řeči za významnou součást dorozumívání, kterou jednak nějak interpretují, jednak sami využívají pro sdělování; prozodie je nepochybný sociálněsémiotický fakt. Aby výsledný obraz mluvené komunikace nebyl v transkriptu zkrácený, je důležité tuto složku v přepisu zachytit.

Náš výběr prozodických jevů pro transkripci je pouze základní. Čtenář transkriptu si má udělat jen hrubou představu o zvukové strukturaci promluvy. Přepisovatel zaznamenává (a) kadenci, (b) pauzy a (c) zdůraznění slabiky nebo slova.

Záznam intonace se v našem transkripčním systému omezuje na zachycení intonační kontury na konci intonačního celku, čili tzv. kadence (Daneš, 1957, 1985). Dříve než přepisovatel zachytí značkou tónový pohyb, musí v replice identifikovat zvukově samostatný úsek řeči: ten určujeme výhradně na základě prozodických kritérií, tzn. bez ohledu na jeho syntakticko-sémantickou strukturu. Konec zvukově samostatného úseku řeči, který nás primárně zajímá, lze charakterizovat kombinovaným výskytem dvou, příp. tří rysů: (1) tónovým pohybem, který jej zřetelně odděluje od předcházející řeči, a (2) pauzou, byť i jen nepatrnou, a/nebo (3) změnou tempa (zpomalením). Zpomalení na konci úseku může být realizováno např. protažením poslední hlásky a stát se tak percepčně velmi výrazným. Pro hranici mezi úseky nemáme speciální značku, v transkriptu je konec úseku zachycen právě jen značkou pro kadenci: vysoké stoupnutí hlasu vyjadřuje otazník, střední stoupnutí hlasu čárka, klesnutí hlasu je označeno tečkou. Pokud je intonace „rovná“, tj. mluví na konci ani nestoupne ani neklesne hlasem, nenásleduje žádná značka (přepisovatel však může vložit značku pro pauzu; viz níže). V nepřipravených, spontánních projevech může být takový zvukový celek i velmi dlouhý, zvláště pokud je tempo řeči rychlé; s dlouhými, rychle pronášenými úseky se často setkáváme v řeči moderátorů:

TK: natálie co ste dělali, když mluví váš manžel o studiích o koleji já vím že ste se sešli na škole?

...

TK: proč ste:, přišli teda právě do české republiky vy ste na začátku říkal sašo že, to nebylo: kam, že to bylo odkud, ale nakonec ste zůstali tady. takže, proč ste nemířili třeba v té době byla hodně populární západní evropa německo a jiné státy

(DIALOG: Sauna, Prima, 3. 4. 2002; TK – Tomáš Krejčíř, moderátor)

V první replice této ukázky jsou dva zvukově samostatné úseky řeči; druhý z nich končí proti očekávání prudkým stoupnutím hlasu na poslední slabice slova *škole*; mluví jako by

na poslední chvíli zapochyboval o správnosti/pravdivosti své informace a původně zamýšlenou oznamovací větu nakonec realizoval pomocí antikadence jako otázku. V druhé replice je konec prvního úseku vyznačen protažením poslední hlásky ve slově *ste* a mírně stoupavou kadencí; realizace tohoto úseku, který jde zcela proti členění syntakticko-sémantickému, byla dána nejspíš důvody formulačními (hezitací).

Pokud je v podobně dlouhých úsecích, pronášených v rychlém tempu, nějaký další náznak kadence, nějaký málo zřetelný tónový pohyb, ať už klesnutí nebo stoupnutí hlasu, a není doprovázen krátkou pauzou nebo zpomalením tempa, značku nevkládáme.

Pauza mezi úseky bývá někdy velmi nezřetelná, příp. málo výrazná v kontextu předchozího zpomalení tempa. V takových případech přepisovatel vystačí se značkami pro tónový pohyb (otazník, čárka, tečka) a nevkládá zvláštní značku pro pauzu. Je-li však pauza delší a výraznější, může pro označení pauzy použít v závislosti na její délce od jedné do tří teček v jednoduchých kulatých závorkách. Délku pauzy stanovuje přepisovatel na základě poslechu celého projevu, bere v úvahu individuální tempo mluvčího a další realizované pauzy. Jak známo, pauzy se v autentických mluvených projevech vyskytují i jinde než na hranicích intonačních celků (srov. typologii pauz v publikacích Müllerová – Nekvapil, 1986; Nekvapil – Müllerová, 1988):

TH: v žádném případě vám nechci kázat, ale když jsem řekl že k vám mluvím jako, také jako kněz, (.) tak chci říci (.) přece jenom, jeden (.) jediný (.) moment (.) z evangelia. v okamžiku kdy (.) ježíš e chválí samaritána který se sklonil k člověku, jiné etni- jiného etnika. a jiného náboženství aby mu, prokázal něco dobrého tak se ho, lidé okolo ptají, kdo je náš bližní kdo je ještě ten bližní, o kterého se mám postarat a kdo už je daleko, že mi může být lhostejný nebo může být nepřítelem. a ježíš odpovídá, ty se učiň bližním. ty (.) buď (.) bližko každému člověku.

(DIALOG: Pro a proti, ČT 2, 14. 6. 1998; TH – Tomáš Halík)

V předchozí ukázce vidíme kromě pauz a intonačních kadencí ještě zachycení dalšího prozodického jevu – nápadného zvukového zdůraznění slova nebo slabiky; v transkriptu je zdůraznění znázorněno podtržením příslušného segmentu. Tímto způsobem zaznamenává přepisovatel ovšem pouze taková zdůraznění, která jsou zřetelná v kontextu celé repliky; v některých případech, např. při velmi klidném nebo velmi krátkém projevu, se zdůraznění nemusí v replice vůbec objevit.

Transkripce prozodické složky mluveného projevu patří patrně k nejnáročnějším úkolům, které před přepisovatele staví náš transkripční systém. Zápis suprasegmentální složky je proto také nejméně spolehlivý a nejvíce postižený subjektivním, individuálním přístupem přepisovatele. Aby výsledky přepisu odpovídaly do jisté míry zvukové realitě, vyžaduje to alespoň minimální trénink.

### 2.2.3. Sekvenční složka

Sekvenční složkou dialogu rozumíme mechanismus střídání účastníků rozhovoru v řeči. Skutečnost střídání mluvčích (*turn-taking*) lze zachytit v transkriptu dvojím odlišným způsobem. První způsob můžeme nazvat řádkovým zápisem (*Zeilenschreibweise*, Kowal – O'Connell, 2000; Selting, 2001). V řádkovém zápisu začínají jednotlivé repliky vždy na novém řádku, časová posloupnost replik je tak zachycena na vertikální ose shora dolů. Tento způsob záznamu střídání mluvčích dobře známe z četby divadelních her a ne-

činí nám při čtení žádné potíže, připadá nám „přirozený“. Druhý způsob se nazývá *partiturní zápis* (*Partiturschreibung*, Ehlich – Rehbein, 1976; *Partiturschreibweise*, Kowal – O’Connell, 2000; Selting, 2001; *score notation*, Ehlich, 1993). V partiturním zápisu se časový průběh interakce zobrazuje podobně, jako je tomu při záznamu hlasů v hudební partituře: každému mluvčímu je vyhrazen zvláštní řádek a tyto řádky, zleva sevřené svorkou do jednoho svazku, vytvářejí v transkriptu tzv. simultánní plochu (Ehlich – Rehbein, 1976); časová osa je zde podobně jako v hudební partituře orientována horizontálně zleva doprava: spustíme-li v kterémkoliv místě horizontální osy pomyslnou svislici, uvidíme na simultánní ploše transkriptu, co se v tomto okamžiku děje v interakci současně.<sup>12</sup>

Zajímá-li nás mechanismus střídání replik mluvčích, je řádkový zápis pravděpodobně lepší formou transkripce. Jak poznamenává M. Seltingová (2001), v řádkovém zápisu je sekvenčnost postavena do popředí: je zde ikonický vztah mezi časovou posloupností replik v reálné komunikaci a posloupností replik v transkriptu (tj. mezi reálným střídáním mluvčích a jeho prostorovým zachycením v transkriptu). V korpusu DIALOG jsme se i proto rozhodli pro řádkový zápis replik. Hlavním důvodem je ale především ten fakt, že na lineárním principu fungují počítačové textové editory a jakékoliv další zpracování je tak na počítači snazší právě při použití řádkového zápisu.

V řádkovém zápisu začíná každá replika na novém řádku a je uvozena iniciálami mluvčího:<sup>13</sup>

ME: a čím byl tatínek profesně

TT: tatínek byl profesí obchodník.

ME: to vám taky předal oběma tenhle dar

(DIALOG: Na plovárně s Tomášem Töpfrem, ČT 1, 24. 4. 2002; ME – Marek Eben, moderátor; TT – Tomáš Töpfer)<sup>14</sup>

Jak zjistili ve své klasické studii H. Sacks, E. Schegloff a G. Jeffersonová (1974), mluvčí se při střídání v řeči snaží minimalizovat pauzy mezi replikami (*gap*) a mluvení přes sebe (*overlap*). Ve většině případů tedy po jedné replice nastupuje replika druhá, která se nepřekrývá s první, ani nezačíná s nápadným zpožděním. Takový případ výměny mluvčích reprezentuje výše uvedená ukázka. Od takového vystřídání mluvčích, které v transkriptu z důvodu bezpříznakovosti neoznačujeme žádnou zvláštní značkou, se liší případy, kdy mluvčí „vpadne“ do partnerovy repliky, překotně na ni naváže apod., tedy kdy vystřídání

<sup>12</sup> Asi nejznámějším transkripčním systémem, který využívá partiturní zápis, je systém HIAT (Halbinterpretative Arbeitstranskriptionen) autorů K. Ehliche a J. Rehbeina (Ehlich – Rehbein, 1976, 1979; Ehlich, 1993). K. Ehlich a J. Rehbein jej vytvořili během svého výzkumu komunikace ve škole, tedy při studiu komunikačních situací, které se vyznačují vysokým počtem účastníků a kdy navíc hrozí, že i několik mluvčích bude mluvit současně. Pro záznam takové komunikační situace je užití partiturního zápisu pravděpodobně výhodnější než využití řádkového zápisu (v podobném duchu se vyjadřuje i Seltingová, 2001).

<sup>13</sup> Součástí každého transkriptu je také záhlaví, v kterém se kromě názvu pořadu, televizního kanálu a data vysílání uvádí také seznam mluvčích s odpovídajícími iniciálami jmen, které je v transkriptu zastupují, příp. i s údaji o sociální identitě mluvčího, která je z hlediska pořadu relevantní (např. VŠ – Vladimír Špidla, předseda vlády České republiky).

<sup>14</sup> Pro názornou představu a pro srovnání ještě uvádíme, jak by přepis téže sekvence vypadal v partiturním zápisu:

ME	[	a čím byl tatínek profesně		to vám taky předal oběma tenhle dar
TT	]	tatínek byl profesí obchodník.		

v řeči proběhne bez jakékoliv, byť jen krátké pauzy. Takové okamžité navázání označujeme v přepisu rovnítkem umístěným před prvním slovem takto zahájené repliky:

VJ: dobrý večer, já bych měl úvodem, krátkou informaci, (.) pro (.) diváky, (.) chtěl bych omluvit nepřítomnost, mého kolegy, jana rejžka, (.) e který byl odvelen za jistě důležitější kulturní událostí než je katovna, a sice udílení oskarů. ne že by snad dostal oskara za, kritiku nebo publicistiku ale,

JPK: =pomalu by si to zasloužil.

VJ: jistě., ale ta., e kategorie tam není mysím zatím, a pan rejžek komentuje, (.) oskara, v době kdy, e udílení oskarů v době kdy natáčíme tuto katovnu.

(DIALOG: Katovna, ČT 2, 4. 4. 2000; VJ – Vladimír Just, moderátor; JPK – Jiří P. Kříž, kritik, publicista)

Situace, se kterou se v autentických rozhovorech běžně setkáváme, je mluvení několika mluvčích „přes sebe“, tj. situace simultánního pronesení replik nebo jejich úseků dvěma i více mluvčími (tzv. *overlap*, viz Sacks – Schegloff – Jefferson, 1974; v němčině se pro tuto situaci používá termín *Überlappung*, viz Selting, 2001). Simultánně pronesené úseky se v transkriptu uzavírají do hranatých závorek a umísťují se pod sebe:<sup>15</sup>

VŠ: samozřejmě že vláda splní svou povinnost a, využije ty prostředky, které má [k dispozici]

MŠ: [můžete být]

konkrétnější

(DIALOG: Naostro, ČT 1, 11. 8. 2002; VŠ – Vladimír Špidla, předseda vlády České republiky; MŠ – Milan Šíma, moderátor)

Z uvedených příkladů je zřejmé, že při transkripci rozhovoru se pracuje s p l o c h o u transkriptu. To se netýká jen případů překrývání replik, ale také vertikálního záznamu střídání v řeči, který je pro analýzu tohoto mechanismu výhodný. V korpusech vytvářených lingvisty se ale veškeré texty l i n e a r i z u j í, badatel pracuje většinou s nálezy ve formě tzv. konkordančních řádků. Bude úkolem programátorů vyvinout takový korpusový manažer, který by umožňoval nejen práci s konkordancemi, ale dokázal by vybrané pasáže zobrazovat v podobě dvojrozměrného transkriptu, jaký se používá při analýze rozhovoru.

#### 2.2.4. Parajazyková a mimojazyková složka

Parajazykovými jevy rozumíme neverbální hlasové projevy, které provázejí mluvený projev. V přepisu zaznamenáváme systematicky veškeré hezitační zvuky (tj. případy, kdy mluvčí přeruší svou řeč zvuky *e*, *e:*, *hm*, *ehm*, *m:* apod.) a responzní zvuky (tj. případy, kdy posluchač reaguje na partnerovu řeč pomocí výrazů *hm*, *ehm*, *mhm*, *m:* apod.). Přepisovatel volí takový přepis, který je realizovanému zvuku nejbližší.

MD: [je (..) je to: tohleto je jedno] jedno z deseti pobytových středisek, a jedno z pěti které neprovozuje přímo (.) stát, ministerstvo vnitra,

TK: =mhm

MD: =ale provozujeme je na základě: nové strategie a sice: e: (.) podle smlouvy s e obecním úřadem, (.) v seči,

(DIALOG: Sauna, Prima, 3. 4. 2002; MD – Miluše Dohnalová, Správa uprchlických zařízení; TK – Tomáš Krejčíř, moderátor)

Transkripční systém umožňuje zaznamenat i další parajazykové jevy, jako je smích, povzdech, zakašlání, hlasitý nádech/výdech apod. Nemáme sice pro ně žádné speciální znač-

<sup>15</sup> Přechnodně se zde tedy uplatňuje varianta partiturního zápisu.

ky (s výjimkou mluvení se smíchem, viz níže), v přísném slova smyslu se tedy ani nejedná o transkripci, přepisovatel však může informaci o těchto jevech zapsat formou komentáře, tj. může vložit do transkriptu na patřičném místě popisnou poznámku. Od ostatního textu se komentář přepisovatele odděluje dvojitými kulatými závorkami. Tyto jevy (s výjimkou smíchu) nezaznamenáváme systematicky, necháváme na rozhodnutí přepisovatele, které jevy považuje z hlediska smyslu komunikátu za relevantní:

JP: jak obstál liberalismus u nás, (.) objevil-li se vůbec,

VB: ((hlasitě vydechnutí)) český liberalismus e ((vydechnutí a mlasknutí)) český liberalismus je problematický. český národ e: nebo česká národní tradice není příliš li- liberální, u nás větší roli sehrály (.) sehrál řekněme socialismus, (..) anebo (.) snad i (.) katolicismus, e (..) e: takové ty (.) e takové ty hodnoty které představuje řekněme masaryk. (..) který byl kritikem liberální liberální filozofie

...

JP: sám říkáte pane profesore bělohradský že úkolem filozofa, ve vašem pojetí nebo možná v obecném pojetí je všechno zpochybňovat. prosím

VB: to ((odkašlání)) toto tuto větu neříkám. to skutečně:

JP: =mám ji tady napsanou

VB: =ne ale to sem (.) tuto větu jsem skutečně já neřek...

(DIALOG: Pro a proti, ČT 2, 14. 6. 1998; JP – Jan Pokorný, moderátor; VB – Václav Bělohradský)

Samostatný smích a pasáže vyslovené se smíchem zaznamenává přepisovatel systematicky. Samostatný smích zapisuje formou komentáře ve dvojitých kulatých závorkách, např. ((*směje se*)), ((*smích všech*)), ((*smích v publiku*)) apod. Pasáže vyslovené se smíchem se v přepisu uzavírají do špičatých závorek:

JB: [ano. nicméně:] e: pan vicepremiér špidla zmínil, že právě vaši poslanci v konsolidační agentuře byli proti, (.) e: konkrétní sanaci konkrétnímu řešení té rokytnice takže vy byste prostě, (.) e: [proda:]

VK: [nezlobte

se] já (.) <já nevím jesi je to pravda> a nemohu to potvrdit já znám zoufalé dopisy, (.) nejrůznějších radních, které putují (.) putovaly mně, putovaly pravděpodobně i vládě, putovaly kamkoliv jinam, já je tady dokonce (.) dokonce mam, ...

(DIALOG: Sedmička, Nova, 9. 6. 2002; JB – Jana Bobošíková, moderátorka; VK – Václav Klaus, předseda Poslanecké sněmovny a předseda ODS)

Mimojazykové jevy jako pohledy, gesta apod. v transkriptu nezaznamenáváme. Analyticky relevantní se ale někdy ukazuje být stříh, zejména u pořadů vysílaných ze záznamu, příp. v předtočených reportážích, které bývají často součástí diskusních pořadů. Podoba repliky, ale např. i následnost dvou replik může být totiž výsledkem práce stříhače. Informaci o stříhu, pokud je z obrazu a zvuku patrná a přepisovatel ji považuje za významnou, lze zaznamenat opět formou poznámky ve dvojitých kulatých závorkách:

VB: do lesa ne:. (.) do přírody ne: (.) já patřím d- na periferii metropole. ze všeho nejradší mám periferii metropole, nejradší se procházím (.) na periferiích prahy nebo (.) eventuálně paříže nebo velkých měst (.) já: spíš sem člověk, (.) perif- z periferie metropolí. tam se cítím nejlíp. ((stříh)) (řád) pobývám na tom rozhraní (..) toho (.) silného města a té poražené přírody která už přece jen ale, (..) přece jen už je, (..) mimo město. ((stříh)) s tomášem halíkem nechám-li stranou jeho: spisy:, e: které sem znal už dřív, sem se setkal tuším hned jak sem přijel do prahy nebo velmi brzy (..) (na) první setkání přímo si nepamatuju, ale setkal sem se s ním velmi brzy (.) on znal (.) setkal sem se s ním na fakultě sociálních věd. poprvé. už si vzpomínám. bylo to v roce devadesát, a bylo to na fakultě sociálních věd. kde on e tenkrát vyučoval. ((stříh)) to se mi právě líbí na halíkovi že (.) on je sice údem církve. (..) ale (.) ty křesťanské ideály realizuje, nejen nějakou politikou. katolickou politikou (nebo). ale sám v sobě a svým



vystupováním ve světě a svým učením. ((konec předtočené reportáže))  
(DIALOG: Pro a proti, ČT 2, 14. 6. 1998; VB – Václav Bělohradský)

Formu komentáře přepisovatele lze využít i k zápisu titulků, které se během vysílání objevují na obrazovce:

TK: ale když ste přišla tak předpokládám že s:te nedělala rovnou v médiích to se vám asi nepodařilo.

GO: to víte že ne.

TK: a co ste dělala

GO: ((titulek na obrazovce: Goranka Oljača (52 let), „Přišla jsem jako host.“)) dělala sem: různý věci na začátku: (.) e: sem byla uklízečka, protože e: ony nepotřebují moc mluvit (.) a: <ta> skupina matek e s dětmi, e: i když vysokoškolačky, neuměla česky. tak. (.) [a]

(DIALOG: Sauna, Prima, 3. 4. 2002; TK – Tomáš Krejčíř, moderátor; GO – Goranka Oljača, redaktorka Českého rozhlasu Regina)

Zbývá se zmínit o značce, kterou nepoužíváme v samotných transkriptech, ale jež se uplatní při prezentaci korpusových dat. Je to značka pro vypuštění části transkriptu; má běžnou podobu tří teček (viz v příkladech výše).

### 3. Využívání korpusu DIALOG

Od doby svého vzniku byl korpus využíván dvěma základními způsoby. Analýzy, které vznikaly na jeho materiálu, byly orientovány (1) směrem k teorii dialogu (dialogismu, dialogičnosti a odpovídající teorii mluvenosti), (2) směrem k popisu současné mluvené češtiny (jenž ovšem samozřejmě předpokládá určité teoretické pochopení mluvenosti). Těmito dvěma souběžnými směry byly vedeny i výzkumy mluveného jazyka, které v Ústavu pro jazyk český předcházely vlastnímu vzniku korpusu DIALOG; nebudeme se proto v následujícím výkladu vyhýbat ani citování některých prací, na něž výzkum opřený o korpus DIALOG navazoval (jde zejména o práce O. Müllerové, J. Hoffmannové, S. Čmejkové vycházející z materiálových zdrojů staršího data, a to jednak z audionahrávek mediálních dialogů z počátku 90. let, a pokud jde o autentickou komunikaci nemediální, pak i z audionahrávek z 80. let).

#### 3.1. Rozvíjení teorie dialogu

Analýza dialogu, dialogismu, dialogičnosti, polyfonie dvou a více hlasů, které jsou ve vztahu impulzu a odezvy, stimulu a reakce, otázky a odpovědi, pochopení, nepochopení, echa, navazování, střetu, konfliktu a – obecně řečeno – vzájemné reflexe, tvoří metodologickou osu analýzy diskurzu rozvíjející se v posledních několika desetiletích. Z perspektivy dialogičnosti se zkoumají nejen projevy mluvené, tj. především rozhovory tváří v tvář, které jsou dialogem v úzkém slova smyslu, ale i texty psané, neboť i ony jsou umístěny na komunikačních osách autorů a adresátů a mají svou interakční a intertextovou povahu. Perspektiva dialogičnosti je s to odhalit v psaných textech rysy, které by pojetí psaných textů jakožto monologických nebylo schopno spatřit, nebo by je odsunulo na okraj zájmu jakožto jevy nerelevantní.

Co přináší zostřený dialogický zřetel k popisu projevů mluvených? Snaží se např. nalézt odpověď na otázku, kterou kladl před české lingvisty už Karel Hausenblas (1984), zda je totiž dialog tvořen dvěma komunikáty, anebo jím vzniká jeden komunikát. Můžeme se také tázat, zda dialog vytváří komunikát spíše homogenní, anebo heterogenní, čím je dána

jeho koherentnost a koheznost; jakou má při tom roli koheze a koherentnost horizontální, tedy návaznost replik jednoho partnera na repliky druhého partnera, a jakou roli má naopak koheze a koherentnost vertikální, tedy návaznost replik jednoho partnera (viz k tomu např. Kraus, 2003, s. 281). Oba typy návaznosti jsou dobře patrné v následujícím úryvku:

JV: dobře ale odpověď na: tu první otázku byla ta že za tu nezaměstnanost může: můžou mohou předchozí vlády

MZ: =obráceně. že za tu nezaměstnanost, e může především klesající (.) dynamika hospodářského (růstu dokonce)

JV: [zaviněná činností]

MZ: záporná dynamika, zaviněná hospodářskou politikou [minulých vlád teď]

JV: [( ) minulých vlád takže]

MZ: například doznívá, hospodářský efekt vašich balíčků.

JV: tak (.) paní předsedkyně

LB: =každopádně samozřejmě že ta narůstající nezaměstnanost souvisí s poklesem hospodářského růstu, můžeme si tady (.) říkat všechny možné příčiny které, způsobily pokles hospodářského růstu: můžeme to rozlišovat na vnější vnitřní vlivy, ale nemůžeme říkat že každopádně za všechno co se v ekonomice odehrává, může vláda. to ať ta vláda minulá, nebo ať ta vaše vláda pane předsedo jestli to chcete takhle vztahovat, tak ten interval za chvíli o kterém ste mluvil za chvíli uplyne a už to bude všechno na vás potom.

MZ: =ano. v pořádku.

LB: =(a) jestli za všechno teda může vláda tak pak za všechno nikoliv že by mohl klaus, ale za všechno ( ) budete moci už pak vy,

MZ: a proč ne ale, aby bylo jasno my jsme ve své zprávě o stavu unie, jak ji říkáme, připustili několik faktorů. například sme tam dali i ten faktor který naprosto právem, uvádíte vy to znamená vysoké úrokové sazby čnb, a vím že pan guvernér tošovský proti tomu (.) dosti výrazně protestoval čili snažíme se být objektivní, a nesvalovat to jenom zdůrazňuji jenom na minulou vládu. ale zase paní předsedkyně (.) nám se například podařilo dosáhnout velmi dobré komunikace s českou národní bankou, ta snížila ty úrokové sazby celkem osumkrát, nebo devětkrát, a dnes už představují zhruba polovinu (.) výše těch předchozích úrokových sazeb. čili (.) netvrdím že za to může jenom (.) hospodářská politika minulé vlády, může za to od roku devadesát šest i česká národní banka. ale vy zase obráceně tvrdíte že za to může jenom česká národní banka, a nepřipouštíte si vlastní [odpovědnost.]

(DIALOG: 7 čili Sedm dní, Nova, 21. 3. 1999; JV – Jan Vávra, moderátor; MZ – Miloš Zeman, předseda vlády a předseda ČSSD; LB – Libuše Benešová, předsedkyně Senátu a místopředsedkyně ODS)

A konečně, někdy je dialog – v případě dominující návaznosti vertikální a takřka nulové návaznosti horizontální – spíše jen střídáním dvou míjejících se monologů, jaké známe např. z dialogů exponovaných v absurdním dramatu.

Desetiletí, po která se teorie dialogu vyvíjí, přinášejí stále zjemňování a rafinování zájmu o dialogické dění. Tam, kde teorie dialogu donedávna vystačila s pojmy otázky či odpovědi, případně otázky zjišťovací, doplňovací a vylučovací, odpovědi kladné a záporné, úplné a neúplné, otázky vyžadující odpověď a otázky nepředpokládající odpověď, tedy tzv. otázky rétorické (k nesnadnému vymezení rétorické otázky srov. Karhanová, v tisku), se při analýze autentických dialogů objevuje kontinuum jemně odstíněných variant výpovědí, jež se sice dají občas zařadit podle zvolených kritérií do té či oné skupiny dialogických nástrojů (otázka vyžadující/nevyžadující odpověď, odpověď jednoznačná/nejednoznačná, úplná/neúplná, přesná/nepřesná apod.), ale nelze vyloučit ani interpretaci odlišnou (uvede-li host v interview nějakou odpověď, kterou sám považuje za dostatečnou, může být tato odpověď tazatelem považována za nedostatečnou apod.).

Posuzujeme-li pak dialogické dění, záleží na tom, zda se jako badatelé budeme pokoušet uplatnit svůj vlastní pohled na toto dění, tedy zároveň určitou lingvistickou zkušenost, badatelský odstup a teoreticko-metodologickou znalost, anebo budeme poměřovat dialogické dění perspektivou jeho bezprostředních účastníků; pak ovšem máme, teoreticky vzato, možnost zaujmout postupně perspektivu jednoho participanta po druhém a pokaždé se nám může dialog a dialogické chování jeho účastníků jevit v jiném světle. Že se tu nevěnujeme nedůležitému metodologickému tápání, ale dotýkáme se závažného faktu s mnoha důsledky, o tom svědčí jak rozdíly v diváckém hodnocení dialogické kompetence jednotlivých dialogických soupeřů, tak hodnocení dialogické kompetence redaktora v jeho interview s hostem: část diváků bude přesvědčena, že redaktor klade otázky nekompetentně, provokativně, špatně a hloupě, jiné části diváků se budou zase jako nedostatečné nebo vyhýbavé jevit odpovědi hosta, srov. následující ukázkou:

MŠ: [(ale) můžete třeba teď divákům] říci že, (.) vylučujete tuto možnost že ó dé es už nic takového neudělá

VK: já já (..) <já prostě [prohlašuji], (.) že]

MŠ: [že by (.) udělala nějakou opoziční] smlouvu třeba se sociální demokracií

VK: =prosím vás, co (.) já prohlašuji, opakovaně, že jsou tři politické subjekty u nás, a že dnes je situace natolik otevřená že volby mohou dopadnout (.) jakýmkoliv způsobem a že žádná z těchto stran nepřesešla (.) definitivně, (.) (ale) já vám odpovídám]

MŠ: [(dobře.) (..) já se ptám na tu opo: na tu novou opoziční] smlouvu

VK: =já vám říkám (.) jednoznačně.

MŠ: =je to možné.

VK: =posloucháte mě nebo ne.

MŠ: =já se ptám abyste jednoznačně ano. abyste jednoznačně řekl, jestli je něco takového možné [že by ó dé es podepsala]

VK: [já jsem řekl]

MŠ: se sociální demokracií nějakou novou opoziční smlouvu po těchto volbách.

VK: chcete (.) slyšet odpověď? (..) tak prosím nechte mě to větu dořici. ano? (..) ne: jinak ji nezačnu jestli to nekejdnete na to.

MŠ: =ale já na <vás kývám. (.) [(pane profesore)>]

VK: [dobře já říkám] že tyto tři politické subjekty, se rozhodly, (.) jasně to říkají že nepřesešly (.) možnost (.) navázání (.) vzájemných vztahů (.) v libovolné (.) kombinaci, které mezi nimi (.) existují. tečka.

MŠ: to není odpověď, pane [předsedo. (..) to není]

VK: [to je stoprocentní odpověď]

(DIALOG: Naostro, ČT 1, 21. 4. 2002; MŠ – Milan Šíma, moderátor; VK – Václav Klaus, předseda ODS)

Jakou perspektivu by měl zaujmout lingvista? Budeme-li posuzovat daný dialog z hlediska vyhovění pravidlům či maximám dialogu a přistoupíme-li při tom jednou na stanovisko tazatele a podruhé na stanovisko dotazovaného, můžeme dospět k odlišným interpretacím dialogického dění a k odlišnému vidění dialogické kompetence obou participantů.

Snad i proto získávají značnou úctu, vážnost i oblibu takové směry analýzy dialogu, jako je etnometodologická konverzační analýza (srov. Hutchby – Wooffitt, 1998; Nekvapil, 1999/2000; Psathas, 1995; ten Have, 1999) nebo interpretativní sociolingvistika (srov. Nekvapil, 2000). Výzkumník je nabádán, aby stál nikoli nad dialogickým děním, ale pozoroval dialogické participanty stojí mezi nimi, a zajímala ho fakta a kritéria dialogického

dorozumívání do té míry, do jaké jsou vnímána, interpretována a formulována samotnými účastníky dialogu. Zájem se soustřeďuje na komunikační přizpůsobování dialogických partnerů a jejich vyjednávání jak obsahových, tak procedurálních stránek vzájemné interakce (k pojmu vyjednávání srov. Weigand – Dascal, 2001).

Analýza mediálních dialogů, které tvoří korpus DIALOG, se může inspirovat jak obecnou anglosaskou metodologií konverzační analýzy a příbuzných disciplín, tak konverzační analýzou aplikovanou právě v zemi svého vzniku na dialogy institucionální a ještě úžeji na mediální interview (srov. Heritage, 1985). Obecně lze říct, že analýza mediálních interview, jakou nacházíme u takových autorit konverzační analýzy, jako jsou J. Heritage (1985), P. Drew (Drew – Heritage, 1992), D. Greatbatch (1998), S. E. Clayman (1991, 1993; Clayman – Heritage, 2002) aj., je zaměřena na identifikaci a formulaci pravidel, principů, opakujících se mechanismů. Ve srovnání s touto tendencí postihují české lingvistky a někdy i lingvisté ve svých analýzách častěji také případy zvláštní, porušování pravidel, nedodržení principů, předpokladů a očekávání. V tomto svém úsilí postihnout také jevy individuální a zvláštní zdaleka není česká lingvistika v rámci světové lingvistiky dialogu osamocena. Obdobné tendence dokázat, že televizní interview či diskuse, stejně tak jako třeba parlamentní debata, je dialogické dění plné napětí, jehož účastníci se snaží vymanit z předem ustaveného – a na individualitách mluvčích nezávislého – dialogického mechanismu, lze najít i u některých badatelů zahraničních; i oni ukazují, že dialog (ať už jde o politické interview nebo vědeckou disputaci), je především střetem lidských individualit, snažících se vtisknout předem danému dialogickému formátu pečeť svých osobností, svého vidění světa a hodnot (Weizman, 1998).

Ani v českém, ani ve světovém kontextu se tak nepřestává rozvíjet teorie dialogu, která využívá nejrůznější inspirační zdroje (srov. Hoffmannová, 2000, 2003). O jednotlivých směrech analýzy dialogu Slovo a slovesnost pravidelně informuje, mj. i ve zprávách o kongresech Mezinárodní organizace pro výzkum dialogu (International Association for Dialogue Analysis, zkr. IADA), jež sdružuje vedle britských lingvistů zejména lingvisty německé, francouzské a italské, ale i lingvisty z jiných evropských zemí (jeden z významných kongresů se konal v r. 1996 v Praze; srov. sborník z konference Čmejrková – Hoffmannová – Müllerová – Světlá, 1998). K tématům probíraným v rámci kongresů této společnosti patří vztah teorie dialogu a teorie řečových aktů, vztah teorie dialogu a teorie rozumění a interpretace, vztah dialogu a argumentace, zdvořilost, nestrannost a manipulativnost, nedorozumění, konflikt, moc, emoce, humor, ironie, smích i jiné projevy neverbální komunikace apod.

Inspirace současnými směry analýzy dialogu, mezi něž dnes bezesporu patří především etnometodologická konverzační analýza, je pro českou teorii dialogu nezbytná. Zrovna tak je ale pro českou teorii dialogu výhodné nezapomenout na tradici funkční lingvistiky, k níž se čas od času vrací i badatelé zahraniční – v současné době např. v souvislosti s oživením lingvistického zájmu o emoce v řečovém projevu a v řečové interakci, o rozlišení expresivity, emotivnosti a emocionálnosti (srov. Daneš, 2004; Čmejrková, 2004). Korpus DIALOG nabízí mj. i k tomuto tématu hodně materiálu, neboť mnohé z diskusních pořadů věnovaných závažným společenským tématům mají, jak jsme se v ukázkách pokusili naznačit, atmosféru hovoru zaujatého až vzrušeného.

### 3.2. Popis současné mluvené češtiny

Pro lingvisty nezajímající se o analýzu diskurzu, ale o analýzu a popis jednotlivých jazykových rovin češtiny (fonologické, morfologické, lexikální, syntaktické atd.), resp. o otázku současných norem, a to jak češtiny spisovné, tak nespisovné, je korpus DIALOG zajímavý tím, že na žánrově vymezeném vzorku řeči usilující o určitou míru sdělnosti a zpravidla také o jistou míru kultivovanosti lze pozorovat typické rysy současného mluveného úzu. Aniž bychom chtěli přeceňovat relevantnost a směrodatnost údajů získaných analýzou právě takto zvoleného vzorku řeči, máme za to, že mediální debaty, interview, besedy a talk-show patří mezi pořady s poměrně vysokou sledovaností, a předpokládáme tedy – celkem v souladu s míněním mediálních odborníků – že plní úlohu jistých řečových vzorů (platí to zejména o řeči moderátorů) a mohou ovlivňovat jazykové chování široké veřejnosti. Třebaže jsme si vědomi toho, že výrazová složka některých mediálních pořadů může někdy vypovídat spíše jen o krátkodobé módnosti některých výrazů nebo tvarů a konstrukcí než o dlouhodobých vývojových trendech češtiny, domníváme se, že rozsáhlý korpus mediálních diskusních pořadů za období jednoho desetiletí umožňuje postihnout jak jevy pomíjivé, tak obecnější tendence mluveného úzu.

Na korpusu DIALOG můžeme např. dobře pozorovat, jak se do aktivní slovní zásoby uživatelů dostávají prostřednictvím médií jednak tolik diskutované anglicismy, jednak nejruznější módní slova česká, z nichž některá jsou dobově podmíněná a užívána třeba jen přechodně (srov. operování výrazy typu *kupónová privatizace, dvojdomek, balíček* apod.), jiná se zřejmě v češtině usadí na delší dobu: *oslovit, vysílat signály, vnímat signály, vyjednávat, ustát, vytunelovat, horizonty, mantinely, relace, formáty* aj. Korpusové vyhledávání může významně pomoci při mapování výskytů, distribuce a frekvence určitého módního slova a popřípadě upozornit na jeho omezenou či kolísající životnost.

Korpusová analýza mluvených mediálních projevů může ukázat současné tendence v konkurenci některých spisovných a nespisovných jevů a kolísání mezi nimi. Frekvenční analýza může např. odhalit, jak mluvčí používají v mluvených veřejných projevech volné morfémy *bychom/by sme* a *bych/by sem*, s jakými problémy se setkávají při skloňování číslovky (*vyhrál dvěšřtřpadesátřpřeti hlasy*) a nakolik do mluveného jazyka pronikají analytické tvary stupňování přídavných jmen (*vřice aktivně*) konkurující tvarům syntetickým, jakým způsobem se mluvčí vyrovnávají s duálovými koncovkami podstatných jmen (*mřt osud ve svých rukou/rukách*), přídavných jmen i číslovek (*s obřema/dvěmi svřimi známřmi / svřima známřma* apod.), jak z mluvené češtiny mizí plurálová shoda neuter (*slova zazněly*), jak vypadá konkurence reflexivního a složeného pasiva v mluvených televizních projevech (*jsme připravenř / jsme připravenř, jsem připraven / jsem připravenř*), proč mluvená řeč tolik svádř k užívání demonstrativ (*ty opatřenř tř vládř*), že některé spojovací výrazy ustoupily jinřm (*aniř / bez toho ře by*), jak se vracř oblřba výrazů považovaných dřve za kniřnř (*nic-mřnř*) apod.

Zajřmá-li nřs vřskyt jednotlivřch – v zásadř izolovaných – jevů, resp. konkurence jevů blřzkřch, synonymnřch apod., je korpus spolehlivřm a rychlřm zdrojem poučenř o frekvenci a distribuci jazykovřch jevů, kterř si pro analřzu vybereme. Zjiřřtřváváním frekvence vřskytů vřak nemusř nřše analřza konřit. Hlubbř analřza korpusu dialogickřch projevů mřže např. ukázat, jak je volba toho ři onoho jazykovřho jevu spjata prřvř s dialogickřm

půdorysem projevu, jak souvisí s dialogickým přizpůsobováním partnerů, apod. Dobře patrné je to na střídání kódů češtiny spisovné a obecné (srov. např. Müllerová – Hoffmannová, 2001). Kód spisovné a obecné češtiny je v kompetenci jednotlivých mluvčích zpravidla v různém poměru, objevují se mluvčí, kteří po celou dobu svého projevu dodržují zásady spisovné češtiny, jiní mluvčí zase záměrně používají i ve veřejném projevu češtinu obecnou, a ze všeho nejčastěji se setkáváme s tím, že mluvčí oba tyto kódy nějakým způsobem střídají; my jako analytici se pak můžeme pokoušet odhalit dialogickou motivaci takového střídání (citace, přizpůsobování se partnerovi, vyvedení z míry, rozhořčení, tápání, ztráta sebekontroly, nabytí sebedůvěry, intimita, naléhavost, odlehčení, humor apod.). V korpusu mediálních textů se vyskytují nejen bezděčné, ale i záměrné překlady z jednoho kódu do druhého, kdy je např. moderátor nucen transponovat hostovu repliku nebo dotaz do kultivovanější podoby (bývá to nezbytné při moderování Kotle), anebo si host převádí moderátorovu otázku do svého stylově nižšího idiolektu apod.

### **Závěrem**

Výzkum volby jazykových prostředků v souvislosti s výzkumem dialogického chování tvoří nejzazší cíl zkoumání, pro nějž bychom si přáli korpus DIALOG využít. Jsme si totiž vědomi toho, že se nám v daném korpusu podařilo shromáždit texty, v nichž je velmi intenzivně zastoupena řeč vedená mluvčím v 1. osobě singuláru, resp. plurálu, a orientovaná na 2. osobu plurálu, resp. singuláru. Korpus se proto dobře hodí na analýzu jazykových projevů spjatých se subjektem mluvčího, vyjádřením jeho stanoviska, hledáním a prosazením vlastního výrazu, často v polemice s adresátem; zároveň se hodí využít i k analýze úsilí o apelativnost a přesvědčivost vzhledem k adresátovi, k analýze souhlasu a nesouhlasu, interakce a intersubjektivitu. Korpus DIALOG je soubor textů, v nichž dominuje osa mluvčí – adresát, a v tom spočívá jeho specifičnost a specializovanost. Texty v něm soustředěné jsou z tohoto hlediska do značné míry komplementární k textům obsaženým v korpusu SYN2000, tedy k textům psaným, informujícím převážně o dění vztahujícím se k 3. gramatické osobě. Proto je naším cílem (1) ověřovat na tomto korpusu alespoň některé lingvistické závěry, k nimž dospěje analýza jazykového materiálu soustředěného v SYN2000, a (2) pokusit se o nárys gramatiky mluveného jazyka, která by byla komplementární ke gramatice jazyka psaného.

Myšlenka postihnout gramatiku mluveného jazyka je pro lingvisty lákavá, předpokládá značnou odvahu a intuici a vyžaduje také modifikaci pojmů, s nimiž pracuje lingvistika psaného jazyka. Pátrat v transkriptech mluvených projevů po dodržení pravidel známých z gramatiky, syntaxe a struktury psaného jazyka znamená nevšímat si opakování, zadržávání, vyšínutí, nedokončení, přerušení... Ovšem zároveň s tím vědomím, že právě tyto jevy jsou projevem vlastní strukturovanosti mluvené řeči. Jak uvádí B. Gasparov (1978) ve své stati o sémiotice mluvené řeči, aby metaforicky podpořil myšlenku její specifické strukturovanosti, i „hromada písku má svou strukturu“ (srov. k tomu Čmejrková, 1993). V popisu stavby mluvených projevů lze navázat na práce existující (z českých prací zejména Müllerová, 1994; Müllerová – Hoffmannová, 1994; Hoffmannová – Müllerová, 1999, 2000) a doufat, že mluvená řeč, jak ji reprezentuje korpus mediálních dialogů, vydá i svá další tajemství (Hoffmannová, 2003, s. 69), jako je už dříve vydala řeč psaná.

**Příloha 1. Přepsané pořady (stav k 1. 8. 2004)**

NÁZEV POŘADU	KANÁL	STOPÁŽ <sup>16</sup>	OBDOBÍ	POČET
„21“	ČT 2	6–10 min. <sup>17</sup>	5/1998 – 9/1999	5
7 čili Sedm dní / Sedmička <sup>18</sup>	Nova	60 min.	6/1997 – 5/2004	68
Aréna	ČT 1, 2	45 min.	1/1998 – 5/1999	5
Bez imunity	ČT 1, 2	54 min.	3/2002	1
Bez kravaty	TV 3	45 min.	3/2002 – 4/2002	2
Čaj pro třetího	ČT 1, 2	27 min.	8/2002 – 6/2003	4
Další, prosím	Prima	55 min.	11/2000 – 12/2000	3
Debata	ČT 1, 2	60 min.	7/1997 – 11/1997	5
Dvojhlas	ČT 1	22 min.	1/2003 – 3/2003	7
Karanténa	ČT 2	60 min.	11/1997 – 7/1998	4
Katovna	ČT 2	54 min.	3/1999 – 6/2000	5
Klub Netopýr	ČT 2	46 min.	12/1996 – 12/2002	6
Kotel	Nova	50 min.	3/2001	1
Krásný ztráty	ČT 1, 2	27 min.	4/2000 – 3/2004	37
Média	ČT 2	26 min.	2/2003 – 10/2003	14
Na hraně	ČT 2	46 min.	12/1997 – 1/2004	17
Na plovárně s...	ČT 1, 2	22 min.	3/1999 – 4/2004	30
Naostro	ČT 1	40 min.	9/2000 – 8/2002	17
Nedělní partie	Prima	45 min.	7/1998 – 5/2004	47
O cestě na Hrad a zase zpátky	ČT 1, 2	52 min.	6/2003	1
Otázky Václava Moravce	ČT 1, 2	55 min.	2/2004 – 3/2004	4
Pohled zvenku	ČT 2	47 min.	5/2001	1
Pro a proti	ČT 2	60 min.	6/1998 – 1/1999	5
Přísně veřejné!?	ČT 1	45 min.	12/1999	1
Respektování	ČT 2	60 min.	4/1997 – 11/1997	6
Rozhovory	ČT 2	27 min.	4/2001 – 11/2002	3
Sauna	Prima	40 min.	4/2002	1
Síto	ČT 2	57 min.	11/2003	1
Sněží	ČT 2	60 min.	9/1997 – 12/2001	6
Špona	ČT 1, 2	40 min.	9/2002 – 2/2003	6
V pravé poledne	ČT 1	60 min.	5/1998 – 7/2000	19
Vertikála	ČT 1, 2	56 min.	12/2003 – 3/2004	4
Z očí do očí	ČT 1, 2	27 min.	12/1997 – 5/2003	5
Za dvěma je A. G.	ČT 2	27 min.	9/1999	1
Ženský hlas	ČT 2	24 min.	8/2002 – 10/2002	7
<b>SOUHRNNĚ</b>	<b>ČT 1, 2, Nova, Prima, TV 3</b>	<b>252 hod. 27 min.<sup>19</sup></b>	<b>12/1996 – 5/2004</b>	<b>349</b>

<sup>16</sup> Výrazem stopáž rozumíme čistý čas trvání pořadu, který je totožný s časem nahrávky (pokud není uvedeno jinak).

<sup>17</sup> Přepsány jsou jen několikaminutové rozhovory s hosty tohoto zpravodajsko-publicistického pořadu, nikoli celý pořad.

<sup>18</sup> V září 2000 byl název 7 čili Sedm dní nahrazen jménem Sedmička. V textu na pořad odkazujeme buď jedním, nebo druhým názvem, podle toho, z kterého období citovaná pasáž pochází.

<sup>19</sup> Údaj vyjadřuje součet čistého času všech transkribovaných pořadů.

## Příloha 2. Notační značky používané v korpusu DIALOG

VM	označení mluvčího
TH: to já [nevím]	simultánně pronesené úseky (tzv.
VB: [v té] knize	<i>overlap</i> )
=	okamžité (bezpauzové) navázání na repliku partnera
<u>zase</u>	zdůraznění slabiky/slova
že: že:: že:::	protažení hlásky (v závislosti na délce)
e e: hm ehm mhm	hezitační a rezponzní zvuky
(.)	krátká pauza
(..)	delší pauza
(...)	dlouhá pauza
?	vysoké stoupnutí hlasu
,	mírné stoupnutí hlasu
.	klesnutí hlasu
mon-	nedokončení slova
<uhodil>	úsek pronesený se smíchem
(ale)	předpokládaný, ne dobře srozumitelný výraz
( )	nesrozumitelný výraz
((směje se))	komentář přepisovatele
...	vypuštění části transkriptu

## LITERATURA

- BACHMANNOVÁ, J. (1999): Zvukový archiv nářečních textů. *Naše řeč*, 82, s. 47–48.
- BITRICH, T. – KONOPÁSEK, Z. (2001): Transcriber – pohodlnější přepisování, a možná i něco navíc. *Biograf*, 24, s. 125–146. Dostupné též online na adrese: <<http://biograf.institut.cz/clanky/clanek.pp?clanek=v2412>>.
- CLAYMAN, S. E. (1991): News interview openings: Aspects of sequential organization. In: P. Scannell (ed.), *Broadcast Talk*. London: Sage, s. 48–75.
- CLAYMAN, S. E. (1993): Reformulating the question: A device for answering / not answering questions in news interviews and press conferences. *Text*, 13, s. 159–188.
- CLAYMAN, S. – HERITAGE, J. (2002): *The News Interview: Journalists and Public Figures on the Air*. Cambridge: Cambridge University Press.
- COUPER-KUHLEN, E. – SELTING, M. (1996): Towards an interactional perspective on prosody and a prosodic perspective on interaction. In: E. Couper-Kuhlen – M. Selting (eds.), *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press, s. 11–56.
- ČERMÁK, F. (1995): Jazykový korpus: prostředek a zdroj poznání. *Slovo a slovesnost*, 56, s. 119–140.
- ČMEJRKOVÁ, S. (1993): Slovo psané a mluvené. *Slovo a slovesnost*, 54, s. 51–58.
- ČMEJRKOVÁ, S. (1999): Televizní interview a jiné duely: mediální dialog jako žánr veřejného projevu. *Slovo a slovesnost*, 60, s. 247–268.
- ČMEJRKOVÁ, S. (2004): Emotions in language and communication. In: E. Weigand (ed.), *Emotion in Dialogic Interaction: Advances in the Complex*. Amsterdam – Philadelphia: John Benjamins, s. 37–57.
- ČMEJRKOVÁ, S. (v tisku): Bychom, nebo bysme? *Naše řeč*.
- ČMEJRKOVÁ, S. – DANEŠ, F. – HAVLOVÁ, E. (eds.) (1994): *Writing vs Speaking: Language, Text, Discourse, Communication*. Tübingen: Gunter Narr Verlag.
- ČMEJRKOVÁ, S. – HOFFMANNOVÁ, J. (eds.) (2003): *Jazyk, média, politika*. Praha: Academia.
- ČMEJRKOVÁ, S. – HOFFMANNOVÁ, J. – MÜLLEROVÁ, O. – SVĚTLÁ, J. (eds.) (1998): *Dialoganalyse VI: Referate der 6. Arbeitstagung Prag 1996 / Dialogue Analysis VI: Proceedings of the 6<sup>th</sup> Conference Prague 1996*. Tübingen: Max Niemeyer Verlag.



- DANEŠ, F. (1957): *Intonace a věta ve spisovné češtině*. Praha: Nakladatelství Československé akademie věd.
- DANEŠ, F. (1985): Intonace v textu (promluvě). *Slovo a slovesnost*, 46, s. 83–100.
- DANEŠ, F. (2004): Universality vs. culture-specificity of emotion. In: E. Weigand (ed.), *Emotion in Dialogic Interaction: Advances in the Complex*. Amsterdam – Philadelphia: John Benjamins, s. 23–32.
- DREW, P. – HERITAGE, J. (eds.) (1992): *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press.
- DU BOIS, J. W. (1991): Transcription design principles for spoken discourse research. *Pragmatics*, 1, s. 71–106.
- EDWARDS, J. A. (1993): Principles and contrasting systems of discourse transcription. In: J. A. Edwards – M. D. Lampert (eds.), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale – Hove – London: Lawrence Erlbaum Associates, s. 3–31.
- EDWARDS, J. A. – LAMPERT, M. D. (eds.) (1993): *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale – Hove – London: Lawrence Erlbaum Associates.
- EHLICH, K. (1993): HIAT: A transcription system for discourse data. In: J. A. Edwards – M. D. Lampert (eds.), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale – Hove – London: Lawrence Erlbaum Associates, s. 123–148.
- EHLICH, K. – REHBEIN, J. (1976): Halbinterpretative Arbeitstranskriptionen (HIAT). *Linguistische Berichte*, 45, s. 21–41.
- EHLICH, K. – REHBEIN, J. (1979): Erweiterte halbinterpretative Arbeitstranskriptionen (HIAT 2): Intonation. *Linguistische Berichte*, 59, s. 51–75.
- Eye dialect (2004). In: *Encyclopædia Britannica*. Encyclopædia Britannica Premium Service. Dostupné online na adrese: <<http://www.britannica.com/eb/article?eu=137499>>. Cit. 30. 7. 2004.
- GASPAROV, B. M. (1978): Ustnaja reč' kak semiotičeskij ob'jekt. In: *Semantika nominacii i semiotika ustnoj reči*. Acta et Commentationes Universitatis Tartuensis, 442. Tartu, s. 63–112.
- GREATBATCH, D. (1998): Conversation analysis: Neutralism in British news interviews. In: A. Bell – P. Garrett (eds.), *Approaches to Media Discourse*. Oxford: Blackwell, s. 163–185.
- GUMPERZ, J. J. – BERENZ, N. (1993): Transcribing conversational exchanges. In: J. A. Edwards – M. D. Lampert (eds.), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale – Hove – London: Lawrence Erlbaum Associates, s. 91–121.
- HAIJČ, J. (2000): Popis morfologických značek – poziční systém. Dostupné online na adrese: <<http://ucnk.ff.cuni.cz/manual/znacky.html>>. Cit. 23. 7. 2004.
- HAUSENBLAS, K. (1984): Text, komunikáty a jejich komplexy (Zamyšlení pojmoslovné). *Slovo a slovesnost*, 45, s. 1–7.
- HERITAGE, J. (1985): Analysing news interviews: Aspects of the production of talk for overhearing audience. In: T. A. van Dijk (ed.), *Handbook of Discourse Analysis*, 3. London: Academic Press, s. 95–117.
- HOFFMANNOVÁ, J. (1992): Metodologie „konverzační analýzy“ a transkripční symboly. In: J. Stachová (ed.), *Symbol v lidském vnímání, myšlení a vyjadřování: Sborník příspěvků*. Praha: Filozofický ústav ČSAV, s. 234–241.
- HOFFMANNOVÁ, J. (2000). K univerzálnosti a specifičnosti pojmu *dialog*. In: Z. Hladká – P. Karlík (eds.), *Čeština – univerzália a specifika 2: Sborník z konference*. Brno: Masarykova univerzita, s. 53–61.
- HOFFMANNOVÁ, J. (2003): Intertextové, metatextové, terminologické problémy české analýzy diskurzu. In: P. Žigo (ed.), *Philologica 56: Zborník Filozofickej fakulty Univerzity Komenského*. Bratislava: Univerzita Komenského, s. 67–72.
- HOFFMANNOVÁ, J. – MÜLLEROVÁ, O. (1999): *Dialog v češtině*. München: Verlag Otto Sagner.
- HOFFMANNOVÁ, J. – MÜLLEROVÁ, O. (2000): *Jak vedeme dialog s institucemi*. Praha: Academia.
- HOFFMANNOVÁ, J. – MÜLLEROVÁ, O. – ZEMAN, J. (1999): *Konverzace v češtině při rodinných a přátelských návštěvách*. Praha: Trizonia.
- HŮRKOVÁ, J. (1995): *Česká výslovnostní norma*. Praha: Scientia.
- HUTCHBY, I. – WOOFFITT, R. (1998): *Conversation Analysis: Principles, Practices and Applications*. Cambridge: Polity Press.

- KARHANOVÁ, K. (v tisku): Rhetorical question in polemical media dialogue (based on material drawn from Czech TV political debates). In: A. Betten – M. Dannerer (eds.), *Dialoganalyse IX: Dialogue in Literature and the Media*. Tübingen: Max Niemeyer Verlag.
- KOWAL, S. – O'CONNELL, D. C. (2000): Zur Transkription von Gesprächen. In: U. Flick – E. von Kardorff – I. Steinke (eds.), *Qualitative Forschung: Ein Handbuch*. Reinbek bei Hamburg: Rowohlt, s. 437–447.
- KRAUS, J. (2003): Argumentation in Czech political debates. In: M. Bondi – S. Stati (eds.), *Dialogue Analysis 2000*. Tübingen: Max Niemeyer Verlag, s. 277–282.
- LAPADAT, J. C. – LINDSAY, A. C. (1999): Transcription in research and practice: from standardization of technique to interpretative positionings. *Qualitative Inquiry*, 5, s. 64–86.
- LEECH, G. (2000): Anotační systémy pro značkování korpusů. In: F. Čermák – J. Klímová – V. Petkevič (eds.), *Studie z korpusové lingvistiky*. Praha: Univerzita Karlova v Praze – Nakladatelství Karolinum, s. 185–197.
- MÜLLEROVÁ, O. (1994): *Mluvený text a jeho syntaktická výstavba*. Praha: Academia.
- MÜLLEROVÁ, O. – HOFFMANNOVÁ, J. (1994): *Kapitoly o dialogu*. Praha: Pansofia.
- MÜLLEROVÁ, O. – HOFFMANNOVÁ, J. (2001): Míšení kódů a jeho hodnotové aspekty v různých typech komunikačních situací. In: *Naše a cizí v internetnické a interpersonální komunikaci / Swój i obcy w internetniczej i interpersonalnej komunikacji językowej*. Ostrava: Ostravská univerzita, s. 11–19.
- MÜLLEROVÁ, O. – NEKVAPIL, J. (1986): Pauzy v mluveném textu. *Slovo a slovesnost*, 47, s. 105–113.
- MÜLLEROVÁ, O. – HOFFMANNOVÁ, J. – SCHNEIDEROVÁ, E. (1992): *Mluvená čeština v autentických textech*. Jinočany: H & H.
- NEKVAPIL, J. (1999/2000): Etnometodologická konverzační analýza v systému encyklopedických hesel. *Češtinář*, 10, s. 80–87. Dostupné též online na adrese: <[http://ulug.ff.cuni.cz/sylaby/konv\\_anal.rtf](http://ulug.ff.cuni.cz/sylaby/konv_anal.rtf)>.
- NEKVAPIL, J. (2000): Teze k utváření interpretativní sociolingvistiky. *Časopis pro moderní filologii*, 82, s. 1–4.
- NEKVAPIL, J. (2002): Transkript v konverzační analýze. In: P. Karlík – M. Nekula – J. Pleskalová (eds.), *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, s. 494.
- NEKVAPIL, J. – MÜLLEROVÁ, O. (1988): K pauzám v komunikačním procesu. *Slovo a slovesnost*, 49, s. 202–208.
- O'CONNELL, D. C. – KOWAL, S. (1995): Transcription systems for spoken discourse. In: J. Verschueren – J.-O. Östman – J. Blommaert (eds.), *Handbook of Pragmatics: Manual*. Amsterdam – Philadelphia: John Benjamins, s. 646–656.
- PALKOVÁ, Z. (1994): *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru*. Praha: Karolinum.
- PSATHAS, G. (1995): *Conversation Analysis: The Study of Talk-in-Interaction*. Thousand Oaks – London – New Delhi: Sage.
- SACKS, H. – SCHEGLOFF, E. A. – JEFFERSON, G. (1974): A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, s. 696–735.
- SELTING, M. (2001): Probleme der Transkription verbalen und paraverbalen/prosodischen Verhaltens. In: K. Brinker – G. Antos – W. Heinemann – S. F. Sager (eds.), *Text- und Gesprächslinguistik: Ein internationales Handbuch zeitgenössischer Forschung, 2 / Linguistics of Text and Conversation: An International Handbook of Contemporary Research, 2*. Berlin – New York: Walter de Gruyter, s. 1059–1068.
- SELTING, M. – AUER, P. – BARDEN, B. – BERGMANN, J. – COUPER-KUHLEN, E. – GÜNTNER, S. – MEIER, CH. – QUASTHOFF, U. – SCHLOBINSKI, P. – UHMANN, S. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte*, 173, s. 91–122.
- STARÝ, Z. (1992): *Psací soustavy a český pravopis*. Praha: Univerzita Karlova.
- STUBBS, M. (2000): Using very large text collections to study semantic schemas: A research note. In: C. Heffer – H. Saunston (eds.), *Words in Context: A Tribute to John Sinclair on his Retirement* [CD-ROM]. University of Birmingham.
- TEN HAVE, P. (1999): *Doing Conversation Analysis: A Practical Guide*. London – Thousand Oaks – New Delhi: Sage.

- WEIGAND, E. (2003): *Sprache als Dialog: Sprechakttaxonomie und kommunikative Grammatik*. Tübingen: Max Niemeyer Verlag.
- WEIGAND, E. – DASCAL, M. (2001): *Negotiation and Power in Dialogic Interaction*. Amsterdam – Philadelphia: John Benjamins.
- WEIZMAN, E. (1998): Individual intentions and collective purpose. In: S. Čmejrková – J. Hoffmannová – O. Müllerová – J. Světlá (eds.) (1998): *Dialoganalyse VI: Referate der 6. Arbeitstagung Prag 1996 / Dialogue Analysis VI: Proceedings of the 6<sup>th</sup> Conference Prague 1996*. Tübingen: Max Niemeyer Verlag, s. 281–290.

---

## ZUSAMMENFASSUNG

### Gesprochenes Tschechisch in Fernsehdebatten: das Korpus DIALOG

Das Korpus DIALOG ist eine von zwei Sammlungen gesprochener Sprache, die im audiovisuellen Studio des Instituts für tschechische Sprache der Akademie der Wissenschaften der Tschechischen Republik archiviert sind. Im ersten Teil des Aufsatzes wird die Entstehung des Korpus im Jahre 1997 im Rahmen des Projektes „Dialog in der Welt der Menschen und Maschinen“ behandelt, der Zweck definiert, mit dem die Datenerhebung für dieses Korpus motiviert wurde, die distinktiven Merkmale dieses spezialisierten Korpus formuliert, das als ein zeitlich, interaktionell, genremäßig und teilweise auch thematisch abgegrenztes gesprochenes Korpus zu betrachten ist. Es wird auch versucht, die Texttypen zu definieren, von denen das Korpus Zeugnis ablegen kann. Der erste Teil informiert ebenfalls über die technischen Details der Archivierung der Audio- und Videoaufnahmen, über die Digitalisierungsarbeiten etc. Der zweite Teil bietet einen Überblick über die allgemeinen Prinzipien der Transkriptionssysteme und bringt eine theoretische Begründung des ausgewählten Transkriptionssystems in Hinblick auf die Notation der sequenziellen, segmentalen, suprasegmentalen und paralinguistischen Erscheinungen, einschließlich der praktischen Beispiele der angenommenen Lösung. Der dritte Teil berichtet über die bisherige Nutzung des Korpus DIALOG in der tschechischen Linguistik, einerseits für die Entwicklung der tschechischen Dialogtheorie, andererseits für die Analyse der spezifischen Merkmale des gesprochenen Tschechisch. Im Schlusskapitel werden die Perspektiven der weiteren Nutzung des Korpus DIALOG aufgezeigt.

*Ústav pro jazyk český AV ČR*  
*Letenská 4, 118 51 Praha 1*  
<cmejrkova@ujc.cas.cz>  
<jilkova@ujc.cas.cz>  
<kaderka@ujc.cas.cz>