



Reflexe proměn užívání češtiny na základě korpusu zpravodajského vysílání Českého rozhlasu v letech 1969–2005¹

Svatava Škodová – Jasňa Pacovská – Václav Lábus

ABSTRAKT:

Příspěvek představuje kvantitativní a kvalitativní proměny mluvené češtiny na základě korpusu automaticky přepsaných zpravodajských pořadů Českého rozhlasu. Použitý korpus obsahuje největší kolekci dokumentů nahraných za posledních 90 let a je na něm možno doložit vývoj vybraných jevů jazyka v tomto období. Pro jazykovou analýzu jsme vybrali zpravodajské relace Českého rozhlasu z období 1969–2005 a zahrnutý materiál pochází z hlavní zpravodajské relace Českého rozhlasu z let 1969–2005 a Československého rozhlasu před rokem 1993. Jmenovitě byly do výzkumu zahrnuty pořady *Rozhlasové noviny* (1969–1993) a *Ozvěny dne* (1993–2005). Do samotných sledovaných jevů dokladujících řečové proměny v daném období jsou zahrnuty nejen jevy v úzkém slova smyslu jazykové, ale i jevy vázané na jazykové vysílání a produkci textu jako takovou. Statisticky dokládáme proměny rychlosti pronesení promluv v celém období; mapujeme proměny ve stylizaci úvodních a finálních frází pořadů, dále proměny v používání subordinačních a koordinačních spojovacích prostředků. Nejrozsáhlejší část práce je věnována proměnám nejfrekventovanějších lexikálních sémanticky samostatných jednotek.

KLÍČOVÁ SLOVA:

mluvená formální řeč, rozhlasové vysílání, automatické rozpoznání řeči

ABSTRACT:

In this paper we present the quantitative and qualitative characteristics of the Czech Radio Broadcast News during a period of significant political and social changes in the Czech Republic (1969–2005). The research is mainly focused on the quantitative features of speech that can be determined from the results of automatic speech recognition system. We describe selected characteristics on the level of language and speech of the Radio Broadcast News; namely the changes in studio vs. out-of-studio speech ratio, distribution of speakers by sex, moderators and guest-speakers, approximate use of phrasal introductory and closing language specific for the time periods, speech speed changes, average silence length, coordinative vs. subordinate conjunctions ratio and the most frequent semantic words.

KEY WORDS:

spoken formal speech, radio broadcast news, automatic speech recognition

1 Tato práce navazuje na projekt č. DF11P01OVV013 Ministerstva kultury ČR ve výzkumném programu NAKI.

S. Škodová na textu pracovala za podpory projektu Univerzity Karlovy Progres č. 4, Jazyk v proměnách času, místa, kultury.

ÚVOD

Jazyk rozhlasového zpravodajství nabízí široký potenciál pro nejrůznější typy zkoumání (Šonková 2008). Vzhledem k tomu, že je těsně spjat s historickými událostmi stěžejními pro období jejich vzniku a realizace, reprezentuje rozhlasové zpravodajství také jazykové a sociální normy proměňující se v čase (Čmejrková-Hoffmannová 2011).

V textu navrhujeme možnosti zkoumání založeného na automatické transkripci jazykových dat rozhlasového zpravodajství. (K automatické transkripci blíže Nouza 2012). Data, o něž se opíráme, byla automaticky zpracována, rozpoznána a zpřístupněna jako součást projektu Zpřístupnění archivu Českého rozhlasu pro sofistikované vyhledávání realizovaného na Technické univerzitě v Liberci.² Předložené statistiky, založené na automatickém zpracování řeči rozhlasového zpravodajství, jsou doplněny o částečně manuálně shromážděný přehled vybraných kvantitativních rysů.³

Systém použitý pro transkripci archivovaných dokumentů je systém rozpoznávání plynulé řeči založený na rozsáhlém slovníku the large-vocabulary continuous speech recognition system (LVCSR), který byl vyvinut na Technické univerzitě v Liberci v roce 2003 a od této doby je stále zdokonalován.

Při výstupu dekodér poskytuje zásadní informace o obsahu dokumentu. Vedle standardní ortografické transkripce (Škodová-Kuchařová-Šeps 2012) systém také uchovává výslovnost každého slova, časy začátků a konců (měřeno v milisekundách), stejně jako neřečových událostí jako je ticho, nestandardní hluk (v rozhlasovém vysílání obvykle v podobě znělky), hlasové šумы, hezitační zvuky atd.

Každému dokumentu přísluší také deskriptory v podobě štítků pro řečovou/neřečovou událost, označení jazyka a označení mluvčího (pokud je známo jméno mluvčího je uvedeno jméno mluvčího, pokud ne, je uveden alespoň rod muž/žena).

V našem příspěvku používáme LVCSR výstup pro většinu demonstrovaných statistik. Jsme si však vědomi faktu, že každý tzv. speech-to-text systém je do jisté míry chybový. Rozsáhlé experimenty ukázaly, že typická chybovost, resp. tzv. word-error-rate (WER) byly okolo 11 % pro současné zpravodajství a okolo 14 % pro zpravodajství z období 1969–1989. Tyto hodnoty se mohou zdát vysoké, avšak v předchozích pracích (Boháč-Nouza-Blavka 2012; Boháč a kol. 2012) zkoumajících chybovost systému tým prokázal, že většinu chyb představuje vynechání krátkých slov (převážně jednografémové nebo dvougrafémové spojky a předložky) nebo chybné rozpoznání koncovek (v závislosti na akustické blízkosti českých sufixů). Z tohoto důvodu pokládáme dopad daného typu chyb na prezentované statistiky za nesignifikantní.

2 Text navazuje na projekt DP 11 P010VV013 podpořený Ministerstvem kultury v rámci projektu NAKI.

3 Za pomoc se statistickým zpracováním dat děkujeme kolegům z Ústavu informačních technologií a elektrotechniky, Technické univerzity v Liberci, M. Kuchařové a L. Šepsovi.





1. DATA

Na vybraných jevech heterogenního charakteru chceme ukázat možnosti zkoumání kvantitativních a kvalitativních proměn rozhlasového zpravodajství.

Materiál podrobený analýze pochází z hlavního každodenního večerního zpravodajství Českého rozhlasu (resp. Československého rozhlasu před rokem 1993) v období 1969 až 2005. Průzkum se týká jmenovitě souhrnu denních zpráv nazvaného *Rozhlasové noviny* (1969–1993) a *Ozvěny dne* (1993–2005), tj. 25–30minutových programů, které shrnují domácí a zahraniční události.

Řečový materiál rozhlasového zpravodajství nemá homogenní charakter; ačkoliv převládají tradiční čtené zprávy nahrávané ve studiu, jsou zde obsaženy také v exteriérech zachycené promluvy, které obsahují připravenou řeč (krátké komentáře nebo nahrávky veřejných projevů), ale i řeč spontánní (reportáže, krátká interview). Z hlediska složení mluvčích tedy v záznamech rozhlasového zpravodajství najdeme jak mluvčí profesionální, tak i neprofesionální.

Objem dat použitých pro statistiky představené v našem příspěvku zahrnuje reprezentativní vzorek časově se vyvíjející řeči v rozsahu 6.580 hodin; 48.721.952 slov.⁴

Pro ilustraci vývojových změn zpravodajství Českého rozhlasu jsme zkoumané časové období rozdělili do osmi segmentů v rozsahu pěti let, vyjma posledního období zahrnujícího dva roky. Pětileté segmenty pokládáme za dostatečně detailní pro popis změn s ohledem na celou periodu 42 let.

2. POMĚR STUDIOVÉ A MIMOSTUDIOVÉ ŘEČI

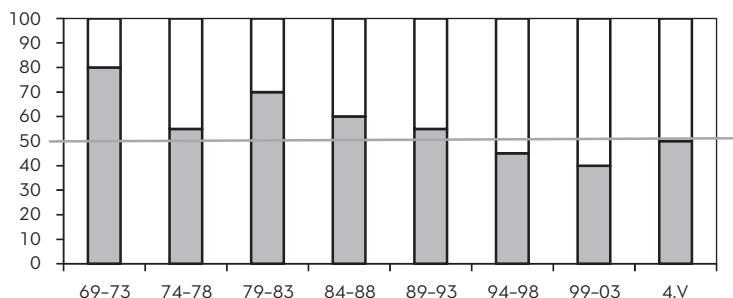
V průběhu celého sledovaného období můžeme pozorovat konstantní nárůst objemu mimostudiové řeči, a naopak pokles objemu řeči studiové (Graf 1). Studiová řeč zahrnuje hlasité čtení, které reprezentuje de facto psaný jazyk, a mimostudiovou řeč, která reprezentuje mluvenou řeč připravenou i spontánní. Jednoznačně převažující objem studiové řeči na počátku období poklesl o 30 % na začátku 90. let. Tento postupný pokles je možné vysvětlit jednak stále se zlepšujícími technickými podmínkami pro záznam a přenos řeči, předpokládá však lze i snahu o oživení a následnou vyšší atraktivitu zpravodajského vysílání.

Zároveň už od 80. let vidíme snahu o objemové vyrovnání studiové i mimostudiové řeči.

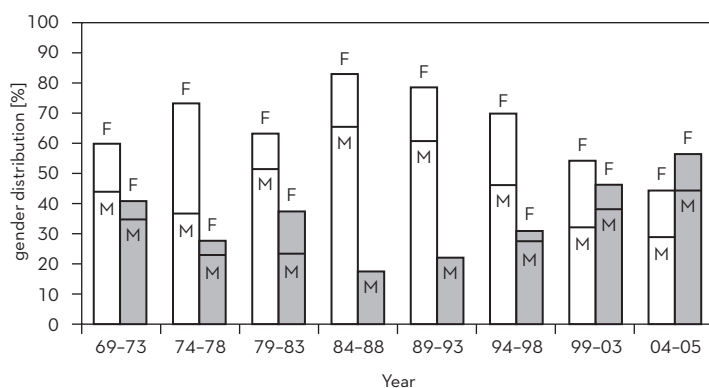
3. GENDEROVÁ DISTRIBUCE ROZHLASOVÉHO ZPRAVODAJSTVÍ

Graf 2 ilustruje poměr mužských a ženských mluvčích v rozhlasovém vysílání. Bílé sloupce reprezentují moderátory, šedé sloupce reprezentují hosty. Vyjma jediného období 1974–1978 převažují mezi moderátory muži. Od roku 1994 lze konstatovat po-

⁴ Slovem rozumíme jeho grafickou realizaci, tj. řadu grafémů (případně i grafém jediný) ohraničenou mezerami.



GRAF 1⁵ — Procentuální poměr studiové (šedá) a nestudiové (bílá) řeči v rozhlasovém vysílání.



GRAF 2 — Genderová distribuce v rozhlasovém vysílání. Male (M) vs. female (F) poměr moderátorů (bílý sloupec) a poměr mluvčích-hostů (šedý sloupec).

měrové vyrovnávání mezi moderujícími muži a ženami, ačkoliv muži v této profesi stále mírně převažují. Daleko překvapivější jsou výsledky zjištění poměru mužů/žen mezi hosty zpravodajství. V období 1984–1994 na tomto zpravodajství žádné hostující ženy neparticipovaly a z daných sledování není možné ani konstatovat tendence k nárůstu jejich účasti.

4. ÚVODNÍ A FINÁLNÍ FRÁZE POŘADŮ

V této části textu se zaměříme na charakteristiku úvodních a finálních pasáží rozhlasového vysílání. Tyto části nejenom rámuji obsahovou složku pořadu, jejímž cílem je předání informace v co možná nejobjektivnější podobě, zároveň také představují kontaktní pasáže, ve kterých se hlasatel nezaměřuje pouze na sdělovanou informaci, ale na adresáta této informace. Naším cílem bylo sledovat, zdali je možné postihnout v těchto částech měnící se postoj k adresátovi, resp. měnící se pojetí ad-

5 Data pro grafy byla zpracována v rámci analýz (viz Kuchařová a kol. 2013).



resáta, a dále snahu ovlivnit přijetí informací předávaných v hlavní části pořadu. Pro jednotlivá období v textu pro ilustraci vybíráme řečové ukázky charakterizující úvody a závěry. V této části nejsou data charakterizována po letech ani v pravidelných obdobích, ale jsou vybírány časové úseky odpovídající proměnám sledovaného řečového jevu.

1969–1976

Rámcové pasáže jsou v tomto období typické krátkostí, stereotypizací a neosobním charakterem, tj. nereflexivním ani osoby posluchače, ani osoby moderátora. Převažují úvody, ve kterých je pouze konstatován následující průběh vysílání, tj. předání informace. Začátky zpráv z historického období normalizace se pohybují ve variantách, které ukazuje příklad (1).

- (1) ZAČ: *vysíláme rozhlasové noviny nejprve stručný přehled hlavních zpráv / přehled dnešních hlavních událostí / úvodem přehled hlavních zpráv/hlavní události dne / vysíláme rozhlasové noviny [n::silence] nejprve v nich uslyšíte nejdůležitější zprávy z domova a ze zahraničí / v jejích úvodu (si poslechněte) přehled hlavních událostí*

V některých případech se v úvodu objevuje také pozdrav či oslovení posluchačů, tato forma však není standardem. Ještě řidší je kombinace obojího *dobry večer, vážení posluchači*, častěji se vyskytuje pouze úvodní pozdrav *dobry večer, vysíláme rozhlasové noviny*.

Převážná většina zakončení rozhlasových zpráv se omezuje na konstatování *Končíme rozhlasové noviny / rozhlasové noviny skončily / a tím končíme dnešní rozhlasové noviny*.

Variantnost a personální aspekt se v rámcových pasážích objevuje jen v mimořádných situacích jako např.: *vítáme vás k poslechu posledních rozhlasových novin v tomto roce*.

1977–1988

K určité změně dochází v roce 1977, od tohoto roku můžeme sledovat pravidelné uplatnění pozdravu v úvodní frázi, tj. *Dobry večer*.

V úvodu vysílání dochází také k nárůstu individualizace reprezentované představením hlasatelů jako doprovodného faktoru pro orientaci posluchače ve zprávách.

- (2) *vysíláme rozhlasové noviny, kterými vás budou provázet hlasatelé Hana Mašková a Viktor Brehovský, úvodem přehled hlavních zpráv*

Finální pasáže se neliší od předchozího období.

1989–1990

Radikální změnou procházejí úvody a závěry v letech politického převratu. Tyto dva roky jsou poznamenány vysokou variabilitou v úvodních pasážích a není možné ho-



vořit o přesném textovém vzorci, který by sledovaly ani o využití stanovených frází. V důsledku absence předepsaného jazyka lze sledovat v rámci autorských hlasatel-
ských inovací rozšíření jak úvodní, tak závěrečné pasáže. Tyto změny jsou posti-
žitelné od května 1989 a jejich nástup je explicitně formulován jako výzva pro pos-
luchače ve společném hledání nové formy; posluchač v tomto období přestává být
pouhým cílem zpravodajské relace, ale stává se komunikačním partnerem.

- (3) *dobrý večer vysíláme rozhlasové noviny [n::silence] tato hlavní zpravodajská relace
československého rozhlasu [h::breath] jak už znělka napověděla [h::breath] bude
mít ode dneška jinou podobu než na jakou jste zvyklí [h::breath] průvodcem se pro
vás stane vždy jeden z redaktorů který spolu s dalšími spolupracovníky [h::breath]
tento pořad připravil [n::silence] naše snaha o kvalitnější obsah [h::breath] si vy-
nutila i změnu formy rozhlasových novin [h::breath] byli bychom rádi kdybyste i vy
naši posluchači [h::breath] přispěli svými názory k tvorbě tohoto pořadu [n::silence]
těšíme se na vaše dopisy [h::breath] a teď už k událostem dnešního dne*

Celkově se úvody a závěry zpravodajských pořadů, jak ukazuje příklad (4), výrazně
orientují na příjemce a simulaci dialogického stylu s ním.

- (4) *jsou pro vás připraveny rozhlasové noviny; jejich dobrý poslech a zároveň příjemný
večer vám přeje XY; XY vás bude provázet; vám neseme rozhlasové noviny; po pří-
jemném dnu kdy počasí po delší době naznačilo že léto přece jenom trvá [h::breath]
jsme pro vás připravili výběr agenturního zpravodajství; při sledování čtvrtěních
rozhlasových novin hlavního zpravodajského pořadu dne vás vítá XY; k vám při-
cházejí rozhlasové noviny; při poslechu rozhlasových novin vás vítají XY; nerušený
poslech vám přejí vaši dnešní průvodci; vás vítáme k poslechu rozhlasových novin;
vítejte u příjemců; jsme pro vás zaznamenali také tyto informace*

Odklon od jediné stereotypní fráze a autorské formulace úvodů ukazují příklady
(5)–(10).

- (5) *připraveno je čtvrtění vydání rozhlasových novin s datem pětadvacátého května
devatenáct set osmdesát devět [h::breath] jaké jsou jejich titulky*
(6) *znělka ohlásila dnešní vydání rozhlasových novin*
(7) *je půl sedmé tedy čas k pravidelné půlhodinové zpravodajské relaci*
(8) *při sledování čtvrtěních rozhlasových novin hlavního zpravodajského pořadu dne
vás*
(9) *dobrý večer i v úterý pátého prosince k vám přicházejí rozhlasové noviny [h::breath]
tedy souhrn zpravodajství z událostí doma i za hranicemi naší vlasti*
(10) *dobrý večer tak jako každý den v tuto dobu začínají i dnes rozhlasové noviny*

Velkou variabilitu vykazuje uvození přehledu zpráv, příklad (11).

- (11) *jaké jsou naše titulky; úvodem několik zpráv telegraficky; jaké nejdůležitější události
přinesl pátek patnáctého září; úvodem rozhlasových novin o tom co přinesl dnešní*



den u nás doma a ve světě; poslechněte si jejich stručný přehled; nejdříve stručně k těm nejdůležitějším; nejprve stručně z jejich obsahu; tak jako vždy i tentokrát nejdříve charakteristika jejich obsahu; úvodem vybíráme z jejich obsahu; nyní přehled zpráv; z jejich obsahu vyjímáme; jaké nejdůležitější události přinesl pátek patnáctého září

Variabilita úvodních frází se pohybuje mezi následujícími dvěma textovými vzorci jako minimálními a maximálními krajními body.

Minimální:

POZDRAV — KONSTATOVÁNÍ PŘEHLEDU ZPRÁV — SAMOTNÉ ZPRÁVY

Maximální:

POZDRAV — LOKALIZACE STANICE — PŘIVÍTÁNÍ K POSLECHU — SPECIFIKACE, KDO JE HLASATELEM — PŘÁNÍ DOBRÉHO POSLECHU V NÁVAZNOSTI NA KONKRÉTNÍ SITUACI — UVEDENÍ PŘEHLEDU ZPRÁV — SAMOTNÉ ZPRÁVY (v různém pořadí)

Finální pasáže pořadů v tomto období vykazují vyšší stereotypičnost než úvody. Ačkoliv se délka závěrečné fráze časově rozšířila stejně jako úvody, lze sledovat textový vzorec, který zaznamenává jen dílčí obměny. Typickým znakem koncových frází tohoto období je snaha o upřesnění pracovníků spolupracujících na zpravodajské relaci. Finální fráze zahrnují výčet všech nejdůležitějších osob, stylisticky se však nejedná o pouhý výčet osob v určitých funkcích, ale je zakomponován personální aspekt odrážející vztah osob k posluchači simulující partnerský vztah, např. *loučí se, přeje hezký zbytek večera, od mikrofonu na slyšenou*.

- (12) *rozhlasové noviny končí vedoucími vydání byla Martina Čiklová zpravodajskou směnu řídil Zdeněk Kuchyňka produkci Zuzana Hudečková [n::rumble] z technického pracoviště se loučí mistr zvuku Miloslav Přerost technička Helena Nejdlová [h::breath] a hezký zbytek večera vám přeje i režisérka Jaroslava Kaucká [h::breath] a od mikrofonu Luboš Lidický [n::rumble] na slyšenou zítra kdy vás rozhlasovými novinami bude provázet Stanislav Benda*

Standardním se stává přání hezkého poslechu a příjemného strávení času po skončení relace, toto přání je opět aktualizováno v návaznosti na reálnou situaci.

- (13) *končí rozhlasové noviny ale nekončí ještě dnešní den [h::breath] abyste i jeho zbytek prožili v letní pohodě vám přeje všichni kdo pro vás [h::breath] hlavní zpravodajskou relaci československého rozhlasu připravovali*

1991–1995

Rokem 1991 začíná období odklonu od autorských obměn jazyka úvodů a závěru zpravodajského vysílání a návrat k unifikovaným rámcovým pasážím, tato unifi-

kace již však není stoprocentní a lze ji pokládat za funkční, objevují se dílčí proměny nesoucí stopy perspektivy posluchače/moderátora. Jednoznačně však dochází k opětovnému zestručnění, resp. minimalizaci obsahového i časového rozpětí úvodní a finální fráze.

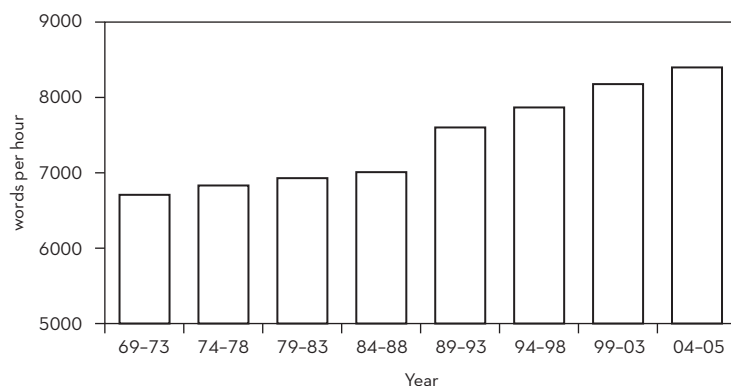
Zestručněny jsou např. výčty spolupracovníků podílejících se na konkrétním vysílání.

- (14) *za zpravodajský kolektiv vám příjemný večer přejí jejich dnešní průvodci Jaroslav Hausner a Věra Čermáková*
 (15) *dnešní relaci připravila a od mikrofonu vás zdraví Jaroslava Cermanová*
 (16) *Ozvěny dne připravila zpravodajská směna Igora Maňoura*

Z pořadů také nemizí kontaktní orientace na posluchače v jejich úvodu vás zavedeme na; vítáme vás u poslechu rozhlasových novin.

5. TEMPO ŘEČI A DÉLKA PAUZ

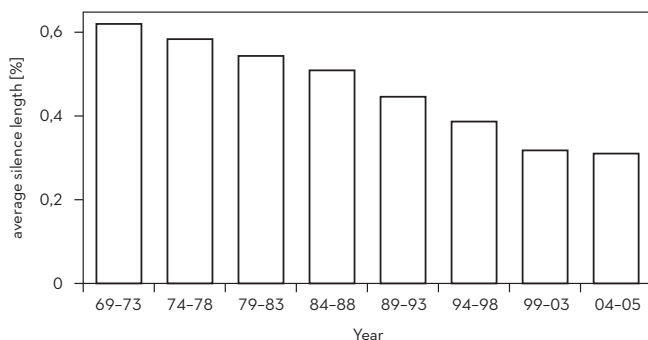
Graf 3 ukazuje kvantitativní změny rychlostního charakteru řeči. Od roku 1969 můžeme sledovat konstantní nárůst slov vyslovených za hodinu vysílání. Nejnižší průměrné množství slov v období 1969–1973 bylo 6.712 slov, v období 2004–2005 se jejich počet zvýšil v průměru na 8.404 slov za hodinu. Toto zvýšení představuje nárůst o 1.692 slov za hodinu v průběhu celého sledovaného období. Zvláště patrný vzrůst rychlosti produkce je patrný po roce 1989. Bylo by zajímavé porovnat vzrůst rychlosti řeči mediálního charakteru s rychlostí řeči v každodenní běžné komunikaci; teprve toto porovnání by ukázalo, zda se jedná o jednotný trend češtiny jako jazyka, či zda jsou rychlostní změny vynuceny pouze komerčními důvody, případně nápodobou zahraničních typů mediálního vysílání.



GRAF 3 — Rychlost řeči měřená ve slovech průměrně vyslovených za hodinu.



Ruku v ruce s tempem řeči jde i délka pauz, které v řeči činíme. Presupozice spojená s akcelerací řečové produkce ukazuje na pravděpodobnost zkracování pauz v řeči. Tato hypotéza se plně potvrdila, což ukazuje Graf 4, i když z tohoto grafu vyplývá, že zkracování ticha je zcela postupné a v celém sledovaném období se nevyskytuje žádný výrazný předěl, jako tomu bylo v roce 1989 při zvyšování rychlosti řeči. Graf 4 ukazuje průměrnou délku pauz v sekundách za dané období, délka pauz se zkrátila téměř o polovinu, tj. z pauz dlouhých průměrně 0,64 sekundy v roce 1969 na pauzy dlouhé 0,32 sekund. Opět se zde nabízí možnost dalšího srovnávání s pauzovou délkou v běžně mluvené češtině a s délkou pauz v mediální produkci v jiných jazycích, zvl. angličtině.



GRAF 4 — Průměrná délka pauz ve vysílání rozhlasového zpravodajství.

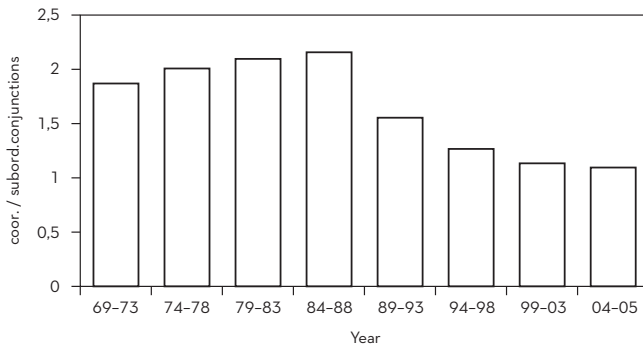
6. UŽITÍ KOORDINAČNÍCH A SUBORDINAČNÍCH SPOJEK

Graf 5 ukazuje změny užití koordinačních a subordinačních spojek v rozhlasovém zpravodajství. K sledování jejich poměru užití nás podnítila monografie H. Kušeho (1998), která se podrobně zabývá využitím koordinačních konjunkcí v přednesech politiků a v kázáních. Kuše ve své monografii ukazuje na možnosti využití koordinace jako prostředku s menší, resp. nejasnější, informační hodnotou, než je informačně jednoznačnější subordinace.

Pro naše sledování jsme zvolili nejčastější české jednoslovné spojky uvedené v publikaci Statistiky češtiny (2009), a to 40 subordinačních a 35 koordinačních spojek. Hodnota 1 na vertikální ose grafu znamená, že množství užití koordinačních spojek je rovno užití subordinačních spojek. Do roku 1989 byl průměrný poměr koordinace a subordinace přibližně 2:1, po roce 1989 se výrazně snížil přibližně na 1:1.

7. NEJČASTĚJI SE VYSKYTUJÍCÍ PLNOVÝZNAMOVÁ SLOVA

Lexikální systém každého jazyka je velmi dynamický na své periférii, kde reflektuje v lexikálním poli změny závislé na politických a sociálních změnách v průběhu času. V celém sledovaném období od roku 1969 do roku 2005 jsme vydělili dvě části, která



GRAF 5 — Poměr užití koordinačních a subordinačních spojek.

jsme označili jako komunistickou a postkomunistickou éru oddělené rokem 1989, ačkoliv jsme si plně vědomi historických a následně i jazykových přesahů mezi těmito dvěma obdobími. Pro daná období jsme vyhledali prvních 150 nejfrekventovanějších výrazů, ze kterých jsme následně vybrali pouze plnovýznamová slova, která odkazují k tématům a skutečnostem diskutovaným v jednotlivých obdobích. Pro nejčastěji se vyskytující slova v psaných korpusech srov. Bartoň a kol. (2009). Tabulka ukazuje, že celkové množství plnovýznamových slov mezi prvními 100 slovy bylo pouze 24 lexémů v období do roku 1989 a 12 slov v období počínajícím daným rokem, slova jsou seřazena sestupně podle frekvence užití, číslo ve sloupci ukazuje, na kolikátém místě se umístila mezi prvními 150 slovy. Pro srovnání jsme u druhého období s nižším užitím plnovýznamových slov mezi prvními 100 nejčastějšími slovy doplnili do tabulky 12 slov podle jejich frekvenčního umístění. (V Tabulce nejsou zahrnuta žádná gramatická a deiktická slova, ani základní číslovky.) Sloupec ukazuje jejich pořadí podle frekvence výskytu. Mnoho z těchto slov reflektuje těsné spojení mezi rozhlasovým zpravodajstvím a příslušným politickým systémem daného období.

1969-1989		1990-2005	
Slovo	Frekv. pořadí	Slovo	Frekv. pořadí
strany	14	dnes	22
dnes	18	České	58
výboru	19	první	70
komunistické	22	řekl	71
ústředního	27	jednání	72
svazu	30	republiky	77
Československa	34	vlády	78
republiky	35	prezident	79
další	42	ministr	81
sovětského	46	rozhlasu	82
socialistické	49	strany	85



1969–1989		1990–2005	
Slovo	Frekv. pořadí	Slovo	Frekv. pořadí
vlády	54	roku	98
práce	58	vláda	101
států	60	český	103
zemí	68	unie	111
jednání	70	korun	112
první	73	stupňů	114
národní	75	zpravodaj	116
rady	79	teploty	120
Praze	80	zahraničí	126
zasedání	83	lidí	127
předseda	94	myslím	130
organizace	95	let	132
spojených	99	evropské	133

TAB. — Nejčastější plnovýznamová slova v obdobích 1969–1989 a 1990–2005.

8. ZÁVĚR A VÝHLEDY

Národní archiv mluvených dokumentů představuje pozoruhodnou databanku pro výzkum mluveného slova. Archiv Českého rozhlasu zpracovaný automatickým rozpoznávacím mluveného slova obsahuje audiální záznam řeči paralelně s automaticky rozpoznávanými grafickými záznamy textů. Toto zpracování dovoluje snadné vyhledávání v mluvené řeči a její následnou analýzu. Dvojitý způsob zachycení informace výrazně rozšiřuje možnosti práce s informacemi.

Naše zkoumání jsme zaměřili na zachycení kvantitativních proměn zpravodajského vysílání Českého rozhlasu a jejich následnou kvalitativní interpretaci.

Pro text jsme zvolili několik rysů zpravodajského vysílání formálního a jazykového charakteru, na nichž jsme pomocí statistických metod ukázali změny v rozhlasovém zpravodajství. Zvl. jsme dokladovali (i) zvyšování spontánní a připravené řeči oproti čtenému zpravodajství; (ii) výrazně menší výskyt žen v rozhlasovém zpravodajství přecházející až do jejich absence v případě žen-hostů.

Statistiky pokrývající řečové změny ukázaly (i) měnící se úzus užití úvodních a zakončovacích frází; (ii) zvyšující se rychlost řeči s výrazným zlomem po roce 1989; (iii) zkracující se pauzy; (iv) snižující se užití koordinačních spojek ve prospěch užití spojek subordinálních; (v) změny ve frekvenci užití plnovýznamových slov spjatých s vládnoucím politickým systémem.

Cílem textu je představit výzkumný potenciál tzv. speech-to-text technologie aplikovaný na historických dokumentech rozsáhlého časového období, které s sebou nese i rozsáhlou materiálovou základnu. Text nepřináší podrobnou studii češtiny užívané v rozhlasovém zpravodajství po dobu více jak čtyř desetiletí, ale poukazuje na exaktní metody zkoumání mluveného slova v archivu Českého rozhlasu.

BIBLIOGRAFIE:

- Bartoň, T. a kol. (2009): *Statistiky češtiny*. Praha: NLN.
- Boháč, M. — Nouza, J. — Blavka, K. (2012a): Investigation on Most Frequent Errors in Large-scale Speech Recognition Applications, in: Sojka, P. a kol. (eds.), *Text, Speech and Dialogue 2012, Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer Verlag, s. 520–527.
- Boháč, M. a kol. (2012): Post-processing of the Recognized Speech for Web Presentation of Large Audio Archive, in: *35th International Conference on Telecommunications and Signal Processing (TSP)*, Praha: IEEE, s. 441–445.
- Čmejrková, S. — Hoffmannová, J. (eds.) (2011): *Mluvená čeština. Hledání funkčního rozpětí*. Praha: Academia.
- Kuchařová, M. a kol. (2013): On the Quantitative and Qualitative Speech Changes of the Czech Radio Broadcasts News within Years 1969–2005, in: Habernal, I. — Matousek, V. (eds.), *Text, Speech, and Dialogue — 16th International Conference*, Berlin/Heidelberg: Springer Verlag, s.360–368.
- Kuše, H. (1998): *Konjunktionale Koordination in Predigten und politischen Reden*. München: Verlag Otto Sagner.
- Nouza, J. a kol. (2012): Making Czech Historical Radio Archive Accessible and Searchable for Wide Public, *Journal of Multimedia* 7. 2, 159–169.
- Škodová, S. — Kuchařová, M. — Šeps. L. (2012): Discretion of Speech Units for the Text Post-processing Phase of Automatic Transcription (in the Czech Language), in: Sojka, P. a kol. (eds.), *Text, Speech and Dialogue 2012, Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer Verlag, s. 446–455.
- Šonková, J. (2008): *Morfologie mluvené češtiny: Frekvenční analýza*. Praha: NLN.

