

Univerzita Karlova v Praze

Filozofická fakulta

Fonetický ústav

Diplomová práce

Eliška Churaňová

**Fonotaktická osnova českého slova
a mluvního taktu**

**Phonotactic framework of the Czech word
and stress-group**

Praha, 2012

Vedoucí práce: doc. PhDr. Jan Volín, Ph.D.

Poděkování

Ráda bych tímto poděkovala doc. PhDr. Janu Volínovi, Ph.D. za příkladné vedení diplomové práce a cenné připomínky a rady v průběhu procesu jejího vzniku. Dále děkuji doc. RNDr. Vladimíru Petkevičovi, CSc. za pomoc s automatickým rozpoznáváním slovních druhů a Mgr. Lence Weingartové za odborné konzultace.

Velký dík patří také Mgr. Robertu Kostnerovi a mé rodině za pochopení a podporu při tvorbě této práce.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 17. 5. 2012

Abstrakt

Tato diplomová práce předkládá zevrubný popis konsonanticko-vokální struktury spisovné mluvené češtiny. Úvodní část je věnována přístupům a poznatkům souvisejícím s kombinatorikou a distribucí zvukových jednotek v řeči: pojednány jsou otázky řečového rytmu, aspekty členění souvislé řeči na promluvové úseky a mluvní takty a problematika fonotaktiky jazyků obecně i konkrétně češtiny. Na základě nahrávek 12 profesionálních mluvčích češtiny o celkovém rozsahu 6639 slov a 5368 mluvních taktů byly vytvořeny soubory dat, díky kterým bylo možné popsat CVCV struktury českých slov a mluvních taktů. Ve výsledcích tohoto výzkumu jsou uvedeny počty slov a taktů a jejich vztahy, četnost slovních druhů, počty hlásek ve slovech a taktech; zohledněny jsou i slabikotvorné likvidy a ráz. Dále je podán přehled nejčastějších CVCV vzorců slov a taktů a jejich variability v rámci slovních druhů. Poslední část práce se zabývá frekvencemi výskytu konsonantů a vokálů a jejich dvojic a trojic na různých místech v jednotce. Výsledky jsou průběžně porovnávány mezi sebou a také jsou srovnávány s výzkumem provedeným na rozsáhlejší psaném korpusu SYN2005 (2009). Tato práce může poskytnout základ pro další bádání v souvisejících jazykových oblastech (např. fonotaktika dialektů, mluvních stylů češtiny apod.).

Klíčová slova: fonotaktika češtiny, konsonanticko-vokální struktura, CVCV vzorce, ráz, rytmus řeči, prozodické členění, mluvní takt

Abstract

This master thesis provides a relatively detailed description of the consonant-vowel structure of standard spoken Czech. The first part covers approaches to and findings on the combinatorial system and distribution of sound units in speech; aspects of continuous speech segmentation into intonation phrases and stress groups, phonotactics of languages in general and Czech in particular, and speech rhythm are also addressed. Recordings of 12 professional speakers of Czech - comprising 6639 words and 5368 stress groups in total - have been used to create data sets that have enabled the author to describe CVCV structures of Czech words and stress groups. The results of this research present frequencies of words and stress groups and their relations to word-class dimension, frequencies of phones in words and stress groups; in this respect, both syllabic liquids and glottal stops have been taken into account. Further, the thesis includes an overview of the most frequent CVCV patterns in words and stress groups and their variability with regard to word classes. The final part focuses on how frequently consonants, vowels and their pairs and trios occur at different places within a unit. The results are continuously compared both between themselves and with the research that used the larger SYN2005 written corpus (2009). This thesis may form a basis for further research on related linguistic areas (e.g. phonotactics in dialects, speech styles of Czech, etc.).

Key words: phonotactics of Czech, consonant-vowel structure, CVCV patterns, glottal stop, speech rhythm, prosodic segmentation, stress groups

Obsah

1. Úvod	11
1.1 Charakteristika a cíl práce.....	11
1.2 Struktura práce.....	12
1.3 Poznámka.....	13
2. Teoretická východiska	15
2.1 Rytmus.....	15
2.1.1 Obecné poznatky.....	15
2.1.2 Rytmus řeči.....	16
2.1.3 Rytmické faktory v jazycích.....	18
2.2 Problematika prozodického členění se zvláštním zřetelem k jednotce mluvního taktu.....	23
2.2.1 Prozodie a intonace.....	24
2.2.2 Prozodické členění.....	25
2.2.3 Jednotka mluvního taktu.....	29
2.2.4 Mluvní takt v češtině.....	33
2.2.4.1 Definice taktu.....	33
2.2.4.2 Přízvučnost v češtině.....	37
2.2.4.3 Délka mluvních taktů.....	43
2.2.4.4 Tendence mluvního rytmu v češtině.....	48
2.2.4.5 Pojetí taktů v této práci.....	49
2.3 Fonotaktika a struktura slabiky.....	51
2.3.1 Obecný úvod k problematice fonotaktiky.....	51
2.3.1.1 Fonotaktika a morfonologie.....	51
2.3.2 Teorie rámce a obsahu.....	51
2.3.3 Fonotaktické poznatky napříč jazyky.....	55
2.3.4 Struktura slabiky.....	57
2.3.5 Struktura slabiky v češtině.....	62
2.3.6 Fonotaktické poznatky v češtině.....	65
2.3.6.1 Výzkum Jiřího Krause a Marie Ludvíkové (1966, 1987).....	65
2.3.6.2 Statistiky češtiny (Bartoň et al., 2009).....	67
2.3.6.2.1 Kombinace C a V ve slovech celkem.....	69
2.3.6.2.2 Kombinace C a V v iniciální pozici.....	70
2.3.6.2.3 Kombinace C a V ve finální pozici.....	72

2.3.6.2.4 Průměrná délka slova a taktu	73
2.3.6.2.5 Struktura slova.....	74
2.4 Shrnutí	77
3. Metoda	79
3.1 Materiál	79
3.2 Postup	80
3.2.1 Příprava materiálu	80
3.2.2 Zpracování materiálu.....	81
3.2.3 Získání a zpracování dat	89
4. Výsledky.....	91
4.1 Poznámky ke struktuře kapitoly.....	91
4.1.1 Poznámka ke klasifikaci jednotek.....	92
4.2 Výsledky a jejich interpretace.....	93
4.2.1 Dysfluence a pauzy.....	93
4.2.2 Slova a takty	93
4.2.3 Slovní druhy.....	94
4.2.4 Hlásky	96
4.2.4.1 Počet hlásek ve slovech a taktech obecně.....	96
4.2.4.2 Počet rázů v materiálu.....	96
4.2.4.3 Počet hlásek ve slovech a taktech.....	97
4.2.4.4 Počet hlásek s ohledem na slovnědruhovou příslušnost slov.....	100
4.2.5 CVCV vzorce	104
4.2.5.1 Celkové počty CVCV struktur	104
4.2.5.2 CVCV vzorce podle frekvence výskytů v korpusu.....	104
4.2.5.3 Nejkratší a nejdelší struktury.....	107
4.2.5.4 Vzorce s likvidami	108
4.2.5.5 CVCV struktury a slovní druhy	110
4.2.6 Kombinace konsonantů, vokálů, likvid a rázů v různých částech jednotky	118
4.2.6.1 Četnost jednotlivých C, V, R a ? v rámci slov a taktů.....	119
4.2.6.1.1 Libovolná pozice v jednotce	119
4.2.6.1.2 Pozice na začátku jednotky.....	120
4.2.6.1.3 Pozice na konci jednotky	122
4.2.6.2 Četnost dvojic C, V, R a ? v rámci slov a taktů.....	123
4.2.6.2.1 Libovolná pozice v jednotce	123

4.2.6.2.2 Pozice na začátku jednotky.....	126
4.2.6.2.3 Pozice na konci jednotky.....	127
4.2.6.3 Četnost trojic C, V, R a ? v rámci slov a taktů.....	129
4.2.6.3.1 Libovolná pozice v jednotce.....	129
4.2.6.3.2 Pozice na začátku jednotky.....	131
4.2.6.3.3 Pozice na konci jednotky.....	133
4.2.6.4 Další poznatky.....	135
5. Diskuse.....	136
6. Závěr.....	145
Literatura.....	147
Přílohy.....	151

1. Úvod

1.1 Charakteristika a cíl práce

Tento výzkum svým zaměřením spadá pod oblast fonotaktiky, disciplíny zabývající se kombinatorikou a distribucí zvukových jednotek v řeči. Klade si za cíl prozkoumat variabilitu fonotaktických skeletů českých mluvních taktů a slov a vytvořit nástin jejich mapy, zhodnotit korespondence mezi oběma rovinami (taktovou a slovní) a přispět tak k základnímu popisu struktury mluvené češtiny. Převedením zpracovávaného materiálu souvislé řeči na konsonantické a vokální prvky získáme velké množství CVCV vzorců slov a mluvních taktů. Naším úkolem bude mj. charakterizovat odlišnosti ve vzorcích slov a vzorcích taktů a také zjistit, zda existuje souvztažnost mezi vzorci určitých slovních druhů.

Spřízněný výzkum byl uskutečněn na rozsáhlejšího korpusu psané češtiny, který zahrnoval subkorpusy beletristické, publicistické a odborné literatury (popsáno v publikaci *Statistiky češtiny* (Bartoň et al., 2009) – viz kapitola 2.3.6.2 v předkládané práci). Toto zpracování probíhalo prakticky výhradně strojově, pouze s dílčími ručními úpravami. Náš výzkum bude proveden na korpusu omezenějším, který by nicméně měl umožňovat srovnání základních tendencí plynoucích z výsledků obou přístupů. Na rozdíl od zmíněného projektu se budeme zabývat jazykem mluveným a budeme brát v potaz nejen otázku CVCV vzorců slov, ale i mluvních taktů.

Jako řečový materiál bylo použito 12 nahrávek mluvených projevů zkušených mluvčích (rozhlisových hlasatelů). Jednalo se o běžné zpravodajské relace Českého rozhlasu. Tento korpus, který tvořilo zhruba 45 minut mluvní aktivity, čítal 6639 slov, 5368 mluvních taktů a více než 37 000 hlásek (s rázem více než 38 000). Vzhledem ke školenosti mluvčích a k prestiži celostátní rozhlasové stanice jsme předpokládali, že projevy budou prosty výraznějších prozodických i hláskových deformací.

Z výše uvedeného vyplývá, že jsme za materiál pro tuto práci zvolili řečové projevy nikoli spontánní, nýbrž čtené. Toto řešení jsme upřednostnili

z několika důvodů: Prvořadým zájmem našeho výzkumu je variabilita fonotaktických vzorců českých slov a mluvnických taktů. Ve spontánním projevu bývá však lexikon značně omezen; jeho nezanedbatelnou část tvoří opakující se frekventované výrazy. V textech zpravodajských relací je oproti tomu slovní zásoba bohatší. Co se týče členění souvislé řeči na prozodické jednotky, můžeme konstatovat, že toto je sice ve spontánním projevu logicky přirozenější, ale na druhou stranu se v takových promluvách může objevit velké množství jevů, které by bránily zobecnění – zde také musíme vzít v úvahu, že náš korpus není příliš rozsáhlý. I ve čteném projevu profesionálních mluvčích se navíc objevují typické tendence frázování v češtině, a je tedy dostačující pro obecnou prozodickou charakteristiku českého mluveného projevu. Je však nutné počítat s faktem, že hlasatelský mluvní styl je dosti specifický, což s sebou nese jisté nesnáze, pokud bychom chtěli naše závěry týkající se prozodického členění absolutizovat na celou populaci rodilých mluvčích češtiny – před takovouto generalizací by bylo nezbytné výzkum rozšířit na rozsáhleji vymezený vzorek mluvčích.

1.2 Struktura práce

Teoretická část práce je rozdělena do tří tematických oddílů – v prvním nastíníme problematiku řečového rytmu, ve druhém se budeme věnovat prozodickému členění mluveného jazyka a ve třetím přejdeme k fonotaktickým tendencím v jazycích obecně a posléze speciálně v češtině. Měli jsme dvě možnosti, jak postupovat – buď směrem od nejnižších úrovní k nejvyšším, nebo zasadit téma do nejširšího kontextu a poté okruh našeho zájmu zužovat. Pro účely této práce jsme zvolili druhou možnost, a ke konkrétnímu předmětu našeho zkoumání, tedy k fonotaktice, se tím pádem dostáváme přes obecnější roviny popisu, jako jsou řečový rytmus a jednotky souvislé řeči. V kapitolách o prozodickém členění řeči již budeme pracovat s určitou koncepcí rytmu řeči a v oddílech o fonotaktice budeme operovat s představou vyšších prozodických jednotek (taktů), a proto považujeme uvedený postup pro cíle této práce za výhodnější.

V prvních kapitolách se budeme zabývat nejobecnější rovinou, která s výzkumem prezentovaným v této práci souvisí – nastíníme zde

problematiku rytmu obecně a rytmu jakožto jevu uplatňujícího se v lidské řeči. V této části budou zmíněny i poznatky týkající se činitelů řečového rytmu a jejich uplatňování se napříč jazyky. V dalším oddílu bude oblast našeho zájmu poněkud zúžena; budeme se zde věnovat jednotkám, do kterých se souvislá řeč člení a na nichž se rytmus realizuje, tedy zejména frázím a mluvním taktům. Nejprve budou uvedeny tendence, které je možné vypožorovat napříč jazyky, a poté se již zaměříme na konkrétní popis mluvené češtiny. Pozornost budeme věnovat i fenoménu taktové předrážky v českém mluveném projevu. V poslední části teoretického úvodu budou akcentovány interlingvální fonotaktické poznatky i aspekty týkající se konkrétně češtiny. V rámci tohoto oddílu se také krátce pozastavíme nad tzv. *teorií rámce a obsahu*.

Ponastínění teoretických východisek pro tento projekt bude následovat podrobný popis postupu naší analýzy, a to včetně výčtu problémů, s nimiž jsme se během zpracovávání materiálu potýkali, a její výsledky, které budou dále interpretovány. V diskusi pak nastíníme další možné cesty, jimiž by se mohl budoucí výzkum v této oblasti ubírat.

1.3 Poznámka

V rámci této práce budeme používat značky C pro konsonant, V pro vokál, případně R pro slabikotvornou likvidu nebo nazálu. Pokud se v některých citovaných studiích objeví zkratky, které by s těmito kolidovaly, mohou být pro potřeby předkládaného textu vhodně odlišeny.¹ V kapitolách o prozodickém členění dále budeme zmiňovat tzv. H (high) a L (low) tóny. S těmito označeními budeme nakládat podobně jako popis intonace, jak jej navrhla Pierrehumbertová (1980).

¹ Ve studii Rochetové a Schwartze (2005) se pracuje s kombinací CL (= koronála, labiála); pro přehlednost bylo C znamenající koronály změněno na K. Ve Statistikách češtiny (Bartoň et al., 2009) autoři operují se značkou O pro tzv. *obojetnou hlásku* (patrně slabikotvornou likvidu); tato zkratka byla v rámci kapitoly pojednávající o uvedeném publikaci ponechána, avšak v našem pojetí byly slabikotvorné likvidy a nazály uváděny pod značkou R (v úzu zakotvenější).

Je nutné podotknout, že v našich analýzách jsme data obsahující ráz oddělovali od těchž dat, do nichž jsme ráz nezahrnovali. Ráz se v češtině objevuje obvykle před slovy začínajícími na vokál a sám nemá fonologický status, nicméně v prozodické rovině je poměrně podstatnou jednotkou a jeho realizace nám v některých případech pomohla řešit sporné situace týkající se členění souvislé řeči na nižší jednotky.

2. Teoretická východiska

2.1 Rytmus

Tato práce se věnuje fonotaktické struktuře a variabilitě českého slova a také mluvního taktu. Na mluvních taktech se realizuje rytmus; ten je založen na typických konfiguracích kontrastů, které definují jazyk – vycházíme z předpokladu, že konkrétní jazyk určuje informaci o taktu a CVCV struktuře a se znalostí této informace je pak umožněna efektivnější percepce. V rámci taktů tedy dochází ke střídání kontrastů a naším úkolem je zjistit, co všechno může „uzly“ obsažené v taktu obsazovat. Proto se teď krátce pozastavíme nad problematikou rytmu. Východiskem nám bude přehledový článek J. Volína (2010).

2.1.1 Obecné poznatky

Rytmus sám je založen na opakování vzorců sestávajících z určitých konfigurací opozic – jde o střídání kontrastů, které je vnímáno jako pravidelné. Nevyskytuje se pouze u lidí a pouze u řeči: jedná se o všeobecný fenomén, který můžeme nalézt v mnoha přírodních sférách či doménách. Zakládá se na skutečnostech, že pravidelné je jednodušší než nepravidelné (a tedy rytmické procesy jsou na realizaci jednodušší než arytmičné) a že tato pravidelnost počítá s případnou součinností akcí. Rytmus je tedy dán periodickým opakováním stejných vzorců.

Ve světě dále existují dva univerzální organizační principy: tendence být součástí komplexních struktur a tendence preferovat procesy, které vyžadují co nejmenší výdej energie. Jednodušší uspořádání vykazuje větší stabilitu; snaha vytvářet složité struktury se realizuje na úkor stability, ale tato skutečnost je kompenzována možností dalšího vývoje. Tyto dva univerzální principy se mohou navzájem neutralizovat, nebo se kombinovat (Volín, 2010: 292-293).

V přírodě vykazují rytmické aspekty všechny živé organismy, a dokonce i neživé předměty, pokud jsou tyto propojeny. Projevy rytmu můžeme vidět u lidských kultur – např. hudba nebo tanec jsou na rytmickém chování

založeny – i v samotné fyziologii člověka: tlukot srdce, pravidelnost dýchání; dokonce i pulzování neuronů v lidském mozku respektuje určité rytmické vzorce. Specifický případ rytmického chování je skandování – zde je nutné spojení energie a synchronizace projevu více jednotlivců, aby bylo dosaženo větší efektivity.

2.1.2 Rytmus řeči

Rytmus je fenoménem projevujícím se i v přirozené řeči. Kontrasty, které se zde vzájemně střídají a které jsou základem vnímání rytmičnosti, jsou tvořeny zejména nukley a hranami slabik a přízvučnými a nepřízvučnými slabikami. Rytmus řeči není popsateľný pouze pomocí akustických charakteristik. Monotónní rytmický vzorec se ale v řeči nevyskytuje, protože pro formulování obsahu potřebujeme systém distinkcí schopných vyjádřit neomezený počet myšlenek; řečové chování je tedy podřízeno obsahu, nicméně i přesto má rytmus poměrně velký vliv na percepci posluchače, a to nejen v případě slovních významů a samotné srozumitelnosti projevu, ale i co se týče dojmu, jaký mluvčí na posluchače dělá.

I řečové chování vyžaduje součinnost procesů, a to zejména proto, že posluchač není pouze pasivním příjemcem signálu od mluvčího, jak předpokládaly tradiční přístupy ke komunikaci, nýbrž akce mluvčího je provázena i aktivitou posluchače. Značný posun v chápání komunikačního procesu znamenalo vytvoření modelu *adaptivní (neurální) rezonance* (Grossberg, 2003), který se zabývá způsobem zpracování informace a jehož základem je kognitivní kódování v mozku. Neuronová síť je schopna do jisté míry rozpoznat předcházející a následující vzorce. Jedná se o sérii synchronizovaných aktivací neuronů, které se objevují, když se očekávané neurální reprezentace setkají s neurálními reprezentacemi vstupního signálu – neurony se tedy aktivují, když je očekávaný objekt ve shodě se vstupním signálem; tím dochází k aktu rezonance. V případě, že je rytmus nepředvídatelný, musejí neurony svoji činnost opakovat, dokud k rezonanci nedojde. Toto dokazuje důležitost predikcí: když si predikce a analýza signálu navzájem odpovídají tak, že vykazují synchronizaci sledu excitací a útlumů, jsme schopni rozpoznat objekt (např. slabiku či slovo) rychle a bez

nadměrného úsilí. Přirozený rytmus řeči má také vliv na srozumitelnost – v experimentech bylo prokázáno, že upravený rytmus zvýšil procento chybovosti; reakční doba také byla kratší u přirozeného rytmu než u upraveného (Huggins, 1979; Buxton, 1983 in Volín, 2010). Upravenost však nemusí znamenat rušivost – v experimentech se pouze některé vokály prodloužily nebo zkrátily, ale signál byl stále percepčně přijatelný; jediné, co se změnilo, byla jeho plynulost. Rychlejší reakce při pravidelném rytmu také dokázali Hugo Quené a Robert F. Port (2005), kteří dále zjistili, že v řeči nehrají roli metrické vzorce (reakce respondentů byly stejné, ať se jednalo o trochej, nebo o jamb; rozdíly se prokázaly, pouze pokud šlo o časování praetury přízvučné slabiky). Tento výsledek není nelogický – o reálné řeči se také nedá prohlásit, že se uskutečňuje v pravidelném metru.

V jiném experimentu (Ghitza & Greenberg, 2009) byla použita třikrát zrychlená přirozená řeč s originální intonací. Věty se rozdělily do částí po 40 ms a po každém řečovém intervalu bylo do signálu vloženo ticho, a to buď vždy stejně dlouhé (20-160 ms), nebo s různými časovými intervaly. Bylo prokázáno, že pokud byla do zrychlené řeči vkládána pravidelná ticha, která netrvala déle než 120 ms, srozumitelnost výrazně vzrostla (na rozdíl od případů s různě dlouhými pasážemi s tichem). Pokud totiž akustická informace přichází v předvídatelných intervalech korespondujících s theta frekvencemi mozku (které jsou v souladu s výskytem jednotlivých slabik v přirozené řeči), může být v percepci dešifrována lépe. Řeč s prediktabilní rytmickou strukturou je jednodušší na vnímání – posluchači sice nekontrolují theta vlny, ale mozek dokáže měnit časování pozornostních pulzů podle aktuálního rytmického kontextu (Volín, 2010: 297-298).

Důležitým aspektem vnímání rytmu řeči je percepční iluze, na niž poukázala Lehistová (1977, 1979 in Volín, 2010) – temporální struktura neřečových stimulů je podle ní lidmi vnímána poměrně spolehlivě, ale stimuly řečové jsou vnímány pravidelněji, než ve skutečnosti jsou. Toto tvrzení experimentálně potvrdili Donovan a Darwin (1979 in Volín 2010), kteří navíc zjistili, že mezi globální ukazatele rytmu nepatří pouze tempo, ale i např. intonace, jejíž akustický korelát, tj. základní frekvence, ovšem taktéž nemá

přímý vztah k její percepci. Dalšími možnými koreláty mohou být dynamická obálka signálu nebo spektrální vlastnosti segmentů.

P. A. Barbosa (2005) v dalších experimentech se synchronizací řeči s metronomem zjistil, že různé slabiky a různá slova se s úderem metronomu srovnávají různě; další rozdíly pak vyvstanou při rozdílném tempu. Řeč v různých tempech tedy nemá jednotný mechanismus, ale naopak funguje při různých tempech v různých modech. Změna tempa řeči není zjednodušitelná na zkrácení nebo prodloužení všech jednotek či všech jednotek stejného charakteru (vokálů a konsonantů). V otázce rytmu řeči existuje velká variabilita dokonce i v rámci jednoho jazyka – různé mluvné styly mají různé rytmické vzorce; stejně tak i různí mluvčí. Dalším aspektem rytmického chování je, že rytmičnost je lidské řeči zřejmě velmi přirozená: pokud je člověk například požádán, aby synchronizoval svůj projev s jiným mluvčím, nečiní mu to velké obtíže. Výsledná synchronizovaná řeč přitom nemusí obsahovat zdůrazněnou prozodickou formu typickou pro sborovou řeč nebo skandování (Cummins, 2009).

Akustická analýza je v popisu rytmických jevů nedostačující, jakkoli např. jevy spojené s trváním promluvy a trváním vokálů nebo směrodatnou odchylkou trvání konsonantů mohou napomáhat pochopení tradičního rozdělení jazyků na taktově, slabičně a morově izochronní² (Ramus et al., 1999) – taková analýza nám však neposkytne informaci o pravé povaze rytmu. Ten, kdo musí rozhodnout, co je rytmické a co ne, je posluchač. Není možné rytmus popsat jednoduše nebo ho omezit pouze na důsledek jediného parametru řečového signálu.

2.1.3 Rytmičné faktory v jazycích

Je možné předpokládat, že správně organizované motorické sekvence vyžadují dostatečně přesné a předvídatelné časování. Jazyk je velmi specifický druh motorického chování, avšak máme důvod se domnívat, že má

² Mluvné rytmy je založen na opakování vzorců v čase. Rozdíl mezi jazyky taktově, slabičně a morově izochronními spočívá v tom, jestli se opakující se vzorce odvíjejí od trvání úseku od jednoho přízvuku ke druhému, nebo od počtu slabik či mor mezi dvěma přízvuky.

rytmickou organizaci srovnatelnou s tou, kterou nalezneme u ostatních motorických dovedností, jako jsou např. chůze nebo psaní (Ramus et al., 1999: 288).

Rytmičké vlastnosti jazyků nejsou arbitrární a tradičně se dělí dle mluvního rytmu na taktově, slabičně a morově izochronní (Abercrombie, 1967; Pike, 1945 in Ramus, 1999). Faktor, na kterém je percepce rytmu založena, je střídání kontrastních jednotek, tj.

„přízvučných a nepřízvučných slabik či vokálních a konsonantických intervalů. Tyto kontrasty a jejich změny jsou základem pro vnímání pravidelností či nepravidelností v proudu řeči.“ (Weingartová, 2011: 29)

Na rozdíl od Abercrombieho je podle Dashera a Bolingera (1982 in Ramus, 1999) dojem různých typů rytmu výsledkem specifických fonologických jevů v daném jazyce. Distinkce mezi taktově a slabičně izochronními jazyky potom není fonologická primitiva, ale spíše produkt jejich individuálních fonologických vlastností. Dauerová (1983 in Ramus, 1999) vysledovala, že taktově a slabičně izochronní jazyky mají řadu různých fonetických a fonologických vlastností, z nichž nejdůležitější jsou poznatky týkající se slabičné struktury a redukce vokálů. Taktově izochronní jazyky mají větší variaci slabičných typů než jazyky slabičně izochronní; inklinují tedy k těžším slabikám. Tento fakt souvisí se skutečností, že v taktově izochronních jazycích spadá přízvuk na nejtěžší slabiky (nukleus je tvořen dlouhým vokálem nebo diftongem a/nebo slabika obsahuje kodu), zatímco ve slabičně izochronních jazycích je přízvuk a váha slabiky nezávislá. Co se týče redukce vokálů, Dauerová konstatuje, že v taktově izochronních jazycích mívají nepřízvučné slabiky redukovaný vokální systém, někdy dokonce omezený pouze na střední středový vokál (schwa) nebo samohlásky dokonce absentují úplně. Tyto dvě skutečnosti vedou k závěru, že některé slabiky jsou významnější než jiné v taktově izochronních jazycích, ale v jazycích slabičně izochronních jsou všechny slabiky významné stejně, což vytváří dojem dvou různých typů rytmu (Ramus et al., 1999: 268). Dauerová nicméně konstatuje, že jazyk nemůže být přiřazen k jedné nebo druhé skupině na základě instrumentálních měření mezipřízvukových intervalů nebo trvání slabik.

Bylo tedy nutné najít efektivnější měření, která by našla akustické parametry odpovědné za percepci řečového rytmu.

Pro rytmické charakteristiky v jazycích nebylo snadné nalézt spolehlivé akustické koreláty. F. Ramus, M. Nespоровá a J. Mehler (1999) proto provedli měření založená na konsonantické a vokální segmentaci v osmi jazycích – angličtině, polštině, nizozemštině, francouzštině, španělštině, italštině, katalánštině a japonštině. Tato měření naznačovala, že

„intuitivní typy rytmu odrážejí specifické fonologické vlastnosti, které jsou signalizovány akustickými/fonetickými vlastnostmi řeči. Data podporují představu rytmických tříd a také umožňují napodobit rozpoznávání jazyků malými dětmi.“
(Ramus, Nespоровá, Mehler, 1999: 265)

Předpoklady autorů také přihlížejí k roli percepce rytmu v osvojování jazyka.

Ramus, Nespоровá a Mehler brali v potaz fakt, že z percepčního hlediska jsou nejvýraznějšími částmi promluv jádra slabik a méně nápadnými jsou konsonantické hrany. Stanovili tedy ukazatele založené na trvání a variabilitě vokálních a konsonantických intervalů a navrhli měřit tři hodnoty: procento vokálních intervalů ve větě, směrodatnou odchylku trvání vokálních intervalů ve větě, což indikuje variabilitu vokálních intervalů, a směrodatnou odchylku trvání konsonantických intervalů. Tyto ukazatele jsou podle autorů přímými fonetickými koreláty fonologických a fonotaktických vlastností jazyků; zvláště procento vokálních intervalů ve větě (tzn. kolik procent promluvy tvoří vokály) a směrodatná odchylka trvání konsonantických intervalů byly vodítky schopnými rozlišovat mezi rytmickými třídami jazyků. Směrodatnou odchylku trvání vokálních intervalů nebylo možné jednoznačně interpretovat. Tyto ukazatele tedy umožňují popsat temporální strukturu na úrovni promluvových úseků či vět, případně vyšší – hodnoty těchto tří proměnných se následně průměrují a výsledkem je jedna hodnota pro mluvčího, nebo dokonce i pro celý jazyk (Weingartová, 2011: 30-31). Ramus a kolektiv tedy ve studii z roku 1999 předkládají akusticko-fonetický model různých typů jazykových rytmů a jejich kategorizace posluchači. Autoři posléze docházejí k závěru, že klasifikace jazyků do tradičních rytmických tříd má své opodstatnění,

nicméně konstatují, že není možná jejich kategorizace pouze do několika stabilních skupin, proto je nutné výzkum provést i na dalších jazycích.

Princip, na kterém se tyto ukazatele zakládají, je presumpce, že rytmus řeči je produkt fonotaktické komplexnosti jazyka. Jazyky obvykle klasifikované jako taktově izochronní vykazují vokalické redukce a složité konsonantické shluky. Předpokládá se, že přítomnost vokalických redukcí v řečovém signálu vede k celkově nižšímu procentu vokaličnosti řeči a že výskyt komplexních konsonantických intervalů přispívá k větší variabilitě trvání konsonantických intervalů (Dellwo, Fourcin, Abbertonová, 2007: 1129).

Existují i jazyky, které jako by nepatřily ani do jedné ze tříd tradiční rytmické klasifikace. Mají např. jednoduché slabiky, ale zároveň i vokalickou redukci (katalánština); stejně tak jsou známy i jazyky s komplexními slabikami bez vokalické redukce při normálním řečovém tempu (polština). Nesporová (1990 in Ramus et al., 2000) navrhla pro tyto jazyky „střední“ rytmickou kategorii; další pojetí pak předkládá pět rytmických tříd jazyků podle příznakovosti využívaných typů slabiky, z nichž tři korespondují s klasickým pojetím (Levelt & van de Vijver, 1998). Dauerová spíše pracuje s předpokladem, že jazyky mohou být rozptýleny do kontinua, jehož konce tvoří na jedné straně úplná slabičná a na druhé úplná taktová izochronie.

Ve studii s percepčními experimenty z roku 2000 Ramus a kolektiv zkoumají rozlišení jazyků za použití resyntetizovaných stimulů. Tato technika byla použita proto, aby bylo jisté, že jediné, co je pro posluchače dostupné, jsou skutečně jen rytmická vodítka. Zkoumanými jazyky byly angličtina, španělština, katalánština a polština, tedy jeden taktově izochronní jazyk, jeden slabičně izochronní a dva jazyky, jejichž kategorizace byla nejasná. Z výsledků vyplynulo, že katalánština se podle rytmu blíží španělštině, a inklinuje tedy více ke slabičné izochronii, zatímco polština je odlišena od angličtiny i od španělštiny. Autoři proto navrhnou rytmus polštiny dále zkoumat a prozatím tento jazyk zařazují do nové, čtvrté skupiny. Počítají spíše se zvýšením počtu tříd, a ne s představou kontinua, jak navrhovala Dauerová, nicméně uznávají, že se otázka diskrétních tříd versus škály může rozhodnout, až když bude prozkoumáno více jazyků (Ramus et al., 2000: 10).

Rozpoznávání jazyků podle rytmu se Ramus snažil doložit i u novorozenců a malých dětí. Za použití stimulů z normální, filtrované i resyntetizované řeči se podařilo prokázat, že novorozenci do jisté míry dokážou rozlišit jazyky různých rytmických skupin. Zdá se, že děti starší dvou měsíců už rozlišují pouze svůj rodný a cizí jazyk, nicméně poté patrně k rozlišování jazyků používají i jiná vodítka než rytmus, např. intonaci a fonologickou nebo fonotaktickou informaci.

Ve výzkumu z roku 2002 týž autor představuje čtyři experimenty s rozlišováním nizozemštiny a japonštiny u novorozenců. Tyto pokusy byly provedeny na přirozené řeči, resyntetizovaných stimulech a dále dvou stimulech s umělou intonací (F0 nahrazena konstantním průběhem), z nichž jeden měl přirozenou fonotaktickou strukturu, kdežto stavba druhého byla omezena na sykavku a vokál [a], aby v podnětu zůstala pouze rytmická vodítka. Děti byly schopny rozpoznat jazyk, pokud šlo o přirozenou řeč nebo resyntetizovaný stimul s původní intonací. Pokud byla intonace nahrazena konstantním průběhem, nebyla diskriminace tak úspěšná a v případě změny přirozené fonotaktické struktury na tříslabičný stimul [sasasa] již jazyky rozpoznány nebyly. Vystala tedy otázka, zda je k percepci rytmu určitá fonotaktická informace nutná a nakolik jsou schopny děti rozpoznat jazyky např. na základě četnosti a pozic souhláskových shluků. Další testy nicméně ukázaly, že novorozenci sice rozliší slova s různým počtem slabik, ale nikoli slova s různou slabičnou strukturou a stejným počtem slabik (např. VCCV nerozliší od CVCV; CVCCCV od CCVCCV apod.); nejsou tedy schopni odlišit ani bimoraická a trimoraická slova (např. CVCV od CVCCV). Zdá se, že s fonotaktickými vzorci patrně pracují až starší děti (kolem 6.-9. měsíce věku), kdežto prozodické vzorce reflektují již novorozenci. Pravděpodobnější interpretace neúspěchu tohoto pokusu je, že stimuly [sasasa] nebyly dětmi pro svou plochost a nízkou fonetickou diverzitu akceptovány jako řečové.

Fakt, že nebyly dětmi rozpoznány čistě rytmické stimuly a že v experimentu, kde fonotaktická informace byla obsažena, ale intonace nahrazena konstantním průběhem, byl efekt rozpoznání poměrně slabý, nás může vést k závěru, že samotný rytmus nelze redukovat na načasování vokálů

a konsonantů – je možné, že pokud se rytmus separuje od ostatních suprasegmentálních podnětů s podobnou funkcí, ztratí něco ze své podstaty. V reálné řeči jsou rytmus a intonace silně propojeny – např. přízvuk bývá signalizován pomocí trvání i výšky slabiky a její intenzitou, podobně fonologický frázový předěl je značen prodloužením slabiky i pohybem výšky. Je tedy možné, že zdánlivě redundantní rytmická vodítka pomáhají samotnému zpracování rytmu a že rytmické pravidelnosti mohou být klíčem k úspěšnému zpracování řečového signálu pouze v integraci s ostatními současnými vodítky (Ramus, 2002: 110-111).

Na Ramusovy ukazatele rytmických tříd – směrodatnou odchylku trvání konsonantických intervalů a procento vokálních intervalů ve větě – navázali Dellwo, Fourcin a Abbertonová (2007), kteří navrhli tyto parametry přepočítat na směrodatnou odchylku trvání neznělých intervalů a procento intervalů znělých. Aby se předešlo možnému zkreslení dat kvůli rozdílným tempům, byl parametr směrodatné odchylky neznělých intervalů normalizován na variační koeficient. Autoři tak získali poměrně silnou korespondenci mezi těmito proměnnými a klasickými rytmickými třídami jazyků (nicméně v případě této studie šlo o výzkum provedený pouze na francouzštině, italštině, angličtině a němčině, a je tedy nezbytně nutné tuto metodu aplikovat na větší počet jazyků). Jejich přístup bere v potaz fakt, že děti – které jsou schopny již v útlém věku rozlišovat mezi třídami – v prenatálním období přijímají zvuky procházející nízkopropustným filtrem a nemají žádné vizuální podněty, a proto by čistě hlasová vodítka měla být nejnápadnějšími (a tedy i výraznějšími než konsonantičnost a vokaličnost řeči).

2.2 Problematika prozodického členění se zvláštním zřetelem k jednotce mluvního taktu

Jednou z důležitých součástí předkládané práce byla segmentace mluvených projevů na reálné mluvní takty. Proto je vhodné se nad chápáním taktu v jazycích obecně a specificky v češtině podrobněji pozastavit.

Pro konkrétnější uchopení mluvních taktů je třeba vymezení této jednotky zasadit do širšího rámce, kterým je prozodické členění a prozodie obecně. Nejprve se budeme zabývat uplatněním prozodických parametrů na členicí funkci a posléze se zaměříme na jednu konkrétní jednotku prozodického segmentování řečových projevů, totiž na mluvní takt a na aspekty s jeho definováním a určením spojenými. Zejména poslední část kapitoly pak budeme vztahovat na mluvenou češtinu. V tomto oddílu budeme čerpat z textů Daniela Hirsta a Alberta Di Crista, Jany Ondráčkové, Zdeny Palkové, Bohuslava Hály, Milana Romportla, Františka Daneše a dalších autorů.

2.2.1 Prozodie a intonace

Hirst a Di Cristo, kteří se ve své práci *A survey of intonation systems* (1998) snaží podat komplexní výklad prozodických jevů pro mezijazyková srovnání, rozlišují pojem intonace v užším a v širším slova smyslu. Intonace v širším smyslu počítá i s lexikálními charakteristikami jako slovním přízvukem, tónem nebo kvantitou. Oproti tomu užší vymezení tyto jevy neobsahuje. Autoři navrhují pro intonaci v širším slova smyslu používat pojem *prozodie* a pro užší vymezení, které zahrnuje pouze supralexikální charakteristiky (např. forma melodických kadencí, deklinace, hraniční jevy apod.), termín *vlastní intonace*. Na fyzikální rovině pojem intonace obvykle znamená jeden nebo více akustických parametrů. Mnoho autorů klade důraz na fakt, že intonace se skládá z řady parametrů, které kromě základní frekvence zahrnují např. změny intenzity a trvání segmentů (Rossi et al., 1981; Beckmanová, 1986 in Hirst & Di Cristo, 1998: 4). Někteří pod termínem intonace rozumějí i aspekty temporální organizace a rytmu, který kromě intenzity a trvání může hrát roli v proměnách spektrálních charakteristik, jako je např. rozdíl mezi plnými a redukovanými vokály (Crystal, 1969 in Hirst & Di Cristo, 1998: 4). Členění souvislé řeči na rytmické jednotky spadá do kategorie prozodické; vymezení promluvového úseku a mluvního taktu pak souvisí jak s fyzikálními, tak s formálními aspekty.

2.2.2 Prozodické členění

V této podkapitole se nejprve budeme věnovat prozodickému členění v jazycích obecně a posléze uvedeme několik faktů o prozodické organizaci výpovědí v češtině.

Fonetické události, které konstituují intonační kadenci, jsou poměrně nezávislé na segmentálním fonetickém materiálu. Jedním z prvních formálních řešení bylo uvažovat v intencích reprezentace písně – čili brát v potaz dvě více či méně nezávislé linie, intonační a segmentální, kde konstituenty intonace byly nazývány např. *prozodémy* (Bloch & Trager), *suprasegmentálie* (Hockett) nebo *dlouhé komponenty* (Harris).

Jednou z možností, jak synchronizovat segmentální a suprasegmentální elementy reprezentace, bylo předpokládat, že intonace má hierarchickou strukturu a že to je tato struktura, která je spojena se segmentálním materiálem. Toto pojetí je typické pro britskou analytickou školu (Palmer, Cruttenden, Couper-Kuhlen), ale prvním, kdo koncepci navrhl, byl Klinghardt ve své práci o intonaci v angličtině (Klinghardt and Klemm, 1920 in Hirst & Di Cristo, 1998: 34), kde tzv. *řečový takt* (v terminologii současného popisu češtiny *promluvový úsek*³) rozložil do těchto komponentů:

Auftakt	Taktkopf	Neuhebung	Tiefton	Abtakt
(předtaktí)	(hlava úseku)	(nový vzestup)	(nízký tón)	(posttakt)

Tato analýza byla s menší modifikací přejata např. Palmerem, Kingdonem a dalšími:

Pre-head	Head	Nucleus	Tail
-----------------	-------------	----------------	-------------

Systematicky hierarchická analýza intonačních kadenčí byla představena Jassemem. Tento model předkládal dvouúrovňovou strukturu, kterou může být kompletní promluva rozčleněna do diskrétních sekvencí *tónových skupin* (*tonal groups*); každá tónová skupina se může posléze segmentovat na

³ Skupina taktů, kterou mluvčí jazyka považuje za jeden percepční celek.

diskrétní sekvence *tónových jednotek* (*tonal units*). Tonální jednotka je vlastně totéž co přízvukná stopa Abercrombieho, kterou posléze převzal pro svůj popis intonace Halliday (1967). Abercrombie a Halliday používali stejnou jednotku pro modelování jak rytmu, tak intonace, ale Jassem popisuje struktury odlišné: *rytmickou* a *tonální jednotku* (Jassem, 1952 in Hirst & Di Cristo, 1998: 34-35). Tento dvouúrovňový popis je používán u mnoha autorů pouze s jinými termíny.

Otázkou je, zda je pro adekvátní popis intonace třeba více než dvě úrovně. Liberman a Prince navrhuje teorii binárního větvení fonologických struktur, které by obsahovaly potenciálně nekonečný počet úrovní (Libermann & Prince, 1977 in Hirst & Di Cristo, 1998: 35). Martinův model intonace povoluje nekonečnou rekurzivitě (Martin, 1977, 1981 in Hirst & Di Cristo, 1998: 35).

Systém navržený Selkirkovou a rozvíjený Nesporovou a Vogelovou počítal s tím, že fonologická reprezentace sestává z vložené struktury fonologických kategorií seřazených od promluvy k segmentu, které zahrnovaly intonační frázi, fonologickou frázi, skupinu s příklonkami, fonologické slovo, stopu a slabiku; každá kategorie se skládala z diskrétní sekvence kategorií další nižší úrovně (Selkirková, 1978; Nesporová & Vogelová, 1986 in Hirst & Di Cristo, 1998: 35).

Někteří autoři považují za nezbytné zavést minimálně jednu prostřední úroveň mezi intonační jednotkou a přízvuknou stopou. Někdy se uvažuje i o jednotkách nad intonační jednotkou; ty se nazývají např. *paratóny* (Fox) nebo *phono-paragraphs* (Misheva, Nikov). Gårdingová a další autoři problém intonační struktury zmiňují v kontinuu diskurzu a dialogu (Gårdingová, 1998 in Hirst & Di Cristo, 1998: 35). V následující tabulce (převzaté z Hirst & Di Cristo, 1998) uvádíme přehled prozodických jednotek podle jednotlivých autorů. Položky v jednom sloupci si nutně nemusejí odpovídat.

Úroveň 1	Úroveň 2	Úroveň 3	autoři
prozodická věta	prozodická fráze	prozodické slovo	Gårdingová, Botinis
prozodická věta	prozodická fráze	přízvučná skupina	Grønnumová
intonační jednotka	prozodické slovo	tonální jednotka	Di Cristo
fráze	syntagma	fonetické slovo	Misheva & Nikov
fráze	řečový takt	stopa	Iivonen

Tabulka 1: Třístupňová hierarchická struktura prozodických jednotek podle různých autorů (Hirst & Di Cristo, 1998: 36).

Co se týče akustických znaků hranic intonačních jednotek, můžeme konstatovat, že předěly bývají v různých jazycích často realizovány jako koncová změna výšky nebo jako přítomnost větného přízvuku či nukleu. Dalšími vodítky mohou být pauzy (ticha i např. hezitace), koncové dlužení, rytmická soudržnost, změna tempa nebo resetování deklinačního spádu.

Už od doby nejstarších studií o intonaci se vedly debaty o syntaktické, sémantické nebo pragmatické determinaci intonačních jednotek. Přestože pragmatické důvody pro rozdělení promluvy nakonec převládly, hraje tu roli i určitý počet syntaktických omezení, která by měla být respektována. Mnozí autoři také předpokládají, že intonační jednotka konstituuje jednotky informační či významové. Také se shodují v tom, že ačkoli neexistuje zcela jasná korelace mezi syntaxí a prozodií, můžeme např. vysledovat tendenci, aby intonační jednotky korespondovaly s větami v souvětí.

Různé jazyky patrně nepoužívají různý počet úrovní intonační struktury a nevykazují ani mnoho rozdílů ve způsobu, jakým pragmatická a syntaktická omezení určují dělení promluvy do intonačních skupin. Při vytváření obecného modelu intonace by ovšem měla být brána v potaz data pokud možno ze všech jazyků.

V češtině se prozodická struktura rovin nižších než věta obvykle člení do dvou úrovní; tradiční popisy zmiňují jako vyšší jednotku *promluvový úsek* a jako nižší jednotku *mluvní takt*. Otázkou zůstává, zda je pro češtinu toto jednoduché rozdělení dostačující a jestli by mezi těmito dvěma jednotkami souvislé řeči neměla existovat ještě jedna prostřední úroveň (např.

promluvvé úseky nižšího řádu), která by zahrnovala fráze s méně výraznými předěly.

Promluvvý úsek bývá v češtině vymezen percepčně, tedy jako sekvence mluvních taktů, které jsou vázány intonačními prostředky tak, že je recipient pociťuje jako jeden zvukový celek. Nejkratším promluvvým úsekem může být jediný takt; nejdelším pak ukončená výpověď (Palková, 2006: 227). Vnímání úseku jakožto celku může být způsobeno intonačním průběhem či pauzou (ve smyslu ticha nebo hezitace či podobných zvuků). Mluvčí rovněž inklinují ke tvoření těchto celků – vliv na prozodickou organizaci má rytmický faktor a také výše uvedená tendence vázat k sobě významové komplexy.

Členění na úseky významně ovlivňuje srozumitelnost sdělení pro posluchače. Promluvvý úsek má poměrně těsný vztah k syntaktické stavbě věty; mezi tyto roviny však nelze položit rovnítko. Vztah mezi textem a frázováním je oboustranný: Účinek frázování na posluchače může být pozitivní i negativní v závislosti na tom, zda je prozodické členění v souladu se stavbou textu; text sám však také může podporovat členění na intonační fráze, nebo mu svou stavbou bránit (Palková, 2006: 227). Nerovnováha v syntaktické hloubce předělů mezi takty následujícími za sebou členění na úseky podporuje; rovnost syntaktických předělů členění naopak oslabuje (Palková, 2006: 238).

V češtině se dále vyděluje i prozodická jednotka nad promluvvým úsekem – věta. Věta jako popisná jednotka je uzavřený syntaktický celek s charakteristickým zvukovým průběhem (tj. realizací některého melodému). Ve větě může být více promluvvých úseků a hranice promluvvého úseku se vždy shoduje s hranicí věty, nikoli však naopak. Věta i promluvvý úsek mohou obsahovat jednu nebo více vyšších prominencí, než je prominence přízvukné slabiky v taktu; v české terminologii se pak obvykle hovoří o *větném* a *kontextovém* přízvuku.

2.2.3 Jednotka mluvního taktu

Pojem mluvního taktu, přízvukové skupiny nebo stopy je úzce spjat s fenoménem přízvučnosti.⁴ Nyní tedy tuto problematiku s ohledem na různé jazyky přiblížíme.

V klasických popisech bývají rozlišovány jazyky s *pevným* a *volným* přízvukem (pevný přízvuk: přízvučná slabika má konstantní vztah k některé ze stavebních vlastností slova, obvykle k mezislovní hranici, nemůže tedy rozlišit znakové jednotky, nicméně může mít delimitativní funkci; volný přízvuk: to, kde se nalézá přízvuk, patří k vlastnostem konkrétního slova; volný přízvuk může rozlišovat lexikální význam (Palková, 1994: 157-158)). Neexistuje ale žádná akustická hodnota, která by mohla korespondovat se skutečností, že v některých jazycích je přízvuk lexikálně distinktivní a v jiných ne; stejně tak není akustická distinkce mezi tónovými a přízvukovými jazyky. Ačkoli je využito stejného zvukového prostředku, může mít tento v každém typu jazyků jinou funkci.

Nedávné práce z oblasti nelineární fonologie předpokládaly, že tón je formálně reprezentován v lexikonu tónového jazyka jako sekvence tónových segmentů (např. H a L) spolu s pravidly specifickými pro každý jazyk, která upřesňují, jak jsou tóny spojovány se segmenty nebo se slabikami slova. Slovní přízvuk je ale reprezentován různými způsoby – jako distinktivní rys segmentů, jako abstraktní symbol spojený s jednou slabikou lexikální jednotky nebo jako hierarchická prozodická struktura, ve které je sekvence slabik seskupována do vyšší jednotky, přičemž jedna z oněch slabik náleží do celku vyššího řádu; tato slabika je pak určena jako nejsilnější prvek neboli hlava této sekvence (s touto koncepcí se pracuje v rámci metrické fonologie).

Jazyky, které jsou brány jako tónové, jsou např. vietnamština (šest tónů), thajština (pět tónů) nebo čínština (čtyři tóny). Kromě tónové charakteristiky mají všechny tři tónové jazyky i rozdíl přízvučných a nepřízvučných slabik, který je ale lexikálně distinktivní pouze v čínštině (Hirst & Di Cristo, 1998: 9).

⁴ V této kapitole budeme zohledňovat pouze slovní/taktový přízvuk. Prominence na vyšších prozodických úrovních (např. „důraz“ nebo „větný přízvuk“) nejsou předmětem našeho zkoumání.

Jazyky, které se nacházejí „mezi“ tónovými a přízvukovými, jsou např. japonština a švédština. Zde tedy k rozlišení na jazyky s přízvukem (angličtina, ruština ad.) a jazyky tónové (čínština, thajština ad.) přistupují ještě jazyky s tzv. *melodickým přízvukem*. Mezi tónovými a přízvukovými jazyky tím pádem není ostrá hranice, nýbrž plynulý přechod. Dále existují rozdíly i v rámci přízvučných systémů – v některých jazycích je důležitá pouze pozice přízvuku, někde hraje roli také jeho přítomnost nebo nepřítomnost. Rozdíl mezi jazyky s melodickým přízvukem a jazyky s přízvukem dynamickým lze vysvětlit takto: V jazycích s dynamickým přízvukem je jedna slabika lexikálně označena za *hlavu* toho kterého slova, zatímco v jazycích s melodickým přízvukem je tou relevantní lexikální charakteristikou přítomnost nebo nepřítomnost (a pokud přítomnost, tak i pozice) lexikálně specifikovaného tónu. V jazycích s dynamickým přízvukem je tedy maximální počet potenciálních kontrastů roven počtu slabik (v praxi je ale často pozice hlavy omezena na nižší počet možných pozic), ale v jazycích s melodickým přízvukem je počet potenciálních kontrastů vyšší než počet slabik, protože se zde mohou objevit i nepřízvučná slova.

Zdena Palková (1994) k problematice melodického přízvuku dodává, že konstantní průběh melodie se v jazycích s melodickým přízvukem uplatňuje také jako prostředek pro rozlišení lexikálního významu. Toto se ale zpravidla týká jen části slovní zásoby.

„Melodický přízvuk vytváří jednodušší, obvykle dvoučlenný kontrast a neváže celé slovo, nýbrž je podmíněn slabičnou strukturou; bývá např. příznakem přízvučné slabiky.“ (Palková, 1994: 159-160)

Distinkce mezi přízvučnými a nepřízvučnými slabikami nevyžaduje odkazování na struktury vyšších úrovní. Přízvuk nicméně plní určitou seskupovací funkci (Bruce, 1985 in Hirst & Di Cristo, 1998: 22) – každá přízvučná slabika se kombinuje různými způsoby se sousedními slabikami nepřízvučnými, aby vytvořila *přízvukovou stopu* (*stress-foot*) nebo *přízvukovou skupinu* (*stress group*). Toto pojetí má poměrně uspokojivý vztah k rytmickým charakteristikám a také se jím může zjednodušit popis intonačních vzorců.

Pokud předpokládáme takovouto strukturu (tj. strukturu skupiny slabik tíhnoucích ke slabikám přízvučným), opakující se melodické pohyby mohou být popsány jako jev, který se objevuje nikoli na přízvučné slabice, nýbrž na jednotce vyššího řádu. V britské angličtině a v dánštině je tato vyšší jednotka explicitně definovaná jako sestávající z jedné přízvučné slabiky a jedné nebo více následujících nepřízvučných slabik: (Sss...)(Sss...)(Sss...), kde „S“ reprezentuje přízvučnou slabiku a „s“ nepřízvučnou. Zde tedy můžeme hovořit o tzv. *sestupné* (*left-headed* – pojem z metrické fonologie) přízvukové skupině (Hirst & Di Cristo, 1998: 22).⁵ Toto je rytmická jednotka, kterou Abercrombie nazýval *stopou* (termín si vypůjčil z tradiční poetiky), čili jednotkou od jednoho přízvuku k druhému bez omezení lexikálními hranicemi. Jelikož v češtině je přízvuk tradičně na první slabice slova/taktu, může se nabízet analogie s Abercrombieho stopou, ačkoli tato koncepce nepočítá s jevy, jako jsou např. předrážky. Nicméně mluvní takt se od stopy liší v tom smyslu, že takt může brát ohled na hranice slov: i když je v některém jazyce přízvuk např. na předposlední slabice, přesto touto slabikou nemusí začínat nebo končit takt; stopa však ano.

Sofistikovanější model prozodické struktury angličtiny byl navržen již dříve Jassemem, který rozlišoval rytmickou a tonální strukturu (viz výše). Tzv. *přízvuková skupina*, jak byla popsána v koncepci Hirsta a di Crista, koresponduje více s tonální strukturou Jasseмова popisu. Podobnou distinkci také vytvořila Grønnumová, která doporučovala, aby prozodické přízvukové skupiny byly analyzovány spíše jako začátek s přízvučnou samohláskou než s přízvučnou slabikou (Grønnumová, 1998 in Hirst & Di Cristo, 1998: 22-23).

Ve francouzštině a italštině je přízvuková skupina popsána jako sestávající z jedné přízvučné slabiky a jedné nebo několika předcházejících slabik:

⁵ S podobným rozlišením pracuje i Ondráčková; pro termíny *left-headed* a *right-headed* jsem zvolila české protějšky *sestupný* a *vzestupný* proto, aby byla tato ekvivalence patrná. Hála pro tutéž kategorizaci používá termíny takt *decrendový* nebo *klesavý*; *crescendový* nebo *stoupavý* (Hála, 1975: 263). Ondráčková i Hála počítají ještě s takty *obstupnými* (*crescendo-decrendovými*), viz dále. Hála kvůli jazykům s volným přízvukem zavádí ještě čtvrtý typ, takt *nepřavidelný*.

(ssS)(ssS)(ssS)... Zde tedy můžeme hovořit o tzv. *vzestupné (right-headed)* přízvukové skupině (Hirst & Di Cristo, 1998: 23). Podle této distinkce můžeme tedy jazyky rozdělit na jazyky s přízvukovými skupinami (příp. takty) *sestupnými* a jazyky s takty *vzestupnými*.

Zajímavý je případ portugalského, kde se explicitně nepoužívá koncept přízvukové skupiny. Něco podobného se ale implicitně např. v analýze Cruz-Ferreirové objevuje (Cruz-Ferreirová, 1998 in Hirst & Di Cristo, 1998: 23). Jedná se o rozdělení intonačního vzorce na tzv. *pre-head, head a nuclear tone*: (sss)(SssSss)(Sss) Tato analýza předpokládá, že nepřízvučné slabiky vždy patří do stejné skupiny jako předcházející přízvuková slabika. Poslední experimentální výzkumy na brazilské portugalské ukazují, že rytmická organizace tohoto jazyka inklinuje k vzestupným taktům – na rozdíl od portugalského evropského, která podle de Carvalha tíhne k taktům sestupným (de Carvalho, 1989 in Hirst & Di Cristo, 1998: 23).

Někteří autoři nezmiňují žádnou tendenci k seskupování nepřízvučných slabik nebo předpokládají, že každé takové seskupování bude respektovat lexikální a/nebo syntaktické hranice. Existuje jistá možnost, že vzestupnost nebo sestupnost taktů je parametr specifický pro ten který jazyk. Kdybychom analyzovali prozodickou klasifikaci indoevropských jazyků s tímto předpokladem, našli bychom určitou korespondenci s jazykovými rodinami – germánské jazyky jako celek spíše inklinují k sestupným stopám; románské jazyky spíše k vzestupným. Germánské jazyky jako nizozemština, němčina, dánština nebo švédština pak také mají přízvuk realizován jako sekvenci LH, stejně jako románské jazyky francouzština, italština a pravděpodobně portugalská. Přízvuk jako sekvenci HL bychom našli v americké angličtině a ve španělštině a *downstep* coby realizaci přízvuku na přízvukové skupině bychom našli např. v britské angličtině a rumunštině. Melodický pohyb, který je zde popsán, se objevuje na přízvukové skupině jako celku, nejen na přízvukové slabice.

2.2.4 Mluvní takt v češtině

2.2.4.1 Definice taktu

V tomto oddíle se zaměříme konkrétně na vymezení taktu v mluvené češtině. V souvislosti s touto jednotkou se budeme také věnovat problematice přízvučnosti v češtině, která je pro určení taktu důležitá, avšak nesnadno uchopitelná akustickými měřeními.

J. Ondráčková ve studii *O mluvním rytmu v češtině* (1954) vychází z předpokladu, že v percepci posluchačů je jednotka taktu vnímatelná. Takt je vymezován jako skupina slabik podřízená právě jednomu přízvuku. Mluvní takt je prozodickou jednotkou v rámci intonační fráze nebo promluvového úseku, nejmenší jednotkou, na které se realizuje rytmus řeči a která je vnitřně strukturovaná.⁶ Autorka uvádí, že takty jsou základními rytmickými jednotkami pro členění souvislé mluvy (Ondráčková, 1954: 24).

Podobně jako Hirst a Di Cristo rozeznává Ondráčková takty podle umístění přízvučné slabiky: pokud je prominence na začátku, jde o takt *sestupný* (čeština), pokud na konci, mluvíme o taktu *vzestupném* (francouzština). Když se prominence realizuje na slabice uprostřed taktu, jedná se o takt *obstupný* (polština). V některých jazycích se vyskytují všechny druhy taktů (ruština, angličtina, němčina). Toto rozdělení také souvisí s rozlišením přízvuku na *pevný* a *volný* (viz výše).

Ondráčková uvádí přehled používání různých termínů pro tuto jednotku souvislé řeči v české i zahraniční odborné literatuře. O *taktu* v české literatuře hovoří např. Chlumský, Hála, Petřík, Šmilauer a dříve Trávníček; jako takt *mluvní* (Sprechtakt) tuto jednotku označují např. Sievers, po něm Beyer ad.; z českých odborníků např. Palková. Jakobson nazýval takt *větným*, Mutchmann zdůrazňoval jeho významovou stránku (Sinntakt). Takt je však v první řadě jednotkou rytmickou, nikoli významovou. Mnoho dalších autorů zdůrazňuje v taktu přízvuček, někteří (např. F. Trávníček) ho proto nazývají

⁶ Na nižší úrovni prozodického členění výpovědi nalezneme slabiku, která je přirozenou jednotkou řeči; na vyšší rovině pak promluvový úsek a promluvu.

přízvukovou jednotkou (*stress-group* - Sweet, *Akzentgruppe* - Beyer). Takt má také další pojmenování, která souvisejí s rytmem. V naší práci dáváme spolu s Ondráčkovou a Palkovou přednost označení tohoto rytmického útvaru jako takt *mluvní*. Proti používání termínu *přízvukový takt* se vymezují Palková, Veroňková, Volín a Skarnitzl (2004), protože tento pojem

„podporuje představu, že takt je přízvukovou slabikou vskutku omezen, vymezen, signalizován atd.“ (Palková et al., 2004: 66)

V poslední době se ukazuje, že jednotka taktu nemusí být opřena o doložitelné zvukové kvality přízvukné slabiky slova. To, že takt je v percepci vnímán jako ucelená jednotka, je způsobeno průběhem zvukových vlastností v sekvenci slabik, díky němuž vzniká dojem samostatnosti slova. Proto autoři doporučují užívat přesnější termín *mluvní takt* (Palková et al., 2004).

Takt jako rytmickou jednotku můžeme chápat dvěma způsoby: Pokud pod tímto pojmem rozumíme úsek od jednoho přízvuku ke druhému, aniž bychom brali v potaz hranice slov (Sievers, Sweet), což je pojetí, které je úzce spjato s Abercrombieho *stopou*, všechny takty se stanou sestupnými nebo vzestupnými podle toho, jestli hranici položíme za přízvuk nebo před něj. Toto řešení je ale diskutabilní, neboť by se posléze všechny jazyky uvedly na společný rytmický základ. Jiné členění bere v úvahu i významovou stránku jednotky, aby nebyla porušena celistvost slova.

Jak už jsme zmínili výše, takt bývá v odborných textech definován jako skupina slabik, která obsahuje právě jednu slabiku přízvuknou a jednu (popř. žádnou) nebo několik slabik nepřízvukných. Např. Sievers nebo Beyer ztotožňují takt se *skupinou výdechovou*. Podle Ondráčkové je ale takt obvykle pouze součástí výdechové skupiny (v dnešní terminologii jde patrně o *promluvový úsek*). Passy rozlišuje skupiny *výdechové* a *silové*; výdechové jsou nadřazeny silovým. Shoda panuje v tom, že hlavní úlohu při spontánním členění souvislé řeči má rytmus, tj. střídání přízvukných a nepřízvukných slabik (Ondráčková, 1954: 27).

Bruneau definoval takt jako *fonetické slovo*, což je slovní skupina spojená jedním přízvukem. Toto pojetí aplikoval na francouzštinu. B. Hála určil takt

podobně: jako rytmický celek, který se skládá z jednoho nebo několika slov spojených jedním přízvukem (Hála & Sovák, 1947 in Ondráčková, 1954: 27). Z této definice tedy mimo jiné vyplývá, že hranice taktu je vždy zároveň hranicí slova, ale hranice slova nemusí nutně být i hranicí taktu. V češtině toto tvrzení platí: hranice mezi takty je vždy zároveň i hranicí mezi slovy, ale nikoli naopak – takt může obsahovat i více slov, zvláště pokud se jedná o slova jednoslabičná. Mluvčí dále inklinují k tomu, aby nebyly rozdělovány významové jednotky. Hranici taktu signalizuje přízvukový vrchol, což je však hodnota spíše percepční (a slovo vrchol je tedy metaforické) – ověří se lépe než z akustického výzkumu z poslechu, a proto tato jednotka není dobře objektivizovatelná (Churaňová, 2010: 10).

Ondráčková ve své studii podává historický přehled pojetí mluvního taktu u českých jazykovědců. Tento zde tedy ve stručnosti uvádím a rozšiřuji ho o další autory.

Josef Dobrovský takty v češtině v potaz nebral, nicméně kladl důraz na přízvuk na první slabice slova, dokonce explicitně zmiňoval, že neexistuje v češtině slovo bez přízvukné slabiky, s přízvukem na předposlední slabice nebo na konci slova.

J. Durdík uvádí, že slova se seskupují ve *významové skupiny* a tvoří jednotu, a doporučuje, aby mluvčí mezi takto spjatými slovy nedělali pauzy. Těmito významovými skupinami však ve skutečnosti myslí syntagmata, jak vyplývá z příkladů (*velmi krásný* apod.); taková by měla být ale v pojetí Ondráčkové i našem samostatnými takty.

A. Frinta členil promluvu na *výdechové skupiny* (v našem pojetí promluvové úseky) ohraničené pauzami a dále i na jednotky podobné taktům; rozlišuje také přízvuk větný a ostatní (hlavní a vedlejší) přízvuky.

O. Zich definoval takt jako skupinu s jedním prvkem s větší vahou než ostatní. Tímto prvkem je pak spojena v jednotu celá skupina. Takto se vytvoří těžká část skupiny, již se podřídí část lehká. Taková skupina obsahuje jedno nebo více celých slov; Zich je nazývá *slovními celky*. Pojmy *váha* nebo *těžká část*

ovšem nejsou nijak zakotveny ve fonetické terminologii; můžeme se však domnívat, že autor měl patrně na mysli percepčně relevantní prominenci.

J. Chlumský uvádí, že takt je skupina slabik ovládaná jedním přízvukem; Jílek zastává názor, že takt je sestupnou jednotkou od jednoho přízvučného vrcholu ke druhému. Mathesiův takt je v jeho terminologii promluvový úsek:

„Je to v podstatě totéž, co Zichův slovní celek, a může se proto definovat jako nejmenší díl promluvového celku, který lze bez újmy smyslového členění oddělit od ostatní promluvy pausami.“ (Mathesius, 1937)

Krátký promluvový celek podle Mathesia nemusí být členitelný na menší jednotky, a může se tedy stát samostatným. U Petříka jde v případě taktu o celek významový.

František Trávníček nejprve definoval takt vágněji, tj. jako spojení slov tvořících jednotu po stránce přízvukové a zpravidla i významové; v pozdějších pracích je popisuje jako *přízvukové jednotky*, tedy spojení dvou nebo více slov majících jeden přízvukový vrchol (hlavní přízvuk), k němuž se seskupují slova, popř. slabiky bez přízvuku nebo s vedlejším přízvukem, a to předklonná a příklonná.

Bohuslav Hála ve svém *Úvodu do fonetiky* (1948) definuje takt následujícím způsobem:

„V souvislé mluvě splývají často slova dohromady a tvoří tak rytmickou jednotku, zvanou takt, který se skládá z jedné slabiky přízvučné a jedné nebo několika slabik nepřízvučných. Poměry jsou po této stránce v jazycích velmi nestejně, v češtině jsou nejoblíbenější takty o dvou až třech slabikách.“ (Hála, 1948: 142)

K této definici se přiklání i Ondráčková a i její následný výzkum ukazuje podloženost tvrzení o nejfrekventovanějších taktech v češtině. V Hálově textu dále najdeme poznámku, že takty s vyšším počtem slabik jsou v jazycích vzácné a že takt může mít i jenom jednu slabiku, což je jev, ke kterému ve spontánní řeči dochází poměrně často.

Milan Romportl používá termín *takt přízvukový*; považuje za něj „spojení několika slabik s jedním přízvukovým vrcholem“ (Romportl, 1985: 111)

a poznamenává, že v češtině taktu odpovídá nejčastěji jedno slovo, ale nikoli bezpodmínečně. Stejně jako předchozí autoři rozlišuje takty vzestupné, sestupné a obstupné a reflektuje i případ *taktové předrážky*, kterou též nazývá *předtaktím*.

V práci Ondráčkové se setkáváme s rozlišováním významového a rytmického členění výpovědi.⁷ Významové členění výpovědi přispívá k realizaci obsahové náplně sdělení; ve zkoumání členění rytmického se zaměřujeme na pozorování umístění přízvučných a nepřízvučných slabik, tempo a pauzy a celkový akustický dojem z mluvního projevu z pohledu posluchače. Při členění významovém docházíme k významovým jednotkám (kólům, promluvovým úsekům), které obsahují většinou skupinu slov sémanticky spojených. Mezi kóly by se neměly nacházet *absolutní pauzy* (Ondráčková, 1954: 145). *Absolutními pauzami* zde autorka patrně míní reálné časové prodlevy, tedy ticha či hezitační zvuky.

Při zaměření na členění rytmické docházíme k nejmenší rytmické jednotce souvislé řeči, k mluvnímu taktu. Při zjišťování rytmických jednotek je třeba rozumět obsahu promluvy, na rozdíl od zkoumání pouze přízvučných a nepřízvučných slabik – ty vnímáme, i když promluvě nerozumíme. Správné prozodické členění promluvy je pak velmi důležité pro srozumitelnost a přirozenost projevu.

2.2.4.2 Přízvučnost v češtině

V dřívějších pracích byl český přízvuk hodnocen jako *dynamický*, silový. B. Hála uvádí, že „základním pojítkem taktu je síla“ (Hála, 1948: 142). Dále zmiňuje, že členění výpovědi závisí na přízvuku a důrazu, přičemž „obojí je projevem síly“ (Hála, 1948: 143); pouze důraz má sílu větší, a tak je výraznějším činitelem frázování. Termín *síla* ovšem nemá oporu ve dnešním fonetickém názvosloví; asi nejpřesnějším ekvivalentem by v tomto případě byla *prominence*. V příručce *Uvedení do fonetiky češtiny na obecně fonetickém základě* (1962) autor výslovně uvádí, že podstatou českého přízvuku je

⁷ Rozdělení na rytmickou a významovou stránku výpovědi používá také Hála. Jako reprezentace jednotek významových uvádí slova, jako reprezentaci jednotek rytmických takty (Hála, 1975: 262).

zesílení hlasu, dosažené zvětšením výdechového tlaku na hlasivky a zvýšením jejich elasticnosti. Vokály přízvukné slabiky jsou tedy sonornější než ve slabikách nepřízvukných; dále je podle Hály pro přízvuknou slabiku typické zesílení a prodloužení artikulace konsonantů; někdy se také na přízvukné slabice zvyšuje základní tón (Hála, 1962: 303). V této i v pozdější práci *Fonetika v teorii a praxi* (1975) ovšem uvádí, že síla přízvuku se neřídí absolutní silovou stupnicí; může se stát, že přízvukná slabika v tiché řeči může znít slaběji než nepřízvukná slabika v řeči hlasité.

„Spokojujeme se proto s komparativním hodnocením, uskutečňovaným v rámci téhož slova nebo taktu.“ (Hála, 1975: 255)

V práci z roku 1962 dokonce výslovně uvádí, že český přízvuk je „mírný“ a že „silový rozdíl mezi slabikou přízvuknou a nepřízvuknou je v češtině nevelký.“ (Hála, 1962: 303) Můžeme tedy konstatovat, že přízvuk v češtině rozhodně není dynamický podle absolutních měřítek, nýbrž je založen na kontrastu.

Z předchozího textu vyplývá, že při rytmickém členění výpovědi je z vlastností zvuku nejdůležitější percepční kontrast silný-slabý. Přízvuk je tedy hlavním činitelem rytmizace promluvy. Ondráčková připomíná, že v mluvních taktech se síla neprojevuje absolutně, nýbrž relativně – vzhledem k ostatním slabikám; výslovně uvádí, že přízvukná slabika může být vzhledem k nepřízvukným slabikám jiného taktu (a dle některých analýz současné češtiny dokonce i vzhledem k nepřízvukným slabikám téhož taktu) i slabší; ani nepřízvukné slabiky nemusejí mít ve všech případech stejnou intenzitu (Ondráčková, 1954: 146). Slabiky přízvukné a nepřízvukné bychom měli srovnávat vždy jen v rámci jednoho taktu; podle autorky je slabika přízvukná vždy silnější (patrně míněno hlasitější) než slabiky ostatní v témže taktu (Ondráčková, 1954: 147). S tímto tvrzením bychom však mohli s přihlédnutím k předcházejícím řádkům polemizovat. Autorka dále uvádí, že to, jak silná přízvukná slabika je, také ovlivňuje délku taktu, nicméně s tímto opět nelze souhlasit: významnost prominence přízvukné slabiky počet slabik v taktu nijak neovlivňuje.

Tzv. *důraz*, kterým v rámci větné prozodie vyjadřujeme svůj postoj nebo důležitost toho kterého slova, se také obvykle realizuje na přízvučné slabice taktu.

Přízvuk v češtině není kvalitativní – kvalita vokálů přízvučných se od nepřízvučných nijak neliší; není tedy nijak omezeno, která samohláska se může objevit v české přízvučné nebo nepřízvučné slabice. Během zpracování materiálu jsme však někdy při rozhodování o přízvučnosti konkrétní slabiky přihlíželi i k případné vokalické redukci, o které se předpokládá, že je typičtější pro nepřízvučné než pro přízvučné slabiky. Přízvuk v češtině nemůžeme nazvat ani dynamickým a také není možné ho považovat za kvantitativní; v úvahu by přicházelo jeho vymezení jakožto melodického – ve starších pracích se k tomuto faktu uvádělo, že jde zpravidla jen o melodické posuny v rámci čtvrttónů, nicméně v současnosti se spíše soudí, že melodické posuny mohou být i vyšší. Jak již bylo zmíněno výše, český takt nemusí být vymezen pomocí měřitelných zvukových charakteristik přízvučné slabiky. Je vnímán spíše na základě hlediska segmentace než prominence.

F. Daneš ve své publikaci *Intonace a věta ve spisovné češtině* (1957) předpokládá, že pod pojmem přízvuk se obvykle rozumí různý stupeň síly, s níž vyslovujeme jednotlivé slabiky nebo hlásky. Tato síla se v různých pramenech definuje různě, např. jako síla výdechového proudu (Jespersen, Jones), síla artikulační (Hála), síla hlasu (Havránek & Jedlička, 1988) – ani toto však vesměs nejsou exaktní veličiny, nýbrž jsou založeny spíše na intuitivním povědomí. F. Trávníček rozlišuje přízvuk výdechový a tónový, protože zahrnuje pod přízvuk i melodii; s podobným rozlišením pak pracuje Jespersen, Trubeckoj a Hála, který rozlišuje „modulaci hlasovou“ na silovou a výškovou, tedy na přízvuk a melodii.

Daneš zmiňuje, že o tom, jestli je silový přízvuk jevem založeným na intenzitě, jsou pochybnosti, a dodává, že nevylučuje ani možnost, že slovní přízvučnost je výslednicí několika činitelů.

„Relevantní není absolutní stupeň síly, nýbrž silový poměr jednotlivých slabik v rámci nějakého úseku řeči.“ (Daneš, 1957: 21-22)

Daneš dále konstatuje, že český přízvuk je na první slabice taktu a má pouze delimitativní, nikoli fonologickou funkci, a předkládá i Trávníčkovu pojetí přízvuku slovního, podle něž nese přízvuk každé slovo. S tímto tvrzením ale autor nesouhlasí, neboť v reálné výpovědi některá slova přízvuk ztrácejí, a proto je namístě hovořit spíše o *přízvuku taktovém* (Daneš, 1957: 22). Fonetický takt je potom dán seskupením nepřízvučných slabik kolem jediného přízvuku slovního (taktového). Není tedy vhodné na této úrovni popisu⁸ používat pojem *slovní přízvuk*, neboť samostatný přízvuk se netýká každého slova, ale i rytmického spojení několika slov.

Ve starších studiích a příručkách (včetně Danešovy práce z roku 1957) se můžeme setkat i s pojmem *přízvuk vedlejší*. Havránek a Jedlička dodávají, že tento fakultativní přízvuk se objevuje při pečlivé a pomalé výslovnosti v delších (alespoň trojslabičných) slovech, vyskytuje se zpravidla na lichých slabikách a bývá slabší než přízvuk hlavní (Havránek & Jedlička, 1988). Přízvuk vedlejší je ale v češtině spíše tradován než skutečně realizován; můžeme se s ním setkat pouze vzácně, např. při slabikování, skandování, vlivem důrazu nebo ve slovech složených. Je fakultativní a Hála ze své zkušenosti poznamenává, že i pohyblivý (a tedy nikoli realizující se na lichých slabikách). Někdy se jako přízvuk vedlejší chápe spojení předpony a kořenu slova, kdy jsou např. v superlativech zdůrazněny obě slabiky. Takto vidí přízvuk vedlejší např. Gebauer, ale Chlumský tyto případy považuje za rozpad slova na dva takty s dvěma samostatnými přízvuky. Vedlejší přízvuk se podle některých autorů vyskytuje někdy i v situaci spojení jednoslabičné předložky a jména. Za extrémní případ je považováno kladení vedlejšího přízvuku na každou nepřízvučnou slabiku taktu, k čemuž dochází vlivem emfáze (*Ne-na-po-ví-dej!*) (Hála, 1962: 306).

V pozdější studii *K analýze přízvučnosti, zvláště v češtině* (1962) Ondráčková uvádí, že otázka české přízvučnosti stále není jednoznačně vyřešena, a proto chce autorka svými metodami (zapisovač hladiny zvuku a indikátoru

⁸ Popisujeme realizaci, nikoli teoretický potenciál každé první slabiky ve slově nést přízvuk.

dynamiky P. Janoty) dokázat komplexní podstatu přízvuku. Je si vědoma úskalí s tímto úkolem spojeného:

„V některých případech se při použití různých přístrojů ukazuje, že se naměřené výsledky mohou jevit odlišně od našeho sluchového dojmu. Je to způsobeno tím, že při zjišťování místa a kvality přízvuku hodnotíme přízvučnost (tohoto termínu užíváme v této souvislosti raději než termínu přízvuk) jako celek, tj. jako komplexní jev. Naproti tomu přístroje registrují sice spolehlivě a v absolutních hodnotách, ale pouze ty složky přízvuku, na které jsou podle své konstrukce schopny reagovat.“ (Ondráčková, 1962: 81-82)

Např. ze záznamu hlasitosti vyplynulo, že intenzita slabiky, která je z poslechu hodnocena jako přízvučná, může být i nižší než intenzita slabiky sousední, jež je hodnocena jako nepřízvučná. Přízvučnost se tedy nemusí automaticky projevit větší objektivní intenzitou. Menší hlasitost přízvučné slabiky může být nahrazena jinými charakteristikami, jako je např. změna výšky tónu aj. (Na důležitost paralelního záznamu dynamiky a melodie řeči poukázal již v roce 1949 Janota.)

Autorka zde hovoří o zkoumání hlasitosti, nicméně podle popisu postupu jsme oprávněni se domnívat, že nešlo o hlasitost, jak ji známe dnes, totiž o percepčně založenou veličinu se sony jakožto jednotkami. Podle užitého zařízení zřejmě zaznamenávala hladinu zvuku (SPL, jednotka dB), což jí umožnilo postihnout pouze aspekty intenzity.

Ondráčková při zkoumání hlasitosti v souvislé mluvě a jejím uplatnění v přízvučnosti vycházela jak ze záznamů, tak z přímých projevů a používala přístroj pro záznam hladiny zvuku (záznam hlasitosti by odpovídal percepci stupňů přízvučnosti jen tehdy, byla-li by inherentní sonorita jednotlivých slabik stejná, nebo v případě, že je v přízvučné slabice vokál s vyšší specifickou hlasitostí, než mají okolní slabiky). Autorka zaznamenávala izolované hlásky, slabiky, slova i výpovědi ve spisovné češtině i v nářečí. Zjišťovala, jak se projevují různé složky přízvuku – v rámci toho analyzovala vždy spektrum stejných samohlásek, z nichž jedna byla stabilně hodnocena jako přízvučná a druhá jako nepřízvučná. V nářečním mluvním projevu byla přízvučná slabika také nejhlasitější, ale v jedné ze dvou zkoumaných jednotek

neměly sousedící slabiky intenzitu menší než 5 dB ve srovnání s nejhlasitější slabikou (slabiky s jádry [i, i, i]); ve druhé byl rozdíl mezi přízvučnou a sousedící nepřízvučnou 13 dB ([a, a:, i:]); to autorka vysvětluje různou specifickou hlasitostí samohlásek [a] a [i]. Kromě toho bylo třeba zohlednit funkční stupeň hlasitosti v souvislé mluvě, který posluchač posuzuje ve vztahu k hlasitosti ostatních slabik rytmické jednotky.

Pokud jsou ve slabikách rytmické jednotky (taktu) samohlásky lišící se kvalitou i kvantitou, jsou vztahy týkající se jejich hlasitosti složitější. Autorka uvádí příklad se slovem *teta*, ve kterém první slabika, ač přízvučná, má nižší stupeň hlasitosti než druhá. V analýze nářečních slov *pisničky* a *kramářský* byla hlasitost přízvučné slabiky největší. Přístroj na záznam hladiny zvuku vypovídá jen o jedné složce přízvučnosti; nelze z něj ale jednoznačně určit její podíl na celkové přízvučnosti. Záznamy hlasitosti jsou tedy naopak důkazem, že stupeň přízvučnosti není totéž co stupeň hlasitosti (Ondráčková, 1962: 85).

Při spektrální analýze byly srovnávány samohlásky na předposlední pozici ve slově; šlo o hlásky z přízvučných slabik *pisničky* a *kramářský* a z nepřízvučných slabik *nebáli* a *modliči*. Po stránce rytmické jsou tedy ve stejném postavení, nemají však stejné souhláskové okolí, nicméně i při vlivu, který na samohlásky sousední konsonanty mají, si vokály musejí zachovat takovou kvalitu, aby byly hodnoceny jako právě ty konkrétní samohlásky. Analýza probíhala pomocí spektrografu; zkoumány byly vrcholné fáze hlásek; podle reprodukováných výseků (prováděno Janotovým segmentátorem) musely být samohlásky sluchem jasně identifikovány.

Oproti očekávání se neobjevovalo tolik vysokých frekvencí u [i] – spektrogramy nezachycují výšku a hlasitost, ale frekvenci a intenzitu, takže ve výsledku se lépe díky větší intenzitě uplatnily formanty [a:]. Slabiky přízvučné vykazovaly větší spektrální bohatost – tak se tedy opravdu projevil rozdíl mezi přízvučnými a nepřízvučnými slabikami v rovině silové a výškové. Kromě těchto dvou aspektů také hraje roli poměr jednotlivých formantů samohlásek. U přízvučného [a:] např. téměř splývaly formanty F1 a F2; u nepřízvučného jsou výrazně odděleny.

Pokud budeme brát v potaz všechna tři [i] ve slově *pisničky*, musíme konstatovat, že nejbohatší spektrum má [i] přízvučné. První [i] je méně zřetelné, poslední má stejné frekvenční formantové oblasti jako přízvučné. Oproti [a:] je zde větší rozestup formantů u hlásky přízvučné než u nepřízvučné. Frekvenční oblasti u nepřízvučného [i] jsou (na rozdíl od [a:]) širší než u přízvučného [i].

Výše zmíněná fakta jsou podle mého názoru poněkud diskutabilní: experiment se prováděl na minimálním počtu vzorků, šlo spíše o jakousi sondu. Jak sama autorka uvádí, nebralo se v úvahu hláskové okolí, které konečné výsledky také mohlo ovlivnit. Navíc (jak je uvedeno výše) nešlo o výzkum subjektivní hlasitosti, ale hladiny zvuku. Nicméně i tak se lze domnívat, že přízvučnost nelze redukovat na pouhou vlastnost přízvučné slabiky: je v první řadě založena na kontrastu, který přízvučné a nepřízvučné slabiky odlišuje. Toto rozlišení také nespočívá pouze v jediné vlastnosti.

2.2.4.3 Délka mluvních taktů

Co se týče délky mluvních taktů podle počtu slabik, Ondráčková usuzuje, že v jazycích obecně se vyskytuje tendence k tvoření taktů poměrně krátkých, což přičítá snaze o úsporu energie při mluvení (Ondráčková, 1954: 147). Krátce se také zabývá sklony k vyrovnávání délky taktů: takty o vyšším počtu slabik pronášíme rychleji než takty krátké, abychom vyrovnali rytmus – např. Chlumský vidí příčinu ve fyziologii; Hála pouze uvádí, že k tomuto jevu dochází „vlivem délky taktu“ (Hála, 1948: 144), a dále si všímá i dlužení konsonantických segmentů a vytváření schwa vlivem důrazu (většinou přízvuku). Podle Hály a Chlumského způsobuje určité dlužení konsonantu také přízvuk.

Dále autorka definuje pojmy *mezitaková pauza* a *mezitaktová hranice* a vymezuje se proti absolutnímu pojetí mezitaktové hranice jako časového intervalu. Taková hranice se jako pauza realizuje jen tehdy, když jde zároveň i o pauzu mezi výdechovými nebo významovými skupinami; jinak pouze dochází ke snížení intenzity fonace. Tato hranice se v češtině projevuje jako kontrast mezi nepřízvučnou poslední slabikou předchozího taktu

a přízvučnou první slabikou taktu následujícího; v jazycích s přízvukem vzestupným jde o kontrast poslední přízvučné a první nepřízvučné slabiky. U taktů obstupných je nutné přihlížet k významové stránce jazyka.

Ondráčková uvádí jako doklad mezitaktové hranice četný výskyt asimilací znělosti a artikulačních zejména v rámci taktu, a nikoli na rozhraní dvou taktů. S tímto tvrzením však při dnešní úrovni znalostí českého mluveného projevu nelze souhlasit – asimilace znělosti jdou přes hranici taktů zcela běžně; vyskytují se i přes hranici úseků.

Členění na takty není pouze subjektivní záležitostí vycházející od mluvčího, ačkoli individuální faktory také mohou hrát jistou roli. V každém jazyce lze vysledovat určité tendence k rytmickému členění na mluvčím nezávislé.

Výzkum Ondráčkové se zaměřoval na analýzu jak textů, tak mluvených projevů. 50 000 taktů pocházelo z rozboru psaných textů českých spisovatelů s následnou kontrolou podle předčítání těchto textů; materiál dále obsahoval 1775 taktů mluvené češtiny (gramofonové desky, zápisky Chlumského); autorka dále odděluje řeč přímou (9613 taktů) a nepřímou (40 387 taktů) a reflektuje, že ke čteným textům je třeba z hlediska členění přistupovat jinak než ke spontánním. Zdůrazňuje, že rytmické členění promluvy lze tímto způsobem zkoumat pouze poslechem, je tedy třeba důkladná kontrola analýzy (např. při čtení textu jinými osobami, čtení pomalé, s časovým odstupem ad.).

Autorka bere v potaz i tzv. taktovou předrážku, ale nepovažuje ji za součást taktu, nýbrž za zvláštní typ taktu (proti tomuto pojetí se vymezuje Daneš – předrážka podle něj nemůže být samostatným taktem, protože jí „chybí konstitutivní znak taktu, totiž přízvuk“ (Daneš, 1957: 23); k tomuto řešení se přikláníme i my). Předrážka se vyznačuje tím, že je pronesena výrazně hlouběji a s menší intenzitou než první slabika taktu – jedná se o odlišnou realizaci rytmicko-melodického zlomu. Někteří čeští teoretici předrážku spojovali s následujícím taktem, což mohlo navodit dojem obstupnosti (Zich); jiní ji považovali za samostatný takt (Hála). Za takt obstupný však nelze takt s předrážkou považovat zejména proto, že ve skutečných obstupných taktech

nezaznamenáme onen dynamicko-melodický zlom (Ondráčková, 1954: 151). Nejčastějšími předrážkami byly v materiálu Ondráčkové spojky *a* a *že*, dále výrazy *co*, *je*, *když*, *jak*, *i*, *to*. Podle našeho pojetí je taktová předrážka jednoslabičné nepřízvučné slovo, jež se přiklání k následujícímu taktu.⁹

Víceslabičná slova obvykle tvoří samostatné takty. Jednoslabičná slova pak mohou k rytmickému členění promluvy přispět několika dalšími způsoby: jednotka může nést přízvuk a být součástí následujícího taktu (např. pravá předložka; spojení dvou jednoslabičných slov; spojení jednoslabičného a víceslabičného slova), může přízvuk ztratit a a) být součástí následujícího taktu (např. spojka) b) být součástí předcházejícího taktu (např. příklonka *se*); může přízvuk zachovat a stát se samostatným taktem (Palková, 1994: 281-282). Dlužno však podotknout, že v mnoha případech by bylo podle tohoto obecného rozlišení těžké rozhodnout, jak by takt mohl být rozčleněn syntakticky nebo významově, a může tedy být dokonce snazší rozdělit promluvu podle percepce: obecně lze říci, že se často nabízí více možností rozčlenění textu, ale realizace mluvčím je vždy jedinečná. V reálném mluveném projevu nicméně dochází k jevům, které v popisu příliš zakotveny nejsou – v našem materiálu se často objevovaly případy z hlediska percepce jednoznačné, nicméně z hlediska ideálního popisu neobvyklé, např. situace, kdy jednoslabičná synsémantická slova (např. spojky, příklonky) nesla přízvuk a byla samostatnými takty. Podrobněji jsou tyto problémy a jejich řešení popsány ve třetí kapitole této práce.

Co se týče počtu slabik v taktu, je v materiálu Ondráčkové nejčastější takt dvouslabičný, po něm tříslabičný; taktů čtyřslabičných, jednoslabičných a více než čtyřslabičných je v materiálu méně. Desetislabičný takt se v korpusu neobjevil. V řeči přímé je předrážek cca 5 %, jednoslabičných taktů 10,5 %, dvouslabičných 41,6 %, tříslabičných 29,5 %, čtyřslabičných 10,9 %, pětislabičných 2,3 %; delších než pětislabičných dohromady cca 0,3 %. V řeči nepřímé je předrážek cca 4,3 %, jednoslabičných taktů 7,2 %, dvouslabičných

⁹ Plnovýznamová jednoslabičná slova přízvuk na rozdíl od neplnovýznamových často mají.

38,1 %, tříslabičných 32,4 %, čtyřslabičných 13,9 %, pětislabičných 3,5 %; delších než pětislabičných celkem cca 0,7 % (Ondráčková, 1954: 152).

Z těchto hodnot tedy můžeme vyvodit závěry, že v nepřímé řeči je silnější tendence k tvoření delších taktů, ale v obou typech projevů přesto naprosto převažují takty krátké, tedy dvouslabičné a tříslabičné. Devítislabičný takt byl v materiálu Ondráčkové pouze jeden a vyskytl se v řeči nepřímé.

V jednom taktu se převážně nacházelo pouze jedno nebo dvě slova. Dvouslabičná slova byla v taktu pouze dvě; častá byla kombinace jednoho dvouslabičného a jednoho nebo více jednoslabičných slov; velmi zřídka kombinace dvou dvouslabičných slov se slovy jednoslabičnými. Kombinace s tříslabičnými slovy se vyskytovaly velmi málo. Nejčastější případ byl ten, že počet slabik byl vyšší než počet slov. Pokud byly tyto hodnoty stejné (což také nebyl zcela neobvyklý případ), pak musela být všechna slova v taktu jednoslabičná, nebo víceslabičná slova kompenzovaly neslabičné předložky. Výjimečně byl počet slabik nižší – k takovým jevům docházelo v případě, že takt obsahoval pouze jednoslabičná slova a alespoň jednu neslabičnou předložku.

Autorka reflektuje i možnost rozpadu slova ve více taktů, což je jev, kterému se teorie příliš nevěnuje, nicméně ve spontánním projevu k tomuto úkazu dochází např. kvůli slabikování, přítomnosti emfáze nebo mechanickému opakování krátkých vět bez rozmyslu (patrně jde o podobný případ, jako je recitace). Tohoto fenoménu si v souvislosti se silným důrazem způsobeným expresivitou povšiml i Hála: „Dnes je nád-**herně**!!“ (Hála, 1948: 143) Chlumský také konstatuje, že u slov citově zabarvených může nastat výskyt dvou přízvuků vedle sebe – jeden je

„na první slabice, druhý na slabice druhé, a ta druhá slabika je při tom vytčena nejen melodicky, nýbrž také silou, čímž se takové důrazné slovo rozděluje na dva takty.“ (Chlumský, 1933: 176)

K příklonkám autorka uvádí, že se mohou objevit v podobě taktové předrážky, ale nejčastěji se spojují s předcházejícím slovem (nebo s předcházejícím i následujícím slovem) v jeden takt. Jako přízvučné se

vyskytují pouze v mluvě argotické (Ondráčková, 1954: 152-153). Toto tvrzení by si v současné češtině rozhodně žádalo bližší přezkoumání, protože v běžné konverzaci, ale ani ve veřejném projevu není nijak výjimečné, že se v kombinaci s jednoslabičnými slovy dostane přízvuk např. příklonce *se*. O spojkách *že, a, i, než* se Ondráčková vyjadřuje v tom smyslu, že většinou bývají součástí taktu, a to zpravidla na jeho začátku nebo na konci. Dnes však převážně považujeme spojky na začátcích taktů za taktové předrážky (v našem pojetí jsou ale předrážky také součástí taktu). Co se týče předložek, shoduje se autorčina koncepce s naší – tedy jednoslabičné původní předložky jsou součástí taktu a přebírají přízvuk; nepůvodní, víceslabičné a složené předložky přízvuk následujícího slova nepřebírají, je-li více než dvouslabičné.¹⁰ Tvoří tedy zpravidla předrážku nebo samostatný takt; někdy mohou být součástí taktu. Jednoslabičná původní předložka přízvuk přebírat nemusí, pokud za ní následuje víceslabičné (podle Romportla a Palkové „příliš dlouhé“) slovo, např. superlativ, je-li po předložce slovo nesklonné, udělá-li mluvčí za předložkou pauzu, setkají-li se dvě předložky nebo je-li předložka součástí taktu v jiné roli než v roli první slabiky (že za ní/chodí). Někdy se ve spontánním projevu setkáváme s tím, že přízvuk nese místo předložky původní první slabika jména. Považujeme to za výslovnost substandardní, avšak nikoli zřídka. Ondráčková toto nazývá výslovností nepřírozenou a přičítá to snaze o rytmizaci promluvy a analogii s předložkami nepůvodními a dvouslabičnými (Ondráčková, 1954: 155). Tento výčet je podobný pojetí Romportla (1985) a Palkové (1994), která dále uvádí, že původní předložka přízvuk mít nemusí, je-li následující slovo silně zdůrazněno (Romportl toto nazývá „vytýkáním do protikladu“ a zmiňuje další mimořádné případy, např. emfázi nebo básnický slovosled (Romportl, 1985: 130)) nebo je-li za ní slovo, které předložka neřídí (Palková, 1994: 339).

¹⁰ Romportl k předložkám nepůvodním zmiňuje, že nemají nikdy hlavní přízvuk a většinou jsou nepřízvučné, a výslovně uvádí předložky jednoslabičné i dvouslabičné (dle 'rady, podle 'tebe). Přízvuk podle něj získávají jen tehdy, následuje-li jednoslabičné nepřízvučné slovo (Romportl, 1985: 130) Pokud bychom ale na toto tvrzení přistoupili, museli bychom uznat existenci dvouslabičných předrážek.

Během zpracovávání materiálu pro tento výzkum jsme zjistili, že dosavadní teoretická řešení problematiky týkající se přízvučnosti jednoslabičných slov nepokrývají dostatečně důkladně všechny možnosti realizace zmíněných jevů ve standardním mluveném projevu. V rámci této práce předložíme různé doklady tohoto fenoménu pro formulaci závěrů k případnému začlenění do popisu podob prozodické jednotky mluvního taktu.

Ondráčková se dále zabývá výzkumem Zichova členění na *slovné celky*. Použila proto stejné texty jako Zich a dala je analyzovat šesti osobám; při rozčlenění na takty vždy vyšel výrazně vyšší počet než v původním výzkumu. Autorka odhaduje, že je tento fakt způsoben tím, že Zich patrně počítal taktové předrážky k následujícímu taktu (což se shoduje i s naším pojetím). Co se týče délky taktů podle počtu slabik, došla Ondráčková k naprosto stejnému pořadí jako Zich, což je zajímavé z toho hlediska, že Zich pracoval s texty od autorů o generaci starších, než byli ti, ze kterých čerpala Ondráčková. K tomu, že nejčastější takty jsou v češtině dvojslabičné nebo trojslabičné, došel i Frinta na základě rozboru textu z Babičky B. Němcové. Hála provedl podobný rozbor Nerudovy povídky Pan manžel a došel ke stejnému závěru jako Ondráčková. K těmto analýzám můžeme poznamenat, že každý z těchto autorů členil promluvy pouze na základě vlastního úsudku.

2.2.4.4 Tendence mluvního rytmu v češtině

Ondráčková se vymezuje proti Zichovým a Oberpfalcerovým názorům, že v češtině existují i takty vzestupné nebo obstupné, neboť celkový ráz češtiny je jasně **sestupný**. V češtině nikdy není přízvučnou slabikou poslední slabika; žádné slovo nemá přirozenou výslovnost s přízvukem na poslední slabice nebo na slabice uprostřed slova. Zichův příklad celistvého taktu *a dala jí* není adekvátní – slabika se může beze změny významu oddělit od slova *dala* dokonce i absolutní pauzou, což u pravých obstupných nebo vzestupných taktů není možné. Jak již bylo uvedeno výše, v pravých obstupných nebo vzestupných taktech nenajdeme dynamicko-melodický zlom typický pro předrážku a takt – slabiky navazují plynule (Ondráčková, 1954: 156).

V závěru práce formuluje autorka tyto tendence mluvního rytmu v češtině:

„Přízvuk je vždy na první slabice taktu. V češtině jsou tedy takty pouze sestupné; vzestupné a obstupné takty se v češtině nevyskytují.

Nejčastěji se vyskytují takty dvouslabičné a tříslabičné.

Takty jednoslabičné převyšují počet taktů čtyřslabičných v řeči hovorové (přímé); taktů čtyřslabičných je více než taktů jednoslabičných v řeči popisné (nepřímé).

Počet taktů o větším počtu než čtyři slabiky klesá prudce s přírůstkem počtu takových slabik.

Zvláštní druh mluvního taktu tvoří taktová předrážka.“ (Ondráčková, 1954: 156)

S výše uvedenými tezemi lze v podstatě souhlasit i dnes; bližší přezkoumání vyžaduje pouze pojetí taktové předrážky jako samostatného mluvního taktu – taková jednotka by postrádala to, co mluvní takt mít v češtině musí, totiž přízvuk. Proto považujeme za vhodnější řešení přiklonit se k názoru, že taktová předrážka se připojí ke slovu následujícímu a s ním bude tvořit jedinou rytmickou jednotku. Takt s předrážkou potom nebude mít přízvuk na první slabice.

Výsledky, ke kterým Ondráčková došla, byly zpracovány objektivně, tzn. na zpracování se podílelo několik různých osob, nicméně vždy je třeba brát v potaz, že takové hodnocení se zakládá na sluchovém vjemu, tedy na percepci posluchače. Intenzita slabik, jakou vnímá lidské ucho, nemusí tedy přesně korelovat s jejími absolutními akustickými hodnotami.

2.2.4.5 Pojetí taktů v této práci

V naší práci jsme zohledňovali nikoli takty ideální, nýbrž reálné; členění promluvy bylo tedy založeno na percepci. Co se týče formální klasifikace předrážek, na rozdíl od Ondráčkové jsme předrážku přiřadili do taktu k následujícímu slovu nesoucímu přízvuk.

Ačkoli podoba ideálních taktů není v češtině příliš rozpracována (literatura se obvykle zmiňuje pouze o spojení původní předložky se jménem a realizaci jednoslabičných slov), ani pravidla, která pro tento jazyk nějakým způsobem formulována jsou, nebyla mluvíci ve zkoumaném materiálu striktně dodržována. V realitě řečového projevu představuje problém, že skutečné

členění promluvy nemusí respektovat syntaktická ani sémantická omezení: významové jednotky mohou být ve spontánním projevu rozčleněny na několik taktů, v přirozené řeči se také velmi často objevuje tendence k vytváření jednoslabičných taktů (a někdy dokonce i promluvových úseků) např. z předložek, které by ideálně měly nést přízvuk a být první slabikou taktu, ze spojek, které zase obvykle tvoří taktovou předrážku, a někdy i z příklonek. V našem materiálu, který sestával pouze z projevů profesionálních mluvčích, se všechny tyto odchylky objevovaly často. (Ve spontánním projevu je situace ještě komplikovanější: není například neobvyklé, že mluvčí zaváhá, hezituje, odmlčí se atd., a z toho se odvíjejí i mnohé nepravidelnosti v rytmickém členění promluvy. Projev je také ovlivněn dalšími idiosynkratickými řečovými faktory.)

Jelikož zkoumáme mluvený jazyk, rozhodli jsme se naše postupy aplikovat na reálné takty, nikoli na ideální – ideální takty sice také mají svůj původní základ v realitě řečového projevu, ale stále mají blíže k teoretickému konstruktu a o plné shodě mezi skutečným a ideálním členěním promluvy rozhodně nelze hovořit. Proto se tedy chceme držet reality, která má větší výpovědní hodnotu o mluveném projevu, ačkoli se v některých případech může bránit zobecnění. I tak je ale naše řešení kompromisem – materiálem jsou nám texty, které jsou čteny profesionálními hlasateli. Je tedy nutné brát v potaz, že kdyby se podobný výzkum prováděl na čistě spontánních projevech, patrně bychom v jeho výsledcích našli jisté rozdíly. V tomto výzkumu není ale členění promluvy naším hlavním zájmem, a proto je možné takovouto střední cestu připustit: Primárně nás zajímá variabilita fonotaktických skeletů slov a mluvních taktů s ohledem na slovnědruhové zařazení a morfologickou komplexnost. Jak jsem již podotkla v úvodu, ve spontánních projevech je slovní zásoba dosti omezená; jejich velkou část tvoří neustále se opakující frekventované výrazy; ve čtených textech zpravodajských relací je oproti tomu lexikon bohatší a členění promluvy, ačkoli není zcela spontánní, má dostačující výpovědní hodnotu pro charakteristiku českého mluveného projevu.

2.3 Fonotaktika a struktura slabiky

2.3.1 Obecný úvod k problematice fonotaktiky

V této části práce podám stručný úvod k problematice fonotaktiky obecně s přihlédnutím k praxi v různých jazycích a specificky v češtině. V první části budu čerpat zejména z textu *The Frame/Content Theory of Evolution of Speech Production* (1998) Petera MacNeilage a z práce Tomáše Dubědy *Jazyky a jejich zvuky* (2005); v kapitolách věnujících se češtině z výzkumu Marie Ludvíkové (in Těšitelová, 1987), z práce Henryho Kučery (1968), ze *Statistik češtiny* (Bartoň et al., 2009) a další literatury.

2.3.1.1 Fonotaktika a morfonologie

Způsoby kombinace fonémů a hlásek spadají pod rámec fonotaktiky. Základním principem řeči je střídání apertury a striktury – jazyk s výlučně tímto ideálním kontrastem by tedy sestával pouze z kombinací typu CV. Tento princip se skutečně v některých jazycích silně uplatňuje – jedná se např. o japonštinu, tahitštinu nebo čínštinu. Ve většině dalších jazyků zpravidla vznikají několikačlenné konsonantické shluky. Vokalické skupiny (pokud nepočítáme diftongy) jsou obvykle řidší a vznikají spíše na hranicích morfémů a slov. Pokud ale v jazyce existuje minimum konsonantů nebo fonémů obecně, nejsou ani samohlásková seskupení nijak výjimečná – příkladem může být např. jazyk rotokas.

2.3.2 Teorie rámce a obsahu

Fylogeneze jazyka se do jisté míry zrcadlí v jeho ontogenezi – například slabičný typ CV je základem pro dětské žvatlání i pro jazykový inventář, přičemž preferovány jsou ty kombinace, kde C i V jsou na stejném místě (přední konsonanty se spojují s předními vokály, veláry se zadními vokály a labiální konsonanty s centrálními vokály – poslední typ slabiky je považován za tzv. čistý rámec); jazyky rovněž inklinují ke střídání místa artikulace pro následující konsonanty v CVCV sekvencích.¹¹ Je však třeba

¹¹ Rochetová a Schwartz v souvislosti s teorií rámce a obsahu a strukturou CVCV zkoumali stabilitu sekvencí, kde se střídá labiální a koronální konsonant v normální a zrychlené řeči. Ukázalo se, že úseky

podotknout, že v těchto případech jde pouze o tendence, nikoli o absolutní pravidlo.

CVCV jsou stabilnější ve zrychlené řeči, pokud jsou LK (labiální-koronální), než když jsou KL (koronální-labiální). Tento fakt by mohl vysvětlovat, proč jazyky preferují spíš LK sekvence. Původ větší stability LK sekvencí patrně souvisí s interartikulátorovou koordinací během KL a LK sekvencí se zrychlujícím se tempem; to pak vede rozmanité CVCV struktury, aby byly produkovány v jediném čelistním cyklu (netýká se ovšem reduplikací).

Strukturu CVCV sekvencí může vysvětlit skutečnost, že když opakujeme stejnou sekvenci jedním orgánem, může být hranice dosažena rychleji a výdej energie bude vyšší, než když tuto strukturu zastanou orgány dva ve fázové opozici – reduplikace je tedy méně efektivní a více energeticky náročná, než kdyby konsonanty variovaly. Autoři studie tuto problematiku přirovnávají ke hře na klavír, kde taktéž hudebník obvykle nepoužívá jeden prst na dva za sebou jdoucí tóny. U variujících sekvencí může interartikulátorová součinnost záviset také na stupni anticipace a fázování mezi artikulátory.

Během opakující se produkce LK nebo KL CVCV struktur v rostoucím tempu byly LK sekvence stabilnější, avšak KL sekvence spíše inklinují k přechodu k inverznímu LK vzorci. Na těchto faktech se tedy dá potvrdit větší stabilita LK sekvencí. Existuje možnost, že toto vyplývá z anticipace posunu špičky jazyka během pohybu rtů – je zde umožněno lepší fázování pro KL sekvence a jejich seskupování v jediném úkonu apertury (Rochetová & Schwartz, 2005: 1013). Základ LK efektu tedy poskytuje motorická ekonomie.

V experimentu Rochetové a Schwartze s reduplikovanými i variujícími CVCV sekvencemi ve zrychlené řeči byly výsledky koherentní pro všechny mluvčí – ukázala se podobná variabilita mluvčích i sekvencí s výsledkem velké stability LK struktur oproti KL. Během fáze zrychlení tempa řeči se snižovalo trvání sekvencí. Po dosažení určitého tempa se čelistní cyklus separoval od trvání slabiky a prodloužil se na dvojnásobek – dvě slabiky pak byly realizovány v jediném čelistním cyklu. Pokud se ovšem jednalo o reduplikace, pak se vždy dvě slabiky realizovaly ve dvou cyklech. V případě, že šlo o variující sekvence, byl v některých případech pozorován jeden cyklus pro obě slabiky a v jiných dva.

Pokud se tedy vztah struktura : čelistní cyklus 1 : 1 změní v 2 : 1, vyžaduje tato modifikace reorganizaci interartikulátorové koordinace. Tato koordinace je potom snazší pro sekvence LK než pro KL - koronální konstrikce může být anticipována během té labiální, ale obrácená tendence je pozorovatelná jen těžko. Taková anticipace je nemožná, pokud se jedná o reduplikovanou CVCV sekvenci: jediný artikulátor nemůže začít nový pohyb, dokud předtím nedokončí ten předchozí. Z toho plyne, že alternující pohyb dvou různých artikulátorů povoluje dosažení rychlejšího tempa produkce s menší spotřebou energie (Rochetová & Schwartz, 2005: 1016). V jazycích tedy existuje obecná preference variujících vzorců CVCV sekvencí nad reduplikovanými a také upřednostňování LK struktur nad KL.

K této studii je ovšem možno podotknout, že se jednalo o experiment s pouhými pěti francouzskými mluvčími a pro definitivní potvrzení těchto hypotéz by bylo potřeba výzkum rozšířit na další jazyky a větší počet zkoumaných subjektů.

Teorie rámce a obsahu tyto poznatky zohledňuje. Tento model patří ke všeobecnému rámci, který se snaží vysvětlit jazykové univerzálie z řečové matérie skrze všeobecné percepčně-motorické principy. Ústní a obličejová komunikace se podle tohoto přístupu vyvíjela původně z pohybů čelisti typických pro mechanismus příjmu potravy – sekvence otevírání a zavírání čelisti tedy vytvořily řečový **rámec**. Řízení motoriky ostatních artikulátorů se později vyvinulo ke globální koordinaci pro generování **obsahu** (Rochetová & Schwartz, 2005: 1013).

Specifickou organizační vlastností řeči je spojitá alternace apertury a striktury, v kterýchžto fázích dochází k artikulaci řečových jednotek. Tento cyklus otevírání a zavírání tvoří slabiku: fáze otevření zastupují vokály a fáze zavření konsonanty. Prvky slabičného „rámce“ a segmentálního „obsahu“ jsou řízeny v řečovém procesu odděleně (MacNeilage, 1998: 499).

Rámcové a obsahové komponenty řeči se mohly postupně vyvinout v samostatné realizace uvnitř dvou obecných záměrů řídicích motorických systémů: s motivací spjatý „vnitřní“ systém – zahrnující přední cingulární kůru neboli ACC (anterior cingulate cortex) a doplňkové motorické oblasti pro spontánní chování – který je zodpovědný za rámce, a laterální „vnější“ systém – zahrnující Brocovo centrum a okolí a Wernickovo centrum – který se specializuje na reakce na externí vstup a je z větší části zodpovědný za obsah (MacNeilage, 1998: 499).

Artikulační systém u savců jiných, než jsou lidé, je typicky používán pouze v otevřené konfiguraci během vytváření volání nebo výkřiku, ačkoli u některých druhů japonských makaků je možné vysledovat rytmické série alternací otevírání a zavírání během vokalizace.

S výjimkou několika slov, která se skládají pouze ze samotného vokálu, prakticky každá promluva každého mluvčího kteréhokoli lidského jazyka zahrnuje alternaci apertury a striktury vokálního traktu. Slabika je určena přítomností nukleu s relativně otevřeným vokálním traktem a okraji, které jsou charakterizovány jeho relativní zavřeností. Modulace tohoto cyklu na sebe u lidí vzala podobu typických odlišných základních jednotek,

konsonantů a vokálů v postupných fázích otevírání a zavírání. Lidská řeč je tedy odlišná od jiné vokalické komunikace typické pro savce.

Soudržnost slabiky a vztah segmentu ke slabice bývá zkoumán pomocí řečových chyb – záměn. Základním faktem, který se váže k řečovým záměnám, je, že se v prakticky všech segmentálních záměnách jednotky přesunou na pozici, kterou zaujímaly v původní slabice – iniciální konsonanty se zaměňují s iniciálními konsonanty, vokály s vokály a finální konsonanty opět s finálními. Na řečové rovině tedy existuje omezení slabičné pozice – v tomto smyslu tedy konsonanty a vokály neinteragují. Např. v angličtině existuje mnoho „zrcadlových“ párových slov (*tea* x *eat*), jejichž jediný rozdíl je pořadí fází, nicméně záměny takového typu nejsou doloženy. V hláskových ortografických systémech se však záměny vokálů a konsonantů dějí běžně, protože na rozdíl od zvukové jazykové produkce nejsou v grafické formě vokály od konsonantů nijak inherentně odlišeny (MacNeilage, 1998: 502).

Podle teorie artikulační fonologie (Browman, Goldstein) nejsou nejmenšími artikulačními jednotkami konsonanty a vokály, ale tzv. *úkony* (*gestures*), akce jednotlivých artikulátorů. Otázkou je, zda nepovažovat za základní jednotky právě je, ale proti tomu hovoří fakt, že pokud v řeči něco zaměníme, je to zpravidla celý segment (nejčastější případ) nebo skupina segmentů. Pokud se původní a vyslovený segment neshodují pouze v jediném úkonu, není pochopitelně možné toto prokázat. MacNeilage však uvádí výsledky výzkumu 72 případů řečových chyb se záměnou, kde se chybné segmenty vždy lišily alespoň dvěma úkony, nicméně pouze ve třech případech se tyto jednotky lišily právě úkonem; ve všech ostatních případech šlo o záměnu celých segmentů. Proto je z hlediska popisu vhodnější uvažovat o hláskách či fonémech jako základních jednotkách. Tento výsledek podle autora plně neprokazuje, že úkony nejsou nezávislými jednotkami v řídicím procesu, ale lze říci, že pokud tato nezávislost existuje, pak pro řeč ani pro popis systému není potřeba.

Otázkou může být, jestli řeč sama vyvinula zcela nový generátor rytmu s vlastními nadřazenými řídicími strukturami, které by mohly reagovat na

požadavky koordinace podobě jako ty, které byly ve starém systému – pokud by to tak ale bylo, nemohli bychom vysvětlit, proč řeč využívá stejný vzorcový generátor v mozkovém kmeni, který používají ingestivní cykličnosti, a proč řídicí struktury pro řeč jsou alespoň částečně sdíleny těmi, které zodpovídají za přijímání potravy (MacNeilage, 1998: 503).

Již jsme zmínili, že ontogeneze řeči je do jisté míry rekapitulací fylogeneze. I ve vývinu řeči u lidského individua se nejprve objevuje slabičný rámeček (zhruba v období prvotního žvatlání, kdy promluvy mají pevný rytmus, ve kterém je slabičný rámeček jednotkou) a až poté segmentální obsah. Vycházející z těchto poznatků předkládá autor hypotézu „prvního jazyka“ právě na základě raných řečových projevů dětí. Ve svém protojazyce počítá s dvěma nazálami, třemi neznělými explozivami, dvěma glidami a třemi krajními vokály [i, a, u] a předpokládá jednoduchou slabičnou strukturu CV, která by stála buď sama o sobě, nebo by byla opakována. Takový jazyk by pak měl podle autora cca 102 slov.

2.3.3 Fonotaktické poznatky napříč jazyky

Výzkum J. H. Greenberga (in Duběda, 2005) ukázal, že ze 104 zkoumaných jazyků jich 90 povoluje souhláskové shluky na začátcích slov, zatímco na jejich koncích pouze 62, což odpovídá tendenci k otevřenosti slabiky a k maximálnímu obsazení praetury (viz níže). Z jeho zkoumání také plyne závěr, že shluky neznělých souhlásek jsou častější než skupiny souhlásek znělých, dále že souhlásky ve skupině bývají co do znělosti homogenní (a asimilace znělosti v češtině tedy není nijak výjimečným úkazem) a že konsonanty na konci slova bývají spíše neznělé (opět i v češtině). Z jeho studie také vyplývá potvrzení skutečnosti, že počet konsonantů v praeture a v kodě narůstá postupně (pokud jazyk povoluje např. tříčlenné shluky, pak i dvoučlenné apod.). Souhláskové shluky podle Greenberga také bývají rozložitelné do kratších řetězců, které se vyskytují v témže jazyce i samostatně. Další princip se týká kombinací hlásek podle způsobu tvoření a podle sonority: v mnoha jazycích platí, že sekvence frikativa + exploziva či exploziva + frikativa jsou častější než kombinace týchž hlásek. Poslední zobecnitelnou univerzálií je fakt, že likvidy obvykle stojí za obstruentem na

začátku slova a před ním na konci slova (což odpovídá principu stoupající sonority slabiky – sonorita vzrůstá od začátku k vrcholu slabiky a klesá od vrcholu ke konci, viz níže).

Z těchto principů pak je možné vyvodit, že souhláskové skupiny podléhají jak asimilačním (asimilace znělosti), tak disimilačním (kombinace explozivy a frikativy) tendencím. Výsledky výzkumu se však nedají absolutizovat, neboť nepojednávaly konsonantické skupiny uvnitř slov ani přes jejich hranice, nicméně i tak zde můžeme pozorovat silnou korespondenci s univerzálními principy slabičné struktury.

Kombinací fonémů a alofonů se vztahem k jednotkám, které nesou význam, tedy k morfémům, se zabývá morfonologie. Obecně platí, že jednotka foném sama o sobě význam nenesou (až na okrajové případy vazby hlásky na význam, např. tendence užívat vysoké a přední hlásky u zdvojnásobení (Ulan, 1978 in Duběda, 2005: 95)). Existuje několik univerzálních či kvaziuniverzálních morfonologických principů, z nichž zmíním např. pravidlo, že

„velikost inventáře a struktura slabiky určují množství kombinací, které jsou při dané délce slova schopny zajistit fonologickou odlišitelnost.“ (Duběda, 2005: 94)

Přítomnost tónů pak další fonologickou rozlišitelnost výrazně zvyšuje (např. čínština má omezenou slabičnou strukturu a morfémy pouze jednoslabičné, ale díky přítomnosti tónů je zajištěna významová rozlišitelnost). Obecně platí, že jazyky využívají svoje fonologické inventáře, aby zajistily rozlišitelnost jednotek nesoucích význam a obvyklou fonologickou redundanci.

V mnoha jazycích není neobvyklým jevem, že často užívaná slova inklinují ke kratší délce než slova vyskytující se v korpusu řídce; stejně tak mají kratší slova obvykle složitější slabičnou strukturu než slova s větším počtem slabik. Stoupající počet slabik tedy zpomaluje nárůst fonémů (Menzerathův zákon).

Rozlišujeme tři oblasti fonologických korelátů morfematické a slovní struktury, a to hraniční signály (v některých jazycích např. aspirace, umístění rázu apod.), znaky fonologické soudržnosti (např. asimilační tendence)

a fonologické procesy na hranicích morfémů a slov – zde se jedná např. o přeslabikování, elize, epenteze či změny segmentů (Duběda, 2005: 95).

2.3.4 Struktura slabiky

Slabika je nejmenší jednotkou souvislé řeči a je založena na kontrastu sonority segmentů, které obsahuje. Základem a povinnou součástí každé slabiky je jádro neboli nukleus – může tedy postrádat praeturu i kodu. Její průběh je charakterizován plynulým pohybem od striktury k apertuře. Primární slabikou je patrně kombinace bilabiální nazály a hlásky [a]; dále se často vyskytují velární hlásky v kombinaci se zadními vokály a alveolární souhlásky se samohláskami předními. Slabika jako základní percepční jednotka je v jazycích univerzální. Co se liší, je hodnocení slabičnosti – to závisí na fonologické struktuře jazyka a na mluvčím, který tímto jazykem vládne, např. vnímání slabikotvorných souhlásek se v různých jazycích liší: v češtině mohou likvidy v jistém kontextu tvořit nukleus slabiky, ale např. mluvčím vietnamštiny by sekvence -CR (*bratr*) na konci slova slabičně vnímána nebyla. Přesto jsou posluchači v delimitaci slabičného jádra obvykle mnohem více konzistentní než ve vymezení hranic periferních součástí, tedy praetury a kody – mluvčí jazyka mohou tedy spolehlivě určit podle počtu nukleů, kolik má daná jednotka slabik; jejich povědomí o slabičných hranicích se však bude rozcházet. Někdy mohou pro vnímání hranice pomoci fonotaktická pravidla daného jazyka (např. slovo *basta* může být rozděleno na slabiky *ba-sta* i *bas-ta*. Ve španělštině ale kombinace -st- není v praeture přípustná, takže zbývá pouze jedna možnost dělení. Takové ovšem nejsou zdaleka všechny případy.). Jak píše ve své práci *Číslo o hláskách* M. Ludvíková, z hlediska přirozeného slabikování umíme bezpečně určit slabiku nebo její hranici jen tehdy, když se souhlásky a samohlásky pravidelně střídají (např. *zá-sa-da*) (Ludvíková, 1987: 104). Ani pravidla pro dělení slov v psaném textu nemohou tuto problematiku bezvýjimečně ošetřit.

Problémy při stanovování univerzálních principů týkajících se slabiky tvoří případy, kdy se podle některých teorií hranice slabiky neshoduje s hranicí segmentu – pak se obvykle počítá se segmentem náležejícím k oběma slabikám, k první jako koda a ke druhé coby praetura. Zvláštním případem je

pak interpretace geminát. Tyto situace se vyskytují spíše v jazycích se složitější slabičnou strukturou.

Dalšími jevy komplikujícími formulování univerzálií jsou tzv. pobočné slabiky (likvidu od jádra odděluje jeden nebo více obstruentů, např. v iniciální pozici u českých pobočných slabik ve slovech *rty*, *lžíce*, *msta*), neznělé slabiky a případy slabikotvorných souhlásek. V češtině na rozdíl např. od chorvatštiny ovšem není možné, aby slabičným jádrem byla likvida bez praetury, a navíc pobočné slabiky nenesou přízvuk, takže se nabízí interpretovat tento fenomén jako neslabičný. Duběda k této problematice poznamenává, že v určitém kontextu mohou i iniciální likvidy v pobočných slabikách vytvářet skutečné slabiky, a to na hranicích slov, srv. spojení *pět lžic* (Duběda, 2005: 132). Do kategorie pobočných slabik je podle některých teorií možné zahrnovat i případy, kdy je vedlejší vrchol neznělý, nebo např. postkonsonantické sonory (francouzština).

Také nemusí platit absolutně, že slabiky jsou vždy sonorní, nebo dokonce znělé – v japonštině dochází systematicky k desonorizaci vokálu situovaného mezi dvěma neznělými konsonanty. To, co zde zachovává dojem slabičnosti, jsou časové proporce jednotek. K podobným jevům dochází v mnoha jiných jazycích mimo standard. Do této skupiny můžeme zařadit i onomatopoická slova sestávající cele z neznělých konsonantů, např. v češtině *pst*, *pšt* apod.

Jako slabičný základ se zpravidla označuje nukleus a koda – praetura do něj zahrnuta není, ačkoli (jak už bylo řečeno) průběh slabiky se pohybuje od striktury k apertuře a nejobvyklejší strukturou slabiky je právě kombinace CV. Mezi jádrem a kodou se však na rozdíl od vztahu praetury a jádra projevují kompenzační strategie, např. v angličtině trvání vokálu v nukleu kompenzuje trvání konsonantu v kodě, takže vokál před neznělým a napjatým konsonantem bývá kratší než před znělým a nenapjatým konsonantem. Důležitost kody a postradatelnost praetury se projevuje i v morových jazycích (japonština, arabština). Ze všech pozic má nicméně nejmenší obsazení právě koda.

Databáze jazyků WALS (Haspelmath, Dryer, Gil & Comrie, 2008) rozlišuje jazyky s tzv. jednoduchou slabičnou strukturou – typu (C)V, dále slabiky středně komplexní – typu (CC)V(C), přičemž druhý konsonant je likvida, a strukturou komplexní – typu (CCC)V(CCCC). Nejvíce jazyků (56 %) má středně komplexní slabiku; tyto jazyky jsou geograficky poměrně rovnoměrně rozvrženy. Druhým nejčastějším typem je slabika komplexní (30 %), která se vyskytuje zejména v oblasti severozápadní Severní Ameriky, v oblasti Eurasie a Papui-Nové Guinei. Nejméně zastoupeným typem je slabika jednoduchá. Pouze takovouto strukturu má jen 12-13 % jazyků v oblastech rovníkové a jižní Afriky, Polynésie, Papui-Nové Guinei a Amazonie. Distribuce konsonantů ve slabikách pak jistým způsobem koresponduje s velikostí konsonantických inventářů: např. komplexní stavba slabiky tedy může znamenat velký inventář konsonantických fonémů (Maddieson, 2003 in Duběda, 2005: 138).

Nukleem je nejčastěji vokál, řidčeji pak slabikotvorná sonorní souhláska, zpravidla pak r-ová či l-ová hláska nebo nazála, nebo dokonce i šumový konsonant. Podle A. Bella má 58 % jazyků umožňujících nevokalické jádro pouze slabičné sonory, 30 % slabičné sonory i obstruenty a 12 % pouze slabičné šumové hlásky. Nejpreferovanějšími nukleji jsou tedy vokály, o něco méně obvyklé jsou sonory a poměrně výjimečné jsou pak slabikotvorné obstruenty. (Bell, 1987 in Duběda, 2005: 133) Slabikotvorné souhlásky pak stojí spíše v nepřízvučných slabikách. V celosvětovém měřítku jsou jako slabikotvorné hlásky častější nazály než likvidy, mezi slabičnými šumovými konsonanty jsou preferovány konstriktivy. Pro češtinu je typická slabikotvornost kromě vokálů u hlásek [r, l], méně často [m] a v cizích slovech a v substandardní, nejčastěji dialektální výslovnosti [n].

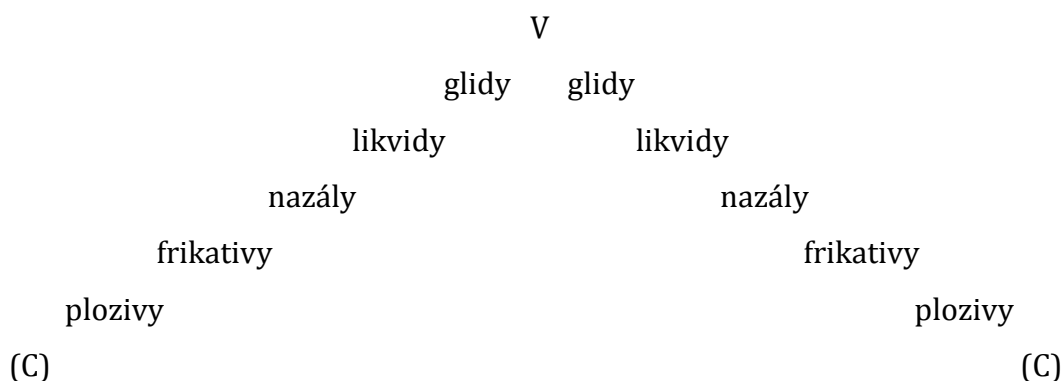
V jazycích se dále může uplatňovat tzv. slabičná váha. Za metricky lehké jsou považovány ty slabiky, jejichž nukleus tvoří krátký vokál a ve kterých absentuje koda; za metricky těžké označujeme slabiky, jejichž nukleus je tvořen dlouhým vokálem nebo diftongem a/nebo které obsahují kodu. Jak je již zmíněno výše, praetura je pro hodnocení slabiky jako těžké nebo lehké irelevantní. Pouze v některých jazycích je však kontrastu těžká/lehká slabika

využito. Asi nejnámějším projevem slabičné váhy je morový princip, který najdeme např. v japonštině, telugu nebo ve slovenštině. V těchto jazycích je mora základní jednotkou, na níž se odehrává vnímání rytmu (Duběda, 2005: 134). V japonštině se má krátká samohláska a koda hodnotu jedné mory, dlouhá samohláska je vnímána jako dvoumorová – v tomto jazyce jsou tedy rytmicky ekvivalentní slova s různým počtem slabik, ale stejným počtem mor (srv. *tanuki* x *tóki*).

Slabika je nejenže nejmenší jednotkou souvislé řeči, ale je pravděpodobné, že předcházela i hláskám, které vznikly až posléze diferenciací a stabilizací části slabiky, přičemž základem byl pohyb od striktury k apertuře. Modifikacemi striktury a apertury se posléze rozlišily jednotlivé segmenty (Hála, 1962: 281). Zde můžeme opět vidět paralelu s ontogenezí řeči, konkrétně s *prozodickou samoorganizací* u batolat – slabiky nemají jasnou podobu, ale jen prozodické obrysy, které postupně zaplňují specifické charakteristiky hlásek. V jazycích je tedy jednotkou univerzální, což podporují fakta, že slabika nepostrádá artikulační a percepční přirozenost a je využitelná v rytmické výstavbě výpovědi.

Nejčastějším a v jistém smyslu ideálním druhem slabiky je typ CV, od něž se odvíjejí všechny struktury složitější nebo jednodušší (Bell & Hooperová, 1978; MacNeilage, 1998 in Duběda, 2005: 136). Praetura je pak přirozeným začátkem slabičné artikulace, pro což svědčí vyšší obsazenost praetury vůči kodě i tzv. princip maximální praetury, který říká, že jazyky se snaží o co největší zaplnění praetury – např. i v češtině jsou častější konsonantické shluky v praetuře než v kodě a jejich maximální počet může být vyšší než u české kody. Neexistuje jazyk, který by nepovoloval slabičné praetury; jazyky se spíše snaží případy slabik bez praetury eliminovat protetickými nebo hiátovými hláskami.

Dalším kvaziuniverzálním slabičným principem je tzv. princip vzrůstající sonority, který praví, že sonorita vzrůstá od začátku k vrcholu slabiky a klesá od vrcholu k jejímu konci:



Např. approximanta se tedy pravděpodobně objeví blíže slabičnému jádru než neznělé obstruenty. Princip sonoritní výstavby ale mohou narušovat pobočné slabiky, slabičné konsonanty, případy, kdy dva sousedící konsonanty mohou být stejného druhu (*který, sféra*), a jiné situace, kdy není respektováno pořadí sonoritní stupnice (např. v češtině je stejně frekventovaná kombinace ploziva + frikativa jako frikativa + ploziva).

Slabika je také základním prvkem výstavby řečového rytmu. Vnitrosylabická alternace pak zahrnuje vesměs pravidelné střídání určitých souhlásek a samohlásek inventáře daného jazyka; intersylabická je potom střídáním různých slabik v rámci vyšších prozodických jednotek (o mluvním rytmu podrobněji níže v oddíle 3). Pozoruhodné je i zjevné kognitivní zakotvení slabičnosti: při metatezi hlásek *dojdi pro něj – projdi do něj* je zachován počet i struktura slabik (Duběda, 2005: 137). Kromě rytmu je slabika (resp. její nukleus) v prozodické rovině důležitá jako nositel intonačních charakteristik promluvy.

Zastoupení různých slabičných typů v 16 jazycích (Roussetová, 2004 in Duběda, 2005: 137) je následující: Nejsilněji zastoupenými slabičnými strukturami jsou typy CV (průměrné zastoupení ve vzorku 53 %) a CVC (31 %). S velkým rozdílem pak následují typy V (5 %), CCV a VC (oba po 3 %), CCVC a CVCC (oba po 2 %) a typ VCC (0,2 %). Ostatní struktury nedosahují ani jednoho procenta ve vzorku.

2.3.5 Struktura slabiky v češtině

V češtině je vztah hlásky na rovině řečové a fonému na rovině jazykové poměrně blízký, toto však neplatí o rovinách vyšších – slabika se zde s morfémem překrývá jen částečně. Na rozdíl např. od čínštiny, kde jsou původní morfémy vždy jednoslabičné, může morfém v češtině pokrýt několik slabik nebo být kratší než jedna slabika. Přesto jisté omezení ve vztahu slabiky k vyšším jednotkám existuje – hranice jednoslabičných a víceslabičných slov, přesněji taktů, se shodují i se slabičnými hranicemi.

Jádro české slabiky tvoří vokály a slabikotvorné sonory, tzn. [r], [l] a také [m] a [n] (v cizích jménech a v substandardní výslovnosti). Nejvyšší počet konsonantů je v praeture čtyři, v kodě tři – čeština tak inklinuje k principu maximální praetury. Tento fakt také může souviset s relativně velkou četností neslabičných předložek. V češtině se též projevuje obecná tendence ke tvoření slabik typu CV. Výskyt konsonantu mezi dvěma vokály signalizuje slabičný předěl, ten je vždy hodnocen jako praetura následující slabiky; v konsonantické skupině mezi dvěma vokály aspoň jeden C tvoří praeturu druhé slabiky. Povědomí o hranicích slabiky nicméně kolísá, jak už bylo zmíněno na začátku této kapitoly.

Pro spolehlivé odlišení praetury a kody bychom se měli podle Henryho Kučery (1968) řídit těmito pravidly: 1) praetura a koda by neměly být odděleny takovým způsobem, aby se jejich získáním zvýšil počet v korpusu nalezených praetur a kod, které se mohou vyskytovat na začátcích a koncích slov; 2) když 1. pravidlo neumožní vhodné dělení nebo jich umožní příliš mnoho, rozhodujeme se podle frekvence jednotlivých možností a upřednostníme tu častější, např.: *mužstvo*: 1. *mušstv.o* (čtyři konsonanty v kodě – nepřípustné); 2. *mušst.vo/mušs.tvo* by zase rozšířilo repertoár kod (*šs/šst* nejsou české kody); 3. *mu.šstvo* by rozšířilo repertoár praetur, a tak zbývá jen jediná možnost – 4. *muš.stvo* (Kučera & Monroe, 1968: 49). S podobnými faktory slabičného dělení pracuje i Zdena Palková (1994). Roli podle ní hraje zvukový kontrast sonority v řetězu hlásek, analogie se strukturou frekventovaných slabik, jejichž samostatnost je nesporná, a případná přítomnost morfologického švu. Henry Kučera se ve svém výkladu

pevně drží morfologického hlediska, nicméně tento přístup plně neodpovídá skutečnosti, protože slabiky se dělením na morfémy řídit nemusejí.

Ve vztahu slabiky a prefixu je nutné vzít v úvahu přítomnost či nepřítomnost tzv. rázu (*be.zot.klad.ný* vs. *bes.ʔot.klad.ný*). V mluvené češtině se také projevuje snaha vložit slabičný šev mezi konsonanty a i tam, kde je zřetelný morfologický šev (záporka *ne-* apod.). Pokud jsou tyto tendence v rozporu, dochází často ke kolísání.

Existuje několik dalších tendencí týkajících se struktury české slabiky: 1) v praetuře na začátku slova je často posledním segmentem před vokálem sonora nebo [r] (viz *ztlumit, střecha, vrána*); 2) časté je toto uspořádání dvousegmentální praetury: obstruent + sonora (*vlak, krása*), u třísegmentální praetury obsazení frikativa + exploziva + sonora; 3) tříčlenné praetury často začínají na hlásku [s]; 4) pokud se ve slabičné struktuře setkají dvě sonory, jde většinou o nazálu a likvidu (*mráz, mrak*); 5) nevyskytuje se kombinace [n] + obstruent; [n] nebo [ŋ] + další sonora nebo kombinace [r] + [ŋ] (Palková, 1994). Kombinace dlouhé praetury a nulové kody se všeobecně snaží slovanské jazyky vyvarovat (Kučera & Monroe, 1968: 45).

V českém mluveném projevu (nikoli lexikonu) jsou podle studie M. Ludvíkové, která ve svém výzkumu zpracovávala dva české mluvené projevy o celkové délce 10 000 slabik, nejčastějšími kombinacemi typy CV (60 %), CVC (17 %) a CCV (10 %). 87 % všech slabik v češtině je tvořeno slabikami těchto tří typů. Slabiky, jako jsou V (5 %), CCVC (4 %) a VC (2 %), pokryjí dalších 11%. Zbývají tedy 2 %, o která se dělí slabičné typy CVCC, CCCV, CCCCV, CCCVC, CCVCC, CVCCC a CCCVCC. Z výše uvedeného je patrné, že jen pět typů slabik je otevřených, kdežto zbylých osm je zavřených; nicméně toto ještě nemusí vypovídat nic o frekvenci jejich užití - zavřené slabiky sice mají více typů, ale v češtině tvoří pouze čtvrtinu ze všech slabik v mluveném projevu. (Ludvíková, 1987).

Na začátku slova se vyskytuje nejvíce slabičných typů: CV (*ne*), CCV (*pro, kte*), V (*a*), VC (*ob*), CVC (*jed*), CCCV (*spra*), CCCCV (*v-zdra*), CCCVCC (*zvlášť*); na střední pozici se může objevit sedm typů: CV (*va*), CVC (*jak*), CCV (*jme*), CCVC

(*sloup*), V (*o*), CCCVC (*střed*), VC (*ar*). Nejméně obsazena je koncová pozice; nalezneme zde jen šest typů: CV (*ho*), CVC (*sím*), CCV (*sto*), CCVC (*klad*), CVCC (*lost*) a CCCV (*kstu*). Pro konec slova jsou však charakteristické pouze dva typy – typ CV (*no*) a CVC (*sit*); tyto jsou v češtině nejběžnější (Ludvíková, 1987: 108).

Na základě práce Ludvíkové bylo zjištěno, že nejčastějšími slabikami ve zkoumaných českých mluvených projevech jsou tyto: /di, ní, to, a, ta, po, je, ňe, ja, na, vá, ta, ne, ka, no, se, pro, tu, ko, o/ (Ludvíková, 1987: 106). Těchto 20 slabik tvořilo 32 % všech slabik textu.

Pokud se zajímáme o četnost slabik v textu, a nikoli v lexikonu, je nutné do úvah zahrnout skutečnost, v jak frekventovaných slovech jsou tyto slabiky zastoupeny. Nejfrekventovanější uvedená slabika patří některým tvarům plurálu substantiv a také částicím typu *tedy* (což je záležitost zejména konkrétního mluveného projevu i mluvího); slabiku /ní/ nalezneme u adjektiv apod. Důležité nejsou jen koncovky, ale i frekventované kořeny. Existují také slabiky charakteristické pro začátek slova, jako např. /te, po, ne, pro, a/, pro jeho střed – /vá, no, ja, jme/ i pro jeho konec – /ní, ci, ho, tu, te/. Zde hrají roli zejména předpony, např. /po, ne, pro/, koncovky slov, frekvence spojek, zájmen a gramatických slov, opakování kořenů. Do těchto úvah je dále třeba zahrnout i další faktory, jako např. komunikační situaci, téma projevu, idiolektické zvyky apod.

K podobnému závěru jako Ludvíková ohledně nejčastějšího typu slabiky dospěli i Kučera a Monroe (1968), když se zabývali frekvencí slabičných typů, kterou zkoumali na jednoslabičných slovech z korpusu čítajícího 10 280 slov a 24 480 fonémů. Typ CV tedy i podle Kučery tvoří většinu ze všech slabik v korpusu, konkrétně cca 45 % slabik (Ludvíková však dochází k 60 % – oproti Kučerovi zahrnuje neslabičné předložky do následující slabiky, a ne do samostatné kategorie, a navíc vychází ze souvislého textu, zatímco Kučera ve svém výzkumu počítá pouze s jednoslabičnými slovy); typ CVC cca 18 %; V 13 %; nejméně pak CVCCC 0,01 % a VCC, který je zastoupen v pouhých 0,02 % případech.

Nejčastější praeturou je podle Kučery v češtině je typ C, který tvoří 63 % všech praetur; dalším obvyklým typem je CC (24 %); slabiky bez praetury tvoří 12 %; delší praetury jsou zřídka - CCC (2 %) a CCCC, která tvoří jen 0,013 %. Nejčastější kodou je koda nulová (73 %), následuje typ C s 26 %; typ CC tvoří jen 1,2 % a CCC 0,004 %. Obvyklé dělení českých slabik je následující: V.CV, VC.CV, VC.CCV, VCC.CCV, VCC.CCCV, VCCC.CCCV, VCCC.CCCCV.

Průměrná délka slabiky s přihlédnutím ke všem rozdílným slabikám je podle Kučery 3,18 fonému; pokud vezmeme v potaz celý slabičný soubor, se kterým autor pracoval, pak je délka slabiky 2,37 fonému, nicméně výsledky zkoumání frekvence užívání slabik poněkud zkresluje fakt, že do jednofonémových slabik autor započítává i neslabičné předložky a předponu /vz-/.

2.3.6 Fonotaktické poznatky v češtině

Kombinace hlásek v jazycích se netvoří náhodně; jejich vzájemné spojování podléhá určitým pravidlům, či spíše tendencím typickým pro konkrétní jazyk. Nyní si tedy přiblížíme určité tendence kombinování hlásek v češtině. Východiskem nám bude zejména část studie M. Ludvíkové (1987) a poznatky uvedené v publikaci *Statistiky češtiny* v oddíle *Fonologie a grafémika*.

2.3.6.1 Výzkum Jiřího Krause a Marie Ludvíkové (1966, 1987)

Podle práce Ludvíkové a Krause (1966) tvoří samohlásky 41,3 % českého textu; souhlásky pak 58,7 %. Lze konstatovat, že nejpoužívanějšími hláskami v češtině jsou krátké vokály (32,12 %), párové souhlásky neznělé (22,95 %) a nepárové souhlásky (22,35 %), o poznání méně frekventovanými se ukázaly párové znělé (13,4 %) a nejnižší frekvenci užívání měly dlouhé vokály (8,29 %) a diftong [ou] (0,89 %). Nejčastějším vokálem je [e]; Ludvíková tento fakt vysvětluje jeho přítomností v často užívaných slovech i v prefixech, pádových koncovkách apod. Nejfrekventovanější konsonant [t] zase tvoří zakončení infinitivu a je součástí mnoha zájmen. (Ludvíková, 1987)

Z počtu všech možných dvoučlenných kombinací (1296) se v rámci slova realizuje pouze 60 % - tato skutečnost souvisí s fyziologií i s jazykovou normou. Krátké vokály se mohou spojovat se všemi konsonanty, mohou stát na začátku i na konci slova, ale nespojují se s dlouhými vokály nebo diftongy. Právě dlouhé vokály a diftongy nevynikají velkou schopností kombinace. Největší kombinační využití mají likvidy a nazály, konkrétně tedy hlásky [m, r, l, n], z těchto má nejširší kombinační spektrum hláska [l] (66 dvojčlenných kombinací) – to tedy znamená, že takřka všechny hlásky mohou před ním a i před [m, n, r] předcházet i za nimi následovat (Ludvíková, 1987: 96). Likvidy a nazály se vyhýbají pouze kombinacím s palatálami, řídké jsou ale i některé kombinace s následujícími neznělými plozivami (*lpět, lkát*) a spojování s hláskami z téže skupiny (*rmoutit se*), což souvisí s všeobecnou snahou o eliminaci pobočných slabik, potažmo pak s principem vzrůstající sonority.

Nejfrekventovanějšími neznělými konsonanty jsou hlásky [t, s, k].; u znělých hlásek bychom nejvíce kombinací našli u hlásky [v]. Specifickými kombinačními vlastnostmi se vyznačují palatály: většina hlásek se může vyskytovat před nimi, za nimi však jen vokály přední řady a souhlásky [k, t, s, řš, j]. Podle Ludvíkové je spojení palatál a zadních samohlásek nečetné a používá se spíše v onomatopoických a emocionálně zabarvených výrazech – toto ovšem platí, bereme-li ohled na morfologické švy slov; v případě, že se vydáme přes jejich hranici, můžeme takovou kombinaci běžně zaznamenat např. u sloves třetí prezentní třídy vzoru *kupuje* a u desubstantivních adjektiv – např. *vyprazdňovat, ujišťovat, povodňový* apod.

Nejčastějšími kombinacemi hlásek v češtině jsou tyto dvojčlenné skupiny: /je, st, ne-, na, po, se, ní, ro, ňe, en, le, em, la, ov, li, to, ko, te, el, pr/ (Ludvíková & Kraus, 1966). I na této malé skupině nejfrekventovanějších kombinací můžeme vidět typické tendence spojování segmentů: CV tvoří 70 % kombinací, VC 20 % a CC 10 % možných kombinací hlásek v češtině. Nejčastějšími kombinacemi souhlásek jsou shluky dvojčlenné (např. Mazlová, 1946; Churaňová, 2010).

Dalším aspektem fonotaktické struktury češtiny je skutečnost, že slova obvykle začínají konsonanty. Jen malá skupina českých slov (12 %) začíná samohláskou; pokud se takový případ vyskytne, pak bude tímto iniciálním vokálem pravděpodobně [a]. (Za konsonant zde nepovažujeme ráz, který se objevuje před iniciálním vokálem českých slov nebo jejich kořených morfémů – ráz v češtině nemá fonologický status.) Konec slova se zpravidla vyznačuje výskytem samohlásky, což souvisí zejména s rozvinutou flexí v češtině. Těmito samohláskami jsou zejména hlásky [e, i, a, i:, o], které představují 56 % všech možných slovních zakončení. V českém textu asi 71 % slov končí samohláskou; zbývajících 29 % končí souhláskou (Ludvíková, 1987: 102).

2.3.6.2 Statistika češtiny (Bartoň et al., 2009)

Publikace *Statistika češtiny* je založena na kvantitativní analýze korpusového materiálu psaného textu, konkrétně se jedná o korpus SYN2005, který obsahuje přibližně 100 milionů textových slov bez interpunkce. Je rozdělen do tří srovnatelně velkých subkorpusů – beletrie, odborná literatura a publicistické texty. Autoři *Statistik češtiny* přihlížejí postupně k různým oblastem jazyka – fonologii, grafémice, morfematicke, morfologii, slovním druhům, slovnědruhovým syntagmatům a nakonec se věnují porovnání dílčích subkorpusů a stylometrii. Data byla zpracovávána zejména softwarovými nástroji automatické analýzy, ale v některých případech se muselo přistoupit k manuálním opravám či k ručnímu zpracování.

Pro naše účely je z této publikace důležitá část zabývající se fonologií. Zde autoři kvantifikují distribuci jednotlivých hlásek, dvojic a trojic hlásek v rámci výskytů slovních druhů v korpusu (*slova-tokeny*) a v rámci lexikonu (*slova-typy* - slovní tvary, které se v korpusech objevily). Psaný text se nejprve převedl na svou přibližnou fonetickou podobu (tzn. brala se v potaz např. i asimilace znělosti přes hranici slov apod.) automatickou transkripční procedurou. Tato transkripce reflektovala standardní inventář českých hlásek; počítala tedy s fonémy a jejich kombinatorními variantami (dlužno však podotknout, že se nekalkulovalo se samostatným segmentem [eu], ačkoli se tento vyskytuje ve slovní zásobě obsahující přejatá slova).

S fakultativními variantami a nesystematickými alternacemi se v přepisu nepočítalo. Autoři dále nevylučují, že v důsledku použití automatické metody mohlo docházet k odchylkám od předpokládané výslovnosti, a to např. u slov cizího původu s výslovností, která není typická pro češtinu – výskyt takových slov je ale v korpusu minimální. Tyto odchylky jsou pro naše účely nepodstatné, jelikož hlavním účelem této práce je pokus o vysledování fonotaktických vzorců slov a taktů, přičemž bereme v úvahu pouze rozlišení C/V. Je ale nutné zmínit, že analýza korpusu SYN2005 byla prováděna výhradně na psaných textech. Pokud se autoři zabývali fonetickými takty, šlo pouze o jejich velmi hrubé rozlišení podle jednoznačných pravidel, která ale v reálné řeči nejsou zdaleka tak bezvýhradná. V naší analýze se také budeme věnovat fonotaktice slova, takže budeme mít mimo jiné možnost uvést v souvislost výsledky ze *Statistik*, týkající se písemného projevu, s daty z projevu mluveného, a také poskytneme rozbor fonotaktických specifik reálného mluvního taktu, jenž je ve *Statistikách češtiny* zpracován silně zjednodušeně.

Ze všech těchto údajů nás pro účely této práce budou nejvíce zajímat poměry konsonantů a vokálů v rámci slovních druhů a pododdíl, který se zabývá strukturou slova. Stručně se ale zmíníme i o frekvenci konkrétních hlásek a skupin hlásek, jež jsou do přehledů zapracovány.

Za důležité považuji před excerpací výsledků statistik zmínit poměr slovních druhů v rámci korpusu (měřeno podle všech výskytů): Největší podíl v korpusu mají substantiva – 30,2 %. Následují slovesa se 17,5 %, adjektiva s 11,7 %, zájmena s 10,9 % a prepozice s 10,7 %. Spojky zastávají 7,8 % všech slov v korpusu a adverbia 6,9 %. Velmi malý podíl v korpusu mají potom číslovky (2,6 %), částice (1,6 %) a citoslovce (0,05 %). Mluvené texty, ze kterých vychází naše práce, jsou zařaditelné pod publicistický styl – pokud se podíváme na subkorpus PUB korpusu SYN2005, zjistíme, že procentuální hodnoty víceméně odpovídají zastoupení slovních druhů v celém korpusu.

Konsonantů celkem bylo v korpusu SYN2005 303 266 416; vokálů 214 749 022. V rámci lexikonu se nacházelo celkem 6 975 620 konsonantů a 4 998 939 vokálů.

2.3.6.2.1 Kombinace C a V ve slovech celkem

V textech měl z **dvousegmentových** kombinací celkem největší zastoupení typ CV (cca 198 milionů výskytů, např. /je/). Nejčastěji se vyskytoval u substantiv, významně zastoupená je tato kombinace i u ostatních slovních druhů. Tento fakt není překvapivý – čeština inklinuje k otevřeným slabikám, takže poměrně pravidelné střídání konsonantů a vokálů je přirozené. Frekventovanými typy se ukázaly být i VC a CC (143 a 74 milionů, např. /ov/, /st/); typ VV (např. /eu/) je méně obvyklý – byl silně zastoupen pouze u substantiv, adjektiv a sloves.

Co do počtu výskytů je v celém korpusu nejfrekventovanější dvojicí spojení /je/, což potvrzuje nález Ludvíkové (viz výše). V lexikonu skupina /je/ ovšem není tolik častá, protože velkou část jejích výskytů pokrýval tvar slovesa *být*.

Co se týče rozložení dvojic hlásek CV/VC/CC/VV v lexikonu, význačně se zde projevil rozdíl četnosti druhů flektivních a neflektivních – flektivní slovní druhy mají díky většímu počtu tvarů zastoupení všech dvojic vyšší než neflektivní; flektivní slovní druhy také obecně bývají delší. Spojení CV je opět nejčetnější; častý výskyt vykazuje u adjektiv (zde je tato dvojice nejfrekventovanější, což patrně souvisí s rozsáhlostí adjektivních paradigmat i s faktem, že v koncovkách přídavných jmen se obvykle nalézá dlouhý vokál), substantiv a sloves; poměrně málo se vyskytuje u zájmen a číslovek; z neflektivních druhů je více zastoupen jen u adverbií. Typ VC je příznačný více pro adjektiva než pro substantiva. Typ CC se vyskytuje zejména u podstatných a přídavných jmen a sloves; méně u příslovcí. Velmi malé zastoupení měl v korpusu u zájmen a neflektivních slovních druhů, zejména pak u konjunkcí. Typ VV se u zájmen a neflektivních slovních druhů mimo adverbia prakticky nevyskytuje.

Nejfrekventovanější **trojice** hlásek co do všech výskytů bývají typu CVC (necelých 127 milionů výskytů, např. /ter/), dále VCV (např. /ova/), CCV (např. /vje/) a VCC (např. /ost/) – u těchto typů bylo zaznamenáno 41-76 mil. výskytů; výrazně méně výskytů bylo v korpusu trojic CCC (např. /str/), CVV (např. /neo/), VVC (např. /eut/) a nejméně zastoupen byl typ VVV, který

u zájmen, sloves, předložek, spojek a částic vykazoval nulový výskyt. Dlužno podotknout, že typ VVV v české domácí slovní zásobě nefiguruje; jeho výskyt v korpusu mohl být tedy způsoben nesystematičností v chápání diftongu /eu/ jakožto dvou segmentů a přítomností cizích vlastních jmen a zkratek. Nelze také vyloučit možnost chyby ve zpracování materiálu.

Pořadí trojic hlásek je u slov-typů (viz výše) identické. Nízké nebo nulové zastoupení všech typů (zvláště pak CVV, VVC, VVV) bylo zaznamenáno u neflektivních slovních druhů (kromě adverbí) a zájmen; typ VVV se také vůbec nevyskytoval u sloves.

2.3.6.2.2 Kombinace C a V v iniciální pozici

V **iniciální** pozici ve slovech v celém textu korpusu se vyskytuje víc konsonantů (86 milionů) než vokálů (přes 13 milionů). Pokud vezmeme v úvahu pouze lexikon, je rozdíl ještě patrnější: konsonanty vykazují 1,2 milionu výskytů na začátku slova, kdežto vokály jen cca 175 000 výskytů.

Nejfrekventovanějším typem **dvojice** hlásek v textu je opět kombinace CV s bezmála 60 miliony výskytů (např. /po/); hojně zastoupená byla ve všech slovních druzích. Totéž je možné prohlásit i o dvojici CC (22,8 mil. výskytů, např. /pr/) a VC (8 mil. výskytů, např. /od/). Dvojice VV byla silněji zastoupena pouze u substantiv a adjektiv (patrně způsobeno přítomností slov začínajících diftongem /eu/ – euforický, eutanazie – nebo jiných případů – eon apod.); vzácný výskyt byl zaznamenán v případě citoslovcí, adverbí a sloves. U ostatních slovních druhů se tato dvojice nevyskytovala v iniciální pozici vůbec.

V rámci lexikonu byla situace co do poměrů podobná – nejsilněji zastoupenou skupinou byl typ CV, který se vyskytoval na začátku všech slovních druhů, nejčastěji však u substantiv, sloves a adjektiv. O něco méně byly zastoupeny skupiny CC a VC. Nízkou frekvenci vykazovala dvojice VV, která se více uplatnila jen u substantiv a adjektiv; u jiných slovních druhů se vyskytovala minimálně nebo vůbec.

Co se týče **trojic** hlásek na začátku slova ve slovech-tokenech, můžeme si povšimnout vyšší variability co do slovních druhů – jsou slovní druhy, ve kterých jsou jisté trojice zastoupeny silně; v jiných mají výskyt nulový. Např. nejfrekventovanější trojice /pro/ je u substantiv, sloves, spojek, adjektiv, adverbíí a prepozic poměrně četná, ale u ostatních slovních druhů vykazuje velmi nízký nebo nulový výskyt. U slov-typů je situace co do poměrů obdobná. Opět je zde nejfrekventovanější trojicí hlásek /pro/.

Zastoupení jednotlivých typů trojic hlásek na začátku slova v rámci všech výskytů v korpusu SYN2005 je následující: Nejčastějším typem je CVC (40,6 mil. výskytů, např. /nep/), který je zastoupen u všech slovních druhů, stejně jako typ CCV (17,3 mil. výskytů, např. /pro/). V případě typu VCC (např. /atr/) je ale situace poněkud odlišná – je sice poměrně častý (necelé čtyři miliony výskytů), ale u citoslovcí se nachází minimálně (272 výskytů) a u zájmen je výskyt nulový. Typ VCV (např. /oko/) je sice méně častý (cca tři miliony výskytů), ale zastoupení má u všech slovních druhů. Typu CCC (např. /vzd/) opět vykazuje nulový výskyt u zájmen, typ CVV (např. /neu/) u zájmen, předložek, spojek a citoslovcí (přičemž výskyt u číslovek je taktéž zanedbatelný), typ VVC (např. /eut/) má silnější zastoupení pouze u substantiv a adjektiv; u adverbíí a sloves byl zaznamenán výskyt minimální a u ostatních slovních druhů nulový. Typ VVV nemá silné zastoupení u žádného slovního druhu – v korpusu se vyskytoval jen u substantiv, adjektiv a citoslovcí; jeho výskyt byl patrně způsoben přítomností cizích slov, zkratk a pojetím diftongu /eu/ jakožto dvou segmentů.

Pokud se jedná o trojice hlásek na začátcích slov-typů, je opět nejčastějším typem CVC s největším zastoupením u substantiv, adjektiv a sloves a adverbíí; podobné zastoupení nalezneme u typu CCV; u typu VCC je nulový výskyt u zájmen a velmi nízký u neflektivních slovních druhů mimo adverbia. Typ CCC je v lexikonu (oproti všem výskytům v korpusu) méně frekventovaný než typ CVV; ten je silněji zastoupen jen u substantiv, adjektiv a sloves; méně se vyskytuje u adverbíí. U ostatních slovních druhů se prakticky neobjevuje. Typ CCC má nulový výskyt pouze u zájmen; u číslovek a neflektivních slovních druhů je jeho výskyt minimální. Typ VVC nalezneme

zejména u substantiv a adjektiv; méně u sloves a adverbíí. Ostatní slovní druhy tento typ nezahrnují. Typ VVV byl opět nejméně četným.

2.3.6.2.3 Kombinace C a V ve finální pozici

Nejvíce **dvojic** hlásek na **konci** slova v korpusu je typu CV (necelé 64 miliony; velké zastoupení u všech slovních druhů, např. /ní/); velmi častá je i kombinace VC (necelé 24 miliony, např. /em/); o něco nižší zastoupení má skupina CC (necelé 3 miliony, např. /st/); skupina VV (cca 695 000 výskytů, např. /ie/ – epentetická hláska se patrně v korpusu nebrala v potaz) je obvyklá pouze u substantiv, méně se vyskytuje u adjektiv, citoslovcí a prepozic; u ostatních slovních druhů se výskyt tohoto typu kombinace pohybuje v řádu desítek nebo je nulový.

U slov-typů je co do pořadí jednotlivých typů podle jejich frekvence situace podobná.

Nejčastějším typem **trojice** hlásek podle výskytů v celém korpusu je typ VCV (přes 31 milionů výskytů, např. /eňí/), následuje typ CVC (přes 17 mil. výskytů, např. /ním/), CCV (cca 16,5 mil. výskytů, např. /st'i/) a VCC (přes dva miliony výskytů, např. /ost/). Všechny tyto typy jsou zastoupeny u všech slovních druhů. Výrazně méně je potom typů CVV (silně zastoupen je tento typ jen u substantiv; u ostatních slovních druhů je jeho výskyt minimální nebo nulový; např. /die/), VVC (nejvyšší zastoupení u substantiv, menší u adjektiv, zanedbatelný u ostatních slovních druhů; např. /uet/), CCC (zastoupení nejsilnější u sloves; zanedbatelný výskyt u všech slovních druhů kromě zájmen a předložek, kde je nulový; např. /mst/) a VVV (nenulový výskyt pouze u substantiv, adjektiv a citoslovcí), kterážto skupina je jako ve výše uvedených případech poněkud sporná.

Pokud bychom zohlednili pouze slova-typy, pořadí jednotlivých kombinací bude stejné. Většina výskytů první skupiny je u adjektiv, substantiv nebo sloves, méně pak u adverbíí, číslovek a zájmen. V řádu desítek se pohybuje výskyt tohoto typu u ostatních neflektivních slovních druhů. Toto je potom možné říct i o skupinách CVC a CCV. Skupina VCC je silněji zastoupena pouze u substantiv, méně u sloves a adjektiv. Ostatní slovní druhy mají velmi nízkou

četnost. U skupiny CVV a VVC je větší podíl pouze u substantiv a adjektiv; u ostatních slovních druhů je výskyt zanedbatelný nebo nulový. U skupiny CCC je významnější podíl pouze u substantiv a sloves; u ostatních slovních druhů je výskyt minimální nebo nulový. Nejméně zastoupena je opět skupina VVV (pouze substantiva a adjektiva).

2.3.6.2.4 Průměrná délka slova a taktu

Počet všech slov-typů v korpusu je 1 351 065, slov-tokenů 99 846 195. Celkový počet hlásek u slov-typů je 12 171 974, u slov-tokenů 531 760 850. Průměrná délka slov-typů je 9,01 hlásky, slov-tokenů 5,33 hlásky – zde je tedy vidět poměrně velký rozdíl, pokud se jedná pouze o lexikon, nebo o veškerý materiál korpusu SYN2005. Tato skutečnost je dána faktem, že frekventovaná slova obecně bývají kratší.

V následující tabulce uvádíme průměrnou délku slov různých slovních druhů (měřenou počtem hlásek) v korpusu SYN2005 a v jeho části publicistických textů:

slovní druh	průměrná délka u slov-typů (hlásky) SYN2005	průměrná délka u slov-tokenů (hlásky) SYN2005	průměrná délka u slov-typů (hlásky) PUB	průměrná délka u slov-tokenů (hlásky) PUB
substantiva	8,31	6,50	8,06	6,58
adjektiva	10,48	8,07	10,28	8,16
zájmena	6,36	3,15	5,98	3,23
číslovky	11,31	5,45	11,16	5,47
slovesa	8,78	5,91	8,67	5,97
příslovce	9,54	5,17	9,13	5,19
předložky	4,62	1,97	4,35	1,98
spojky	4,93	2,89	4,96	2,87
částice	5,74	4,91	5,74	5,22
citoslovce	4,86	3,86	4,45	3,71
neurčené	8,62	6,95	8,15	6,66

Tabulka 2: Průměrná délka slov různých slovních druhů podle hlásek – typy a výskyty v celém korpusu SYN2005 a v jeho oddílu publicistických textů (Bartoň et al., 2009: 106-107).

Ve *Statistikách češtiny* se objevuje snaha věnovat pozornost i vyšší prozodické jednotce než hlásce nebo slabice, a to fonetickému taktu, konkrétně jeho délce. Zpracování mluvního taktu v kvantitativní analýze zahrnovalo shlukování slov do vyšších jednotek. Seskupování se týkalo pouze

spojení s primárními (slabičnými i neslabičnými) předložkami a příklonkami. Z toho vyplývá, že nutně muselo docházet k částečnému zkreslení výsledků (např. kvůli nereflexivní předložce nebo seskupování plnovýznamových jednoslabičných slov).

Počet takových taktů-typů byl tedy v celém korpusu 2 035 359, tedy větší než slov-typů. Toto lze vysvětlit přítomností předložek a příklonek v taktu (když se stejná předložka připojila k různým slovům do jednoho taktu, vznikly tak nové typy taktů). Taktů-tokenů bylo 93 161 553 – z rozdílu mezi počtem všech výskytů slov a tímto údajem můžeme vyčíst množství předložek a příklonek, které nebyly považovány za samostatné takty. Počet hlásek v taktech-typech byl 19 146 700 a taktech-tokenech 531 754 285. Průměrná délka taktu-typu byla 9,41 hlásky, taktu-tokenu 5,71 hlásky. I zde se tedy projevil rozdíl mezi celkovým materiálem a lexikonem korpusu v závislosti na omezené délce frekventovaných výrazů.

2.3.6.2.5 Struktura slova

Nyní si uvedeme několik základních typů kombinací konsonantů, vokálů a tzv. obojetných hlásek (tímto termínem autoři patrně označovali slabikotvorné likvidy; značeny jako O), které tvoří slova v korpusu (počítáme se všemi výskytů v SYN2005). Tyto údaje mohou být někdy mírně zkresleny výskytem zkratk, které byly interpretovány jako ten slovní druh, který zastupovaly. V této části práce zmíním deset nejobvyklejších kombinací a dále se budu věnovat některým specifickým případům.

Nejfrekventovanější kombinací hlásek, která tvoří samostatné slovo, je typ CV (necelých 17 milionů výskytů; nejsilnější zastoupení u zájmen, předložek a konjunkcí, např. *na*); nulový výskyt byl zaznamenán pouze u číslovek. Další frekventovanou kombinací je CVCV (cca 8,7 mil. výskytů, např. *vana*); silně zastoupena je tato skupina zejména u substantiv, sloves, spojek a zájmen; nulový výskyt v tomto případě není v korpusu u žádného slovního druhu. Následuje skupina CVC (cca 5,7 mil. výskytů, např. *můj*), která se vyskytuje v případě všech slovních druhů; nejvíce pak u adverbíí a zájmen. Skupina CVCCV (necelých 5 mil. výskytů, např. *podle*) je taktéž schopna tvořit všechny

slovní druhy; nejvíce se pak objevuje u substantiv, adjektiv a adverbíí. Typ CVCVCV (necelých 4,5 mil. výskytů, např. *nedělá*) je nejsilněji zastoupen u substantiv a sloves, typ CCVCV (přes 4 mil. výskytů, např. *stojí*) u substantiv a zájmen. V pořadí sedmý je typ C (přes 4 mil. výskytů, např. *k*); prakticky všechny jeho výskyty jsou v rámci předložek (3,9 milionu výskytů). U ostatních slovních druhů je podíl výrazně nižší (substantiva, adjektiva, zájmena) nebo zanedbatelný (či nulový u číslovek). Kombinace CVCVC (3,7 mil. výskytů, např. *došel*) je typická zejména pro substantiva a slovesa; typ CVCCVCV (2,8 mil. výskytů, např. *dojdeme*) také nejčastěji tvoří substantiva a slovesa. Desátým typem v pořadí je typ CCVC (cca 2,5 mil. výskytů, např. *proč*), pro který platí, že se nejčastěji vyskytuje v rámci sloves, substantiv a konjunkcí.

Dalšími velmi frekventovanými (1 mil. výskytů v korpusu a více) typy kombinací hlásek tvořících samostatná slova jsou např. CCV (např. *pro*), CVCVCCV (např. *nedojde*), CCVCCV (např. *stavme*), CVCVCVCV (s nulovým výskytem u konjunkcí, např. *rachotina*), CVCCVC (např. *tančit*), CCVCVCV (nulový výskyt u částic a zanedbatelný u předložek; např. *druhový*), CVCVCVC (nulový výskyt u předložek, zanedbatelný u interjekcí; např. *dodávat*), CVCCVCCV (nulový výskyt u konjunkcí, velmi nízký u zájmen a interjekcí; např. *poštačka*), CCVCVC (např. *prodej*) a CVCCVCVCV (nulový nebo zanedbatelný výskyt u neflektivních slovních druhů kromě adverbíí a u zájmen; např. *zaprodaný*).

V prvním stu nejčastějších kombinací konsonantů, vokálů a slabikotvorných hlásek nenajdeme ani jeden vzorec, který by byl typický právě pro jeden slovní druh. Můžeme obecně konstatovat, že všechny kombinace z první stovky nejfrekventovanějších jsou zastoupeny u substantiv a většina i u adjektiv (až na typ CVCCVCO, který je jinak silně zastoupen jen u sloves, a to patrně kvůli l-ovým participiím, např. *nesvedl*; méně u substantiv; typ CVCVCO má u adjektiv právě jeden výskyt; jinak je silně zastoupen jen u sloves, např. *dořekl*; méně u substantiv); u sloves nejsou zastoupeny pouze typy CVCVV (ten se častěji vyskytuje pouze u substantiv, např. *režie*; zanedbatelný podíl výskytů mají adjektiva a interjekce), CVCVCVV (častěji jen

u substantiv, např. *melodie*; u adjektiv a adverbíí zanedbatelně) a CVCVVC (četnější jen u substantiv, např. *manuál*; u adjektiv a konjunkcí výskyt zanedbatelný).

Z tabulky na str. 108-109 můžeme vyvodit nejčastější typy u každého slovního druhu podle výskytu v korpusu:

substantiva: CVCV, CVCVCV, CVCCV, CVCVC, CCVCV

adjektiva: CVCCV, CVCVCCV, CVCCVCCV, CVCVCV, CVCV

zájmena: CV, CVC, CVCV, CCVCV, CCV

číslovky: CCV, CVCCV, CVCVC, CCVCV, COCCV

slovesa: CVCV, CV, CVCVCV, CCVC, CVCVC

příslovce: CVC, CVCCV, CVCV, CCVCV, CCVCCV

předložky: CV, C, CCV, CVC, CCVC

spojky: CV, CVCV, CCVC, CVC, CVCVC

částice: CVC, CVCV, CV, CVCVC, CVCCVC

citoslovce: CVC, CV, CVCV, CCVCVCV, CVCCV

Jisté procento ze všech vzorců také tvořily nerozpoznané slovní druhy. Jejich statistickými údaji se zde však nezabýváme.

Dále je důležité konstatovat, že v prvním stu nejfrekventovanějších kombinací se slabikotvorné sonory vyskytují toliko v kombinacích CVCO (např. *řekl*), COCCV (např. *srnčí*), COCV (např. *vrní*), CVCVCO (např. *netekl*), COCVCV (např. *mrkání*), CVCCVCO (např. *podlehl*) a COCVC (např. *mrzák*) a ani jeden z těchto vzorců nedosáhl v korpusu na milion výskytů.

Další oddíl týkající se fonologické kvantitativní analýzy se věnuje počtům slabik v rámci jednotlivých slovních druhů. Slabikou jsme se již zabývali v předchozí kapitole, proto uvádím jen některá stručná fakta:

Oproti analýze všech výskytů slovních druhů byla u analýzy přihlížející pouze ke slovním tvarům zjištěna tendence k posunu k víceslabičným jednotkám – můžeme z tohoto faktu tady vyvodit závěr, že velký výskyt kratších jednotek v korpusu je dán tím, že se často opakují stále stejná frekventovaná slova, ale pokud se berou v potaz různé jednotky, je podíl co do slabičnosti více vyvážený a také méně výrazně preferuje krátké jednotky (k témuž závěru jsme dospěli také v případě, že se jednalo o průměrnou délku slova nebo taktu vymezenou počtem hlásek).

Celkem bylo rozpoznáných slovních jednotek 96,4 miliony; největší podíl měla slova dvouslabičná (30,96 %). Následovala slova jednoslabičná (29,1 %), trojslabičná (21,93 %), čtyřslabičná (10,23 %), neslabičná (4,1 %) a pětislabičná (2,93 %). Slova delší než pětislabičná měla v korpusu podíl nižší než jedno procento.

Celkem rozpoznáných slov-typů bylo 881 013. Největší podíl měla slova čtyřslabičná (32,08 %) a slova trojslabičná (29,74 %). Následovaly jednotky pětislabičné (17,21 %), dvouslabičné (10,97 %), šestislabičné (6,29 %) a sedmislabičné (1,94 %). Slova neslabičná nebo delší než sedmislabičná měla v korpusu podíl nižší než jedno procento (Bartoň et al., 2009: 110-114). Z těchto údajů tedy můžeme vyčíst, že vysoký výskyt jednoslabičných slov v korpusu je patrně založen na opakování stejných slov, zejména předložek a spojek.

Výstupem tohoto zkoumání by měl být nástin fonotaktické mapy českých slov a skutečně realizovaných taktů. V závěru práce se tedy pokusíme srovnat naše výsledky s těmi, ke kterým došli autoři *Statistik češtiny*, a zjistit korespondence, nebo naopak možné vzniknuvší rozpory mezi těmito analýzami.

2.4 Shrnutí

V těchto kapitolách jsme podali přehled základních poznatků o rytmu řeči, jednotkách popisu, které pomáhají k jeho uchopení, a dále jsme se věnovali fonotaktickým aspektům v jazycích obecně i v konkrétně v češtině. V rámci teoretické části této práce jsme také zmínili známé přístupy k rytmu a k CVCV

problematice. Tím jsme vymezili tematický okruh, do kterého tento projekt spadá, a zasadili jsme náš výzkum do teoretického kontextu. Zároveň můžeme konstatovat, že jsme si díky prostudování současných postupů analýz prozodie a fonotaktiky češtiny vědomi možných úskalí, na něž bychom během zpracování materiálu mohli narazit. Nyní tedy přejdeme k praktické části výzkumu.

3. Metoda

3.1 Materiál

Využitým materiálem byly mluvené projevy rodilých mluvčích češtiny, profesionálních hlasatelů Českého rozhlasu. Jednalo se o vybrané čtené texty běžných zpravodajských relací. Texty čítaly celkem 6641 slov. Šlo o analogové nahrávky z FM, které byly posléze digitalizovány na frekvenci 22 500 Hz a později byly zpracovány v programu Praat (Boersma & Weenink, 2010). Pocházely z Pražského fonetického korpusu (Skarnitzl, 2010).

Analyzovali jsme projevy dvanácti školených mluvčích, pěti mužů a sedmi žen. Jednotlivá zpravodajství se pohybovala v rozsahu 413-764 slov. Vzhledem k profesionalitě a vzdělání mluvčích jsme předpokládali, že projev bude kultivovaný a bez větších hláskových nebo prozodických deformací. Jazyk všech nahrávek byl spisovný a neobsahoval nářeční prvky.

Následující tabulka 3 obsahuje informace o nahrávkách jednotlivých mluvčích:

mluvčí	NÚ	počet slov	počet taktů	trvání (min)
MSA03	45	460	379	3,576
CMA01	54	479	390	3,665
GVA05	63	620	503	4,523
JPC04	47	548	449	3,854
ODC05	65	690	564	4,556
SSA02	55	496	395	3,29
VOZ01	51	620	519	4,184
ULR05	60	527	415	3,761
PRK01	35	425	326	2,645
KOB02	60	597	487	4,088
STR04	70	764	600	4,546
WSS05	52	413	341	2,882
celkem	657	6639	5368	45,57

Tabulka 3: Nahrávky jednotlivých mluvčích: počet nádechových úseků (NÚ), počet slov, počet taktů, trvání v minutách.

3.2 Postup

3.2.1 Příprava materiálu

Všechny nahrávky byly rozděleny do samostatných souborů podle nádechových úseků; k nim byly vytvořeny kolekce popisných objektů (dále textgridů) k anotaci, očíslovaných podle odstavců v textu. Všech nádechových úseků bylo 657. Text byl následně pomocí souboru programů Prague Labeller (Pollák, Volín & Skarnitzl, 2007) rozdělen na hlásky a slova. Převedení textu na hlásky se snažilo respektovat ortoepická pravidla, avšak vyskytovalo se zde určité množství nepřesností a chyb, a to zejména v oblasti asimilací znělosti, přepisu zkratk a cizích slov. Ani hranice hlásek nebyly stanoveny se zvláštní přesností. Chyby v asimilacích jsme nijak neupravovali, neboť pro další zpracování nebylo podstatné, zda je hláska znělá, nebo neznělá, nýbrž jen skutečnost, zda jde o konsonant nebo vokál – v tomto hrubém rozlišení pak obvykle k chybám nedocházelo. Oproti tomu nedostatky v převodu cizích slov nebo vlastních jmen a zkratk na hlásky korigovány byly – v takových případech nemusely souhlasit počet ani podoba hlásek. Přesnému umístění hranic segmentů nebylo potřeba se příliš věnovat, neboť pro naše účely byla důležitá přítomnost hlásky (potažmo její hranice), nikoli však pozice její hranice nebo její trvání.

Jak již bylo zmíněno v úvodu práce, data s rázem, který v češtině není považován za fonologický prvek, jsme brali v potaz odděleně od ostatních dat. V rámci jeho zohlednění jsme brali v potaz fakt, že slova, jejichž první hláska je vokál, obvykle ve standardní podobě začínají konsonantickým prvkem, konkrétně glotální plozivou – rázem. Kvůli jeho problematické fonologické identitě jsme ho však neznačili C, ale ?. Je však nutné podotknout, že ačkoli ráz není systémovou jednotkou, hraje jistou roli v rovině prozodické. Např. v následné segmentaci na reálné takty nám často pomohla realizace rázu v rozhodování, zda jde o přízvučnou, nebo nepřízvučnou slabiku (viz níže).

Pomocí dalšího příkazového souboru pro Praat jsme přidali do textgridů novou vrstvu intervalů pro takty a vrstvu pro následné převedení hlásek na značky konsonantů C a vokálů V.

Tímto postupem tedy bylo dokončeno první, „mechanické“ stadium: příkazové soubory pro Praat zajistily rozdělení nahrávek na nádechové úseky a přidání dalších vrstev do textgridu; výše uvedený softwarový nástroj Prague Labeller (Pollák, Volín & Skarnitzl, 2007) rozdělil text na hlásky a slova. V této fázi bylo naším úkolem zkontrolovat korespondence mezi textgridy a zvuky a případné nedostatky opravit.

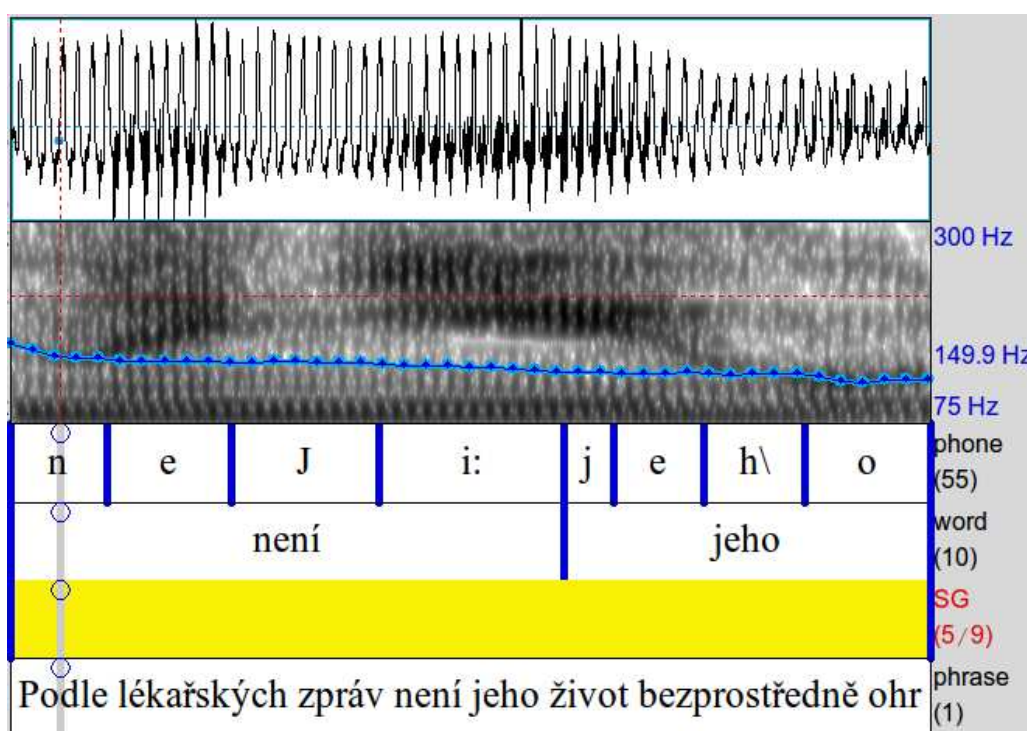
3.2.2 Zpracování materiálu

Jelikož nás zajímala problematika CVCV struktur nejen v rámci lexikálních jednotek, ale i mluvních taktů, bylo nutné provést segmentaci zvuku na reálné takty. Toho jsme docílili tak, že jsme si každý nádechový úsek pečlivě poslechli a do speciální vrstvy v textgridu vložili značku taktové hranice podle toho, jak byl takt vnímán v naší percepci. Dlužno podotknout, že toto rozdělování řeči na reálné takty se neobešlo bez přítomnosti problematických jevů, které byly často od ideálního popisu poměrně dalece vzdáleny.

Při určování hranic taktů jsme se rozhodovali podle několika faktorů. Vedle jasných akustických a percepčních znaků předělu, jako jsou např. pauza, hezitace, změny tempa nebo ukončující melodém, jsme brali ohled i na změny průběhu f_0 a změny dynamiky; v některých případech jsme přihlíželi i k tomu, zda jádro sporné slabiky je realizováno plně, nebo je jeho výslovnost poněkud redukována. U slov začínajících samohláskou jsme se také rozhodovali podle přítomnosti a typu realizace rázu a jeho trvání. Konečný verdikt o přítomnosti, popř. nepřítomnosti hranice byl založen na konsenzu dvou názorů školených fonetiků. Při určování hranic jsme měli jednu značku pro všechny hranice; nezabývali jsme se tedy tím, jak silné ony předěly byly – náš zájem byl omezen na jejich pouhou přítomnost, či absenci. Pro účely této práce nebyla totiž zřetelnost předělů zásadní, proto jsme využili jednodušší řešení.

V našem materiálu se pravidelně objevovalo několik jevů souvisejících se segmentací řečového signálu, které stály na okraji tradičního popisu spisovné mluvené češtiny nebo se mu takřka vymykaly. Nyní tyto případy stručně představíme.

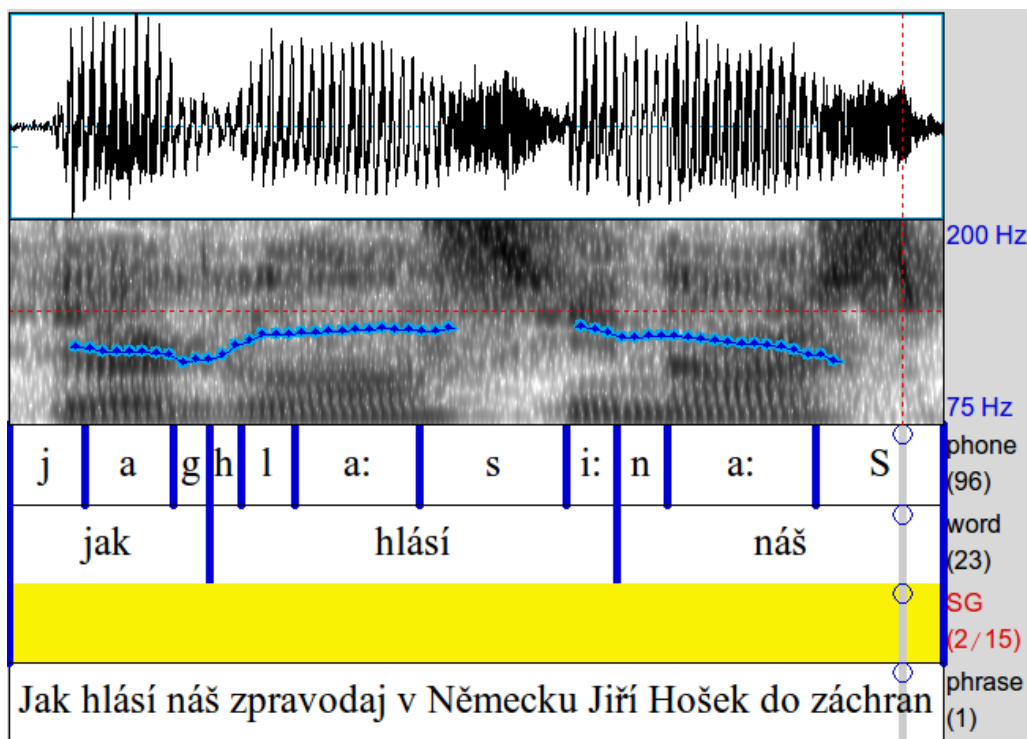
Z méně obvyklých situací, jež se v našem korpusu objevily, bylo několik případů realizace dvou víceslabičných slov (zpravidla se jednalo o dvě dvouslabičná slova, v materiálu se ale vyskytovaly i případy kombinace tříslabičného a dvouslabičného slova), která společně tvořila jeden takt (obr. 1).



Obr. 1 Zobrazení taktu z nádechového úseku MSA03-06-02 v editačním okně programu Praat. Nejvyšší vrstva obsahuje oscilogram, druhá spektrogram zvuku; nejvyšší anotační vrstva je vrstvou hlásek, druhá je vrstvou slov, ve třetí jsme určovali hranice mluvních taktů a ve čtvrté je kompletní text nádechového úseku. V tomto konkrétním taktu jde o méně obvyklé členění – jsou zde spojena dvě dvouslabičná slova v jeden takt. Modrá linie ve spektrogramu zobrazuje průběh základní frekvence, který je zde velmi jednoduší a neobsahuje žádný skok.

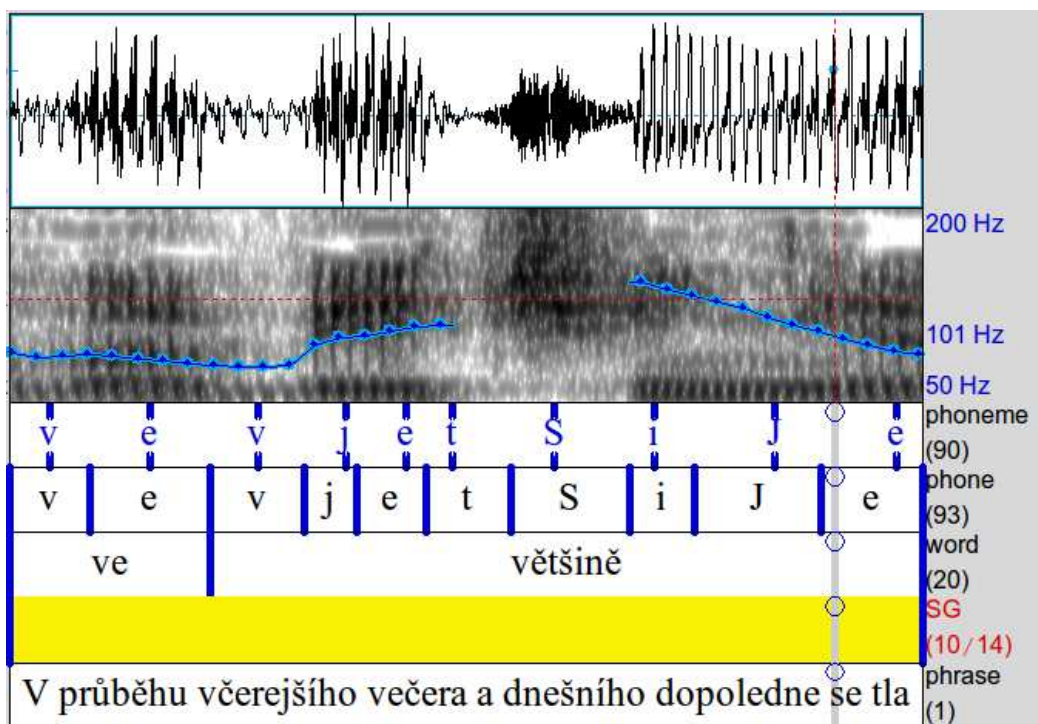
Další nejasnosti v segmentaci řeči vznikaly při výskytu několika jednoslabičných – často neplnovýznamových a frekventovaných – slov, která následovala za sebou. V těchto situacích nebylo vždy zcela zřejmé, kolik daná slova tvoří dohromady taktů; někdy docházelo i k případům, kdy slovo sestávající z jedné slabiky bylo samostatným taktem. Ve zkoumaném

materiálu se také vyskytovaly kombinace několika jednoslabičných a dvouslabičného nebo víceslabičného slova v rámci jednoho mluvnického taktu (obr. 2).

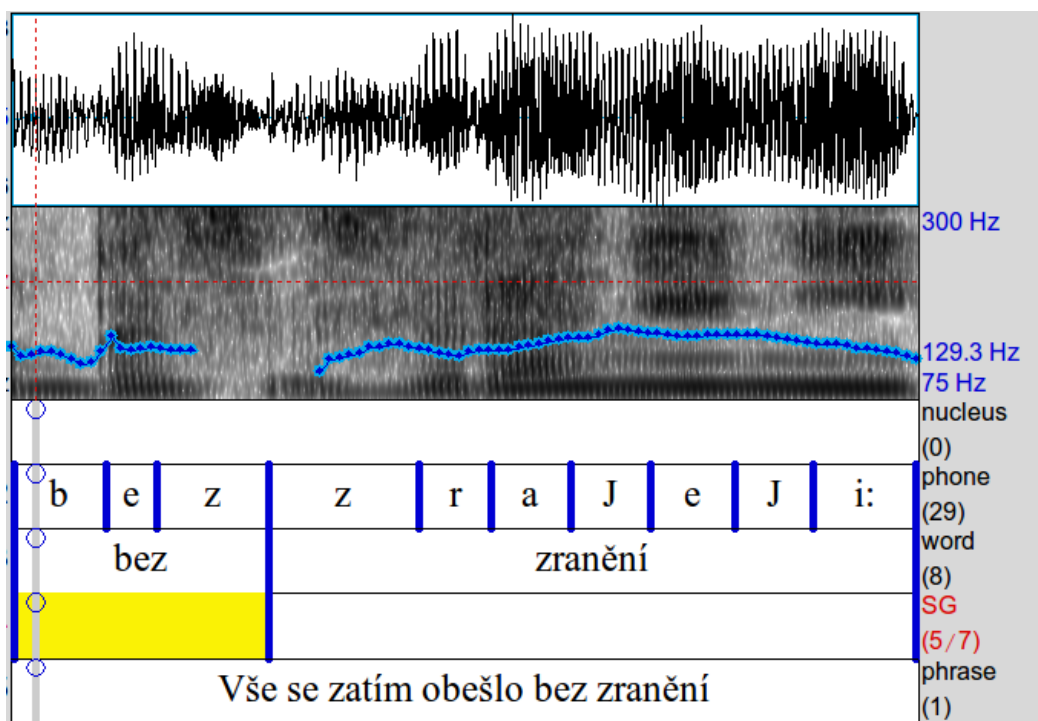


Obr. 2 Zobrazení taktu z nádechového úseku JPC04-04-04 v editačním okně programu Praat. Zde jsou spojena dvě jednoslabičná a dvouslabičné slovo v jeden takt. Průběh f0 naznačuje stoupavoklesavou konturu typickou pro český takt.

Jednou z největších obtíží, se kterou jsme se během značkování hranic reálných mluvnických taktů potýkali, byly jevy spjaté s výskytem taktové předrážky neboli předtaktí. Často totiž nebylo snadné rozhodnout, zda jednoslabičné slovo uvnitř nádechového úseku je předrážkou, tedy slabikou na začátku taktu, která nenesé přízvuk, nebo samostatným taktem, či jestli se váže k předcházející jednotce. Tyto problematické situace nastávaly u všech zpracovávaných projevů. Mluvní (ač profesionálové) se vždy nedrželi ortoepických pravidel, a tak např. z jednoslabičných původních prepozic v některých případech tvořili nepřízvučné anakruze (obr. 3), nebo docházelo k segmentaci, ve které jednoslabičná předložka vytvářela samostatný takt (obr. 4).

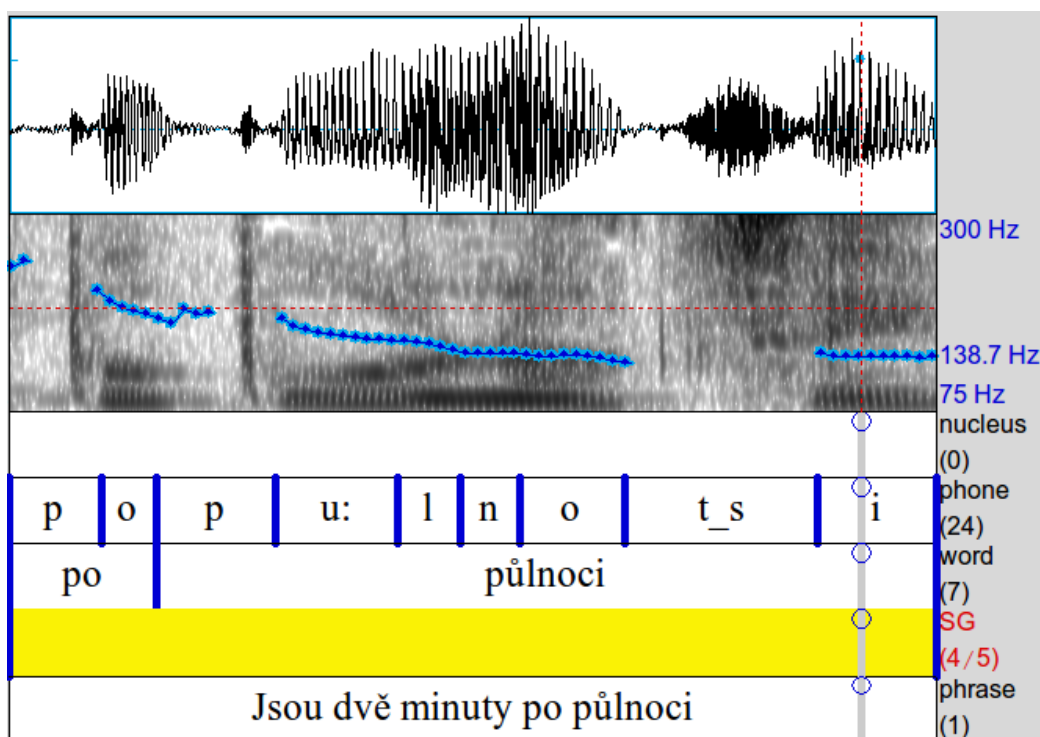


Obr. 3 Zobrazení taktu z nádechového úseku ULR05-04-07 v editačním okně programu Praat. Předložka zde tvoří předdrážku taktu; ve standardní češtině jde o netypické členění. Na spektrogramu si můžeme povšimnout změny dynamiky (stupně šedé) v rámci taktu: předložka má menší energii. Také je dle zobrazeného průběhu f_0 zřejmé, že předložka byla realizována s nižší základní frekvencí.

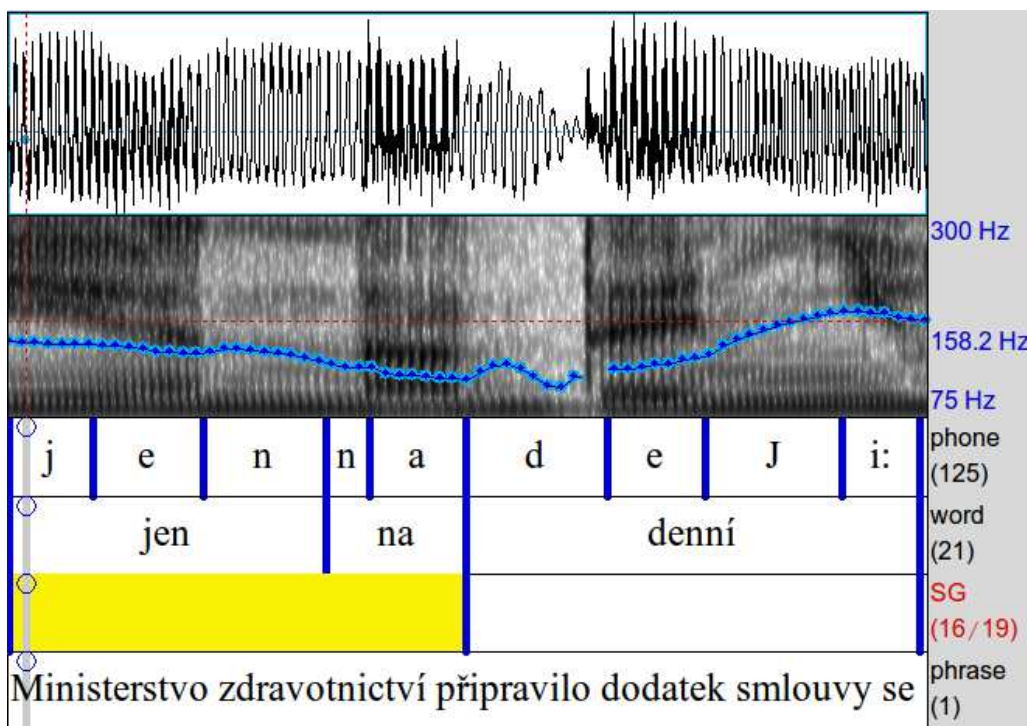


Obr. 4 Zobrazení dvou taktů z nádechového úseku CMA01-05-04 v editačním okně programu Praat. Předložka zde tvoří samostatný takt, což se např. projevilo i v pečlivé artikulaci a ve výrazném oddělení dvou stejných hlásek na hranici předložky a plnovýznamového slova. Ve standardní češtině jde o atypické členění.

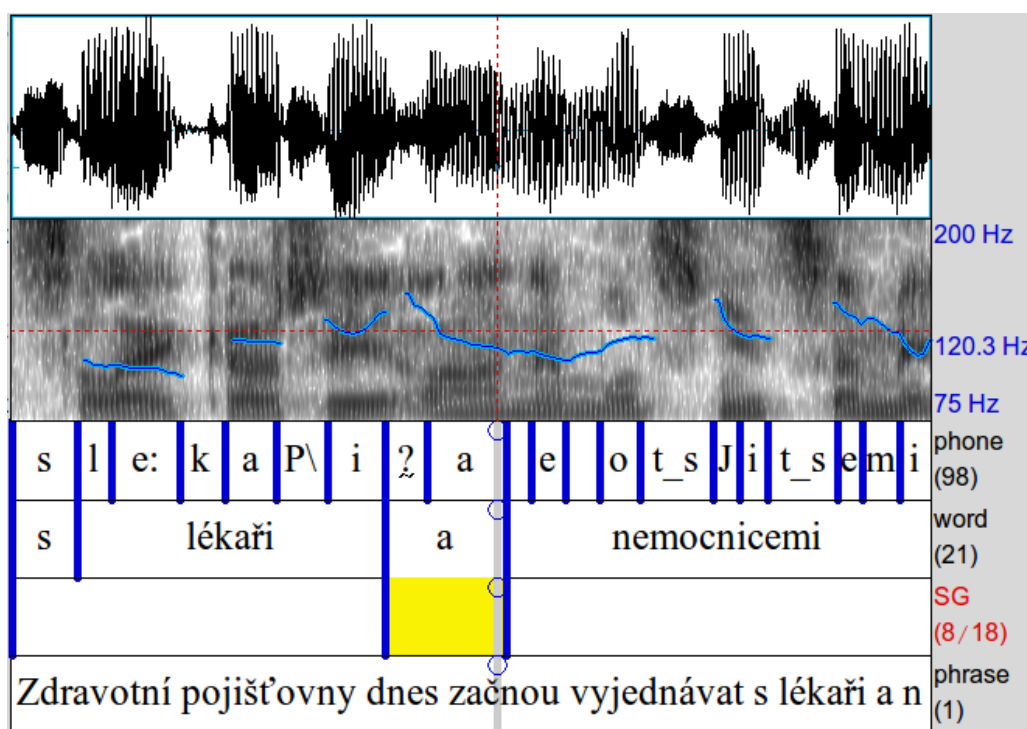
Pokud se jednalo o skutečnou předrážku, potom se však z našeho hlediska nejednalo o problém, neboť jsme značili pouze taktové hranice, nikoli přízvuky; pokud byl přízvuk jinde než na první slabice, nehrálo to tedy v tomto projektu roli – a předložka, ať přízvučná jako první slabika víceslabičného taktu, nebo jako nepřízvučná předrážka, byla vždy součástí jednoho taktu. Takt typu přízvukovaná předložka + plnovýznamové slovo (obr. 5) tím pádem vypadal v našem značení stejně jako takt složený z nepřízvučné předrážky a plnovýznamového slova. Rozdíl nastal v případě, kdy předložka nenesla přízvuk, nebyla předrážkou a byla připojena k předcházejícímu taktu (obr. 6). Samostatné takty tvořily poměrně často i jednoslabičné spojky (např. *že*, *a*), které bychom z hlediska ideálního popisu označili za předrážky (obr. 7). Stejně tak se výjimečně vyskytovaly i případy, kdy nesla přízvuk příklonka *se*, nebo dokonce byla samostatným taktem (obr. 8).



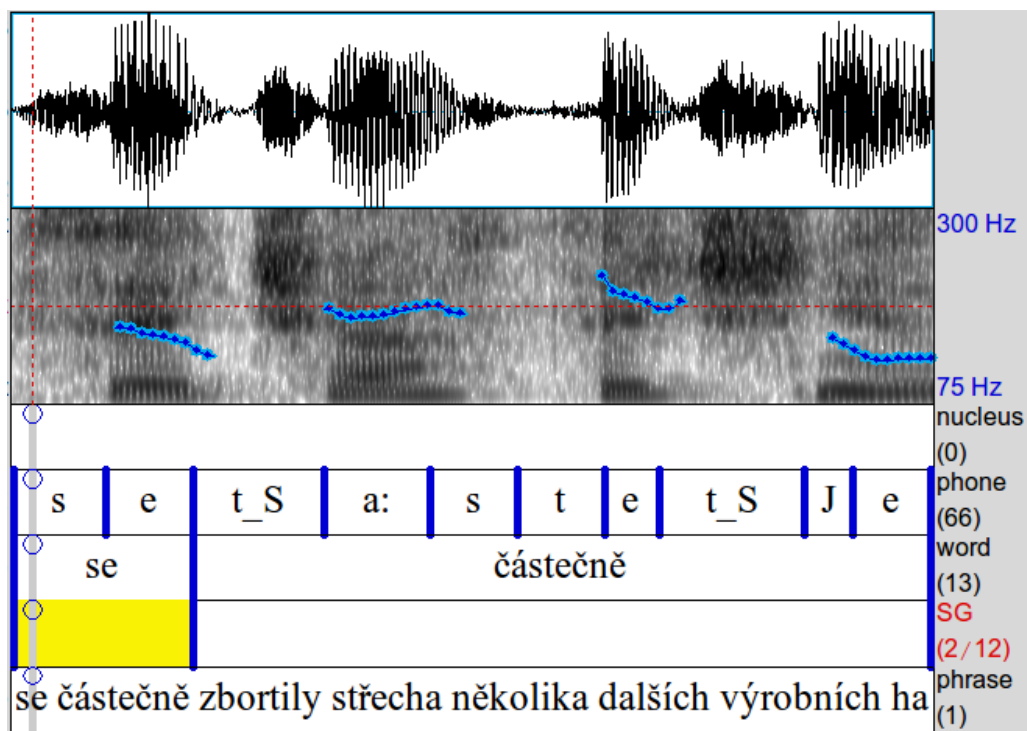
Obr. 5 Zobrazení taktu z nádechového úseku CMA01-01-01 v editačním okně programu Praat. Předložka je zde přízvukovaná a tvoří první slabiku čtyřslabičného taktu. Průběh základní frekvence je klesavý od začátku do konce. Pro češtinu je toto členění typické.



Obr. 6 Zobrazení dvou taktů z nádechového úseku ODC05-05-02 v editačním okně programu Praat. Předložka zde ztratila přízvuk a připojila se k předcházejícímu taktu. Průběh f0 pozvolna klesá k předložce a stoupá na začátku následujícího taktu. Ve standardní češtině je toto členění atypické.



Obr. 7 Zobrazení výřezu z nádechového úseku JPC04-03-01. Spojka *a* zde tvoří samostatný takt. Tento fakt se projevuje zejména v trvání celé samohlásky – tento vokál je z časového hlediska srovnatelný jenom s dlouhým [e:] v taktu předcházejícím. Také je nutno přihlídnout k přítomnosti rázu, který je realizován pomocí třepené fonace. Ve standardní češtině je toto členění atypické.

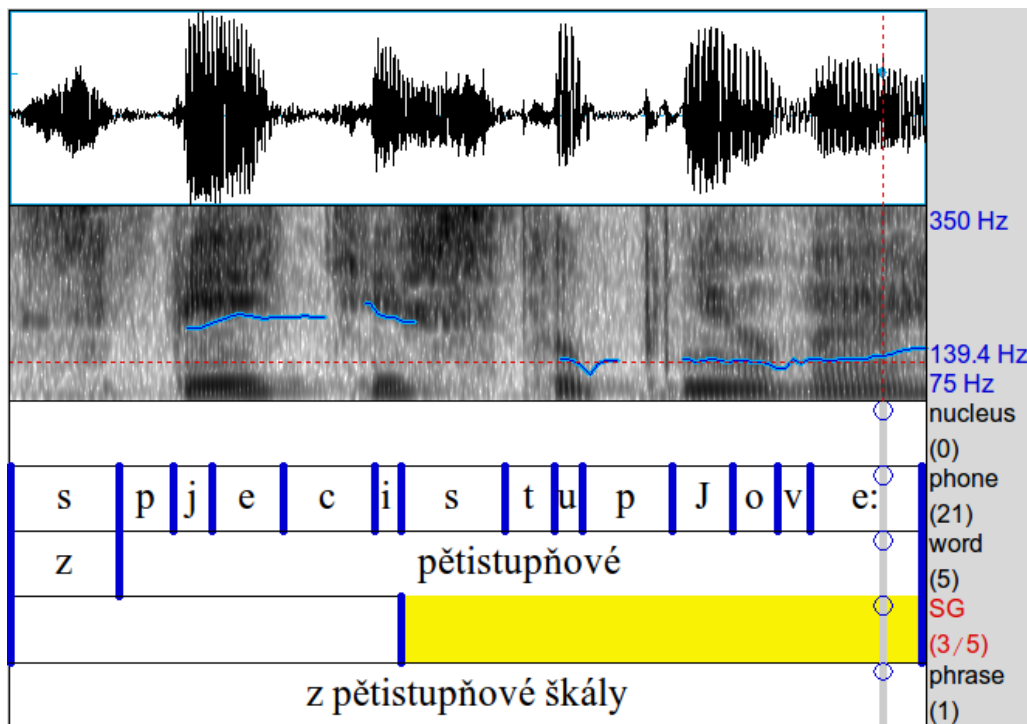


Obr. 8 Zobrazení dvou taktů z nádechového úseku CMA01-05-03. Klitikon *se* zde tvoří samostatný takt – to můžeme demonstrovat např. na poměrně důkladné realizaci vokálu z hlediska trvání (kromě dlouhého [a:] v následujícím taktu zde nenajdeme delší samohlásku) i z hlediska dynamiky (stupně šedé ve spektrogramu). Ve standardní češtině je toto členění atypické.

Pokud bychom přijali řešení, že předrážka může mít přízvuk, museli bychom pak uvažovat dva přízvuky v rámci jednoho taktu, což je však nepřijatelné již z definice předrážky (anakruze tvoří *nepřízvuknou* slabiku před přízvukem, jenž následuje za ní) i s přihlédnutím k dosavadnímu popisu, který mluvní takt obvykle definuje jako skupinu slabik podřízenou *jednomu* přízvuku. Proto jsme se rozhodli upřednostnit řešení se samostatnými jednoslabičnými taktů, byť z hlediska prozodie ani toto řešení není z důvodu, že v případě češtiny není zvykem předpokládat možnost dvou za sebou jdoucích taktových přízvuků jinde než na konci vyšší prozodické jednotky, zcela bezproblémové.

Dále se vyskytly ojedinělé případy slov, většinou složenin, která byla rozdělena na dva takty. Bylo nutné si položit otázku, zda nejde o tzv. *vedlejší přízvuk*, který je v současnosti považován spíše za teoretický konstrukt než za realitu. Dospěli jsme však k závěru, že ve všech případech, vyskytnuvších se v našem korpusu, byly předěly příliš zřetelné, a tak jsme se rozhodli

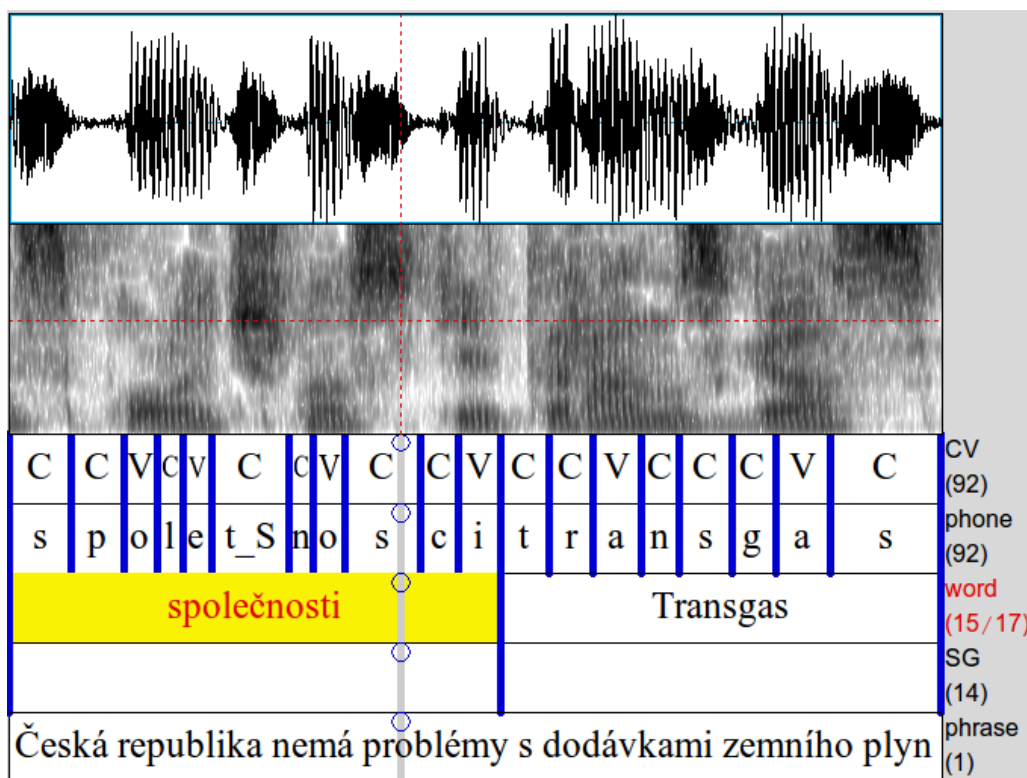
respektovat realitu a rozdělit slovo na dva takty (obr. 9), ačkoli tradiční popis mluvené češtiny s touto variantou příliš nepočítá.



Obr. 9 Zobrazení dvou taktů z nádechového úseku CMA01-05-07. Jedno slovo je v realitě řeči rozděleno na dva samostatné takty, což se projevilo ve výrazné změně průběhu základní frekvence i ve změně dynamiky. Ve standardní češtině jde o netypické členění.

V textech se také místy objevovaly přerážky a jiné dysfluence. Všechny postižené takty byly – podobně jako pauzy – patřičně odlišeny nebo odstraněny, aby nevnášely zkreslení do konečných výsledků.

Nakonec jsme vytvořili další příkazový soubor pro Praat, který převedl hlásky z vrstvy hlásek na vrstvu novou, ve které byly zohledněny pouze rysy *konsonant, vokál, slabikotvorná likvida, ráz*. V nejvyšší vrstvě jsme tedy měli vypsány pouze značky C, V, R a ? v souladu s vrstvou hlásek. Z této vrstvy jsme pomocí nového příkazu získali vzorce CVCV struktur pro slova a mluvní takty, které jsme poté dále vyhodnocovali. Všechny vzorce jsme pro každého mluvčího zpracovávali ve dvou variantách: první obsahovala CVCV struktury, v nichž ráz nebyl zahrnut, a ve druhé ráz figuroval jako značka ?. Náhled konečné verze editačního okna programu Praat uvádíme na obr. 10.



Obr. 10 Náhled editačního okna programu Praat. Tato verze zvuků s textgridy již obsahuje všechny vrstvy, které byly ručně i pomocí příkazových souborů náležitě vyplněny, a je připravena k dalšímu zpracování.

3.2.3 Získání a zpracování dat

Tento výstup jsme posléze exportovali do tabulkového kalkulátoru sady MS Office (MS Excel) a dále s ním pracovali.

Přesné prepisy všech nahrávek jsme upravili gramaticky a převedli do prostého textu. Tento prostý text jsme dále postoupili ke zpracování speciálnímu programu na rozpoznávání slovních druhů, který poskytl Ústav teoretické a počítačové lingvistiky FF UK. K našemu výstupu jsme tedy mohli přidat i informace o slovnědruhové příslušnosti jednotlivých slov.

Vzorci CVCV v našich datech musely být taktéž podrobeny ruční kontrole. Nejprve bylo nutné zkontrolovat, zda se v nich nenacházejí chyby vyplývající z nesprávného SAMPA přepisu nebo z přidání mezery v původních textgridech. Části signálu, kde figurovaly pouze dysfluence nebo pauzy, byly z analýz vypuštěny. Se slovy a takty, které částečně obsahovaly dysfluence, jsme nakládali dvojím způsobem: pokud byla hezitace či jiný rušivý zvuk na

začátku nebo na konci slova, ale nijak nenarušovala jeho strukturu, vypustila se z analýz pouze hezitace a se vzorcem pak bylo zacházeno stejně jako s ostatními; jediný rozdíl spočíval v tom, že postižená jednotka byla označena jako „obsahující dysfluenci“. Pokud rušivý zvuk narušoval strukturu slova nebo taktu (např. se kvůli němu vynechaly/připojily některé hlásky), byl celý vzorec z analýz vypuštěn.

Manuální opravě dále podléhaly znaky pro střední středový vokál – pokud bylo jeho využití funkční (tzn. byl přítomen např. v anglických jménech – *Denson*), byla tato hláska přepsána jako vokál. Pokud byla tímto způsobem zaznamenána hezitace nebo preglotalizace, byla z analýz vypuštěna. Stejným způsobem jsme zacházeli s rázem, který se ve výjimečných případech vyskytl před nebo za konsonantem ve slově – jelikož se nejednalo o systémovou jednotku, byl tento z analýz vypuštěn.

Dalšími nutnými manuálními opravami byly případy některých likvid. Označení hlásek v textových souborech Praatu jako R vycházelo z příkazu, že pokud se v materiálu vyskytne hláska [r], [l] nebo [m] uvnitř shluku dvou konsonantů nebo po souhlásce na konci slova, pak se jedná o slabikotvornou likvidu. Tento postup ale nebyl stoprocentně přesný – jako slabikotvorná byla označena např. hláska [m] v konsonantických skupinách, jako je např. ve slovech *smlouva*, *zřejmě* apod.

Po vyčištění dat jsme přistoupili k samotné analýze, která z excelovských funkcí vyžadovala zejména použití filtrů a kontingenčních tabulek z našich dat. Z dat slov i taktů jsme vytvořili tři pracovní základy: tabulku obsahující všechny informace, a to i s vzorci s dysfluencemi, tabulku s vzorci bez dysfluencí a tabulku s vzorci bez dysfluencí a bez speciálního značení slabikotvorných likvid – tyto byly ve zmíněné tabulce označeny jako V. Většina našich analýz vychází z druhého zdroje, tedy z tabulky bez informací o hezitacích (resp. s dysfluencemi zaznamenanými pouze v poznámce) a s informacemi o likvidách. Pro srovnání s výzkumem prezentovaným v publikaci *Statistiky češtiny* jsme taktéž v některých případech provedli analýzu dat, ve kterých byl ráz považován za běžný konsonant.

4. Výsledky

4.1 Poznámky ke struktuře kapitoly

Cílem tohoto výzkumu je podat informaci o konsonanticko-vokalické struktuře mluvené češtiny, zhodnotit variabilitu fonotaktických skeletů českých mluvních taktů a slov a prozkoumat korespondence mezi oběma rovinami. Zde se postupně budeme zabývat všemi uvedenými otázkami.

V této kapitole uvedeme výsledky fonotaktických analýz. V pododdílech budou podle jednotlivých hledisek pojednávána zároveň slova i takty, aby bylo možné přímé srovnání těchto dvou jednotek. V první části se budeme zabývat obecnými informacemi o nahrávkách. Dále upřesníme své postupy co do klasifikace hlásek jako konsonant (C), vokál (V), ráz (?) a likvida (R). Poté se budeme zabývat počty hlásek u slov i taktů, počty CVCV vzorců, nejčastějšími a nejvzácnějšími vzorci. V dalším oddíle se zaměříme na vzorce se slabikotvornými likvidami. Také se pozastavíme nad vzorci vykazujícími jistou atypičnost (vysoký počet rázů, více vokálů/konsonantů vedle sebe apod.). Poté budeme věnovat pozornost samostatným segmentům (C, V, R, ?) a jejich dvoučlenným a trojčlenným skupinám (např. CV, CR, ?VC, CVC, VCR apod.): vyhodnotíme, kterými jednotkami a kterými skupinami jednotek slovo nebo takt obvykle začíná a končí a které skupiny slova a takty nejčastěji obsahují. Podotýkám, že všechny uvedené analýzy pro slova i takty byly provedeny dvakrát: jednou pro data, v nichž byl zohledněn ráz (a byl značen speciálním symbolem), a podruhé pro táž data, ve kterých ale ráz nefiguroval. Toto opatření jsme přijali zejména z důvodu, že ve fonotaktických analýzách češtiny nebyla dosud rázu věnována potřebná pozornost. Ráz sice není fonologickou jednotkou, nicméně je podstatný např. v rovině prozodické a je legitimní součástí českého mluveného projevu, a to jak spontánního, tak připraveného. Pokud bychom ale provedli výzkum pouze se zahrnutím rázu, mohli bychom se později potýkat s problémem srovnatelnosti tohoto výzkumu s jinými, které ráz nezohledňují.

Společné analýzy taktů a slov se netýká vyhodnocení slovnědruhového zařazení. Slovnědruhová příslušnost byla hodnocena pouze u vrstvy slov.

4.1.1 Poznámka ke klasifikaci jednotek

V analýzách jsme používali čtyři druhy jednotek: konsonant (C), vokál (V), ráz (?) a slabikotvornou likvidu či nazálu (R). Za konsonant jsme zde považovali všechny členy standardního inventáře souhláskových fonémů a alofonů češtiny (ve slovní zásobě domácích i přejatých slov), tj. hlásky [m, n, ɲ, ŋ, ɸ, b, t, d, c, ʃ, k, g, ts, dz, tʃ, dʒ, f, v, s, z, ʃ, ʒ, x, h, ɣ, r, ɾ, ɾ°, j, l],¹² a jejich substandardní formy. Za vokál jsme považovali všechny samohlásky standardního vokalického inventáře češtiny včetně hlásek ve slovech přejatých a cizích, tj. [ɪ, u, e, a, o, i:, u:, e:, a:, o:, ã, ẽ, õ]. Za standardní vokály se tedy taktéž považovaly diftongy [ã, ẽ, õ]. Dlouhé vokály od krátkých nebyly nijak odlišeny. Pokud se v materiálu vyskytl vokál, který se svou kvalitou blížil hlásce schwa, bylo s ním nakládáno podle toho, zda plnil funkci vokálu, nebo v projevu figuroval jen jako forma krátkého či delšího hezitacího zvuku. Pokud mluvčí vyslovil střední středovou samohlásku místo jiné samohlásky nebo ji použil v cizím slově, byla považována za vokál; pokud byla jen hezitací, byla z analýz vypuštěna. Ráz jsme v datech reflektovali podle skutečnosti – zaznamenali jsme ho tedy pouze tam, kde byl opravdu realizován. Pokud se v datech ale objevil např. před konsonantem nebo na konci slova před pauzou jako forma zaváhání, byl z dalšího zpracování (podobně jako hezitací schwa) vynechán. Za slabikotvornou hlásku jsme považovali jednotky [r, l, m, n], pokud tvořily jádro slabiky. Likvidy a nazály se v těchto případech značily R.

Pokud není uvedeno jinak, pracujeme s daty, která jsou vyčištěna od dysfluencí a obsahují zvláště značené slabikotvorné likvidy. Pokud používáme pouze hrubá data, nebo naopak korpus, ve kterém jsou slabikotvorné likvidy a nazály považovány pouze za jádra slabik, a pracovní jsou tedy značeny jako V, v textu tuto informaci výslovně zmiňujeme.

¹² Uvedené hlásky tvoří inventář IPA pro češtinu. Přepisy slov v předkládané práci se však pro zjednodušení a lepší čitelnost zpravidla uvádějí v ortografické podobě.

4.2 Výsledky a jejich interpretace

4.2.1 Dysfluence a pauzy

U mluvčích bylo zjištěno celkem 42 hezitací a 44 pauz. Zatímco pauzy byly z dalšího zpracování vypuštěny všechny, z jednotek s hezitacemi mohlo být 27 dále v analýzách použito; postižené jednotky byly pouze označeny jako „obsahující dysfluenci“. Podrobný rozpis dysfluencí vztažený k jednotlivým mluvčím je uveden v tabulce 4:

mluvčí	pauzy	samostatné hezitace	dysfluence v použitém materiálu	dysfluencí celkem
STR04	2	6	15	21
ULR05	7	0	1	1
WSS05	0	1	2	3
GVA05	3	0	0	0
CMA01	6	0	0	0
KOB02	6	2	3	5
SSA02	3	0	1	1
VOZ01	2	0	1	1
JPC04	4	0	0	0
ODC05	1	4	3	7
MSA03	6	0	0	0
PRK01	4	2	1	3
celkem	44	15	27	42

Tabulka 4: Pauzy a dysfluence u jednotlivých mluvčích. Sloupec „samostatné hezitace“ zahrnují i významné narušení celého taktu/slova – toto byly dysfluence, které v datech nebyly dále zpracovávány. Dysfluence v použitém materiálu značily zaváhání mluvčího na začátku nebo na konci slova/taktu, aniž by toto narušovalo strukturu jednotky – z takto postižených jednotek byly dysfluence vynechány a slova a takty se dále zpracovávaly.

4.2.2 Slova a takty

Jak jsme již zmínili v předchozí kapitole, materiál sestával z nahrávek 12 profesionálních mluvčích. Tyto nahrávky měly celkem 45,57 minuty a sestávaly z 657 nádechových úseků, 6639 slov a 5368 mluvních taktů. Z původních 6641 slov byla dvě slova zcela vynechána kvůli přeřeknutí narušujícímu jejich strukturu (KOB02-01-01 *Radiožurnál* a KOB02-12-01

měl); stejně tak byly vynechány takty obsahující tyto jednotky (KOB02-01-01 *Radiožurnál* a KOB02-12-01 *by se měl*).

V následující tabulce vidíme rozdělení taktů podle počtu slov. Nejčastější byly takty jednoslovné (76,84 %); významné bylo i množství taktů zahrnujících dvě slova (20,86 %). V korpusu se nicméně vyskytovaly i případy, kdy se v taktu objevila slova tři (1,58 %). Velmi vzácný byl případ, kdy bylo jedno slovo rozděleno do několika taktů (0,73 %). Tato situace nastávala většinou při výslovnosti zkratk (např. JPC04-05-02 *KCNA*, STR04-04-01 *ODS* apod.), kdy byla poslední slabika zřetelně přízvukovaná. V některých případech jsme zaznamenali případ nekorektního přízvukování v rámci jednotek, které nebyly zkratkami: jednalo se např. o slova v úsecích JPC04-04-01 *Reichenhal* a KOB01-04-02 *pětapadesát*. V prvním případě byl rozpad slova na dva takty zaznamenán u cizího jména a ve druhém u složeniny. Ačkoli se v češtině nejedná o častý fenomén a popis ho obvykle nezahrnuje, je evidentní, že v některých situacích může tento případ nastat i u profesionálních českých mluvčích.

počet slov v taktu	taktů	%
0	39	0,73
1	4124	76,83
2	1120	20,86
3	85	1,58
celkem taktů	5368	100

Tabulka 5: Počet výskytů taktů o určitém počtu slov a přepočtení na procenta. 0 = slovo rozděleno na více taktů. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Lichý počet je u rozdělených slov z důvodu, že na dva takty bylo rozděleno i slovo s neslabičnou předložkou (CMA01-05-07 *z pětistupňové*) – předložka potom tvořila jedno slovo v prvním taktu, i když ho nepokrývala celý, a druhá část plnovýznamového slova pak tvořila druhý samostatný takt.

4.2.3 Slovní druhy

Zpracování slovních druhů bylo provedeno pouze na korpusu slovním, nikoli taktovém, jelikož takty často zahrnují několik slov různé slovnědruhové příslušnosti. Byly zohledněny soubory dat se zahrnutím rázu mezi hlásky i bez jeho zohlednění.

Nejfrekventovanějším slovním druhem byla zcela jednoznačně substantiva – z 6638¹³ slov jich bylo celkem 2607. Je nutné podotknout, že za substantiva byla považována také cizí i česká vlastní jména a příjmení. Některá (patrně obecně známější) podstatná jména byla rozpoznána automaticky a zbytek byl analogicky zařazen manuálně. Takto jsme získali prakticky kompletně rozpoznáný korpus – nerozpoznaných jednotek bylo pouze 8 (šlo např. o *de* ve francouzských jménech a názvy internetových stránek). Na tomto místě by bylo možné nám vytknout, že jsme cizí slova samozřejmě zařadili mezi česká. Jako protiargument bychom v takovém případě uvedli, že sice šlo o původně nečeské jednotky řeči, které ale byly zpravidla realizovány podle českých doporučení k výslovnosti cizích slov, a také bychom namítli, že ačkoli tyto výrazy jednoznačně nepatří do domácí české slovní zásoby, objevují se v profesionálních i spontánních a mluvených i psaných českých projevech dnes a denně, a jejich výskyt tedy nelze ignorovat.

slovní druhy	počet výskytů	%
substantiva	2607	39,27
adjektiva	915	13,78
předložky	861	12,97
slovesa	831	12,52
příslovce	381	5,74
zájmena	348	5,24
spojky	322	4,85
číslovky	304	4,58
částice	63	0,95
nerozpoznané	6	0,09
celkem	6638	100

Tabulka 6: Zastoupení slovních druhů v rámci korpusu a přepočítání na procenta ze všech slov. Slovní druhy jsou řazeny podle frekvence výskytu. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa.

Dalšími jednotkami, jejichž výskyt byl poměrně častý (13,78 % z korpusu), byla adjektiva a s odstupem necelého jednoho procenta následovaly předložky (12,97 % z korpusu). Slovesa kupodivu tvořila pouze 12,52 %

¹³ Rozdíl jednoho slova je způsoben nesrovnalostí ve způsobu anotace jména Saint-Exupéry. V analýze vzorců je toto jméno vnímáno jako dvě jednotky, kdežto ve zpracování slovních druhů jako jedna (a to substantivum). V této analýze byly počty hlásek a hláskové vzorce těchto jednotek manuálně sloučeny.

všech jednotek v materiálu. Příslovce a zájmena byla v datech zastoupeny podobně (příslovce 5,74 %; zájmena 5,24 % z korpusu). Spojky a číslovky tvořily méně než 5 % z celého korpusu; částic bylo necelé jedno procento (0,95 % z korpusu). Poslední slovní druh – citoslovce – se v materiálu nevyskytoval vůbec, což není příliš překvapivé, neboť se jednalo o připravené oficiální projevy profesionálních mluvčích (rozhlasových hlasatelů) a pro tento typ projevů není zpravidla expresivita typická. Přesné podíly jednotlivých slovních druhů ve zkoumaném korpusu a jejich přepočty na procenta vzhledem ke všem výskytům slov (6638) jsou zobrazeny v tabulce 6.

4.2.4 Hlásky

4.2.4.1 Počet hlásek ve slovech a taktech obecně

V rámci slov se vyskytovalo 37 360 hlásek, v rámci taktů 37 356. Tento rozdíl byl způsoben tím, že jedno ze slov obsahujících dysfluenci, které bylo vypuštěno, netvořilo samostatný takt, a tudíž se muselo ze souboru taktů vypustit spolu s ostatními slovy, se kterými takt sdílelo. (Konkrétně šlo o slovo v úseku KOB02-12-01 a takt *by se měl*; deformovanou jednotkou potom bylo slovo *měl*.) Pokud se mezi hlásky započítal i ráz, bylo hlásek v rámci slov 38 243 a v rámci taktů 38 239. Průměrný počet hlásek v jednotce byl tedy u slov nezahrnujících ráz 5,63 hlásky, se započítáním rázu 5,76 hlásky; u taktů bez zahrnutí rázu byl průměrný počet slabik 6,96 hlásky a se započítáním rázu 7,12 hlásky. Mluvní takty tedy byly ve shodě s našimi předpoklady co do hlásek delší než slova, a to cca o 1,3 hlásky.

4.2.4.2 Počet rázů v materiálu

Rázů bylo v korpusu celkem 883. Slova obsahujících rázy bylo celkem 867 a taktů s rázy 854. Je tedy zřejmé, že se v datech vyskytovaly slova i takty, ve kterých figuroval více než jeden ráz. Jejich přesné rozvržení vidíme v následující tabulce:

počet rázů	slova (celkem 6639)	%	takty (celkem 5368)	%
0	5772	86,94	4514	84,09
1	853	12,85	827	15,41
2	12	0,18	25	0,47
3	2	0,03	2	0,04
celkem jednotek s rázy	867	13,06	854	15,91

Tabulka 7: Rozložení rázů ve slovech a taktech v korpusu a převedení těchto hodnot na procenta ze všech slov a taktů. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa.

V korpusu se vzácně objevily i takty a slova s více než dvěma rázy. Tato situace byla zaznamenána ve dvou případech: ve zkratkách *OSN* [ʔo:ʔesʔen] a *USA* [ʔu:ʔesʔa:], a to u taktů i u slov.

Také jsme zaznamenali určité nesystematické použití rázu – jednalo se o situace, kdy před slovem začínajícím na souhlásku byla krátká hezitace a před ní ráz. Zejména mluvčí STR04 se podobných dysfluencí dopouštěla velmi často. Výjimečně se vyskytl samotný ráz před souhláskou (ODC05-03-06; před zkratkou *www*) a ráz za slovem (ODC05-11-05; za výrazem *deštěm*). Tyto případy byly považovány za dysfluence, které ale nenarušovaly strukturu slova, a tak byly tyto rázy (případně hezitace ve formě střední středové samohlásky) z analýz vypuštěny. Takovýchto nesystematických užití rázu bylo v korpusu 13.

4.2.4.3 Počet hlásek ve slovech a taktech

V rámci slov se vyskytovaly jednotky o 1-24 hláskách (se zohledněním rázu i o 25 hláskách). Nejvíce slov sestávalo z 5 hlásek; velmi frekventovaná byla i slova o 6, 7, a 4 hláskách. V rámci taktů byly výsledky obdobné – opět byly nejfrekventovanější jednotky o 5, 7 a 6 hláskách (se započítáním rázu bylo nejčastějších 7 hlásek v taktu a posléze 5 a 6). Pokud se bral v potaz ráz u slov, byly výsledky poněkud odlišné – nejčastější byla stále slova o 5 hláskách, ale vysokou frekvenci měly i jednotky sestávající ze 2 hlásek – to je způsobeno poměrně četným výskytem dvouhláskových prepozic, krátkých tvarů slovesa být (*je, by*), některých zájmen (*to, se*) a zde i jednohláskových spojek s rázem. V rámci slov bez započítání rázu byla poměrně frekventovaná i slova sestávající z jediné hlásky – oproti taktům, v nichž se takový případ vyskytoval vzácně, pokud se nebral v potaz ráz; v případě, že se ráz do analýz

započítával, jsme takt o jedné hlásce dokonce nezaznamenali vůbec. Je tedy možné, že ráz musí být přítomen, pokud slovo sestávající z jediné hlásky (obvykle spojka) aspiruje na samostatnou existenci jako takt; v případě, že ráz absentuje, není podle našeho korpusu pravděpodobné, že by se daná jednohlásková jednotka stala samostatným taktem.

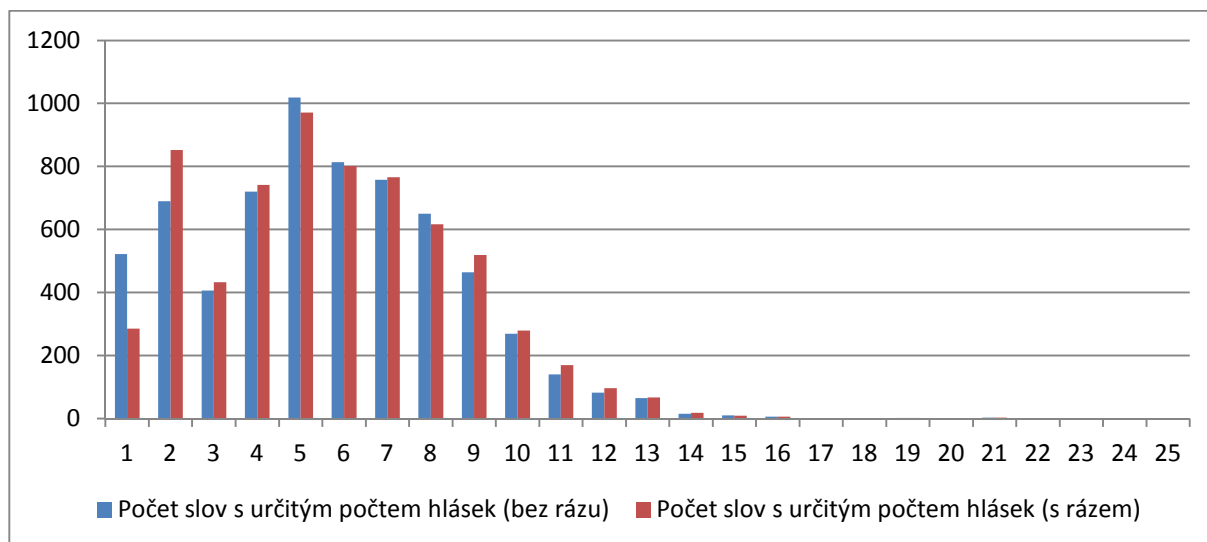
Přesné počty hlásek ve slovech a taktech bez započítání rázu i s rázem vidíme v tabulce 8:

SLOVA						TAKTY					
počet hlásek	počet výskytů	%	počet hlásek + ?	počet výskytů	%	počet hlásek	počet výskytů	%	počet hlásek + ?	počet výskytů	%
5	1019	15,35	5	971	14,63	5	861	16,04	7	829	15,44
6	814	12,26	2	852	12,83	7	809	15,07	5	806	15,01
7	758	11,42	6	801	12,07	6	764	14,23	6	724	13,49
4	720	10,85	7	766	11,54	8	694	12,93	8	653	12,16
2	690	10,39	4	741	11,16	9	567	10,56	9	618	11,51
8	650	9,79	8	616	9,28	4	539	10,04	4	559	10,41
1	522	7,86	9	519	7,82	10	363	6,76	10	369	6,87
9	464	6,99	3	433	6,52	11	210	3,91	11	253	4,71
3	406	6,12	1	285	4,29	3	162	3,02	12	150	2,79
10	269	4,05	10	279	4,2	12	121	2,25	3	139	2,59
11	140	2,11	11	170	2,56	2	82	1,53	2	119	2,22
12	82	1,24	12	96	1,45	13	74	1,38	13	84	1,56
13	65	0,98	13	67	1,01	1	63	1,17	14	28	0,52
14	15	0,23	14	18	0,27	14	25	0,47	15	15	0,28
15	10	0,15	15	9	0,14	15	15	0,28	16	8	0,15
16	6	0,09	16	6	0,09	16	6	0,11	17	3	0,06
21	3	0,05	21	3	0,05	17	4	0,07	18	3	0,06
19	2	0,03	19	2	0,03	21	3	0,06	21	3	0,06
17	1	0,02	17	1	0,02	19	2	0,04	19	2	0,04
20	1	0,02	18	1	0,02	22	2	0,04	22	2	0,04
22	1	0,02	20	1	0,02	18	1	0,02	25	1	0,02
24	1	0,02	22	1	0,02	24	1	0,02			
			25	1	0,02						

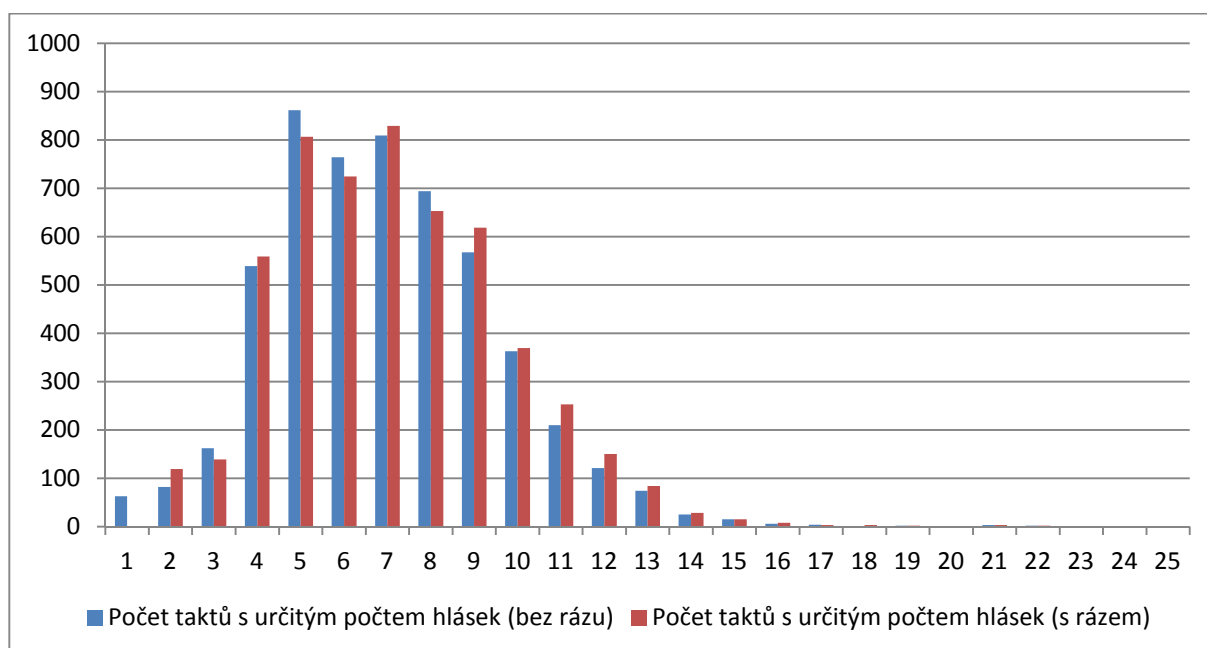
Tabulka 8: Počet výskytů slov a taktů o určitém počtu hlásek (bez rázu, s rázem) a jejich přepočtení na procenta ze všech slov i taktů. Řazení je od nejfrekventovanějšího počtu hlásek v jednotce k nejméně frekventovanému. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa.

V grafu 1 a grafu 2 vidíme vizualizaci počtů jednotek podle počtu hlásek, které jednotky obsahují. V grafu 1 si můžeme povšimnout zejména rozdílů v počtu slov jednohláskových a dvouhláskových, pokud bereme v potaz ráz, nebo ne. Jednohlásková slova, pokud do počtu hlásek započítáváme i ráz, jsou značně omezena (je jich cca o polovinu méně než v datech bez rázu). Jedná se prakticky výhradně o předložky; vzácně se vyskytne spojka tvořená samohláskou, která byla mluvčím realizována bez rázu. Stejně tak vidíme

poměrně velký rozdíl u jednotek dvouhláskových, kde je výrazně více slov při zohlednění rázu. Zde jsou totiž přítomna právě ona slova sestávající z jedné samohlásky a rázu. V dalších počtech jsou slova s rázem a bez rázu poměrně vyrovnaná. Zajímavý je také omezený počet tříhláskových slov. Z grafu je patrné, že výskyt jednohláskových taktů je vzácný a v datech s rázy tyto takty absentují. Pěti- a sedmihláskové takty jsou oproti slovům vyrovnané (u slov jsou pětihláskové jednotky od ostatních frekvenčně odděleny výrazněji).



Graf 1: Počet slov s daným počtem hlásek; bez rázu a s rázem. Na ose x počet hlásek ve slově, na ose y počet výskytů jednotek s tímto počtem hlásek.



Graf 2: Počet taktů s daným počtem hlásek; bez rázu a s rázem. Na ose x počet hlásek v taktu, na ose y počet výskytů jednotek s tímto počtem hlásek.

4.2.4.4 Počet hlásek s ohledem na slovnědruhovou příslušnost slov

Jak už bylo zmíněno výše, všechna slova našeho materiálu obsahovala celkem 37 360 hlásek, do nichž jsme nezapočítávali ráz, a 38 243 hlásek, pokud jsme ráz také k hláskám přiřadili. V této kapitole uvádíme souvislosti rozdělení hlásek s ohledem na slovnědruhovou příslušnost jednotek.

V tabulce 9 vidíme celkové absolutní počty hlásek v rámci daných slovních druhů, jejich procentuální podíl na celkovém počtu hlásek a průměrnou délku slova daného slovního druhu v hláskách. Pokud tuto tabulku srovnáme s tabulkou 2 v kapitole 2.3.6.2, která uvádí procentuální hodnoty počtu hlásek ve vztahu ke slovním druhům v korpusu SYN2005 a v jeho publicistické části, zjistíme určité rozdíly. Až na číslovky a spojky byl průměrný počet hlásek v našem materiálu vždy o několik setin až desetín procenta vyšší, pokud bereme v potaz ráz. Pokud ráz do analýz nezahrnujeme, budou rozdíly až na případ spojek méně patrné. I v případě, že za srovnávací základ považujeme publicistické texty korpusu SYN2005, si můžeme všimnout, že data v našem materiálu v tomto směru víceméně odpovídají datům korpusu SYN2005.

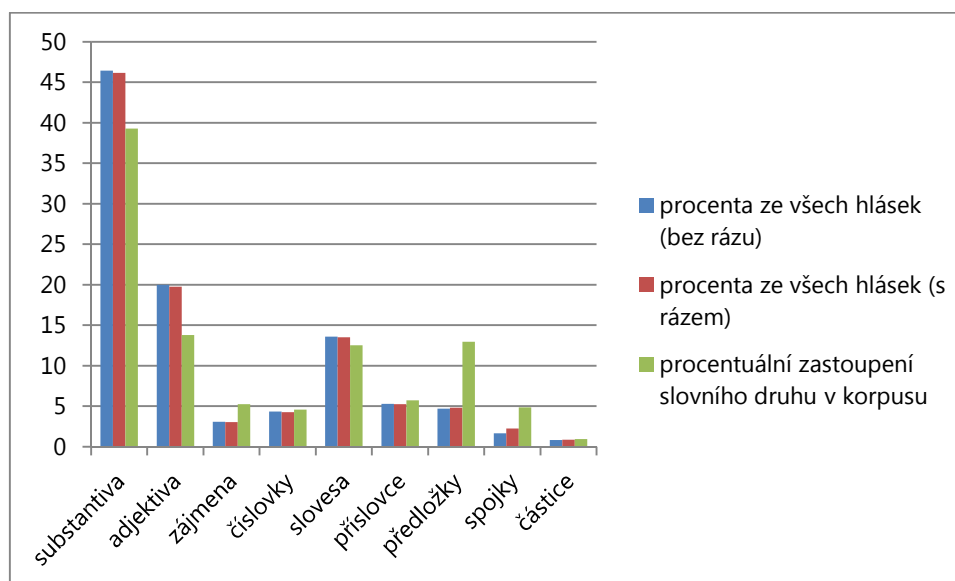
slovní druh	počet hlásek (bez rázu)	%	průměrný počet hlásek	počet hlásek (s rázem)	%	průměrný počet hlásek
substantiva	17347	46,43	6,65	17653	46,16	6,77
adjektiva	7466	19,98	8,16	7557	19,76	8,26
slovesa	5074	13,58	6,11	5163	13,5	6,21
příslovce	1972	5,28	5,18	2013	5,26	5,28
předložky	1753	4,69	2,04	1844	4,82	2,14
číslovky	1627	4,35	5,35	1637	4,28	5,38
zájmena	1160	3,1	3,33	1160	3,03	3,33
spojky	623	1,67	1,93	860	2,25	2,67
částice	312	0,84	4,95	329	0,86	5,22
nerozpoznané	26	0,07	4,33	27	0,07	4,5
celkem	37 360	100		38 243	100	

Tabulka 9: Počet hlásek v rámci všech výskytů daných slovních druhů, přepočten na procentuální hodnoty vzhledem k celkovému počtu hlásek a průměrná délka slova určitého slovního druhu vyjádřená počtem hlásek v jednotce. Procentuální hodnoty i hodnoty průměru jsou zaokrouhleny na dvě desetinná místa.

Na hodnotách průměrné délky slova daného slovního druhu vidíme, že – měřeno počtem hlásek – je průměrně nejdelší přídavné jméno s 8,16 (příp.

8,26) hlásky; následuje podstatné jméno (6,65; příp. 6,77 hlásky) a sloveso (6,11; příp. 6,21 hlásky). Naopak nejkratší průměrnou délku měly logicky předložky (2,04; příp. 2,14 hlásky) a spojky (1,93; příp. 2,67 hlásky). Poměrně značný rozdíl mezi průměrnou délkou spojek v souborech s rázem a bez rázu přičítáme častému výskytu spojek, které začínají vokálem (a tedy v produkci řeči většinou i rázem).

Absolutní hodnoty počtu hlásek u jednotlivých slovních druhů souvisejí s celkovým zastoupením těchto slovních druhů v našem materiálu, proto není nijak překvapivé, že nejvíce hlásek bylo základem pro substantiva; nezanedbatelný podíl měla také adjektiva a slovesa. Předložky, které v korpusu tvořily 12,97 % všech slov, a bylo jich tedy více než sloves, však obsahovaly výrazně menší počet hlásek než slovesa. Stejně tak jsme našli rozdíl mezi počtem hlásek u příslovce a zájmen, ačkoli početně byly v korpusu oba slovní druhy zastoupeny podobně. Tyto rozdíly názorně ilustruje graf 3:



Graf 3: Zastoupení slovních druhů v materiálu vyjádřené procenty (viz tabulka 6) a procentuální zastoupení hlásek v rámci určitých slovních druhů (viz tabulka 9). Hodnoty jsou uvedeny pro soubor hlásek bez rázu i s rázem.

Tabulky 10a a 10b poskytují hlubší vhled do souvislostí mezi počtem hlásek a slovnědruhovou příslušností; v tabulce 10a není do analýz zahrnut ráz, tabulka 10b ráz započítává mezi hlásky. Vidíme, že všechna jednohlásková slova byly buď předložky (např. *s*, *k*, *v*) nebo spojky (např. *a*, *i*). Spojky se při započítání rázu posléze zařadily mezi dvouhláskové jednotky. Mezi

dvouhláskovými slovy byla výrazně vzácná substantiva (jednalo se např. o výrazy *aut*, *eur* apod.); při započítání rázu absentovala v kratších než trojhláskových kombinacích podstatná jména úplně. Adjektiva jsme zaznamenali pouze čtyřhlásková a delší. Dvojháskové nebyly také číslovky; částice pouze bez započítání rázu. Nejdelšími jednotkami byla adjektiva – v korpusu se vyskytovaly dlouhé výrazy jako *jedenáctimiliardový*, *hydrometeorologického* a *sedmdesátimilionového*; nejvíce hlásek mělo slovo *osmašedesátimilionového*.

Co se týče dat bez rázu, podstatná jména byla nejvíce zastoupena slovy sestávajícími z šesti hlásek; značnou část substantiv také tvořila slova pětihlásková, sedmihlásková, čtyřhlásková a osmihlásková. Nejdelší substantivum mělo 17 hlásek. K podobným výsledkům jsme dospěli i při zahrnutí rázu; v tomto případě ale mělo nejdelší substantivum hlásek 18 (jednalo se o jednotku *Uherskohradištsku*). Adjektiva byla zastoupena nejvíce slovy sedmihláskovými; s malým rozdílem následovaly jednotky o osmi, pěti a devíti hláskách. Podobných výsledků dosáhneme i se započítáním rázu. Nejdelší přídavné jméno mělo 24 hlásek, s rázem 25 (*osmašedesátimilionového*). Prepozic bylo nejvíce dvouhláskových a jednohláskových (v datech s rázem i bez rázu). Nejdelší prepozice měly 8 hlásek (*vzhledem*, *navzdory*, *nedaleko*). Sloves bylo nejvíce sedmihláskových a čtyřhláskových. Mezi daty bez rázu a s rázem byly co do počtu hlásek pouze minoritní rozdíly. Nejdelší sloveso čítalo 13 hlásek (*sterilizovány*, *destabilizuje*). Adverbia byla nejčastěji pětihlásková, častý byl i výskyt čtyřhláskových. Nejdelší příslovce v korpusu mělo 13 hlásek (*bezprostředně*). Zájmena byla spíše dvouhlásková, nejdelší zájmena byla osmihlásková (*některých*, *některým*). Spojky byly v datech bez rázu nejčastěji jednohláskové, v datech s rázem pak dvouhláskové. Nejdelší spojky pak byly osmihláskové (*přestože*, *zatímco*). Nejfrekventovanější byly pětihláskové číslovky; nejdelší číslovka měla hlásek 14, se započítáním rázu 15 (*sedmadvacátého*). Částice obsahovaly v datech bez rázu nejčastěji tři, sedm nebo dvě hlásky, v datech s rázem bylo výrazně více trojhláskových částic než ostatních. Nejdelší částice měla devět hlásek (*například*). Citoslovce se v korpusu neobjevila vůbec.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	22	24	Σ	
N		4	86	333	462	500	398	328	221	129	69	31	33	7	5		1							2607
A				27	129	95	159	145	126	86	51	44	27	7	5	6		2	1	3	1	1		915
R	342	356	56	25	68	10	1	3																861
V		79	38	121	108	106	125	117	69	47	14	5	2											831
D		24	45	86	103	42	17	31	25	4	3		1											381
P		159	47	55	59	15	8	5																348
J	180	55	50	24	3		8	2																322
C			69	42	79	41	29	19	14	2	3	2	3	1										304
T		11	14	6	7	4	12		9															63
X		2		1	1	1	1																	6
Σ	522	690	405	720	1019	814	758	650	464	268	140	82	66	15	10	6	1	2	1	3	1	1		6638

Tabulka 10a

Tabulka 10a a 10b: Slova s určitými počty hlásek s ohledem na slovnědruhovou příslušnost jednotek, zpracováno pro data bez rázu (10a) a s rázem (10b). V řádcích jsou počty výskytů slovních druhů vzhledem k počtu hlásek uvedenému v prvním řádku tabulky. Slovní druhy jsou řazeny podle frekvence jejich výskytu. N = substantivum (podstatné jméno), A = adjektivum (přídavné jméno), R = prepozice (předložka), V = verbum (sloveso), D = adverbium (příslovce), P = pronomen (zájmeno), J = konjunkce (spojka), C = numerale (číslovka), T = partikule (částice), X = nerozpoznaný slovní druh.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25	Σ	
N			86	313	432	489	422	323	234	137	88	35	34	8	5			1							2607
A				27	129	93	151	120	149	82	59	50	26	10	4	6	1		2	1	3	1	1		915
R	282	391	79	27	65	12	2	3																	861
V		79	38	121	101	99	115	110	95	47	17	6	3												831
D		10	59	86	93	49	18	33	16	11	4	1	1												381
P		159	47	55	59	15	8	5																	348
J	3	211	33	62	2	1	8	2																	322
C			68	40	82	40	29	18	16	2	1	4	3	1											304
T			22	9	7	2	13	1	9																63
X		2		1	1	1		1																	6
Σ	285	852	432	741	971	801	766	616	519	279	169	96	67	19	9	6	1	1	2	1	3	1	1		6638

Tabulka 10b

Nejčastějším slovem celkově byla předložka *na* (178 výskytů). Následovala předložka *v* (165 výskytů) a spojka *a* (144 výskytů). Výraz *se* jakožto zvrtné zájmeno nebo součást reflexiv tantum měl 89 výskytů. Dalšími frekventovanými slovy byly předložky *s/se* (48 + 11 výskytů), *do* (52

výskytů), o (52 výskytů). Předložka z se v korpusu objevila dvačtyřicetkrát, předložka *podle* jednačtyřicetkrát. Až na desátém místě v pořadí podle frekvence slov se umístil tvar slovesa *být je* (39 výskytů).

4.2.5 CVCV vzorce

V této podkapitole se budeme zabývat výskytem a četností jednotlivých CVCV struktur u slov i taktů a u slov bude zohledněna i jejich slovnědruhá příslušnost. Využita budou data se započítáním rázu i bez něj a výjimečně i soubor dat, ve kterých je ráz považován za konsonant.

4.2.5.1 Celkové počty CVCV struktur

Pokud bychom považovali za základ hrubý slovní korpus bez odstranění dysfluencí a likvid, našli bychom v něm 399 různých CVCV vzorců slov bez zahrnutí rázu a 433 vzorců slov se započítáním rázu mezi hlásky. Vzorců taktů by v hrubém korpusu bylo 509 bez rázu a 580 s rázem. Data vyčištěná od hezitací a bez značených likvid obsahují 339 vzorců slov a 435 vzorců taktů bez rázu a 369 vzorců slov a 501 vzorců taktů s rázem. V této práci však pracujeme obvykle s korpusem, který je **vyčištěn od dysfluencí** a **obsahuje speciálně značené slabikotvorné likvidy/nazály** – zde nalezneme 384 různých vzorců slov a 493 vzorců taktů bez rázu a 413 vzorců slov a 559 vzorců taktů se zohledněním rázu.

4.2.5.2 CVCV vzorce podle frekvence výskytů v korpusu

Nejčastějších 50 fonotaktických struktur slov a taktů v našem korpusu (včetně konkrétního počtu výskytů každého ze vzorců a přepočtu na procentuální hodnoty vzhledem k celkovému počtu výskytů slov i taktů) je uvedeno v následující tabulce 11. Tabulka je uspořádána podle četnosti jednotlivých CVCV vzorců od nefrekventovanějších struktur ke vzácněji se vyskytujícím. Tato tabulka vychází ze souboru dat bez vyznačených dysfluencí a se speciálně značenými slabikotvornými likvidami (popř. nazálami).

SLOVA						TAKTY					
vzorec bez ?	výskytů	%	vzorec s ?	výskytů	%	vzorec bez ?	výskytů	%	vzorec s ?	výskytů	%
CV	603	9,08	CV	603	9,08	CVCV	382	7,12	CVCV	380	7,08
CVCV	481	7,25	CVCV	481	7,25	CVCCV	297	5,53	CVCCV	296	5,51
CVCCV	354	5,33	CVCCV	354	5,33	CVCVCV	230	4,28	CVCVCV	229	4,27
C	278	4,19	C	278	4,19	CCVCV	223	4,15	CCVCV	223	4,15
CVCVCV	271	4,08	CVCVCV	271	4,08	CVCVC	209	3,89	CVCVC	209	3,89
CCVCV	267	4,02	CCVCV	267	4,02	CVCCVCV	171	3,19	CVCCVCV	169	3,15
CVCVC	265	3,99	CVCVC	265	3,99	CCVCVCV	150	2,79	CCVCVCV	150	2,79
V	244	3,66	ʔV	237	3,57	CVCVCVC	132	2,46	CVCVCVC	132	2,46
CVC	211	3,18	CVC	211	3,18	CVCVCVCV	129	2,4	CCVCCV	125	2,33
CVCCVCV	178	2,68	CVCCVCV	178	2,68	CVCVCCV	126	2,35	CVCVCCV	124	2,31
CVCVCCV	148	2,23	CVCVCCV	148	2,23	CCVCCV	125	2,33	CVCVCCV	123	2,29
CCV	141	2,12	CCV	141	2,12	CVCCVC	97	1,81	CVCCVC	97	1,81
CCVCCV	139	2,09	CCVCCV	139	2,09	CCVCVC	85	1,58	CCVCVC	85	1,58
CCVC	130	1,96	CCVC	130	1,96	CVCVCCVCV	80	1,49	CVCVCCVCV	79	1,47
CCVCVCV	124	1,87	CCVCVCV	124	1,87	CCVC	74	1,38	CCVC	74	1,38
CVCVCVCV	119	1,79	CVCVCVCV	119	1,79	CVCCVCVCV	73	1,36	CVCCVCVCV	72	1,34
CVCCVC	111	1,67	CVCCVC	111	1,67	CCVCCVCV	69	1,29	CCVCCVCV	69	1,29
CVCVCVC	105	1,58	CVCVCVC	105	1,58	CVCVCVCCV	69	1,29	CVCVCVCCV	69	1,29
CCVCVC	99	1,49	CCVCVC	99	1,49	CVCCVCCV	68	1,27	CVCCVCCV	67	1,25
VC	87	1,31	CVCCVCCV	82	1,24	V	63	1,17	ʔV	63	1,17
CVCCVCCV	82	1,24	ʔVC	75	1,13	CCVCVCVCV	62	1,15	CCVCVCVCV	61	1,14
CVCCVCVCV	73	1,1	CVCCVCVCV	73	1,1	CVC	59	1,1	CVC	59	1,1
CCVCCVCV	66	0,99	CCVCCVCV	66	0,99	CV	55	1,02	CV	55	1,02
CVCVCCVCV	65	0,98	CVCVCCVCV	65	0,98	CCVCVCCV	53	0,99	CCVCVCCV	53	0,99
CVCVCVCCV	56	0,84	CVCVCVCCV	56	0,84	CCV	52	0,97	CCCVCV	52	0,97
CCCVCV	54	0,81	CCCVCV	54	0,81	CCCVCV	52	0,97	CCV	52	0,97
VCV	53	0,8	CCVCVCVCV	48	0,72	CCVCVCVC	51	0,95	CCVCVCVC	51	0,95
VCCVCV	49	0,74	ʔVCV	47	0,71	VCCVCV	50	0,93	ʔVCCVCV	47	0,88
VCVCV	48	0,72	CCVCVCVC	46	0,69	VCVCV	48	0,89	ʔVCV	44	0,82
CCVCVCVCV	48	0,72	ʔVCCVCV	45	0,68	VCV	47	0,88	ʔVCVCV	43	0,8
CCVCVCVC	46	0,69	CCVCVCCV	43	0,65	CVCCVCVC	43	0,8	CVCCVCVC	42	0,78
VCCVCVCV	44	0,66	ʔVCVCV	42	0,63	CCVCCVC	41	0,76	CCVCCVC	41	0,76
CCVCVCCV	43	0,65	ʔVCCVCVCV	41	0,62	VCCVCVCV	40	0,75	ʔVCCVCVCV	39	0,73
CVCCVCVC	33	0,5	CVCCVCVC	33	0,5	CVCVCCVCCV	37	0,69	CVCVCCVCCV	37	0,69
CVCVCVCVCV	31	0,47	CVCVCCVCCV	31	0,47	CVCVCVCVCV	34	0,63	CVCVCVCVCV	34	0,63
CVCVCCVCCV	31	0,47	CVCVCVCVCV	31	0,47	CVCCVCV	28	0,52	CVCCVCV	26	0,48
CVCVCCVC	30	0,45	CVCCCV	30	0,45	VCVCVCV	27	0,5	ʔVCVCVCV	26	0,48
CVCCCV	30	0,45	CVCVCCVC	30	0,45	CVCVCCVC	26	0,48	CVCVCCVC	25	0,47
CVCR	28	0,42	CCVCCVC	28	0,42	VC	26	0,48	CCVCCCV	24	0,45

CCVCCVC	28	0,42	CVCR	28	0,42	CCVCCCV	24	0,45	CVCCCV	24	0,45
CCVCCCV	26	0,39	CCVCCCV	26	0,39	CVCCCV	24	0,45	ʔVC	24	0,45
CVCCCVCV	26	0,39	CVCCCVCV	26	0,39	VCVCCCV	23	0,43	CCVCCVVC	23	0,43
CVCC	23	0,35	CVCC	23	0,35	CCVCCVVC	23	0,43	CVCR	22	0,41
VCVCCV	21	0,32	ʔVCVCCV	21	0,32	VCCVCCV	23	0,43	CVCCVCCV	21	0,39
VCCVC	19	0,29	CVCCVCCVC	19	0,29	CVCCVCCVC	23	0,43	CVCCVCCVC	21	0,39
CVCCVCCVC	19	0,29	CVCVCCCV	18	0,27	VCVCCV	22	0,41	ʔVCVCCV	21	0,39
VCCVCCV	18	0,27	ʔVCCVVC	18	0,27	CVCR	22	0,41	ʔVCVVC	21	0,39
CVCVCCCV	18	0,27	CRCCV	18	0,27	VCCVC	22	0,41	ʔVCCVVC	20	0,37
VCCVVC	18	0,27	CCVCCVCCV	17	0,26	VCVVC	21	0,39	ʔVCVCCV	20	0,37
CRCCV	18	0,27	ʔVCCVC	16	0,24	CVCCVCCV	21	0,39	ʔVCCVCCV	20	0,37
			CRCV	16	0,24						
ostatní (334)	1140	17,17	ostatní (362)	1166	17,56	ostatní (443)	1437	26,77	ostatní (509)	1485	27,66
celkem	6639	100	celkem	6639	100	celkem	5368	100	celkem	5368	100

Tabulka 11: 50 nejčastějších CVCV vzorců českých slov a taktů v našem korpusu a jejich přepočtení na procenta vzhledem k celkovému počtu výskytů slov/taktů. Vzorce jsou řazeny od nejčastějších k vzácnějším. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa.

Z tabulky 11 můžeme vyvodit několik závěrů: Na první pohled je patrný rozdíl mezi prvními příčkami u slov a u taktů. Nejčastěji se opakujícím vzorcem slov je kombinace C + V, zatímco nejčastější takt tvoří vzorec CVCV. Vzorec CV u slov je navíc poměrně výrazně odlišen od ostatních vzorců: čítá 603 výskytů; druhý nejčetnější vzorec, tedy CVCV, má pouze 481 výskytů. S ještě větším odstupem následuje struktura CVCCV se 354 výskytů. Čtvrtým vzorcem v pořadí je samotný konsonant (278 výskytů), který již od ostatních struktur, jež následují, není tak jasně oddělen. Pořadí a počty těchto vzorců jsou stejné pro vzorce s rázem i bez rázu. Preferenci vzorce CV můžeme přičíst zejména vysoké frekvenci jednoslabičných předložek, výrazu *se* apod. (viz výše). Pokud bychom ráz nepovažovali za speciální jednotku, ale za konsonant, mohli bychom k CV vzorcům ještě přičíst 237 výskytů struktury ʔV – celkově by tak vzorec CV vykazoval 840 výskytů a činil by 12,65 % všech slovních vzorců. Slova složená ze segmentů CVC by v tomto případě vykazovala 211 + 75 výskytů a činila by 4,31 % všech vzorců. Stejně tak by bylo možné upustit od hodnocení slabikotvorné likvidy jako zvláštního druhu jádra slabiky a zabývat se slovem CVCR jako jednotkou CVCV – pak by k původním 481 přibýlo 28 výskytů a struktura CVCV by tvořila 7,67 %.

Mezi takty je, jak již bylo zmíněno, nejčtenějším vzorcem v korpusu struktura CVCV se 382 výskyty. Je patrné, že mezi takty nejsou tolik preferovány dvousegmentální, tedy jednoslabičné jednotky. Vzorec CV nicméně v nejčastějších 50 strukturách figuruje a s 55 výskyty tvoří 1,02 % z celkového počtu 5638 taktů. Pokud bychom ráz považovali za souhlásku, pak by se 117 výskyty tvořila struktura CV 2,18 %. Z tohoto faktu můžeme tedy vyvodit závěr, že struktura CV jako samostatný takt sice není zcela nejfrekventovanějším vzorcem, nicméně nejedná se ani o extrémně vzácný jev a je možné se s ním poměrně běžně setkat i v projevech profesionálních mluvčích.

Dalším poznatkem je, že mezi nejčastějšími fonotaktickými vzorci taktů nejsou co do výskytů tak velké rozdíly jako u vzorců slov; ani jeden rozdíl nepřesáhl 100 výskytů. Největší odstup mezi sebou má první a druhá a druhá a třetí struktura (84 a 67 výskytů). Počet taktů v korpusu byl nižší než počet slov, ale typů vzorců taktů bylo více než typů CVCV struktur slov, což není nijak překvapivé vzhledem k faktu, že v taktech se mohou kombinovat různé vzorce slov. Větší kombinatorické možnosti pak vedou i k větší roztržitosti typů vzorců. V korpusu slov např. vykazovalo jediný výskyt 151 vzorců (39,32 % ze všech typů vzorců slov a 2,27 % ze všech výskytů slov) bez započítání rázu; s rázem pak 169 (40,92 % ze všech typů vzorců slov a 2,55 % ze všech výskytů vzorců slov), zatímco v korpusu taktů bylo jedinečných vzorců 228 bez zahrnutí rázu (tj. 46,25 % ze všech typů vzorců taktů a 4,25 % ze všech výskytů taktů) a 276 s jeho započítáním (tj. 49,37 % ze všech typů vzorců taktů a 5,14 % ze všech výskytů taktů). Ačkoli výskytů taktů bylo 5368 a výskytů slov 6639, jedinečných vzorců taktů bylo kvůli zmíněné vyšší možnosti kombinace více než jedinečných vzorců slov.

4.2.5.3 Nejkratší a nejdelší struktury

Ze slovních vzorců bez rázu o jednom segmentu byly frekventovanější ty, které sestávaly z konsonantu (278 výskytů) než z vokálu (244 výskytů). Z vzorců slov s rázem bylo 237 struktur skládajících se z rázu a vokálu. Samotná samohláska bez rázu se v korpusu objevila pouze sedmkrát. V doméně taktů byla situace odlišná – jelikož slovo sestávající z konsonantu

není schopno tvořit slabiku, nemůže tvořit ani takt. V datech se nicméně vyskytlo 63 případů, kdy takt tvořil vokál (resp. ráz a vokál). V korpusu se započítáním rázu se neobjevil ani jeden případ, kdy by takt tvořila sama samohláska bez rázu. V tomto případě bychom mohli absenci rázu považovat za ukazatel nesamostatnosti takové jednosegmentální jednotky.

Nejdelším vzorcem byla struktura (?)VCCVCVCVCVCVCVCVCVCVCVCV, a to zároveň jako vzorec taktu i slova. Tato jednotka se v datech vyskytla pouze v jediném případě.

4.2.5.4 Vzorce s likvidami

Vzorců se slabikotvornými sonorními souhláskami coby jádry slabik bylo v korpusu celkem 158. Samotných slabikotvorných likvid (popř. nazál¹⁴) bylo rovněž 158. Z těchto informací plyne, že v každý vzorec s likvidou, který se v našich datech objevil, obsahoval právě jednu likvidu. Vzorců bylo stejně u taktů a slov a u vzorců bez rázu i s rázem. Všechny tyto vzorce tvořily 2,38 % ze všech výskytů vzorců slov a 2,94 % ze všech výskytů vzorců taktů. Nejfrekventovanější vzorec s likvidou (CVCR) netvořil ani 0,5 % z celkového počtu vzorců. Typů vzorců slov s likvidou bylo 49 se zohledněním rázu (11,84 % ze všech typů slovních vzorců) i bez něj (12,73 % ze všech typů slovních vzorců). Typů vzorců taktů obsahujících likvidu bylo 64 bez zahrnutí rázu (12,96 % ze všech typů taktových vzorců) a 66 s jeho započítáním mezi hlásky (11,79 % ze všech typů taktových vzorců).

V tabulce 12 jsou uvedeny CVCV vzorce taktů a slov obsahujících slabikotvornou likvidu, které vykazovaly více než jeden výskyt ve zpracovávaném materiálu. Struktury jsou seřazeny podle četnosti výskytu od nejfrekventovanějších k vzácnějším. Vedle každého vzorce je uveden počet jeho výskytů a přepočet na procenta vzhledem k počtu slov a taktů zahrnujících slabikotvorné likvidy (ve všech případech 158).

¹⁴ Vzhledem k převaze likvid coby slabikotvorných souhlásek jsou na některých místech v předkládaném textu všechny slabikotvorné konsonanty označovány za likvidy.

SLOVA						TAKTY					
vzorec bez ?	v.	%	vzorec s ?	v.	%	vzorec bez ?	v.	%	vzorec s ?	v.	%
CVCR	28	17,72	CVCR	28	17,72	CVCR	22	13,92	CVCR	22	13,92
CRCCV	18	11,39	CRCCV	18	11,39	CRCCV	10	6,33	CRCCV	10	6,33
CRCV	16	10,13	CRCV	16	10,13	CRCV	9	5,7	CRCV	9	5,7
CRCCVCV	8	5,06	CRCCVCV	8	5,06	CRCCVCV	8	5,06	CRCCVCV	8	5,06
CVCRCV	5	3,16	CVCRCV	5	3,16	CCRCCV	7	4,43	CCRCCV	7	4,43
CVCCRCVC	5	3,16	CVCCRCVC	5	3,16	CCRCV	7	4,43	CCRCV	7	4,43
CVCVCCR	5	3,16	CVCVCCR	5	3,16	CVCRCV	5	3,16	CVCVCCR	5	3,16
CRCVCVCV	5	3,16	CRCVCVCV	5	3,16	CVCVCCR	5	3,16	CVCRCV	5	3,16
CRCVC	4	2,53	CRCVC	4	2,53	CRCVCVCV	5	3,16	CRCVCVCV	5	3,16
CVCRCVCVC	3	1,9	CVCRCVCVC	3	1,9	CVCRCVCVC	4	2,53	CVCRCVCVC	4	2,53
CCCRCVCV	3	1,9	CCCRCVCV	3	1,9	CRCVC	4	2,53	CRCVC	4	2,53
CCCRCV	3	1,9	CCCRCV	3	1,9	CVCVCCRCVC	3	1,9	CVCVCCRCVC	3	1,9
CVCRC	3	1,9	CVCRC	3	1,9	CCCRCVCV	3	1,9	CCCRCVCV	3	1,9
CRCVCCCV	3	1,9	CRCVCCCV	3	1,9	CVCVCR	3	1,9	CRCVCCCV	3	1,9
VCRCV	2	1,27	CCRCV	2	1,27	CRCVCCCV	3	1,9	CVCVCR	3	1,9
VCCRCVCVCV	2	1,27	CVCVCR	2	1,27	CCVCR	2	1,27	VCCRCVCVCV	2	1,27
CVCVCR	2	1,27	CCRCVCV	2	1,27	CVCRCVCV	2	1,27	CRCVCCCV	2	1,27
CCRCVCV	2	1,27	CRCVCV	2	1,27	CVCRC	2	1,27	CCCRCV	2	1,27
CVCRCVCV	2	1,27	CRCVCCVCV	2	1,27	VCCRCVCVCV	2	1,27	CCRCVCV	2	1,27
CVCCRCVCV	2	1,27	CCVCR	2	1,27	CCCRCV	2	1,27	CRCVCV	2	1,27
VCRCVCCV	2	1,27	CVCRCVCV	2	1,27	CRCVCV	2	1,27	CCVCR	2	1,27
CRCCVCVC	2	1,27	CCVCVCR	2	1,27	CCRCVCV	2	1,27	CVCRC	2	1,27
CVCVCR	2	1,27	VCCRCVCVCV	2	1,27	VCVCR	2	1,27	VCCRCVCVCV	2	1,27
CCVCR	2	1,27	CVCCRCVCV	2	1,27	VCRCV	2	1,27	VCRCV	2	1,27
CCRCV	2	1,27	CVCVCR	2	1,27	CRCVCCVCV	2	1,27	ostatní (42)	1	26,58
CCVCVCR	2	1,27	VCRCV	2	1,27	CVCCRCVCVCV	2	1,27			
CRCVCCVCV	2	1,27	VCRCVCCV	2	1,27	ostatní (38)	1	24,05			
CRCVCV	2	1,27	CRCCVCVC	2	1,27						
ostatní (21)	1	13,29	ostatní (21)	1	13,29						
celkem	158	100	celkem	158	100	celkem	158	100	celkem	158	100

Tabulka 12: Fonotaktické vzorce slov a taktů obsahující likvidy a jejich procentuální zastoupení vzhledem k celkovému počtu vzorců slov a taktů s likvidami. Struktury jsou seřazeny od relativně častých k vzácným. Tabulka obsahuje všechny vzorce s likvidami, jejichž výskyt byl větší než 1. V. = počet výskytů. Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa.

Nejfrekventovanějším vzorcem obsahujícím slabikotvornou likvidu byla v korpusu v případě slov i taktů a se zohledněním rázu i bez něj struktura CVCR. Poměrně neočekávaně však nebylo hlavním důvodem četnosti tohoto vzorce l-ové participium mnohých frekventovaných sloves (*řekl, mohl*), ale zejména častý výskyt podstatných jmen tvořených tímto segmentem. Z 28

výskytů u slov tvořilo 15 z nich substantivum *vítr*. Tato situace tedy nastala vlivem materiálu, jímž byly zpravodajské relace obsahující i předpovědi počasí. Je pravděpodobné, že při zkoumání vzorku z jiného tematického okruhu bychom došli k poněkud odlišným výsledkům. Jak již bylo zmíněno, i tento nejčastější vzorec však tvoří pouze necelých 0,5 % ze všech výskytů vzorců v materiálu.

Dalšími relativně častými vzorci s likvidami byly s určitým odstupem CRCCV (18 výskytů u slov, 10 u taktů) a CRCV (16 výskytů u slov, devět u taktů). Všechny ostatní vzorce se v korpusu vyskytovaly méně než desetkrát u taktů i u slov. U vzorců slov bylo ze 158 vzorců 21 s jediným výskytem (13,29 % ze všech vzorců slov s likvidami s rázem i bez rázu); u vzorců taktů bylo v případě nezapočítávání rázu jako samostatné jednotky jedinečných výskytů 38 (24,05 % ze všech výskytů vzorců taktů bez rázu) a při zahrnutí rázu jsme jedinečných vzorců zaznamenali 42 (26,85 % ze všech vzorců taktů se zahrnutím rázu). Stejně jako v analýze všech vzorců dohromady je zde tedy zjevná tendence taktů k větší variabilitě způsobené vyšším počtem kombinatorických možností.

Vzorce sestávající z jednoho segmentu se případu slabikotvorných likvid netýkají, neboť v češtině neexistuje samostatná jednotka tvořená hláskami [r, l, m] nebo [n]. V češtině také nenajdeme samostatnou dvojčlennou kombinaci sestávající z likvidy a konsonantu (o výjimce by se dalo uvažovat jen v případě některých citoslovcí); v kombinaci likvidy s vokálem není jádrem likvida, ale vokál. Nejkratšími vzorci jsou tedy v češtině struktury tříčlenné, tyto v korpusu zastupuje jediný výskyt, a to struktura VCR výrazu *osm*. Pokud ovšem zohledníme ráz, hovoříme již o jednotce čtyřčlenné. V případě taktů byly nejkratší čtyřčlenné kombinace. Nejdelší vzorec byl zaznamenán u taktové i slovní jednotky *sedmdesátimilionového*; tedy u vzorce CVCRCVCVCVCVCVCVCVCV.

4.2.5.5 CVCV struktury a slovní druhy

V této části práce se budeme věnovat fonotaktickým vzorcům s ohledem na slovní druhy. Tabulka 13 obsahuje absolutní počty výskytů nejčastějších

CVCV struktur s jejich slovnědruhovou příslušností. Zde předkládáme pouze data pro vzorce bez rázu. Vzorce se zohledněním rázu se liší pouze v jediném řádku tabulky, a to v případě 8. nejčastější struktury, kterou je kombinace ?V, jež tvoří 177 případů u spojek a 60 případů u předložek, a tedy má celkem 237 výskytů. Sedm výskytů tvoří vokál bez rázu coby samostatné slovo. V níže uvedené tabulce je tedy těchto sedm výskytů samohlásky bez rázu připojeno k 237 slovům sestávajícím z rázu a vokálu.

vzorce bez ?	N	A	P	C	V	D	R	J	T	X	celkem
CV			159		79	1	330	32		2	603
CVCV	215	26	43	5	106	49	12	20	5		481
CVCCV	107	94	10	22	38	42	41				354
C							278				278
CVCVCV	171	27	3	21	42	5	1			1	271
CCVCV	113	18	38	21	46	16	13	1	1		267
CVCVC	149	15	11	25	12	34	11	1	6	1	265
V							64	180			244
CVC	65		37	7	22	35	24	9	11		210
CVCCVCV	88	41	6	4	29	7	1	1	1		178
CVCVCCV	80	47			13	4		3	1		148
CCV	13		11	61	16	10	28	2			141
CCVCCV	78	26	6		10	18	1				139
CCVC	49	1	11	12	3	35	13	4	1	1	130
CCVCVCV	58	22	2	8	29			5			124
CVCVCVCV	67	11		4	26	10	1				119
CVCCVC	62	25		2	13	9					111
CVCVCVC	73	6		10	14	2					105

Tabulka 13: Vzorce CVCV s ohledem na výskyt v rámci různých slovních druhů. Uvedené hodnoty platí pro vzorce bez rázu. Ve sloupcích je zobrazen počet výskytů daného vzorce jakožto N = substantiva, A = adjektiva, P = zájmena, C = číslovky, V = slovesa, D = příslovce, R = předložky, J = spojky, T = částice, X = nerozpoznaného slovního druhu. Citoslovce nejsou v korpusu obsažena. Uvedeny jsou ty vzorce, jejichž frekvence byla vyšší než 100 výskytů v materiálu. Souborem vzorců byly struktury bez hezitací a se speciálně značenými likvidami. Vzorce vyskytující se u daného slovního druhu ve 35 a více případech jsou označeny zeleným pozadím buňky. Kurzívou a tučně je vyznačen řádek, kterým se tato tabulka liší od tabulky pro vzorce se zohledněním rázu – ta obsahuje 60 výskytů vokálu s rázem jako předložky, 177 jako spojek, celkem má tedy slovo o jedné samohláске a rázu 237 výskytů.

Podobné srovnání nejfrekventovanějších fonotaktických vzorců provedl Bartoň a kol. v publikaci *Statistiky češtiny* (2009: 108-109). Stejně jako v korpusu SYN2005, obsahujícím cca 100 milionů slov, byla nejčastější slovní

fonotaktická struktura CV (603 výskytů). Na rozdíl od SYN2005 jsme však nedetkovali jediný výskyt substantiva, adjektiva nebo částice s touto strukturou. Dlužno podotknout, že i ve *Statistikách* byl u zmíněných slovních druhů výskyt tohoto vzorce řidší než např. u předložek nebo zájmen, kteréžto slovní druhy byly tvořeny touto strukturou nejčastěji (330 a 159 výskytů v tomto pořadí). CVCV struktura (481 výskytů) byla taktéž nejčastější u nás i v SYN2005 a vyskytovala se u všech slovních druhů; v našem materiálu zejména u podstatných jmen (215 výskytů) a sloves (106 výskytů). V dalším vzorci se náš výzkum od *Statistik češtiny* odlišuje: v SYN2005 byl na třetím místě dle četnosti vzorec CVC, který je v našem korpusu až devátý. Naším třetím nejčastějším vzorcem je CVCCV se 354 výskyty (ve *Statistikách češtiny* je čtvrtý v pořadí), který je zastoupen u všech slovních druhů (zejména substantiv a adjektiv) kromě spojek a částic. Následuje struktura tvořená samotným konsonantem (ve *Statistikách češtiny* sedmá nejčastější), která se vyskytovala pouze jako předložka (278 výskytů). V SYN2005 se samotná souhláska také často objevuje jako substantivum nebo adjektivum, což je patrně způsobeno výskytem zkratk zastupujících slova náležející k těmto slovním druhům. Na pátém místě co do frekvence slov je v našem materiálu i v SYN2005 struktura CVCVCV (271 výskytů), která je zastoupena u všech slovních druhů kromě spojek a částic a s jedním výskytem u předložky; výrazně nejfrekventovanější je u substantiv (171 výskytů). Šestým vzorcem v pořadí je CCVCV (shodně i v SYN2005) se 267 výskyty, který je zastoupen u všech slovních druhů (u spojek a částic se jednalo o jeden výskyt); nejfrekventovanější je u substantiv (113 výskytů). Další častou fonotaktickou strukturou je CVCVC s 265 výskyty (ve *Statistikách češtiny* je tento vzorec na osmém místě) – vidíme, že rozdíl mezi frekvencí vzorců C, CVCVCV, CCVCV a CVCVC je velmi malý. CVCVC tvoří opět nejčastěji substantiva (149 výskytů) a objevuje se v rámci všech slovních druhů v korpusu; pouze u konjunktív vykazuje jen jediný případ. Následuje struktura tvořená vokálem s 244 výskyty (příp. s rázem a vokálem s 237 výskyty), která je zastoupena jen u spojek (180 výskytů) a předložek (64 výskytů). V pořadí devátý vzorec CVC (210 výskytů) tvoří v našem korpusu všechny slovní druhy kromě adjektiv; nejfrekventovanější byl u substantiv (65 výskytů). Struktura CVCCVCV

(v SYN2005 devátá) se objevila v korpusu ve 178 případech a u všech slovních druhů, nejčastěji pak u podstatných jmen (88 výskytů). Pouze jeden výskyt měla u předložek, spojek a částic. V prvních deseti nejčastějších vzorcích se oproti *Statistikám češtiny* objevila struktura V (?V) a naopak zde nefiguroval vzorec CCVC (v našem korpusu až na 14. místě).

Ve *Statistikách češtiny* není jednoznačně uvedeno, zda je ráz považován *vždy* za souhlásku (v úvodu ke kapitole Fonologie je zmíněn pouze jako „neznělá laryngální explozíva předcházející vokál po pauze“ (Bartoň et al., 2009: 25) a např. v analýze frekvence hlásek a dvojic a trojic hlásek ve slovech obecně ho autoři berou v potaz zcela jistě; ve zpracovávání hlásek a jejich dvojic a trojic na začátku a konci slova však oproti tomu nikde ráz uveden není, a to ani v příkladech konkrétních hlásek a skupin hlásek). Jelikož je SYN2005 korpus psané češtiny, není také zcela jasné, podle jakých pravidel přesně byly případně rázy do textu dosazovány – v našem materiálu jsou rázy zaznamenány pouze tam, kde se v mluvených projevech skutečně vyskytly.

V souhrnné tabulce nejfrekventovanějších vzorců SYN2005 není explicitně řečeno, jak a jestli je zde ráz považován za konsonant a značen C, ale na základě faktu, že se ve výčtu nejčastějších vzorců neobjevila slova sestávající z vokálu (resp. z rázu a vokálu), je pravděpodobné, že v případě této analýzy byl ráz skutečně započítán mezi souhlásky. Kvůli této možnosti tedy ještě pro srovnání uvádíme tabulku 14 – tato tabulka čerpá z našeho materiálu a ráz je v ní značen jako konsonant (C). Obsahuje všechny vzorce, jejichž výskyt přesáhl 100 případů v našem korpusu. Využitými vzorci byly struktury ze souboru dat bez vyznačených hezitací a se speciálně odlišenými likvidami.

Oproti vzorcům, kde byl ráz považován za specifickou jednotku, nebo kde se ráz nebral v potaz, vidíme několik rozdílů. Nejsilněji zastoupeným vzorcem je opět CV, tentokrát s 840 výskytů – ke vzorcům CV se totiž připojily předložky a spojky sestávající z rázu a samohlásky. Druhou strukturou v pořadí je CVCV (528 výskytů), jejíž počet výskytů zvýšila zejména přítomnost spojek složených z rázu a struktury VCV. Třetím v pořadí zůstává vzorec CVCCV (367 výskytů). Poté se již pořadí podle frekvence mění – následují struktury CVCVCV (316 výskytů) a CVC (285 výskytů), které obě předcházejí slova,

jejichž hláskovou stavbu tvoří samotný konsonant (278 výskytů). S malým početním rozdílem následují struktury CVCVC (275 výskytů) a CCVCV (267 výskytů). Na devátém a desátém místě v pořadí jsou vzorce CVCCVCV (223 výskytů) a CVCVCCV (164 výskytů). V tomto výčtu nejčastějších fonotaktických vzorců našeho materiálu tak pochopitelně zcela absentuje struktura V a v prvních deseti nejčastějších strukturách se až na jednu výjimku objevily všechny vzorce, které byly nejfrekventovanější v SYN2005. Touto výjimkou je struktura CCVC, která byla v našem materiálu (pokud se ráz ve vzorcích považoval za konsonant) na 13. místě.

vzorce (? jako C)	N	A	P	C	V	D	R	J	T	X	celkem
CV			159		79	1	390	209		2	840
CVCV	219	26	43	5	106	49	14	58	8		528
CVCCV	117	94	10	25	38	42	41				367
CVCVCV	209	27	3	21	46	5	4			1	316
CVC	69		37	7	22	49	49	30	22		285
C							278				278
CVCVC	159	15	11	25	12	34	11	1	6	1	275
CCVCV	113	18	38	21	46	16	13	1	1		267
CVCCVCV	124	42	6	5	34	7	2	1	2		223
CVCVCCV	88	48			17	7		3	1		164
CVCVCVCV	83	12		4	31	10	1				141
CCV	13		11	61	16	10	28	2			141
CCVCCV	78	26	6		10	18	1				139
CCVC	49	1	11	12	3	35	13	4	1	1	130
CVCCVC	68	25		2	14	17		1			127
CCVCVCV	58	22	2	8	29			5			124
CVCVCVC	83	6		10	19	2			1		121
CVCCVCVCV	55	14			44	1					114

Tabulka 14: Vzorce CVCV s ohledem na výskyt v rámci různých slovních druhů. Ráz byl v tomto případě považován za konsonant a značen C. Ve sloupcích je zobrazen počet výskytů daného vzorce jakožto N = substantiva, A = adjektiva, P = zájmena, C = číslovky, V = slovesa, D = příslovce, R = předložky, J = spojky, T = částice, X = nerozpoznaného slovního druhu. Citoslovce nejsou v korpusu obsažena. Uvedeny jsou pouze ty vzorce, jejichž frekvence byla vyšší než 100 výskytů v materiálu. Vzorce vyskytující se u daného slovního druhu ve 35 a více případech jsou označeny zeleným pozadím buňky.

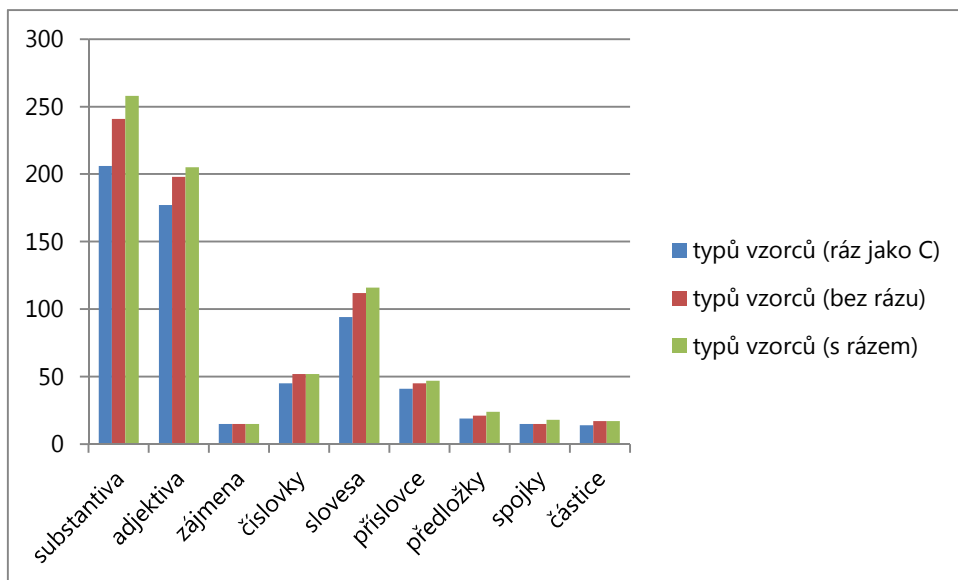
Různé slovní druhy měly také rozdílnou variabilitu CVCV vzorců. V tabulce 15 uvádíme, kolik jednotlivých typů vzorců se vystřídalo u každého slovního druhu (souborem vzorců byly struktury bez hezitací a se speciálně značenými likvidami).

slovní druh	typů vzorců (ráz jako C)	typy : výskyty	typů vzorců (bez rázu)	typy : výskyty	typů vzorců (s rázem)	typy : výskyty
substantiva	206	0,079	241	0,092	258	0,099
adjektiva	177	0,193	198	0,216	205	0,224
zájmena	15	0,043	15	0,043	15	0,043
číslovky	45	0,148	52	0,171	52	0,171
slovesa	94	0,113	112	0,135	116	0,140
příslowce	41	0,108	45	0,118	47	0,123
předložky	19	0,022	21	0,024	24	0,028
spojky	15	0,047	15	0,047	18	0,056
částice	14	0,222	17	0,270	17	0,270
nerozpoznané	5	0,833	5	0,833	5	0,833

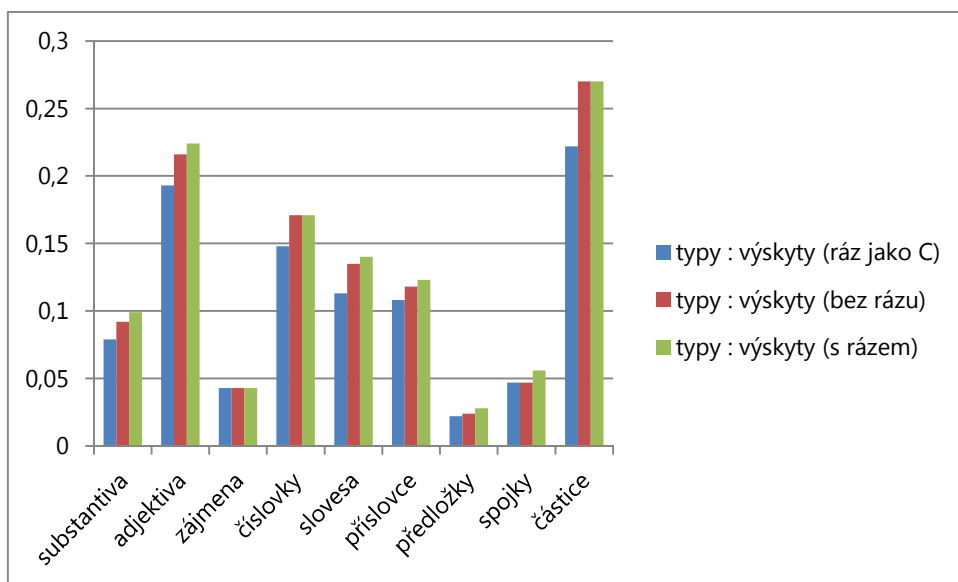
Tabulka 15: Fonotaktická variabilita slovních druhů. Ve sloupcích je zobrazen počet typů CVCV struktur vzhledem ke slovnědruhové příslušnosti. Citoslovce nejsou v korpusu obsažena. Tabulka obsahuje hodnoty pro tři soubory vzorců téhož materiálu: v prvním souboru se ráz považoval za konsonant, ve druhém se nebral v potaz vůbec a ve třetím se ráz ve vzorcích značil zvláštní značkou. Sloupce označené *typy : výskyty* obsahují poměrné vyjádření variability struktur vzhledem k zastoupení jednotlivých slovních druhů v korpusu (viz tabulka 6); uvádějí, kolik typů vzorců připadá na jeden výskyt vzorce daného slovního druhu. Čím blíže je daná hodnota číslu 1, tím je variabilita vyšší. Hodnoty jsou zaokrouhleny na tři desetinná místa.

Tabulka 15 a grafy 4 a 5 srovnávají fonotaktickou variabilitu v absolutních počtech specifických typů vzorců u jednotlivých slovních druhů a posléze tyto počty vztahují na celkové zastoupení daných slovních druhů v korpusu. Na první pohled vidíme poměrně značnou rozdílnost těchto hodnot: zatímco podle absolutních počtů jsou nejvíce variabilní substantiva, podle poměrného vyjádření mají variabilitu poměrně nízkou. Podstatných jmen je v korpusu 2607 a tvoří 39,27 % všech slov v materiálu. Přídavných jmen je oproti tomu v korpusu výrazně méně, totiž 915, a tvoří 13,78 % slov v korpusu. Typů vzorců u podstatných jmen je 206 a u jmen přídavných 177 – můžeme si povšimnout, že rozdíl mezi těmito absolutními počty není zdaleka tak velký, jaký bychom čekali, pokud vezmeme v potaz zastoupení substantiv a adjektiv v korpusu, a toto se odráží i na poměru typů vzorců vůči všem výskytům daného slovního druhu – podstatná jména vykazují hodnotu 0,079 a přídavná jména 0,193 – variabilita adjektiv co do fonotaktických struktur je tedy podle tohoto vyjádření výrazně vyšší. Nejvyšší variabilitu vykazovaly částice, které však v korpusu tvořily necelé 1 % všech slovních druhů. Nízkou variabilitu struktur v absolutních počtech vykazovala zájmena (5,24 % korpusu), předložky (12,97 % korpusu) a spojky (4,85 % korpusu) a tento fakt potvrdil i výpočet poměru typů proti výskytům. Oproti tomu číslovky (4,58 %

materiálu) a příslovce (5,74 % materiálu) měly v absolutních počtech variabilitu nízkou; v poměrném vyjádření se však ukázala jako vyšší. Slovesa (12,52 % materiálu) byla v absolutním počtu specifických vzorců na třetím místě, avšak v poměrném vyjádření zvláště vysokou variabilitu nevykazovala.



Graf 4: Fonotaktická variabilita slovních druhů – absolutní počty. Na ose x jsou jednotlivé slovní druhy, na ose y počet specifických typů vzorců, ve kterých byl ráz považován za konsonant, nebo nebyl do analýz zahrnut vůbec, nebo byl považován za specifickou jednotku.



Graf 5: Fonotaktická variabilita slovních druhů – poměrné vyjádření variability typů vzorců vzhledem k celkovým počtům výskytů struktur jednotlivých slovních druhů v materiálu (tabulka 6). Na ose x jsou slovní druhy, na ose y poměrná variabilita vzorců vzhledem k zastoupení daných slovních druhů v korpusu. Zpracovávány byly tři soubory struktur: první, ve kterém byl ráz považován za konsonant, druhý, v němž nebyl ráz do analýz zahrnut vůbec, třetí, kde byl ráz považován za specifickou jednotku.

V případě, že bychom srovnali nejfrekventovanější vzorec pro každý slovní druh v korpusu psané češtiny SYN2005 s naším materiálem, který byl omezenější, dospěli bychom vždy k podobným výsledkům. Největší množství shod jsme našli, pokud jsme výsledky z korpusu SYN2005 porovnávali s naším souborem dat, ve kterých byl ráz považován za konsonant (takto patrně postupovali i autoři *Statistik češtiny*; v publikaci však tento fakt není explicitně uveden).

Nejčastější vzorce pro jednotlivé slovní druhy v psaném korpusu SYN2005 a v materiálu zpracovávaném v rámci této práce uvádíme v tabulce 16. Východiskem nám jsou data s rázem považovaným za konsonant, data bez zahrnutí rázu mezi hlásky a data, v nichž byla pro ráz použita speciální značka ?. Našimi podklady byly soubory dat se speciálně vyznačenými slabikotvornými likvidami a bez zohlednění hezitačních zvuků.

slovní druh	nejfrekventovanější vzorce – <i>Statistiky češtiny</i> (2009)	nejfrekventovanější vzorce – ráz jako C	nejfrekventovanější vzorce – bez rázu	nejfrekventovanější vzorce – s rázem
substantiva	CVCV	CVCV	CVCV	CVCV
adjektiva	CVCCV	CVCCV	CVCCV	CVCCV
zájmena	CV	CV	CV	CV
číslovky	CCV	CCV	CCV	CCV
slovesa	CVCV	CVCV	CVCV	CVCV
příslovce	CVC	CVCV, CVC	CVCV	CVCV
předložky	CV	CV	CV	CV
spojky	CV	CV	V	?V
částice	CVC	CVC	VC, CVC	?VC, CVC
citoslovce	CVC	-	-	-

Tabulka 16: Nejčastější vzorce pro dané slovní druhy; srovnání prezentovaného zkoumání s výzkumem uvedeným v publikaci *Statistiky češtiny* (Bartoň et al., 2009). Pokud je stejný počet výskytů u dvou vzorců, uvádíme oba dva. Vzorce ze *Statistik češtiny* jsou porovnávány s daty z našeho korpusu – zvláště se souborem struktur s rázem značeným C, bez započítání rázu a s rázem značeným speciálním symbolem. Citoslovce nebyla v našem korpusu obsažena.

Z tabulky je patrné, že ačkoli byl náš korpus podstatně omezenější než SYN2005, vykazují oba výzkumy většinou shodu. Pokud jsme považovali ráz za konsonant, zjistili jsme pouze jeden rozdíl, a to u příslovcí, kde byla podle *Statistik češtiny* nejčastější struktura CVC, avšak u nás kromě této i vzorec CVCV. CVCV strukturu představovaly zejména jednotky *také* a *dále*; CCV potom *pak*, *jak*, *už* apod. Ostatní soubory vykazují rozdílů více – přetrvává rozdíl u příslovcí, který je vyhraněnější, protože nejčastějším vzorcem

zůstává samotná struktura CVCV – CVC zde nefiguruje, neboť bez započítání rázu mezi konsonanty se nyní nepočítá s posílením tohoto vzorce příslovcem *už*. Nejfrekventovanějším vzorcem u spojek v souborech s rázem a bez rázu jsou struktury ?V a V. Ze *Statistik češtiny* v podstatě vyplývá totéž, jelikož autoři patrně považovali ráz za běžný konsonant. Náš výzkum pracuje s jemnějším rozlišením rázu, a tedy máme možnost zaznamenat rozdíl např. mezi nejčastější strukturou předložky a spojky. Pro částice v našem materiálu platí, že nejvíce vzorců tvořilo (?)VC a CVC – zde vidíme vyrovnaný podíl jednotek s rázem a bez rázu, které výzkum na SYN2005 nebyl schopen zachytit. Citoslovce v našem materiálu, sestávajícím z připravených projevů profesionálních mluvčích, nefigurovala, a tak zde srovnání provést nelze.

K uvedeným shodám a rozdílům je nutné podotknout, že korpus použitý v tomto výzkumu byl mnohem menší než SYN2005, a tak v některých případech dochází k odlišnostem. Abychom tyto rozdíly potvrdili, nebo vyvrátili, museli bychom výzkum rozšířit na větší materiál. Považujeme je tedy za podněty, které mohou poskytnout vodítka k dalšímu pokračování výzkumu.

4.2.6 Kombinace konsonantů, vokálů, likvid a rázů v různých částech jednotky

V následujících podkapitolách uvedeme počty konsonantů, vokálů, likvid a rázů a jejich dvojčlenných a trojčlenných kombinací na různých místech v jednotce. Nejprve se budeme věnovat výskytu jednotlivých hlásek a v rámci jednotky a jejich krajním pozicím v jednotce; poté přejdeme k dvojicím a trojicím hlásek a jejich výskytu a pozici v jednotce. Materiálem nám je soubor CVCV vzorců se speciálně značenými likvidami a s odstraněnými hezitacemi – tento korpus obsahoval 6639 slovních jednotek, 5368 taktů, 384 typů slovních fonotaktických vzorců bez započítání rázu a 413 s jeho zahrnutím mezi hlásky, 493 typů taktových fonotaktických vzorců bez zohlednění rázu a 559 typů struktur s rázem. Data zpracováváme pro soubor se zohledněním rázu i bez něj a rámcově srovnáme naše výsledky s poznatky uvedenými v publikaci *Statistiky češtiny* (Bartoň et al., 2009).

4.2.6.1 Četnost jednotlivých C, V, R a ? v rámci slov a taktů

4.2.6.1.1 Libovolná pozice v jednotce

Ve zkoumaném materiálu, který zahrnoval 384 typů slovních vzorců bez zahrnutí rázu, jsme našli stejný počet typů vzorců, kde se vyskytoval konsonant i kde se vyskytoval vokál – de facto šlo o všechny struktury s výjimkou těch, které byly tvořeny samotnou samohláskou nebo souhláskou. Výskytů bylo ale celkově více u vzorců s konsonanty (což se ukázalo podobně i ve výzkumu na korpusu SYN2005 v roce 2009). Pokud jsme zohlednili ráz, zahrnovaly vzorce s vokály o jeden typ více než struktury se souhláskou – konsonant se nevyskytoval kromě slov složených z jedné samohlásky také v jednotkách, které byly tvořeny rázem a vokálem. Slabikotvorné likvidy byly součástí 49 typů vzorců a 158 struktur celkově. Rázy tvořily 111 typů vzorců a vyskytovaly se celkem v 867 slovech. Nejfrekventovanějšími vzorci byly CV pro konsonanty a vokály, CVCR pro likvidy a ?V pro rázy.

Tutéž analýzu jsme provedli i pro soubor dat, který byl rozčleněn na takty. Mezi takty byla celkově vyšší četnost typů jednotek, jak jsme již zmínili výše. Soubor taktů bez rázu i s rázem vykazoval nejvyšší frekvenci vokálů v typech taktů – samohláska byla v každém typu taktů; náš korpus tedy nezahrnoval ani jeden takt, kde by nebyl přítomen vokál. Zatímco taktem mohl být samostatný vokál nebo ráz s vokálem, u samotného konsonantu tato možnost nebyla, a tedy bylo struktur se souhláskami vždy o dvě méně; objevily se však ve více než 99 % ze všech typů. Výskytů vzorců s vokály bylo oproti slovním vzorcům více než struktur obsahujících konsonanty. Typů vzorců s likvidami bylo o něco více než ve slovním korpusu; jejich procentuální zastoupení vůči všem typům taktů však bylo velmi podobné. Výrazně více jsme však našli vzorců s rázy – jejich typů bylo v korpusu taktů 207. Nejčastějšími vzorci byly pro konsonanty a vokály CVCV, pro likvidy CVCR a pro ráz ?V – můžeme si povšimnout, že ačkoli je ve slovních jednotkách nejčastější vzorec CV, v taktových jsou slova s takovou strukturou spíše přidružována k dalším jednotkám, než aby stála samostatně. Samotný vzorec CV má mezi takty 55 výskytů, což ovšem poukazuje na fakt, že ačkoli v češtině existuje tendence ke

sdužování slabik do vyšších rytmických jednotek, nelze pomíjet možnost vytváření taktů jednoslabičných, a to ani v projevu profesionálních mluvčích, kteří čtou připravený text.

Přesné hodnoty četností C, V, R a ? v libovolné pozici ve slovech a taktech uvádíme v tabulkách 17a a 17b:

SLOVA											
bez rázu						s rázem					
	typů	% ze všech typů	výskytů	% ze všech slov	vzorec		typů	% ze všech typů	výskytů	% ze všech slov	vzorec
C	383	99,74	6395	96,32	CV	C	411	99,52	6395	96,32	CV
V	383	99,74	6361	95,81	CV	V	412	99,76	6361	95,81	CV
R	49	12,76	158	2,38	CVCR	R	49	11,86	158	2,38	CVCR
						?	111	26,88	867	13,06	?V

Tabulka 17a

Tabulka 17a a 17b: Četnost výskytu souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) v libovolné pozici ve všech slovních (17a) a taktových (17b) typech i výskytech. Procenta ze všech typů jsou počítána ze základů 384 a 413 (17a) a 493 a 559 (17b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (17a/17b) z celkového počtu jednotek – 6639 (17a) a 5368 (17b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	% ze všech typů	výskytů	% ze všech taktů	vzorec		typů	% ze všech typů	výskytů	% ze všech taktů	vzorec
C	491	99,59	5303	98,79	CVCV	C	557	99,64	5303	98,79	CVCV
V	493	100	5368	100	CVCV	V	559	100	5368	100	CVCV
R	64	12,98	158	2,94	CVCR	R	66	11,81	158	2,94	CVCR
						?	207	37,03	853	15,89	?V

Tabulka 17b

4.2.6.1.2 Pozice na začátku jednotky

V pozici na začátku jednotky byly výrazně nejvíce zastoupeny konsonanty – i v tomto faktu se shodujeme se *Statistikami češtiny* (Bartoň et al., 2009);

vzorci začínající souhláskou tvořily cca 75 % typů vzorců a cca 85 % všech slov bez započítání rázu; vokály na začátku jednotky byly pouze v cca 25 % slovních typů bez rázu a v 6 % s jeho zahrnutím mezi hlásky; pokud ráz zohledníme, většina slov, která v předchozím případě začínají samohláskou, nyní mají jako počáteční hlásku právě ráz (cca 23 %); nejfrekventovanější vzorec pro vokál v předchozím případě, tedy samotný V, se ve většině situací stane jednotkou ?V. Procento výskytů taktů, které i při zahrnutí rázu začínají vokálem, je velmi nízké – takové struktury tvoří pouze cca 1 % všech slov. Kromě těchto jevů se nesetkáme ani s jedním slovem, které by začínalo slabikotvornou likvidou.

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
C	289	75,26	5728	86,28	CV	C	295	71,43	5728	86,28	CV
V	95	24,74	911	13,72	V	V	25	6,05	68	1,02	VC
R	0	0	0	0		R	0	0	0	0	
						?	93	22,52	843	12,7	?V

Tabulka 18a

Tabulka 18a a 18b: Četnost výskytu souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na začátku všech slovních (18a) a taktových (18b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (18a) a 493 a 559 (18b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (18a/18b) z celkového počtu jednotek – 6639 (18a) a 5368 (18b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
C	356	72,21	4574	85,21	CVCV	C	396	70,84	4574	85,21	CVCV
V	137	27,79	794	14,79	V	V	32	5,72	55	1,02	VCVCV
R	0	0	0	0		R	0	0	0	0	
						?	131	23,43	739	13,77	?V

Tabulka 18b

Takty v našem materiálu také začínají většinou souhláskou, procentuální zastoupení typů s konsonantem a vokálem na začátku vykazuje nepatrně menší rozdíl, ale přesto můžeme konstatovat, že konsonantem začíná vždy více než 70 % typů struktur, a to jak u slov, tak u taktů a se započítáním rázu mezi hlásky i bez něj. Při zohlednění rázu začíná samohláskou pouze necelých 6 % taktových typů a 1,02 % všech taktů. Pozoruhodný je ovšem fakt, že nejfrekventovanějším taktovým vzorcem je pro V bez zohlednění rázu samotné V a pro vzorce s rázem ?V, a to nejen u slov, ale i u taktů – zde tedy opět vidíme realizaci možnosti vytváření jednoslabičných taktů. Mezi slovy, která byla tvořena pouze vokálem a neobsahovala ráz (sedm případů), nenajdeme ani jedno, které by bylo samostatným taktem; zdá se, že realizace nebo nerealizování rázu na začátku slova má vliv na vnímání taktových hranic u jednoslabičných slov. Co se týče likvid, ani v rámci taktů nenajdeme jedinou jednotku, která by začínala slabikotvornou likvidou.

4.2.6.1.3 Pozice na konci jednotky

Na koncích jednotek u slov i taktů byly celkově (tj. v typech i výskytech) četnější vokály, ale jejich převaha byla slabší než u konsonantů na začátcích jednotek. Typy slov i taktů bez zahrnutí rázu i s jeho započítáním mezi hlásky končí na vokál v cca 62-63 % případů; konsonant na koncové pozici tvoří cca 36-37 % všech typů vzorců. Likvidy jsou součástí cca 2 % vzorců taktů i slov s rázem i bez něj.

Větší rozdíl než mezi typy vzorců je mezi celkovými výskyty; vokály jsou v 71-72 % slov a taktů, konsonanty v cca 29 % vzorců slov a 27 % vzorců taktů (vše s rázem i bez něj) a podíl typů vzorců končících likvidami se pohybuje kolem 0,6-0,7 % všech výskytů vzorců daných jednotek. Nejfrekventovanější strukturou slova zakončeného konsonantem je samotná souhláska – z tohoto faktu je zřejmé, že neslabičné předložky měly v našem korpusu nezanedbatelný podíl. Přesné údaje o zastoupení konsonantů, vokálů a slabikotvorných likvid na konci jednotek jsou uvedeny v tabulkách 19a a 19b.

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
C	140	36,46	1896	28,56	C	C	148	35,84	1896	28,56	C
V	237	61,72	4703	70,84	CV	V	258	62,47	4703	70,84	CV
R	7	1,82	40	0,6	CVCR	R	7	1,69	40	0,6	CVCR
						?	0	0	0	0	

Tabulka 19a

Tabulka 19a a 19b: Četnost výskytu souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na konci všech slovních (19a) a taktových (19b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (19a) a 493 a 559 (19b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (19a/19b) z celkového počtu jednotek – 6639 (19a) a 5368 (19b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
C	178	36,11	1451	27,03	CVCVC	C	196	35,06	1451	27,03	CVCVC
V	305	61,87	3880	72,28	CVCV	V	352	62,97	3880	72,28	CVCV
R	10	2,03	37	0,69	CVCR	R	11	1,97	37	0,69	CVCR
						?	0	0	0	0	

Tabulka 19b

4.2.6.2 Četnost dvojic C, V, R a ? v rámci slov a taktů

V tomto oddíle shrneme četnosti dvojic hlásek v rámci slov a taktů a na jejich začátku i na konci. V tabulkách nebudou uváděny kombinace hlásek vykazující nulový výskyt v našem materiálu.

4.2.6.2.1 Libovolná pozice v jednotce

Nejfrekventovanější dvojicí hlásek u slov bez zahrnutí rázu byla kombinace CV, která se vyskytovala v 99 % všech typů vzorců. Podobně častou kombinací byla dvojice VC (97 % typů vzorců). Třetí silně zastoupenou kombinací byla skupina CC (86 % typů vzorců). Mezi těmito třemi dvojicemi

byl však velký rozdíl ve frekvenci kombinací u všech výskytů: Zatímco CV a VC se objevily v 91 a 80 % případů, shluk CC byl přítomen pouze v 55 % všech slov.

Kromě těchto tří frekventovaných dvojic se v korpusu objevily dvojice VV, CR a RC, které byly zastoupeny podobně – zahrnovaly 11-13 % typů vzorců slov a cca 1-2 % ze všech vzorců slov bez rázu. Z těchto měla nejvyšší zastoupení co do výskytů dvojice CR (158 výskytů). Pořadí prvních tří dvojic co do výskytů je shodné i se *Statistikami češtiny* (2009), jako čtvrtá v pořadí v SYN2005 se ale objevila kombinace VV. Je však nutné zmínit, že tamní údaje neuvádějí podrobnosti o kombinacích se slabikotvornou likvidou a ráz je ve zmíněné publikaci patrně považován za souhlásku.

Se zahrnutím rázu byla situace podobná; nižší výskyt však zaznamenala dvojice VV – v rámci zkratk (např. názvů některých organizací, politických stran apod.) totiž často ráz tvořil mezi vokály předěl. Kromě zmíněných kombinací se u slov ještě objevily dvojice ?V (ve 27 % všech typů slovních vzorců a 13 % všech slovních vzorců), C? (např. ve zkratkách *ČSSD*, *KCNA* nebo na morfematickém švu – *nejočekávanější*) a V? (tentýž případ). Z kombinací s rázem měla nejvyšší frekvenci dvojice ?V.

Podobné hodnoty frekvence dvojic hlásek jako u slov bez rázu nalezneme i u taktů bez zohlednění rázu. Jsou zde ale i některé rozdíly – např. dvojice CC, která se v rámci všech slov vyskytuje pouze v 55 % vzorců, je zde přítomna v 69 % taktových struktur.

Kromě již zmíněných vzorců zde najdeme i poměrně překvapivě dvojici RV, která ale může v taktu vzniknout zcela přirozeně, pokud jedno slovo v taktu končí na slabikotvornou likvidu a další slovo, které je obsaženo v témže taktu, začíná na vokál. Tento výskyt však byl v korpusu pouze jeden; šlo o spojení *sedm až* v úseku STR04-15-03. Při zohlednění rázu tento vzorec ale nenalezneme, jelikož se mezi jednotkami vyskytoval ráz; v tomto souboru je ale tedy možné nalézt ojedinělou dvojici R?. Opět zde pozorujeme dvojici V? a úbytek typů i výskytů obsahujících strukturu VV.

Přesné informace o četnosti dvojic kombinací v libovolné pozici v jednotce a přepočet na procenta vzhledem ke všem typům vzorců a všem výskytům slov a taktů jsou uvedeny v tabulkách 20a a 20b.

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	380	98,96	6029	90,81	CV	CV	405	98,06	6027	90,78	CV
VC	373	97,14	5322	80,16	CVCV	VC	401	97,09	5322	80,16	CVCV
VV	49	12,76	91	1,37	CVCVVCV	VV	35	8,47	61	0,92	CVCVVCV
CC	329	85,68	3631	54,69	CVCCV	CC	347	84,02	3631	54,69	CVCCV
CR	49	12,76	158	2,38	CVCR	CR	49	11,86	158	2,38	CVCR
RC	42	10,94	118	1,78	CRCCV	RC	42	10,17	118	1,78	CRCCV
						7V	111	26,88	867	13,06	7V
						C7	6	1,45	7	0,11	CV7VC7VCCV
						V7	21	5,08	30	0,45	7VCV7V, CV7VCVCCV

Tabulka 20a

Tabulka 20a a 20b: Četnost výskytu dvojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (7) v rámci všech slovních (20a) a taktových (20b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (20a) a 493 a 559 (20b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (20a/20b) z celkového počtu jednotek – 6639 (20a) a 5368 (20b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	490	99,39	5277	98,3	CVCV	CV	552	98,75	5274	98,25	CVCV
VC	481	97,57	5155	96,03	CVCV	VC	546	97,67	5155	96,03	CVCV
VV	88	17,85	160	2,98	CVCVCV, CVCCVCVVCV	VV	43	7,69	69	1,29	CVCVVCV
CC	426	86,41	3696	68,85	CVCCV	CC	469	83,9	3696	68,85	CVCCV
CR	64	12,98	158	2,94	CVCR	CR	66	11,81	158	2,94	CVCR
RC	53	10,75	120	2,24	CRCCV	RC	54	9,66	120	2,24	CRCCV
RV	1	0,2	1	0,02	CVCRVC	7V	207	37,03	853	15,89	7V
						C7	39	6,98	50	0,93	C7VCVVCV
						V7	59	10,55	91	1,7	CV7VCV
						R7	1	0,18	1	0,02	CVCR7VC

Tabulka 20b

4.2.6.2.2 Pozice na začátku jednotky

Počátek slova i taktu je co do kombinací hlásek poněkud omezenější – u slov ani taktů nenajdeme jediný případ, kdy by jednotka začínala slabikotvornou likvidou. V rámci souboru dat se zohledněním rázu také nenalezneme slovo ani takt začínající dvojicí VV (bez zohlednění rázu se v rámci slov tento případ objevil dvakrát – v úsecích KOB02-08-02 a KOB02-09-02 ve zkratkách OSN a USA v tomto pořadí).

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	168	43,75	3836	57,78	CV	CV	174	42,13	3836	57,78	CV
VC	92	23,96	665	10,02	VC	VC	24	5,81	61	0,92	VC
VV	2	0,52	2	0,03	VVCVC, VVCV	CC	105	25,42	1548	23,32	CCVCV
CC	105	27,34	1548	23,32	CCVCV	CR	15	3,63	66	0,99	CRCCV
CR	15	3,91	66	0,99	CRCCV	?V	93	22,52	843	12,7	?V

Tabulka 21a

Tabulka 21a a 21b: Četnost výskytu dvojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na začátku všech slovních (21a) a taktových (21b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (21a) a 493 a 559 (21b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (21a/21b) z celkového počtu jednotek – 6639 (21a) a 5368 (21b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	211	42,8	3047	56,76	CVCV	CV	225	40,25	3013	56,13	CVCV
VC	124	25,15	716	13,34	VCCVCV	VC	32	5,72	55	1,02	VCVCV
VV	12	2,43	15	0,28	VCCVCVCV	CC	136	24,33	1481	27,59	CCVCV
CC	134	27,18	1481	27,59	CCVCV	CR	11	1,97	46	0,86	CRCCV
CR	11	2,23	46	0,86	CRCCV	?V	131	23,43	739	13,77	?V
						C?	24	4,29	34	0,63	C?VCVCVCV

Tabulka 21b

U slov i taktů je výrazně nejfrekventovanější dvojicí na začátku jednotky CV (40-44 % typů a 56-58 % všech slovních a taktových vzorců). Dalšími častými dvojicemi jsou typy CC (24-27 % u slov i taktů) a VC (24-25 % ze všech typů u slov i taktů bez rázu; tvoří pouze cca 1 % slov a taktů s rázem) a ?V (cca 23 % typů vzorců u slov i taktů s rázem). Slabší zastoupení pak mají dvojice VV a CR u jednotek bez zahrnutí rázu a VC, CR a C? (v rámci taktů; např. výrazy *s občasnými, v Afghánistánu*) s jeho zohledněním. *Statistiky češtiny* (2009) uvádějí co do četnosti dvojic hlásek na začátku slova pořadí CV – CC – VC – VV. Pořadí prvních tří dvojic odpovídá i naší analýze, ale pokud od vokálů odlišíme likvidy, na čtvrtém místě se objeví dvojice CR, a pokud zohledníme ráz, dvojice VV se v našem materiálu na začátku slova nebo taktu neobjeví vůbec. Podrobnější informace o slabikotvorných likvidách však publikace neobsahuje. Přesné údaje naší analýzy uvádíme v tabulkách 21a a 21b.

4.2.6.2.3 Pozice na konci jednotky

Na koncích slov i taktů se převážně objevovala dvojice CV (cca 60-61 % typů vzorců a 67-71 % všech výskytů slov i taktů), poměrně frekventovaná byla i kombinace VC (29-30 % typů vzorců a 23-25 % všech výskytů slov i taktů). Všechny ostatní dvojice byly výrazně méně zastoupeny. Dvojice CC se objevila na konci 5-6 % typů vzorců a cca 1-2 % všech výskytů slov a taktů; při zohlednění rázu větší výskyt vykazovala také dvojice ?V; tato byla zajímavá proto, že byla obsažena pouze v pěti typech vzorců slov a v 10 typech vzorců taktů, ale na konci slov se celkem objevila 243krát – tento jev připisujeme častému výskytu spojek a předložek složených právě z této kombinace (mj. je nejčastějším vzorcem spojek právě ?V). Taktů s touto dvojicí na konci je méně, celkem 76 výskytů, nicméně stále podíl takových taktů přesahuje 1 % v korpusu. Pořadí CV – VC – CC bylo podobné i v korpusu SYN2005, ale opět není shoda ve čtvrté nejfrekventovanější dvojici – v našem výzkumu je u slov bez zohlednění rázu čtvrtou kombinací v pořadí dvojice CR, v SYN2005 VV. Dvojice CR má na konci slov 40 výskytů a na konci taktů 37; typů slov s touto kombinací na konci je sedm a typů taktů 10 bez zohlednění rázu a 11 s rázem. Ostatní dvojice, tedy RC a VV, jsou v korpusu

zastoupeny minimálně. Kombinaci RC reprezentovaly jednotky typu *veletrh, návrh, souhrn* apod.; dvojici VV obsahovaly převážně zkratky a slova cizího původu (např. *ODA, Korea, videa* apod.). Podrobné údaje jsou obsaženy v tabulkách 22a a 22b.

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	233	60,68	4451	67,04	CV	CV	250	60,53	4448	67	CV
VC	113	29,43	1526	22,99	CVCVC	VC	120	29,06	1526	22,99	CVCVC
VV	3	0,78	8	0,12	CVCVV	VV	2	0,48	5	0,08	CVCVV
CC	24	6,25	87	1,31	CVCC	CC	25	6,05	87	1,31	CVCC
CR	7	1,82	40	0,6	CVCR	CR	7	1,69	40	0,6	CVCR
RC	2	0,52	5	0,08	CVCRC	RC	2	0,48	5	0,08	CVCRC
						7V	5	1,21	243	3,66	7V

Tabulka 22a

Tabulka 22a a 22b: Četnost výskytu dvojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na konci všech slovních (22a) a taktových (22b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (22a) a 493 a 559 (22b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (22a/22b) z celkového počtu jednotek – 6639 (22a) a 5368 (22b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CV	295	59,84	3801	70,81	CVCV	CV	338	60,47	3798	70,75	CVCV
VC	147	29,82	1362	25,37	CVCVC	VC	165	29,52	1362	25,37	CVCVC
VV	9	1,83	16	0,3	CVV	VV	4	0,72	6	0,11	CVCVV
CC	29	5,88	84	1,56	CVCC	CC	29	5,19	84	1,56	CVCC
CR	10	2,03	37	0,69	CVCR	CR	11	1,97	37	0,69	CVCR
RC	2	0,41	5	0,09	CVCVCRC	RC	2	0,36	5	0,09	CVCVCRC
						7V	10	1,79	76	1,42	7V

Tabulka 22b

4.2.6.3 Četnost trojic C, V, R a ? v rámci slov a taktů

V této kapitole se budeme věnovat četnosti trojic hlásek v rámci slov a taktů, na jejich začátku i na konci. V tabulkách nebudou uváděny kombinace hlásek vykazující v našem korpusu nulový výskyt.

4.2.6.3.1 Libovolná pozice v jednotce

Nejfrekventovanější trojicí hlásek v rámci slov byla kombinace CVC (přes 90 % typů vzorců a cca 77 % všech slov s rázem i bez rázu. Dalšími silně zastoupenými trojicemi byly VCV (74-76 % typů vzorců a cca 57 % všech výskytů) a CCV (cca 79-81 % typů vzorců a zhruba 54 % všech slov); vysokou frekvenci měla také kombinace VCC (cca 76 % typů vzorců a 39 % všech výskytů slov). Vzorec CCC měl poměrně velkou četnost, pokud jde o samostatné typy vzorců (kolem 34-36 %); co do výskytů však byl jeho podíl nižší (cca 8 %). Ostatních kombinací bylo výrazně méně co do typů i výskytů; jejich pořadí podle četnosti odpovídá rozložení kombinací C a V ve *Statistikách češtiny* (2009), kde ale nejsou zahrnuty kombinace s likvidami. Nejobvyklejší kombinací s likvidou byla trojice CRC, která se v našem materiálu objevila v cca 10-11 % typů struktur a zhruba ve 2 % ze všech slov. Nejčastějším vzorcem obsahujícím ráz byla struktura ?VC (cca 27 % typů a 10 % výskytů). Vzácné byly skupiny C?V, VC? (např. *KCNA*, *nejočekávanější*) a ?V? (*OSN*, *USA*, *BIS*). Přesné údaje jsou uvedeny v tabulce 24a.

V datech bez zohlednění rázu najdeme dva výskyty trojice VVV. V obou případech se jedná o zkratku *BIS*. Mezi vzorci s rázem již tuto kombinaci nenajdeme, neboť mezi jednotlivými vokály byly přítomny rázy. Skupinu VVV nenalezneme ani u taktů – u obou realizací této zkratky byla přízvukována kromě první i třetí slabika, a tudíž se slovo rozpadlo na dva takty.

V rámci taktů byla situace v mnohém podobná četnostem kombinací hlásek u slov – opět byla nejfrekventovanější trojice CVC (obsahovalo ji přes 90 % typů vzorců i výskytů taktů); silně zastoupeny byly i kombinace CCV (kolem 80-82 % typů vzorců taktů a cca 68 % všech výskytů), VCV (75-77 % typů vzorců a 74 % všech taktů) a VCC (76-78 % typů taktů a 53 % výskytů).

Trojice CCC byla silně zastoupena u typů vzorců (cca 33-35 %), u všech výskytů slaběji (cca 11 %). Kombinace VVV se v rámci taktů neobjevila vůbec.

Nejčetnější skupinou s likvidou byla opět trojice CRC (10-11 % typů vzorců a zhruba 2 % všech taktů). Skupina ?VC vykazovala zastoupení ve 36 % typů vzorců a v 15 % všech taktů. Kombinace RVC v souboru taktů bez započítání rázu a R?V a CR? s jeho zohledněním se v korpusu objevily jednou, a to v taktu obsahujícím výrazy *sedm až* (viz výše).

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	47	12,24	89	1,34	CVCVVCV	CVV	35	8,47	61	0,92	CVCVVCV
CVC	352	91,67	5138	77,39	CVCV	CVC	374	90,56	5135	77,35	CVCV
CCV	310	80,73	3565	53,7	CVCCV	CCV	328	79,42	3565	53,7	CVCCV
VCV	290	75,52	3797	57,19	CVCV	VCV	305	73,85	3792	57,12	CVCV
VCC	292	76,04	2601	39,18	CVCCV	VCC	310	75,06	2601	39,18	CVCCV
CCC	138	35,94	527	7,94	CCCVCV	CCC	141	34,14	527	7,94	CCCVCV
VVV	1	0,26	2	0,03	CVVVC	VVC	33	7,99	56	0,84	CVCVVCV
VVC	46	11,98	83	1,25	CVCVVCV	CRC	42	10,17	118	1,78	CRCCV
CRC	42	10,94	118	1,78	CRCCV	CCR	14	3,39	30	0,45	CVCVCCR, CVCCRCVC
CCR	14	3,65	30	0,45	CVCVCCR	RCC	7	1,69	32	0,48	CRCCV
RCC	7	1,82	32	0,48	CRCCV	VCR	20	4,84	62	0,93	CVCR
VCR	20	5,21	62	0,93	CVCR	RCV	33	7,99	81	1,22	CRCV
RCV	33	8,59	81	1,22	CRCV	C?V	6	1,45	7	0,11	CV?VC?VCCV
						?V?V	21	5,08	30	0,45	?VCV?V, CV?VCVCCV
						VC?	6	1,45	7	0,11	CV?VC?VCCV
						?V?	3	0,73	4	0,06	CV?V?VC
						?VC	110	26,63	630	9,49	?VC
						CV?	19	4,6	28	0,42	?VCV?V, CV?VCVCCV

Tabulka 23a

Tabulka 23a a 23b: Četnost výskytu trojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) v rámci všech slovních (23a) a taktových (23b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (23a) a 493 a 559 (23b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (23a/23b) z celkového počtu jednotek – 6639 (23a) a 5368 (23b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	76	15,42	145	2,7	CVVCVC, CVVCCVVCV	CVV	43	7,69	69	1,29	CVCVVCV
CVC	460	93,31	5036	93,82	CVCV	CVC	513	91,77	5029	93,68	CVCV
CCV	404	81,95	3630	67,62	CVCCV	CCV	447	79,96	3630	67,62	CVCCV
VCV	378	76,67	3953	73,64	CVCV	VCV	420	75,13	3944	73,47	CVCV
VCC	384	77,89	2823	52,59	CVCCV	VCC	425	76,03	2823	52,59	CVCCV
CCC	174	35,29	586	10,92	CCCVCV	CCC	182	32,56	586	10,92	CCCVCV
VVC	79	16,02	144	2,68	CVVCVC, CVVCCVVCV	VVC	39	6,98	63	1,17	CVCVVCV
CRC	53	10,75	120	2,24	CRCCV	CRC	54	9,66	120	2,24	CRCCV
CCR	21	4,26	45	0,84	CCRCV, CCRCCV	CCR	21	3,76	45	0,84	CCRCV, CCRCCV
RCC	10	2,03	32	0,6	CRCCV	RCC	10	1,79	32	0,6	CRCCV
VCR	32	6,49	67	1,25	CVCR	VCR	34	6,08	67	1,25	CVCR
RCV	41	8,32	83	1,55	CRCV	RCV	42	7,51	83	1,55	CRCV
RVC	1	0,2	1	0,02	CVCRVC	C7V	39	6,98	50	0,93	C7VCVVCV
						V7V	59	10,55	91	1,7	CV7VCV
						VC7	15	2,68	16	0,3	CV7VC7VC
						7V7	12	2,15	15	0,28	7V7VCCVVCV
						7VC	200	35,78	781	14,55	7VCCVCV
						CV7	47	8,41	76	1,42	CV7VCV
						CR7	1	0,18	1	0,02	CVCR7VC
						R7V	1	0,18	1	0,02	CVCR7VC

Tabulka 23b

4.2.6.3.2 Pozice na začátku jednotky

Pozice na začátku jednotky byla u slov i taktů omezenější, nesetkáme se tedy s takovým množstvím kombinací jako v předchozím případě. I zde byla nejfrekventovanější trojice CVC, a to jak u slov, tak u taktů. Touto trojčlennou kombinací začínalo cca 36-39 % typů slov a cca 48 % všech slov, 33-37 % typů vzorců taktů a cca 54 % všech taktů. Dále byly na začátku taktů i slov poměrně četné trojice CCV (18-21 % typů vzorců slov i všech slov a cca 18-19 % typů taktů a 24 % všech taktů), 7VC (typů slov začínajících touto kombinací bylo cca 22 % a typů taktů 21 %, ale všech výskytů bylo poměrně málo; tvořily pouze zhruba 9 % všech slov a 12 % všech taktů). Spojení VCC zahrnovalo 13 % typů vzorců slov bez rázu a 14 % typů taktů bez rázu, ale pouze cca 4-6 % všech výskytů jednotek začínajících touto kombinací.

V rámci typů vzorců taktů bez zahrnutí rázu byla poněkud silněji zastoupena trojice VCC (cca 10 % typů taktů). Na začátku jednotky se nikdy nevyskytovala trojice VVV. Relativně vzácné byly skupiny VVC a ?V? (výrazy *OSN, USA*), CCR (*shrnuje, čtrnáctý*), VCR (*osmnáct, udrží*) a VC? u taktů (spojení *od okolních*). Spojení CCR bylo u taktů četnější než u slov, a to díky spojení plnovýznamových slov začínajících na CR- s neslabičnými předložkami (*v Brně, v Plzni* apod.). Pořadí kombinací C a V podle četnosti je shodné s pořadím uvedeným ve *Statistikách češtiny (2009)*; v našich datech však kombinaci CVV předchází CRC. Z kombinací s likvidou byla v našem materiálu dle očekávání nejfrekventovanější trojice CRC (2-4 % typů slov i taktů a cca 1 % všech výskytů).

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	18	4,69	35	0,53	CVVCVCCV	CVV	11	2,66	17	0,26	CVVC, CVVCVCCV
CVC	149	38,8	3198	48,17	CVCV	CVC	150	36,32	3198	48,17	CVCV
CCV	75	19,53	1378	20,76	CCVCV	CCV	75	18,16	1378	20,76	CCVCV
VCV	35	9,11	274	4,13	VCV	VCV	10	2,42	24	0,36	VCVCV, VCV
VCC	49	12,76	295	4,44	VCCVCV	VCC	13	3,15	25	0,38	VCCVCCVCV, VCCVCV
CCC	26	6,77	164	2,47	CCCVCV	CCC	26	6,3	164	2,47	CCCVCV
VVC	2	0,52	2	0,03	VVCVC, VVCV	CRC	15	3,63	66	0,99	CRCCV
CRC	15	3,91	66	0,99	CRCCV	CCR	4	0,97	6	0,09	CCRCVCV, CCRCV
CCR	4	1,04	6	0,09	CCRCVCV, CCRCV	?V?	2	0,48	2	0,03	?V?VC?VC, ?V?VC?V
VCR	7	1,82	9	0,14	VCRCVCCV, VCRCV	?VC	90	21,79	604	9,1	?VC
						CV?	12	2,91	18	0,27	CV?VCVCCV

Tabulka 24a

Tabulka 24a a 24b: Četnost výskytu trojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na začátku všech slovních (24a) a taktových (24b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (24a) a 493 a 559 (24b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (24a/24b) z celkového počtu jednotek – 6639 (24a) a 5368 (24b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace, jež jsou zastoupeny v 50-100 % typů vzorců nebo výskytů jednotek, jsou označeny červeně; kombinace vyskytující se v 10-50 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	29	5,88	72	1,34	CVVCVC, CVVCCVCVCV	CVV	12	2,15	20	0,37	CVVCCVCVCV
CVC	181	36,71	2920	54,4	CVCV	CVC	184	32,92	2886	53,76	CVCV
CCV	96	19,47	1296	24,14	CCVCV	CCV	98	17,53	1296	24,14	CCVCV
VCV	51	10,34	345	6,43	VCVCV	VCV	12	2,15	22	0,41	VCVCV
VCC	68	13,79	340	6,33	VCCVCV	VCC	18	3,22	30	0,56	VCCVCCVCV, VCCVC
CCC	30	6,09	164	3,06	CCCVCV	CCC	30	5,37	164	3,06	CCCVCV
VVC	11	2,23	13	0,24	VVCCVCVCV	CRC	11	1,97	46	0,86	CRCCV
CRC	11	2,23	46	0,86	CRCCV	CCR	8	1,43	21	0,39	CCRCV, CCRCCV
CCR	8	1,62	21	0,39	CCRCV, CCRCCV	C7V	24	4,29	34	0,63	C7VCVCVCV
VCR	4	0,81	5	0,09	VCRCV	V?V	1	0,18	1	0,02	V?VVCVCCVC
						7V7	12	2,15	15	0,28	7V7VCCVCVCV
						7VC	118	21,11	661	12,31	7VCCVCV
						CV7	28	5,01	52	0,97	CV7VCV

Tabulka 24b

4.2.6.3.3 Pozice na konci jednotky

Na konci slov i taktů se nejčastěji objevovala jednotka VCV (30-32 % typů vzorců slov i taktů, cca 36 % všech slov a 44 % všech taktů); silně zastoupeny byly i trojice CVC a CCV (26-28 % typů vzorců slov i taktů; cca 21-22 % všech slov a zhruba 25 % všech taktů). Všechny ostatní kombinace se v taktech i slovech objevovaly výrazně méně; skupinu VCC jsme zjistili u cca 5-6 % typů slov i taktů a u cca 1-2 % všech slov nebo taktů). Na konci jednotky se nikdy nevyskytla trojice VVV. Co se týče vzorců s likvidou, nejfrekventovanější byla trojice RCV se zhruba 2 % typů vzorců slov i taktů a cca 0,5-0,6 % všech výskytů jednotek a trojice VCR, která se vyskytovala v 1-2 % typů vzorců a v 0,5-0,6 % všech jednotek v našem materiálu. Poměrně vzácné byly skupiny CVV (*Korea, ODA*), CCC (*vojsk, Hammings*), VVC (většinou v cizích slovech nebo zkratkách – *duel, Moan, Izrael*), CRC (*veletrh, návrh*), CCR (*ministr*); u vzorců se zohledněním rázu se také výjimečně vyskytly skupiny C7V (*KCNA, USA, BSE*) a V?V (*ODA*). Kombinace RVC se mezi takty bez započítání rázu objevila jednou (spojení *sedm až*). Pořadí kombinací konsonantů a vokálů podle jejich frekvence v porovnání se SYN2005

vykazuje drobné odchylky – podle celkového obsahu korpusu se liší pořadí trojic CVV a VVC; ve *Statistikách češtiny* (2009) je jako čtenější uváděna trojice VVC. (Je ale nutné podotknout, že v našem materiálu se jedná o odchylky v řádech setin procenta.) Opět zde navíc nejsou brány v potaz slabikotvorné likvidy. Pokud srovnáváme náš materiál s publicistickou částí korpusu SYN2005, můžeme si také povšimnout rozdílného pořadí skupin CVC a CCV – v našem materiálu byla čtenější trojice CVC, nicméně procentuální rozdíly byly, stejně jako v publicistických textech ze SYN2005, velmi malé. Podrobné informace o četnosti trojic hlásek na konci jednotek uvádíme v tabulkách 25a a 25b:

SLOVA											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	3	0,78	8	0,12	CVCVV	CVV	2	0,48	5	0,08	CVCVV
CVC	106	27,6	1428	21,51	CVCVC	CVC	110	26,63	1427	21,49	CVCVC
CCV	103	26,82	1419	21,37	CVCCV	CCV	113	27,36	1419	21,37	CVCCV
VCV	122	31,77	2398	36,12	CVCV	VCV	129	31,23	2395	36,07	CVCV
VCC	22	5,73	85	1,28	CVCC	VCC	23	5,57	85	1,28	CVCC
CCC	2	0,52	2	0,03	CVCVCCC, CVCCC	CCC	2	0,48	2	0,03	CVCVCCC, CVCCC
VVC	6	1,56	11	0,17	VCVVC, CVVC	VVC	3	0,73	5	0,08	CVVC
CRC	2	0,52	5	0,08	CVCRC	CRC	2	0,48	5	0,08	CVCRC
CCR	2	0,52	6	0,09	CVCVCCR	CCR	2	0,48	6	0,09	CVCVCCR
VCR	5	1,3	34	0,51	CVCR	VCR	5	1,21	34	0,51	CVCR
RCV	7	1,82	31	0,47	CRCV	RCV	7	1,69	31	0,47	CRCV
						C?V	3	0,73	3	0,05	CVCV?VC?V, ?V?VC?V, CV?VC?V
						V?V	1	0,24	3	0,05	?VCV?V
						?VC	6	1,45	82	1,24	?VC

Tabulka 25a

Tabulka 25a a 25b: Četnost výskytu trojic souhlásek (C), samohlásek (V), slabikotvorných likvid (R) a rázů (?) na konci všech slovních (25a) a taktových (25b) typů i výskytů. Procenta ze všech typů jsou počítána ze základů 384 a 413 (25a) a 493 a 559 (25b) typů vzorců bez rázu a s rázem a procenta ze všech slov/taktů (25a/25b) z celkového počtu jednotek – 6639 (25a) a 5368 (25b). Procentuální hodnoty jsou zaokrouhleny na dvě desetinná místa. Kombinace vyskytující se ve více než 10 % typů vzorců nebo výskytů jednotek jsou označeny modře; kombinace zastoupené v méně než 10 % typů vzorců nebo výskytů jednotek mají pozadí bílé. Ve sloupci „vzorec“ je uveden nejfrekventovanější vzorec daného typu. Tabulka obsahuje zpracování pro data bez zahrnutí rázu i s jeho započítáním mezi hlásky.

TAKTY											
bez rázu						s rázem					
	typů	%	výskytů	%	vzorec		typů	%	výskytů	%	vzorec
CVV	8	1,62	14	0,26	CVV	CVV	4	0,72	6	0,11	CVCVV
CVC	136	27,59	1318	24,55	CVCVC	CVC	148	26,48	1315	24,5	CVCVC
CCV	135	27,38	1346	25,07	CVCCV	CCV	149	26,65	1346	25,07	CVCCV
VCV	149	30,22	2370	44,15	CVCV	VCV	178	31,84	2367	44,09	CVCV
VCC	27	5,48	82	1,53	CVCC	VCC	27	4,83	82	1,53	CVCC
CCC	2	0,41	2	0,04	CVCVCCC, CVCCC	CCC	2	0,36	2	0,04	CVCVCCC, CVCC
VVC	9	1,83	17	0,32	CVCVVC	VVC	5	0,89	7	0,13	CVVC
CRC	2	0,41	5	0,09	CVCVCRC	CRC	2	0,36	5	0,09	CVCVCRC
CCR	2	0,41	6	0,11	CVCVCCR	CCR	2	0,36	6	0,11	CVCVCCR
VCR	8	1,62	31	0,58	CVCR	VCR	9	1,61	31	0,58	CVCR
RCV	10	2,03	30	0,56	CRCV	RCV	10	1,79	30	0,56	CRCV
RVC	1	0,2	1	0,02	CVCRVC	C?V	3	0,54	3	0,06	CVC?V, ?V?VC?V, CV?VC?V
						V?V	6	1,07	10	0,19	CV?V
						?VC	11	1,97	38	0,71	?VC

Tabulka 25b

4.2.6.4 Další poznatky

Pro zajímavost zmiňujeme i výskyt komplexních souhláskových shluků: kombinace CCCC se v korpusu slov s rázem i bez rázu objevila ve 34 případech (v 0,51 % všech vzorců slov); typů vzorců s touto kombinací bylo 17 (objevila se tedy ve 4,43 % typů vzorců slov bez rázu a ve 4,12 % typů vzorců slov s rázem). Nejčastějším vzorcem byla struktura CVCVCCVCCCCV (*ministerstvo*). V případě taktů bylo typů vzorců s touto kombinací 23 (4,67 % typů vzorců taktů bez rázu a 4,11 typů vzorců taktů s rázem) a výskytů celkově 39 (0,73 % ze všech výskytů taktů). Na začátku jednotky se kombinace CCCC vyskytla ve dvou případech (*vstřelil*, *vzhledem*) u slov i taktů v rámci vzorců s rázem i bez rázu. Na konci jednotky se skupina čtyř konsonantů neobjevila nikdy, což bylo v souladu s faktem, že čeština neobsahuje čtyřčlenné slabičné kody.

Kombinaci CCCCC jsme našli v jediném případě u slov i taktů – jednalo se o slovo *čerstvě* (CVCCCCCV) v úseku JPC04-04-02.

5. Diskuse

Náš výzkum se pokusil podat obecný a relativně komplexní obraz fonotaktické struktury slov a mluvních taktů češtiny s využitím 45minutového materiálu českého textu čteného profesionálními mluvčími. Tento mluvený korpus byl sice ve srovnání s psanými databázemi poměrně omezený, nicméně byl dostatečně obsáhlý, aby poskytl vhled do fonotaktické problematiky mluvené češtiny. Je možné namítnout, že větší výpovědní hodnotu by měl výzkum provedený na „běžně mluveném jazyku“, např. na spontánních projevech. Tímto bychom ale nezískali tak komplexní a bohatý obraz o typech kombinací konsonantů a vokálů v rámci slov a taktů, neboť běžně mluvená čeština se vyznačuje velkým množstvím často opakovaných slov; navíc by bylo problematičtější vymezení taktu a odstraňování hezitací. Tato práce měla za úkol podat souhrnnou informaci o spisovné mluvené češtině. Běžně mluvený jazyk a různé mluvní styly a dialekty proto stále čekají na podobné zpracování a poskytují tak možnost dalšího výzkumu v této oblasti.

Co se týče členění mluvené češtiny na jednotky souvislé řeči, zjistili jsme, že není neobvyklým jevem, že i profesionální mluvčí vytvářejí jednoslabičné takty z neplnovýznamových slov – např. z původních předložek, které by ideálně měly nést přízvuk a stát na začátku taktu obsahujícího předložku a následující slovo nebo slova, a spojek, které by podle tradičních popisů měly být spíše předdrážkami, tedy nepřízvučnými slabikami před přízvuknou slabikou na začátku taktu. Někteří mluvčí k tvoření jednoslabičných taktů z předložek inklinovali více než jiní, můžeme se tedy domnívat, že tato tendence je mj. i výsledkem idiolektu jedince. Z hlediska ideálního popisu se však v českém neutrálním spisovném projevu obvykle nevyskytují dva přízvuky vedle sebe, pokud se nejedná o konec vyšší prozodické jednotky. V případě, že bychom ale např. přízvukovanou jednoslabičnou spojku prohlásili za předdrážku, popřeli bychom tím definici předdrážky i taktu, v němž by se v tom okamžiku ocitly dva přízvuky, a proto se nabízí výhodnější možnost – připustit, že i ve standardních českých projevech mohou stát dva přízvuky (ve dvou samostatných taktech) vedle sebe.

Dále jsme v našem materiálu několikrát zaznamenali rozpad jednoho slova na více taktů, což je možnost, kterou se obvykle popisy nezabývají (výjimkou je např. Ondráčková, 1954), nicméně ačkoli tento jev nelze označit v našich datech za zvlášť frekventovaný (na dva takty se rozpadlo celkem 20 slov), nelze si opět nepovšimnout faktu, že se rozpad slova na více taktů se může vyskytnout i ve čteném projevu profesionálního mluvčího; pokud se jedná o zkratky, můžeme zde dokonce hovořit o jisté pravidelnosti. Také není scestné se domnívat, že tento úkaz bude mít širší zastoupení v běžně mluveném jazyce.

Naším cílem bylo zkoumat kombinatoriku konsonantů a vokálů v rámci slov a taktů v mluvené (resp. čtené) češtině. Nepopisovali jsme konkrétní hlásky v kombinacích, což je oblast, která by si také zasloužila zevrubné prozkoumání – tím spíše, jaké bohatství typů kombinací souhlásek čeština oproti jiným (např. germánským) jazykům nabízí, jak vyplývá ze studií Ludvíkové a Krause (1966) nebo Novotné-Hůrkové (1974).

Náš materiál obsahoval 6639 slov a 5368 mluvních taktů. V rámci analýzy bylo nutné vyřešit problém daný klasifikační rázou, který v popisu češtiny není považován za hlásku, nicméně je důležitým hraničním signálem: zpracovávali jsme vždy po dvou souborech dat z jednoho korpusu – jeden se zohledněním realizovaného rázu a jeho speciálním značením a druhý bez jeho zahrnutí. Korpus měl tedy 37 360 hlásek v rámci slov a 37 356 hlásek v rámci taktů bez zahrnutí rázu; pokud jsme ráz zohlednili, obsahoval korpus 38 243 hlásek v rámci slov a 38 239 v rámci taktů. Rázů bylo v korpusu 883. U slov i taktů se objevily jednotky, které obsahovaly dva, nebo dokonce i tři rázy – tyto takty nebo slova byly převážně zkratky (např. *USA*).

Takty nejčastěji sestávaly z jednoho slova (tvořily cca 75 % všech jednotek), nezanedbatelnou část tvořily i takty dvouslovné (cca 20 % všech taktů). V korpusu se vyskytly i takty složené ze tří slov (cca 1,6 %) a několikrát také nastal případ, kdy slovo bylo rozděleno do více než jednoho taktu (0,7 %). Průměrně se v taktu nacházelo cca sedm hlásek; ve slově necelých šest hlásek. Nejvyšší výskyt vykazovaly u taktů i slov bez započítání rázu i s rázem jednotky pětihláskové, avšak takty byly nejčastější sedmihláskové – v rámci

taktů byly počty pětihláskových, šestihláskových a sedmihláskových jednotek na rozdíl od domény slov, kde se jednoznačně vyskytovalo nejvíce jednotek složených z pěti hlásek, poměrně vyrovnané, což přisuzujeme faktu, že se často spojovala dvě slova v jeden takt – obvykle šlo o jednotku kratší (tj. složenou z jedné, dvou nebo tří hlásek), zpravidla synsémantickou, která se připojila k delší, většinou autosémantické. Z tohoto důvodu se v materiálu objevilo velmi málo jednohláskových taktů; poměrně vzácný byl i výskyt dvouhláskových a tříhláskových taktů (oproti slovům, kde byly jednohláskové i dvouhláskové jednotky relativně časté). Nejdelší slovo (a zároveň takt) mělo 25 hlásek.

Dále jsme se věnovali konkrétním CVCV vzorcům u slov i taktů. Celkově bylo v korpusu méně taktů než slov, nicméně typy jejich vzorců byly rozmanitější: takty totiž často vznikaly kombinacemi několika vzorců slov. Ve zpracovaném korpusu s daty bez dysfluencí a se speciálně značenými slabikotvornými hláskami jsme našli 384 různých vzorců slov bez rázu a 493 typů vzorců taktů bez rázu a 413 vzorců slov a 559 vzorců taktů se zvláště značeným rázem.

Podle našeho očekávání byly nejfrekventovanějšími jednotkami poměrně jednoduché struktury s nanejvýš dvojčlennými souhláskovými shluky: Pro slova s rázem i bez započítání rázu byla nejčastější jednotkou struktura CV; následovaly vzorce CVCV, CVCCV, C a CVCVCV. Pro takty byla typická víceslabičnost – vzorců sestávajících z jediné slabiky nebylo tolik jako u slov. Vzorce složené z jediného konsonantu mezi takty pochopitelně nefigurovaly vůbec. Nejčastějšími strukturami taktů s rázy i bez rázu byly vzorce CVCV, CVCCV, CVCVCV, CCVCV a CVCVC, přičemž mezi celkovým počtem výskytů jednotlivých vzorců u taktů nebyly tak markantní rozdíly jako u slov, kde byly rozestupy v četnosti výskytů jednotlivých struktur v některých případech tvořeny i 100 a více výskytů. Nejkratšími strukturami byla slova tvořená jediným vokálem, popř. jediným konsonantem. Frekventovanějšími byly v tomto případě struktury konsonantické než vokalické. Vzorce s rázem zahrnovaly 237 struktur složených z rázu a vokálu a sedm pouze z vokálu. V rámci taktů se objevilo 63 případů, kdy takt tvořil vokál (resp. ráz a vokál).

v našem korpusu jsme porovnali s výsledky výzkumu provedeném na korpusu psané češtiny SYN2005, prezentovaném v publikaci *Statistiky češtiny* (Bartoň et al., 2009), a konstatovali jsme, že rozdíly v nejfrekventovanějších vzorcích různých slovních druhů byly způsobeny převážně odlišným pojetím rázu (ve *Statistikách češtiny* je ráz zřejmě považován za konsonant). Mezi fonotaktickými strukturami mluveného a psaného jazyka jsme však našli i jisté odlišnosti, např. v rámci substantiv jsme v našem korpusu nenalezli jednohláskovou strukturu, ačkoli tato se ve výše uvedeném výzkumu vyskytovala – je pravděpodobné, že tato situace vznikla psáním jmen a jiných slov formou jednopísmenných zkratk, což je pro psaný projev typické; v projevu mluveném však tato strategie není obvyklá. Také lze předpokládat, že pokud by mluvčí opravdu zkratku jména přečetl nebo vyslovil, pravděpodobně by nešlo o jednoduchou hlásku, ale spíše o skupinu CV nebo ?V.

V souvislosti se slovními druhy jsme se rovněž věnovali srovnání variability CVCV vzorců různých slovních druhů – co se týče absolutních počtů, zahrnovala nejvíce typů struktur substantiva, následovala adjektiva, slovesa, číslovky a příslovce; nízkou variabilitu vykazovaly předložky, spojky, částice a zájmena. Když jsme s absolutními počty typů CVCV vzorců jednotlivých slovních druhů v korpusu srovnali **poměrnou** variabilitu jejich CVCV vzorců (typy vzorců vůči všem výskytům daného slovního druhu), dospěli jsme k zajímavým výsledkům – nejvyšší variabilitu vykazovaly částice a velmi vysokou adjektiva, ačkoli částice tvořily necelé jedno procento korpusu a adjektiva necelých 14 %. Nezanedbatelná byla i poměrná variabilita číslovek, sloves a příslovcí. Překvapivě nízkou variabilitu vykazovala právě substantiva. Nižší variabilitu než substantiva měly pouze spojky, zájmena a předložky.

Dále jsme se zabývali frekvencí samotných hlásek C, V, R a ? a jejich dvojčlenných a trojčlenných kombinací v různých pozicích ve slově. Vokál zahrnovaly všechny slovní vzorce kromě toho, který byl tvořen samostatným konsonantem; stejně tak konsonant se vyskytl v každém vzorci kromě struktury složené ze samostatného vokálu či rázu a vokálu. (Oproti tomu ráz

nebo likvida nikdy nestály jako samostatná jednotka slova nebo taktu.) Slabikotvorné likvidy se vyskytovaly v cca 13 % typů vzorců slov i taktů s rázem i bez rázu a v cca 3 % ze všech slovních a taktových jednotek. Poměrně četné byly rázy – objevily se ve čtvrtině vzorců slov, ve třetině typů vzorců taktů a v cca sedmině všech slov a taktů. Vokál obsahovaly všechny takty v našem korpusu – ačkoli v češtině existují slova se slabikotvornými likvidami a bez vokálu, která by byla schopna takt tvořit, ve zkoumaném materiálu se nevyskytly. Na začátku slov s rázem i bez rázu byl nejčastěji konsonant (začínalo jím vždy více než 70 % typů struktur a více než 85 % všech výskytů jednotek). Vokálem začínalo slov i taktů výrazně méně (kolem 25 % typů vzorců slov i taktů bez rázu a kolem 15 % všech výskytů slov i taktů bez rázu); pokud se do analýz zahrnul ráz, počáteční vokál obsahovalo pouze kolem 6 % typů slov i taktů a jen 1 % všech slov i taktů – všechny ostatní jednotky začínaly rázem. Takty ani slova neměly na začátku slabikotvornou likvidu – tento jev pro češtinu není typický. Na konci jednotky byla situace jiná – většina typů i výskytů slov i taktů končila vokálem (kolem dvou třetin typů slov i taktů a tři čtvrtiny všech výskytů slov i taktů, vše s rázem i bez rázu); významná část jednotek korpusu (kolem třetiny typů vzorců slov i taktů a zhruba čtvrtina všech výskytů jednotek, vše s rázem i bez rázu) končila konsonantem. Vzácně se na konci slova či taktu objevila likvida (cca ve 2 % typů vzorců slov i taktů a v necelém procentu všech výskytů jednotek, vše s rázem i bez rázu). Ráz se dle předpokladu na konci jednotek nevyskytoval nikdy.

Nejfrekventovanějšími dvojicemi hlásek byly v rámci slov i taktů CV (přes 90 % typů i výskytů vzorců slov i taktů s rázem i bez rázu) a VC (80 % výskytů vzorců slov a přes 97 % všech typů vzorců slov a všech typů i výskytů taktů, vše s rázem i bez rázu), významný byl také podíl kombinace CC. Dalšími dvojicemi, které se častěji objevovaly v libovolné pozici v rámci slov, byly např. dvojice VV, CR, RC pro vzorce bez rázu a CR, RC a ?V pro vzorce s rázem. V rámci taktů měly také vyšší podíl dvojice VV, CR a RC (bez rázu) a CR, ?V a V? (s rázem). Podíl dalších dvojčlenných kombinací byl co do typů i výskytů menší než 10 %. Ačkoli by nebylo neopodstatněné očekávat, že se vedle sebe neobjeví slabikotvorná likvida či nazála a vokál, v jednom

případě se tak v našem materiálu stalo. Jednalo se o spojení dvou slov, z nichž první končilo likvidou a druhé začínalo rázem a vokálem (výrazy *sedm až*). Je tedy zřejmé, že vokál se v našem materiálu vedle slabikotvorného konsonantu objevil pouze tehdy, když jsme pracovali s daty bez rázu. Na začátku jednotky se vyskytovala u slov i taktů zejména dvojice CV (přes 40 % typů vzorců slov i taktů a přes 55 % všech výskytů u slov i taktů, vše s rázem i bez rázu); četné byly i kombinace VC a CC pro slova i takty bez rázu a CC a ?V pro slova a takty s rázem. Kombinace VC se vyskytovala i v rámci slov a taktů s rázem, ale její zastoupení bylo slabší. Mezi slovy a takty s rázem se nacházela i poměrně neobvyklá dvojice C? – realizovala se především jako součást zkratky nebo konsonantické předložky a slova začínajícího na ráz a vokál (*ČSSD, s občasnými, v Afghánistánu*). Na konci jednotek se objevovala nejčastěji dvojice CV (ve dvou třetinách všech typů i výskytů vzorců slov i taktů); poměrně obvyklá byla i kombinace VC (necelá třetina typů vzorců slov i taktů a cca čtvrtina všech výskytů slov i taktů, vše s rázem i bez rázu). Ostatní dvojice tvořily méně než desetinu vzorců korpusu; v datech s rázem ještě na konci výjimečně figurovala skupina ?V. Tato byla obsažena pouze v pěti typech vzorců slov, ale na konci slov se celkem objevila 243krát – tento fakt připisujeme hojnému výskytu spojek a předložek složených právě z této kombinace (?V je nejčastějším vzorcem spojek).

Z trojic hlásek byly nejčastější kombinace CVC, CCV, VCV, VCC, a to u slov i taktů. Nejfrekventovanější kombinací byla trojice CVC (přes 90 % typů i výskytů vzorců taktů a typů vzorců slov a cca 77 % všech výskytů slov, vše s rázem i bez rázu). Tento fakt patrně souvisí se zjištěním, že v českých hlasatelských projevech je 45 % všech konsonantů umístěno mezi dvěma vokály nebo mezi vokálem a koncem nádechového úseku (Volín & Churaňová, 2010). Poměrně silně zastoupenými trojicemi byly dále kombinace CVV, CCC, VVC a CRC pro slova i takty bez rázu a CCC, CRC, ?VC pro slova s rázem a V?V, ?VC a CCC pro takty s rázem. Ostatní jednotky se vyskytovaly v méně než 10 % všech typů i výskytů vzorců. U souboru slov bez zohlednění rázu se vyskytly dva vzorce s kombinací VVV. V obou případech se jedná o zkratku *BIS*. Mezi vzorci se zohledněním rázu tuto kombinaci nenajdeme, protože mezi samohláskami byly rázy přítomny.

Trojice VVV absentuje i u taktů – u obou realizací této zkratky byla přízvukována kromě první i třetí slabika, a tudíž se zde slovo rozpadlo na dva takty. V rámci taktů bez rázu se v jednom případě vyskytla také kombinace RVC, v rámci taktů s rázem CR? a R?V (ve všech případech spojení *sedm až*, viz výše).

Na začátku slov i taktů bez zahrnutí rázu stály nejčastěji trojice CVC, CCV, VCC a v rámci taktů také VCV – nejfrekventovanější byla trojice CVC (cca třetina všech typů slovních i taktových vzorců a cca polovina všech výskytů – platí i pro slovní a taktové vzorce s rázem); v rámci vzorců slov i taktů s rázem byla kromě CVC a CCV častá kombinace ?VC. Z trojic obsahujících slabikotvornou likvidu byla nejčtenější struktura CRC, což není nijak překvapivé vzhledem k faktu, že tři čtvrtiny typů vzorců slov i taktů začínají konsonantem a zároveň platí, že v rámci slov nemůže stát slabikotvorná hláska vedle vokálu. Na konci slov i taktů s rázem i bez rázu se objevovaly převážně skupiny CVC, CCV a VCV – nejčastější kombinací byla trojice VCV (cca třetina typů vzorců taktů i slov a všech výskytů slov a téměř polovina všech výskytů taktů, vše s rázem i bez rázu). Všechny ostatní kombinace hlásek byly zastoupeny vzácně.

V našem materiálu se dále objevilo 34 případů kombinace CCCC (17 typů vzorců) a jedna skupina CCCCC v rámci slov i taktů (výraz *čerstvě*). Můžeme se ptát, jestli by zde nebylo vhodnější považovat spojení [je] za skupinu diftongického charakteru [j̥e] – tradiční popis fonetiky češtiny však počítá pouze s dvojhláskami *klesavými*, v nichž jádro diftongu předchází polovokálu (Palková, 1994). Rovněž je otázkou, jak by pětičlenný (a obecně vícečlenný) konsonantický shluk obstál např. ve spontánním projevu nebo v rychlé řeči, resp. jak by byl redukován.

Pole návaznosti na náš výzkum je velmi široké – od kombinatorické analýzy jednotlivých hlásek spisovné češtiny přes popis fonotaktiky různých mluvních stylů a dialektů a jejich vzájemného porovnání a srovnání se spisovným jazykem k rozšíření zkoumání prozodické struktury češtiny a případně i úpravě jejího tradičního popisu. V navazující práci se budeme

zabývat temporálními konfiguracemi českého taktu s přihlédnutím k jeho fonotaktické stavbě. Pod tuto problematiku spadají zejména charakteristiky zahrnující uspořádání akustických událostí v čase a vysledování možných korespondencí mezi načasováním událostí a fonotaktickou strukturou taktu. Náplň plánovaného projektu bude tvořit zejména zjišťování, zda jsou tyto frekventované fonotaktické struktury, jak byly nastíněny v předkládané práci, nositeli rozdílných rytmických vzorců, a to jak z hlediska produkčního, tak i percepčního.

Možnosti využití našich výsledků jsou tedy poměrně rozsáhlé a poskytují značný prostor pro navázání na náš výzkum a jeho využití v bádání v souvisejících jazykových oblastech.

6. Závěr

V první části této práce jsme zevrubně zhodnotili dosavadní popis fonotaktické problematiky i všech těsněji souvisejících témat, jako jsou např. rytmické aspekty mluvené řeči a její členění. V českém prostředí byla dlouhou dobu pozornost věnována především struktuře slabiky a konkrétním hláskám (např. výzkum M. Ludvíkové (1987) a H. Kučery (1968) a jejich kombinacím). Výzkum, který se věnoval fonotaktice češtiny ve smyslu, v jakém ji pojmáme i my, byl proveden na korpusu SYN2005 a publikován v roce 2009. Korpus SYN2005, ačkoli velmi obsáhlý (cca 100 milionů textových slov), je ovšem korpusem češtiny psané, a ačkoli se autoři ve svém výzkumu pokoušeli se mluvené češtině přiblížit např. vložím rázu na začátky slov začínajících vokálem po pauze (je ovšem těžké určit, podle jakých pravidel odhadovali, před kterými slovy by byla v mluveném projevu pauza), v obecném úvodu ke své publikaci poznamenávají, že

„datová základna však odráží psaný jazyk a předpokládaná povaha dat mluveného jazyka, který zde v zásadě pokrytý není (...), je do značné míry jiná.“ (Bartoň et al., 2009: 15)

Jejich publikace tedy poskytuje pouze řadu kvantitativních dat o kombinacích konsonantů a vokálů v rámci českých slov bez jejich interpretace nebo srovnání s jinými jednotkami, např. mluvními takty. Navíc často nebyl přesně popsán postup, který autoři při zpracování dat využili.

Cílem této práce bylo podat obecnější popis fonotaktické struktury mluvené spisovné češtiny a poskytnout fonotaktickou mapu českých slov a mluvních taktů. K tomu jsme využili materiál od 12 profesionálních českých hlasatelů, který obsahoval 657 nádechových úseků, 6639 slov a 5368 mluvních taktů. V rámci slov se vyskytovalo 37 360 hlásek a v rámci taktů 37 356. Pokud se mezi hlásky započítal i ráz, bylo hlásek v rámci slov 38 243 a v rámci taktů 38 239. Mluvený text jsme v programu Praat rozdělili na nádechové úseky, slova, takty a hlásky a nakonec přidali vrstvu pro označení konsonantů, vokálů a rázů. Pomocí příkazových souborů jsme poté získali data, která byla následně zpracovávána v programu MS Excel a statisticky vyhodnocena.

U slov byly za použití programu poskytnutého Ústavem teoretické a počítačové lingvistiky FF UK rozpoznány slovní druhy. Všechna data byla podrobena manuální kontrole, při které se odstraňovaly chyby vzniknuvší v některých momentech procesu dávkového zpracování.

Postupně jsme se v tomto výzkumu zabývali počtem slov a taktů a jejich vztahy, četností slovních druhů, počtem hlásek ve slovech (s ohledem na jednotlivé slovní druhy) a taktech; speciální pozornost jsme věnovali jednotce rázu. V další části práce jsme uvedli nejčastější CVCV vzorce slov i taktů: Nejfrekventovanějšími konsonanticko-vokálními strukturami slov byly vzorce CV, CVCV a CVCCV a nejčastějšími taktovými skelety byly CVCV, CVCCV a CVCVCV. Nezanedbali jsme ani jednotky se slabikotvornými likvidami. Posléze jsme přihlédli k četnosti jednotlivých vzorců u různých slovních druhů. Nakonec jsme zhodnotili frekvence výskytu samohlásek a souhlásek a jejich dvojic a trojic (s rázem i bez rázu) na různých místech v jednotce. Výsledky jsme porovnávali mezi sebou a také jsme je průběžně srovnávali s výzkumem provedeným na rozsáhlejší psaném korpusu SYN2005 (2009) a zmínili shody i rozdíly.

Poskytli jsme zde tedy relativně komplexní mapu slov a taktů spisovné mluvené češtiny, která umožňuje navázání v této i v dalších oblastech zkoumání mluvené češtiny, pokusili jsme se porovnat výsledky našeho výzkumu se zkoumáním na korpusu psaného spisovného českého jazyka a nastínili jsme cesty, kterými by se mohla ubírat další zkoumání na poli fonetiky češtiny.

Literatura

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Barbosa, P. A., Arantes, P., Meireles, A. R. & Vieira, J. M. (2005). *Abstractness in speechmetronome synchronization: p-centres as cyclic attractors*. In *Proceedings of Interspeech 2005*, pp. 1441-1444. Lisbon: ISCA.
- Bartoň, T., Cvrček, V., Čermák, F., Jelínek, T., & Petkevič, V. (2009). *Statistiky češtiny*. Praha: Nakladatelství Lidové noviny.
- Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Počítačový program], verze 5.3.05, získáno z <http://www.praat.org/>
- Daneš, F. (1957). *Intonace a věta ve spisovné češtině*. Praha: Academia.
- Dellwo, V., Fourcin, A. & Abberton, E. (2007). *Rhythmical classification of languages based on voice parameters*. In: *Proceedings of the XVIth ICPhS*, 1129-1132. Saarbrücken: Organizing Committee.
- Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica* 66, 15-28.
- Duběda, T. (2005). *Jazyky a jejich zvuky. Univerzálie a typologie ve fonetice a fonologii*. Praha: Karolinum.
- Ghitza, O. & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66, 113-126.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31(3-4), 423-445.
- Haspelmath, M., Dryer, M., Gil, D., & Comrie, B. (2008). *The world atlas of language structures online*. Munich: Max Planck Digital Library. Získáno z <http://wals.info/>.
- Hála, B. (1948). *Úvod do fonetiky*. Praha: Melantrich.

- Hála, B. (1962). *Uvedení do fonetiky češtiny na obecně fonetickém základě*. Praha: Nakladatelství Československé akademie věd.
- Hála, B. (1975). *Fonetika v teorii a praxi*. Praha: Státní pedagogické nakladatelství.
- Havránek, B. & Jedlička, A. (1988). *Česká mluvnice*. Praha: Státní pedagogické nakladatelství.
- Hirst, D. & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. di Cristo (Eds.), *Intonation systems: a survey of twenty languages* (pp. 1-43). Cambridge: Cambridge University Press.
- Churaňová, E. (2010). *K problematice konsonantických skupin v češtině: Souhláskové shluky v rámci vyšších jednotek* (Klauzurní práce). FF UK.
- Chlumský, J. (1933). Poznámky k Weingartově Zvukové kultuře českého jazyka. *Naše řeč*, 17(6-7), 161-178.
- Kučera, H. & Monroe, G. K. (1968). *A comparative quantitative phonology of Russian, Czech and German*. New York: American Elsevier Publishing Company.
- Levelt, C. & van de Vijver, R. (1998). Syllable types in cross-linguistic and developmental grammars. Paper presented at the Third Biannual Utrecht Phonology Workshop (11-12/06/1998).
- Ludvíková, M. & Kraus, J. (1966). Kvantitativní vlastnosti soustavy českých fonémů. *Slovo a slovesnost*, 27, 334-344.
- Ludvíková, M. (1987). Číslo o hláskách. In M. Těšitelová et al., *O češtině v číslech* (pp. 91-108). Praha: Academia.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and brain sciences*, 21, 499-546.
- Mathesius, V. (1937). Mluvní takt a některé problémy příbuzné. *Slovo a slovesnost*, 3(4), 193-199.

- Mazlová, V. (1946). Jak se projevuje zvuková stránka češtiny v hláskových statistikách. *Naše řeč*, 30(6,7,8), 101-111, 146-151.
- Novotná-Hůrková, J. (1974). K výslovnosti některých souhláskových skupin a tzv. rázu v češtině. *Slovo a slovesnost*, 35, 113-120.
- Ondráčková, J. (1954). O mluvním rytmu v češtině. *Slovo a slovesnost*, 15, 24-29, 145-157.
- Ondráčková, J. (1962). K analýze přízvučnosti, zvláště v češtině. In: *AUC Philologica - Slavica Pragensia IV*, 81-88.
- Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.
- Palková, Z.; Veroňková, J.; Volín, J. & Skarnitzl, R. (2004). *Stabilizace některých termínů pro fonetický popis češtiny v závislosti na nových výsledcích výzkumu*. In T. Duběda (Ed.), *Sborník z konference česko-slovenské pobočky ISPhS 2004* (pp. 65-74). Praha: Karolinum.
- Palková, Z. (2006). *Textové dispozice pro členění na intonační fráze v češtině*. In Janoušková, J., Palková, Z. (Eds.), *Kapitoly z fonetiky a fonologie slovanských jazyků* (pp. 227-239). Praha: FF UK.
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation* (Dizertační práce). M.I.T. (publikováno Indiana University Linguistics Club, Bloomington).
- Pollák, P., Volín, J. & Skarnitzl, R. (2008). *Phone Segmentation Tool with Integrated Pronunciation Lexicon and Czech Phonetically Labelled Reference Database*. In *Proceedings of 6th International Conference on Language Resources and Evaluation*, 1, 1-5. Paris: ELRA.
- Quené, H. & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, 62(1), 1-13.
- Ramus, F., Nespors, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.

- Ramus, F., Dupoux, E., Zangl, R. & Mehler, J. (2000). *An empirical study of the perception of language rhythm*. [Preprint]
- Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2, 85-115.
- Rochet-Capellan, A. & Schwartz, J. (2005). *The labial-coronal effect and CVCV stability during reiterant speech production: an articulatory analysis*. In *Proceedings of the 9th International Conference on Speech Communication and Technology (InterSpeech)*, 1013-1016. Lisbon: ISCA.
- Romportl, M. (1985). *Základy fonetiky*. Praha: Karolinum.
- Skarnitzl, R. (2010). Prague Phonetic Corpus: status report. In: R. Skarnitzl (ed.), *AUC Philologica - Phonetica Pragensia XII*, 65-67. Praha: Karolinum.
- Volín, J. (2010). On the significance of the temporal structuring of speech. In M. Malá & P. Šaldová (Eds.), *... for thy speech bewrayeth thee* (p. 289-305). Praha: FF UK.
- Volín, J. & Churaňová, E. (2010): Probabilities of consonantal sequences in continuous Czech texts. In: *AUC Philologica 1, Phonetica Pragensia XII*, 49-62.
- Weingartová, L. (2011). *Ukazatele identity mluvího v oblasti temporálních modulací řečového signálu* (Diplomová práce). FF UK. Získáno z Digitálního univerzitního repozitáře (číslo záznamu 000011195).

Přílohy

Přílohy k této práci tvoří 11 tabulek (soubory typu *.xlsx) s extrahovanými daty nahrávek i mluvčích. Soubory jsou umístěny na přiloženém CD ve složce *Přílohy*.

1. Data_mluvcih.xlsx

Zde se nacházejí informace o nádechových úsecích a počtech slov a taktů v projevech jednotlivých mluvčích a údaje o trvání nahrávek.

2. Slova_takty_vysledky-slova.xlsx

V tomto souboru se nalézají kompletní tabulky slov i taktů s rázem i bez rázu včetně informací o slovních druzích slov. Dále jsou zde tabulky dat slov se zvlášť značenými slabikotvornými likvidami a bez hezitací, tabulky bez speciálně značených likvid a hezitací a příslušné kontingenční tabulky.

3. Vysledky_takty.xlsx

Kompletní tabulky taktů s rázem i bez rázu; tabulky dat taktů se zvlášť značenými likvidami a bez hezitací, tabulky bez speciálně značených likvid a hezitací a příslušné kontingenční tabulky.

4. Vytah_KT_slova.xlsx

Tabulky s důležitými daty extrahovanými z kontingenčních tabulek v souboru *Slova_takty_vysledky-slova.xlsx*.

5. Vytah_KT_takty.xlsx

Tabulky s důležitými daty extrahovanými z kontingenčních tabulek v souboru *Vysledky_takty.xlsx*.

6. Vysledky_mluvci+cor_slova.xlsx

Soubor obsahuje tabulky s rozdělením dat z tabulky slov (bez hezitací, s likvidami) podle jednotlivých mluvčích a list provedených oprav.

7. Vysledky_mluvci+cor_takty.xlsx

Soubor obsahuje tabulky s rozdělením dat z tabulky taktů (bez hezitací, s likvidami) podle jednotlivých mluvčích a list provedených oprav.

8. Slova_razy.xlsx

Pomocné tabulky slov s rázy značenými jako X pro usnadnění zpracování dat (pracovně je ve většině souborů ráz značen jako otazník, což znesnadňuje vyhledávání v datech s rázem). Obsahuje i příslušné kontingenční tabulky.

9. Takty_razy.xlsx

Pomocné tabulky taktů s rázy značenými jako X pro usnadnění zpracování dat. Obsahuje i příslušné kontingenční tabulky.

10. ST_komb_obsah+zacatky+konce

Soubor obsahuje počty všech jednoduchých výskytů konsonantů a vokálů a jejich dvojčlenných a trojčlenných kombinací ve slovech i taktech v různých pozicích v jednotce a přepočet na procenta. Základem pro tento soubor byla data slov i taktů bez hezitací a se speciálně značenými likvidami.

11. Pomocny_sldr

Pomocné tabulky ke zpracování slovních druhů; obsahují např. data s rázem značeným jako C a celkové tabulky a grafy variability vzorců s ohledem na slovní druhy.