

## **ANOTACE**

Hlavním cílem této práce je návrh řečového korpusu ukrajinštiny. V procesu zkoumání tohoto problému jsme se seznámili s problematikou syntézy řeči z textu (TTS) a přípravou inventáře řečových jednotek. Pro účely výstupního řečového korpusu byla v rámci této práce navržena fonetická abeceda ukrajinského jazyka. Další problematikou řešenou v této práci je fonetika ukrajinštiny, která je obecně rozebírána se zřetelem na její využití v syntéze ukrajinské řeči. Na základě probrané fonetiky byla vytvořena sada obecných fonetických pravidel, kterou jsme úspěšně použili pro přepis psaného ukrajinského textu na fóny.

Řečový korpus jsme nahráli a popsali na slovní a fonetické úrovni, kde jsme použili navržená pravidla fonetické transkripce. Po nahrávání byla úspěšně provedena segmentace řečového korpusu a ověřena kvalita výstupního korpusu.

## **ABSTRAKT:**

The main goal of this work is the proposal of the Ukrainian speech corpus. In the process of examining this problem, we introduced the issue of Speech Synthesis (TTS) and the preparation of speech unit inventory. For the purpose of the output speech corpus, the alphabet of the Ukrainian language was proposed in this work. Another issue solved in this thesis is the phonetic part of Ukrainian, which is generally solved with regard to its use in the synthesis of Ukrainian language. Based on the issued phonetic part, we created a set of general phonetic rules that we successfully used to overwrite the written Ukrainian text on the phone.

We have recorded and described the speech corpus at the verbal and phone levels where we used the proposed phonetic transcription rules. After the recording, the speech corpus segmentation was performed successfully and the quality of the output corpus was verified.

## **KLÍČOVÁ SLOVA**

Ukrajinský jazyk, počítačové zpracování, fonetická transkripce, konkatenační syntéza řeči.

## **KEY WORDS:**

Ukrainian language, computer processing, phonetic transcription, concatenative speech synthesis.

## Obsah

|  |    |
|--|----|
| <b>Obsah</b> .....                                       | 2  |
| 1. Úvod.....   | 3  |
| 1.1. Úvod do pojmu syntéza řeči.....                     | 3  |
| 1.2. Historie vývoje syntézy řeči.....                   | 3  |
| 1.3. Principy syntézy řeči.....                          | 4  |
| 1.4. Praktické využití TTS.....                          | 6  |
| 2. Reprezentace řeči.....                                | 8  |
| 2.1. Akustická reprezentace.....                         | 8  |
| 2.2. Fonetická reprezentace.....                         | 8  |
| 2.3. Fonologická reprezentace.....                       | 9  |
| 2.4. Fonetická abeceda.....                              | 9  |
| 2.4.1. Fonetická abeceda IPA.....                        | 9  |
| 2.4.2. Fonetická abeceda SAMPA.....                      | 10 |
| 2.5. Fonetická abeceda ukrajinštiny.....                 | 10 |
| 3. Úvod do fonetiky ukrajinštiny.....                    | 12 |
| 3.1. Samohlásky.....                                     | 13 |
| 3.1.1. Artikulační vlastnosti samohlásek.....            | 14 |
| 3.1.1.1. Artikulační vlastnosti předních samohlásek..... | 14 |
| 3.1.1.2. Artikulační vlastnosti zadních samohlásek.....  | 15 |
| 3.1.2. Akustické vlastnosti samohlásek.....              | 16 |
| 3.1. Souhlásky.....                                      | 17 |
| 3.2.1. Artikulační vlastnosti souhlásek.....             | 17 |
| 3.2.2. Akustické vlastnosti souhlásek.....               | 18 |
| 4. Fonetická transkripce ukrajinštiny.....               | 20 |
| 5. Příprava řečového korpusu ukrajinštiny.....           | 25 |
| 5.1. Stažení a úprava textů.....                         | 25 |
| 5.2. Výběr vět.....                                      | 26 |
| 5.3. Nahrávání vět.....                                  | 28 |
| 5.4. Ortografické a fonetické anotace.....               | 29 |
| 5.5. Segmentace řečového korpusu.....                    | 29 |
| 6. Závěr.....  | 31 |
| 7. Literatura.....                                       | 32 |

# 1. Úvod

## 1.1. Úvod do pojmu syntéza řeči

Existuje pět základních forem lidské komunikace, které mohou sloužit pro přenos informací: zraková, dotyková, pachová, zvuková a chuť. Pokud zvuková komunikace nese význam a je tvořena pomocí hlasových orgánů, označuje se jako řeč. Mluvená řeč je složitý proces, který zahrnuje nejen mnoho orgánů člověka, ale vyžaduje i roky praxe pro její dokonalé zvládnutí a správné použití.

S rozvojem technologií nabývá automatické generování řeči na popularitě i v životě běžného člověka. Syntéza řeči je umělou reprodukcí lidské řeči. S uplatněním řečové syntézy se můžeme setkat v rozmanitých zařízeních, které používáme každý den. Zmíněná technologie nejen zjednodušuje život, ale zároveň pomáhá lidem s omezenými možnostmi. Široké spektrum odvětví, kde by bylo možné využít řečovou syntézu, dokazuje, že její rozvoj je důležitým úkolem pro vědce, vývojáře a nadšence v oblasti umělé inteligence.

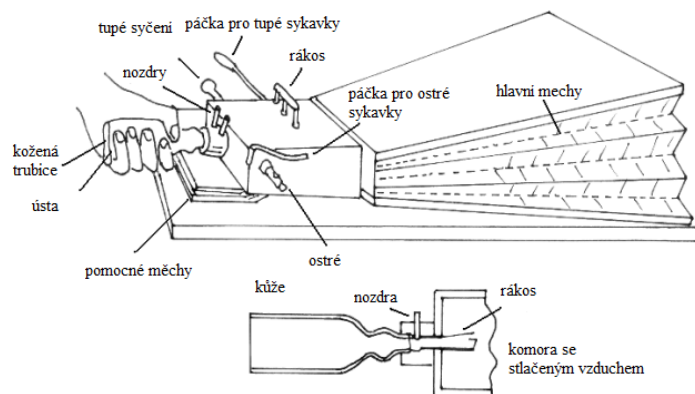
## 1.2. Historie vývoje syntézy řeči

Syntéza řeči se začala rozvíjet v období průmyslové revoluce a tento proces pokračuje i v současné době elektronických systémů.

Na konci XVIII. století vytvořil dánský vědec Christian Kratzenstein, člen Ruské akademie věd, model lidského hlasového ústrojí, který produkoval pět dlouhých samohlásek (a, e, a o, u). Model měl tvar systému akustických rezonátorů různých tvarů, vydávajících samohlásky pomocí vibrací rákosu vyvolaných proudem vzduchu.

V roce 1778 doplnil rakouský vědec Wolfgang von Kempelen Kratzensteinův model o modely jazyka a rtů a prezentoval akusticko-mechanický mluvicí stroj, který je schopen reprodukovat určité zvuky a jejich kombinace.

V roce 1837 představil vědec Charles Wheatstone vylepšenou verzi zařízení von Kempelena, které umělo přehrávat většinu samohlásek a souhlásek. Struktura tohoto mechanismu je ukázána na obrázku 1.



Obr. 1: Vylepšená verze zařízení von Kempelena vytvořené Charlesem Wheatstonem [1].

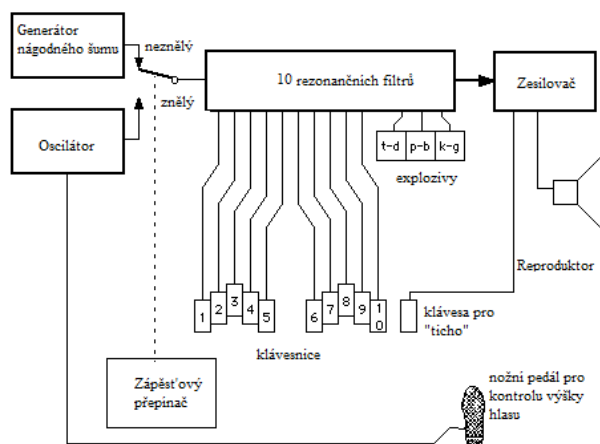
A v roce 1846 představil Joseph Faber své řečové zařízení Euphonia, které syntetizovalo nejen řeč, ale i zpěv.

Na konci XIX. století vyvinul slavný vynálezce Alexandr Bell svůj vlastní „mluvící“ mechanický model, který měl svůj vlastní design velmi podobný Wheatstoneovu stroji.

S nástupem XX. století začala éra elektrických obvodů a vědci byli schopni používat generátory akustických vln a na jejich základně stavět modely.

V roce 1930 zaměstnanec Bell Labs Homer Dudley, pracující na výzvě k nalezení způsobů, jak snížit šířku pásma požadovanou pro telefon s cílem zvýšit jeho přenosovou kapacitu, vyvinul tzv. vocoder (zkratka z

anglického slova pro hlas – voice, srov. anglické kódér – encoder), jenž se ovládal pomocí elektronického analyzátoru klávesnice a hlasového syntezátoru. Vylepšená verze vokodéru Dudley, VODER, byla představena v na světové výstavě v New Yorku v roce 1939. VODER byl první elektrický syntezátor řeči. Schema VODERu je na obrázku 2.



Obr. 2 Schéma řečového syntetizéru VODER [1].

Po demonstraci VODERu vědecký svět už více zajímala technologie řečové syntézy. Nakonec bylo prokázáno, že lze uměle vytvořit srozumitelnou řeč. Základní struktura a myšlenka VODERu je velmi podobná moderním systémům, které jsou založené na modelu řeči se zdrojovým filtrem.

První syntezátory řeči na bázi výpočetní techniky se začaly objevovat na konci 50. let 20. století a první syntezátor „text-to-speech“ byl vytvořen v roce 1968.

### 1.3. Principy syntézy řeči

Existuje několik možností z pohledu použití technologie generování umělé řeči, a to pomocí softwarové a hardwarové realizace.

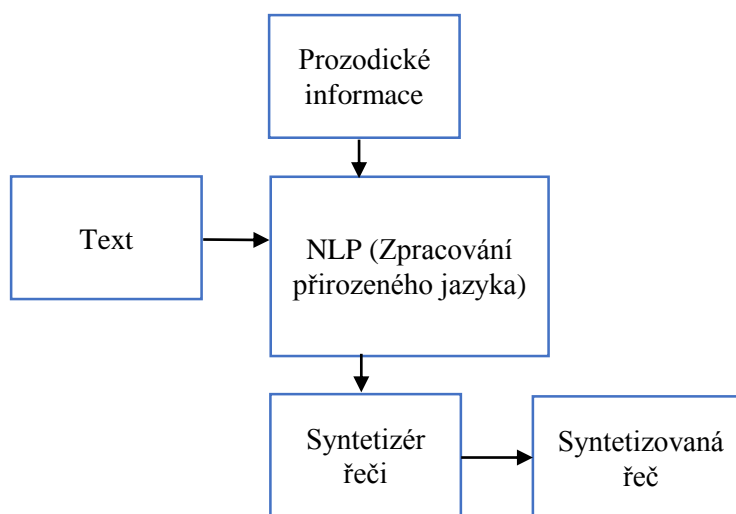
Z pohledu softwarové realizace existuje tzv. syntetizér řeči, program, který provádí konverzi textu na zvukový signál a je hlavní součástí všech generátorů umělé řeči. Systém TTS (angl. text-to-speech) je program, který má za úkol generovat řeč na základě libovolného vstupního textu, například napsaného na klávesnici nebo rozpoznaného po skenování, a systém by měl následně vygenerovat mluvenou řeč.

Proces TTS syntézy obsahuje dva nejdůležitější kroky. Prvním je analýza vstupního textu, který má nějakou lingvistickou reprezentaci. Tato reprezentace je většinou transkripce textových grafémů na fonémy a prozodickou informaci o tom, jakým způsobem by měl být text přečten. Prozodie popisuje zvukové vlastnosti jazyka, kterými jsou například přízvuk, tón, intonace, frázování, rytmus atd. Konverze vstupního textu lze rozdělit na tři hlavních fází: předzpracování nebo normalizace, tvorba lingvistických údajů pro správnou výslovnost, analýza prozodických charakteristik pro správnou intonaci, přízvuk atd.

Normalizace textu je proces procházení textem a mazání nebo přepíšu textových prvků, jako jsou čísla, data, časy, zkratky a speciální znaky (symboly měn atd.). Tvorba lingvistických údajů je fáze sestavení pravidel přepíšu mluveného textu, který často se liší od psaného. Analýza prozodické informace je nutná kvůli tomu, že jedna věta může být čtena různými způsoby v závislosti na významu, který chce donést, osoba, která mluví (je to jedna z nejtěžších problémů řečových syntezátorů).

Druhým krokem je generování zvukových vln podle zadané reprezentace řeči. Kvalita syntetické řeči se většinou hodnotí podle dvou základních kritérií: srozumitelnosti (jak dobře je syntetické řeči rozumět) a přirozenosti (jak přirozeně syntetická řeč zní). V obou případech se využívají poslechové testy, při nichž posluchači hodnotí sledované atributy umělé řeči. Kvalita řečové syntézy je nevyšší, jestliže není možné

rozlišit uměle vygenerovanou a přirozenou lidskou řeč. Schéma algoritmu TTS syntézy je ukázána na obrázku. 3.



Obr. 3: Proces TTS syntézy.

V současné době existuje dva základní přístupy generování umělé řeči: konkatenční a statistická parametrická syntéza.

Konkatenční přístup je způsob syntézy řeči, který je v dnešní době nejpoužívanější a nejjednodušší. Základní princip tohoto přístupu spočívá v řetězení řečových jednotek. Pro tuto syntézu zdrojové řečové prvky slovníku jsou předem nahrané. Je zřejmé, že obsah syntézovaných zpráv je určen slovníkem.. Hlavním problémem při konkatenční syntéze je objem paměti pro uložení slovníku. Kvůli tomuto používají se různé metody kompresi nebo kódování řečového signálu.

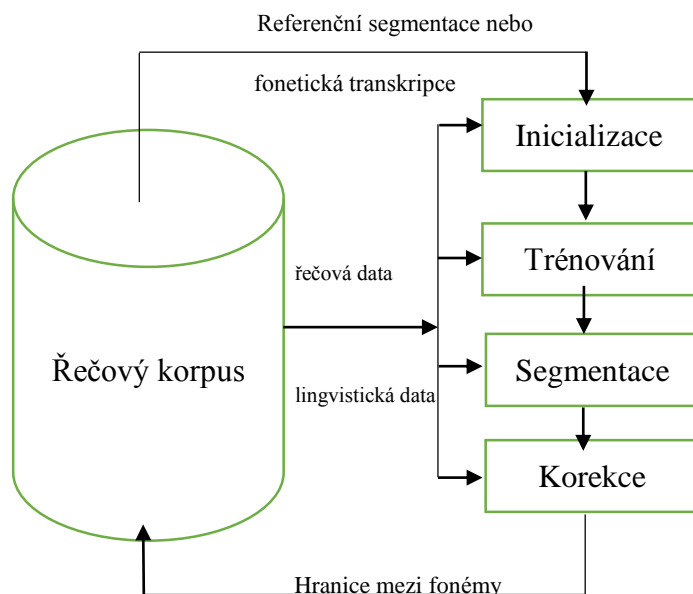
Pomocí konkatenční technologie syntézy může být vytvořena přirozenější řeč. Nicméně rozdíly mezi přirozenými variacemi řeči a povaze automatizovaných technik k segmentaci zvuků někdy vede ke slyšitelným chybám na výstupu. Existují dva hlavní typy konkatenčního přístupu: zřetězení difonů (segmenty obsahující zvukovou informaci ze dvou konkrétních sousedících fonémů.) a korpusově orientovaná metoda.

Systémy řečové syntézy založené na korpusu jsou schopné produkovat přirozenější řeč. Takové systémy jsou založené na dynamickém výběru jednotek a na velkém množství řečových dat. Tato metoda je také známá jako korpusová syntéza. Metoda získala popularitu kvůli vysoce kvalitnímu syntetickému hlasu na výstupu, díky využití přirozené řeči jako jednotek pro zřetězení. Hlavní charakteristikou TTS metody založené na korpusu je využití rozsáhlé databáze nebo korpusu řečových jednotek. Hlavním problémem během přípravy korpusu je anotace, která vždy vyžaduje značné množství lidské práce. V rámci zpracování této práce jsme navrhovali databáze řečových jednotek zejména pro korpusově orientované přístupy syntézy řeči, ale existuje možnost jejího využití i pro ostatní techniky.

Výběr jednotek (angl. Unit Selection) je metodou konkatenčního korpusově orientovaného přístupu řečové syntézy. Její podstatou je to, že počítač bude sestavovat syntetizovanou řeč ne z databáze speciálně nahraných prvků (alofonů, difonů, slabik atd), z nichž každý je představen v jedinečném tvaru, ale z nahraných vět přirozeného jazyka, a pro každý se zvolí nejvhodnější variantu

Alternativou je použití statistických technik. Statistický a konkatenční přístup jsou oba závislé na shromažďování nebo analýze dat. Pro konkatenční přístup data jsou zapamatovaná, zatímco ve statistickém přístupu se snažíme pochopit obecné vlastnosti dat. Dvě výhody statistických modelů, spočívají v tom, že potřebujeme menší objem paměti pro uložení parametrů modelu, a, že můžeme různými způsoby modifikovat model. Výsledná řeč je generována z modelů. Existuje více technik statistické syntézy ale nejpoužívanější je použití skrytých Markovových modelů (angl. HMM=hidden Markov models) nebo (hlubokých neuronových sítí)

HMM proces skládá se ze dvou hlavních částí: trénování a syntézy. Řečový signál je představen jako sada parametrů (např. keprálních koeficientů) a parametry modelů jsou nastavovány automaticky pomocí trénovacích algoritmů, které jsou založené na technikách strojového učení (obrázek 4).



Obr. 4: Zjednodušená schéma HMM procesu [1].

Posloupnost fonémů se převede na posloupnost kontextově závislých jednotek. Následně se provede zřetězení kontextově závislých jednotek. Další krok je generování trvání stavu každého modelu, generování posloupnosti vektorů spektrálních parametrů, generování buzení. Výstupní řeč se generuje z natrénovaných modelů pomocí "syntetického filtru".

Výhodami HMM metody je stabilní kvalita (díky generování řeči nedochází k poklesu kvality jako pro metodu řetězení signálů), možnost změny řečových charakteristik (např. změny hlasu) pomocí změn parametrů modelů a výrazně nižší paměťová náročnost (ukládají se pouze parametry modelů). Nevýhodou je celkově nižší kvalita řeči na signálové úrovni způsobená především statistickým „průměrováním“ během modelování řeči a potřebou generovat řeč z parametrů.

#### 1.4. Praktické využití TTS

Systémy TTS syntézy nacházejí využití v libovolné oblasti, kde je potřebná řeč nebo komunikace uživatele a stroje. Tyto systémy mohou být prakticky nasazeny v telekomunikačních, vzdělávacích, průmyslových, společenských technologiích atd. Možné využití TTS je například v programech zaměřených na automatický překlad textů nebo v systémech ozvučování multimediálních systémů jako videí nebo animací. Občas používané jsou ve sférách spojených s komunikací se zákazníkem jako například informační centra veřejných podniků, internetové obchody, banky. V takovém případě TTS ve spojení s ASR systémy (tj. hlasovými dialogovými systémy) usnadňují práci operátorů a šetří čas klientů, protože ještě před spojením zákazník prostřednictvím mluvené řeči zadává údaje potřebné pro řešení svého požadavku.

Jednou z dalších možností využití TTS systémů je náhrada řečového aparátu u lidí se zdravotním omezením (postižením). Nejznámějším člověkem, který používá danou technologii, je Stephen Hawking, britský teoretický fyzik, který kvůli amyotrofické laterální skleróze od roku 1985 mluví jen pomocí uměle produkované řeči. Systémy mohou být doplněny o výstupy jako řádek s Braillovým písmem, který pak slouží nevidomým uživatelům. Kromě toho existuje možnost předčítání elektronických knih, dopisů, novin pro tuto kategorie postižených lidí.

Pomáhá i pro vidoucí s povoláním, které vyžaduje intenzivní zrakové zatížení. TTS je možné v takových případech využít, aby mohli zadávat zvukovou informaci (například při pilotování letadel, řízení nákladních aut atd.).

Systemy umělého generování řeči posouvají možnosti i ve vzdělání. V této době už není překvapující prezentace, která samostatně čte své snímky. Využití syntézy řeči lze by také využít při samostudiu cizích jazyků a při práci s chytrými systémy.

Přínosy TTS systémů jsou čtení dynamických, předem nepřipravených textů, provádění audio komunikace a poskytování odpovědí na určitý požadavek, hlasová upozornění/varování apod

## 2. Reprezentace řeči

### 2.1. Akustická reprezentace

Akustika je věda zkoumající zvuk, jeho fyzikální vlastnosti, přenos zvuku, specifika jeho šíření, percepce a vlivu. Akustika je oblast mechaniky, která studuje pružné kmity a vlny od nejnižších frekvencí (kolem 0 Hz) po nejvyšší frekvence. Zvuk je přenášen pomocí mechanických kmitů a je představen svou zvláštní akustickou reprezentací. Akustické parametry zvuku jsou rytmus, výška, délka, síla, tónbr a formantová struktura.

Kmity, které tvoří zvuky, mohou být rytmické, nebo periodické, kdy následující kmit se rovná předchozímu. Rytmické kmity jsou vytvářeny vibrací hlasivek a nazývají se tóny. Existují nerytmické (neperiodické) kmity, kdy následující kmit nerovná se předchozímu. Tyto kmity se nazývají hluky. Periodické kmity se nazývají tonální. Mezi tonální zvuky (tóny) patří vokály a šumové (atonální zvuky) zvuky jsou konsonanty.

Při artikulaci tonálních zvuků se jejich výška téměř nemění, zatímco výška atonálních zvuků se výrazně proměňuje. Výška zvuku je určena frekvencí. Výška zvuku se měří v hertzech (1 Hz je jeden kompletní kmit za sekundu). Průměrný mužský hlas má frekvence v intervalu přibližně 85–200 Hz a ženský hlas 160–340 Hz. Během mluvení se výška zvuku neustále mění, což určuje intonace.

Síla zvuku je akustický parametr definovaný amplitudou, konkrétně amplitudou zvukových vibrací hlasivek. Čím větší jsou amplitudové výkyvy, tím intenzivnější je zvuk. V důsledku toho je hlasitost mluvení úměrná síle zvuku. Nejvyšším stupněm intenzity je křik. V rámci ukrajinského jazyka má největší sílu přízvukový zvuk. V jednom řečovém aktu různé mají zvuky různou intenzitu. Zejména samohláska [a] má větší sílu výslovnosti než samohláska [i]. Jednotkou síly zvuku je decibel (dB).

Trvání nebo délka zvuku je určena časem jeho artikulace. Jednotkou měření je milisekunda (ms). V ukrajinštině trvání zvuku nepatří k důležitým charakteristikám popisujícím zvuk. V mluvené řeči se stejným rytmem je přízvuková samohláska obecně delší než nepřívuková, a pokud se vokály nacházejí v nepřívukové pozici, pak je foném [i] delší než [a]. Mezi nejdelší souhlásky [ʃ] patří afrikáty ([ʃʃ], [ʃʃ], [ʃʃ], [ʃʃ], [ʃʃ/ʃʃ]), a nejkratší je [p].

Tónbr nebo barva tónů je další akustickým parametrem zvuku, který mu dodává specifičnost. V porovnání s ostatními akustickými vlastnostmi je nejdůležitější pro rozpoznávání zvuků mluvené řeči. Tónbr se liší u různých zvuků a záleží na obsahu dodatečných šumů a tónů, které se přidávají k základnímu tónu tvořenému hlasivkami.

Formantová struktura zvuku je vlastnost, která se určuje na základě zvukového spektra. Formant je pásmo zesílení zvukové energie. Formanty jsou doplňkové tóny zvuku. Spektrum, počet a struktura každé formantové pásma jsou různé. Ukrajinská lingvistka Nina Totskiy studovala spektrogram ukrajinských samohlásek [ʃ] a zjistila, že například zvuk [a] má širší formantovou zónu 254–4060 Hz s největší intenzitou 806–1280 Hz a zvuk [i] má dvě odlišné pásma: 202–508 Hz a 2030–4060 Hz, s nejvyšší intenzitou pro každou zónu 254 Hz, resp. 2030 Hz. Formantová struktura určuje kvalitu zvuku. Každý zvuk je díky svým akustickým parametrům specifickým fenoménem.

### 2.2. Fonetická reprezentace

Základní jednotka fonetiky je hláska Jakákoli řečová jednotka (morfém, slovo, věta) je přenášena zvukovým projevem. Zvuková podoba jazyka je předmětem fonetiky. Fonetika (z řeckého φωνή phōnē – hlas, zvuk) je věda, která zkoumá zvuky jazyka.



Předmětem studia fonetiky jsou nejen jednotlivé zvuky, ale také zákony jejich kombinací (kompatibilita), fonetické procesy (vliv pozice zvuku ve slově a jiných zvuků v sousedství na jeho znění) povaha a struktura slabiky, přízvuk a intonace.

Rozlišuje se fonetika obecná a specifická. Obecná fonetika zkoumá charakteristiky obecných znaků zvuků všech jazyků. Specifická fonetika zkoumá zvuky stanoveného jazyka nebo skupiny jazyků.

Fonetika má velký význam v praxi. Znalosti fonetiky jsou potřebné k vytvoření abecedy, zlepšení grafiky a pravopisu, ke studiu ortoepie různých jazyků (studium cizího jazyka obvykle začíná výukou fonetiky), k výuce a porozumění znakové řeči (ASL), automatickému rozpoznávání řeči počítačem.

### **2.3. Fonologická reprezentace**

Fonologie je část jazykovědy, která zkoumá funkcionální a strukturní zákonitosti zvukové struktury jazyka. Tím se fonologie liší od obecné fonetiky, proto bývá fonologie někdy označována jako funkcionální fonetika.

Základní jednotkou fonologie je foném. Foném vystupuje jako jednotka, která zobecňuje nějakou zvukovou množinu a charakterizuje jazyk jako společenský jev. Foném a zvuk jako fyzikální jev nejsou ve fonologii stejnými pojmy. Na rozdíl od fonémů nabývají zvuky různých osobitostí v závislosti na mluvčím.

Fonologie a morfologie jsou těsně spojeny, protože fonémy slouží pro rozlišování morfémů, jsou způsobem vyjádření a rozlišení gramatického smyslu slov (рыка—рыку), změna gramatické formy slova způsobuje střídání fonémů (рыка—рыці, рух—рушати, ніс—носа).

Existuje obecná fonologie (zkoumá obecné zákony funkcí fonémů v různých jazycích), charakteristická (zkoumá zákony fungování fonémů v určitém jazyce nebo skupině jazyků) a historická (úkolem je zkoumání změn fonémů jazyka během určitého období vývoje nebo jeho existence).

Historické změny fonologie přímo souvisí s fonetickými a jsou na nich založené. V důsledku těchto změn mohou některé fonémy zaniknout. Například u Slovanů zanikly v průběhu dějin nosové samohlásky. Samohlásky [ъ] [ъ], [Ѣ] zmizely ve staroruském období. Foném [y] se v ukrajinštině vyvinul v důsledku spojení fonémů [ы] a [і] [např. съмъ > syn].

### **2.4. Fonetická abeceda**

Fonetická abeceda je standardizovaný způsob zápisu výslovnosti jazyka. Je určena pro účely daného jazyka nebo organizace.

#### **2.4.1. Fonetická abeceda IPA**

Fonetická abeceda IPA je jedná z fonetických abeced (systém znaků pro zápis fonetické transkripce). Zobrazují specifičností tvoření zvuků všech světových jazyků na základě speciálních znaků. Mezinárodní fonetická asociace za účelem popularizace praktického využití fonetiky vytvořila IPA (International Phonetic Alphabet) a dosud ji koordinuje.

IPA je široce využívána učiteli cizích jazyků, studenty, lingvisty, logopedy, zpěváky, herci, překladateli atd. Hlavní funkcí fonetické abecedy IPA je předávání ústních jazykových zvláštností, například fonémů, intonací, rozdělení slov na slabiky apod. Pro ukrajinský jazyk je abeceda sestavena na základě ukrajinského cyrilického písma s využitím dodatečných liter a nad-řádkových symbolů. Celkem ukrajinská IPA obsahuje 6 samohláskových fonémů a 32 souhláskových fonémů (viz tabulka 1.).

## **2.4.2. Fonetická abeceda SAMPA**

Speech Assessment Methods Phonetic Alphabet je systém zápisu znaků fonetické transkripce určený pro počítačové zpracování textu a založený na mezinárodní abecedě IPA. V abecedě je použito kódování ASCII s objemem paměti sedm bitů pro jeden psaný symbol.

SAMPA byla vyvinuta na konci 80. let dvacátého století v rámci projektu ESPRIT, financovaného Evropským hospodářským společenstvím (EHS) pro šest světových jazyků. Cílem bylo usnadnit výměnu e-mailových dat a zpracování transkripce v oblasti fonetiky a řečové techniky. V současné době SAMPA je rozpracována pro 29 světových jazyků a pro každý z nich existuje samostatná SAMPA, což neumožňuje její univerzální využití. Proto byla v roce 1995 vytvořena rozšířená verze, která má název X-SAMPA (angl. Extended Speech Assessment Methods Phonetic Alphabet). Hlavním cílem zakladatele X-SAMPA Johna C. Wellse bylo spojení všech abeced SAMPA v jednu a také její rozšíření o chybějící symboly. Fonetická abeceda ukrajinštiny.

## **2.5. Fonetická abeceda ukrajinštiny**

Fonetická abeceda nebo také fonetický inventář ukrajinské mluvené řeči je sada fonémů ukrajinského jazyka reprezentovaných pomocí specifických značek. Fonetickou abecedu je možné popsat pomocí mezinárodních abeced jako IPA a SAMPA (viz kapitoly 2.4.1, 2.4.2.). Kvůli tomu, že IPA je příliš složitá abeceda pro počítačové zpracování a není úplně vhodná pro popis ukrajinštiny, jako slovanského jazyku což způsobuje to, že není často používanou v praxi. Pro úlohy automatického zpracování řeči se častěji používají abecedy, se kterými je možné snadněji pracovat a zobrazovat na počítači (například SAMPA). Další abeceda, kterou jsme použili pro popis ukrajinských fonémů je EPA (ERIS phonetic alphabet). Tato sada symbolů je interní abeceda Západočeské univerzity, která byla speciálně přizpůsobená pro ukrajinštinu. Fonetický inventář je založen na bázi jednoduchých symbolů latinské abecedy, které odpovídají jednotlivým ukrajinským fónům. V této práci budeme používat abecedu EPA pro všechny příklady fonetické transkripce ukrajinštiny.

Při návrhu řečového korpusu ukrajinštiny jsme používali navrženou interní fonetickou abecedu EPA, kterou jsme porovnali v tabulce 2. s ostatními fonetickými abecedami. Je zřejmé, že všechny fóny ukrajinštiny se dělí na skupiny, které vyznačují a zevšeobecňují jejich lingvistické osobitosti. V tabulce 2. je vidět klasifikace fónů, která je podrobněji rozebrána v kapitole 3.

|         | EPA | IPA | SAMPA | Ukrajinské příklady | Česky     |           | EPA | IPA       | SAMP A         | Ukrajinsk é příklady | Česky    |
|---------|-----|-----|-------|---------------------|-----------|-----------|-----|-----------|----------------|----------------------|----------|
|         | Fón | Fón | Fón   | Příklad:            | Příklad:  |           | Fón | Fón       | Fón            | Příklad:             | Příklad: |
| Vokály  | i   | i   | i     | очи                 | oči       | Frikativy | f   | f         | f              | форма                | forma    |
|         | I   | i   | "i    | діту                | děti      |           | F   | fʃ        | f'             | фізика               | fyzika   |
|         | Y   | I   | "l    | туh                 | ohrada    |           | v   | v         | v              | звук                 | zvuk     |
|         | y   | I   | l     | пруходуту           | přicházet |           | V   | vʲ        | v'             | віза                 | vízum    |
|         | a   | ɛ   | a     | мова                | jazyk     |           | s   | s         | s              | сун                  | syn      |
|         | A   | a   | "a    | сова                | sova      |           | S   | sʲ<br>(ɛ) | s'             | сіно                 | seno     |
|         | o   | o   | o     | носуту              | nosit     |           | z   | z         | z              | знату                | vědět    |
|         | O   | ɔ   | "o    | поле                | pole      |           | Z   | zʲ<br>(z) | z'             | зірка                | hvězda   |
|         | u   | ʊ   | u     | угода               | dohoda    |           | w   | ʃ         | S              | школа                | škola    |
|         | U   | u   | "u    | стипа               | moždír    |           | W   | ʃʲ        | S'             | щика                 | štika    |
|         | e   | e   | e     | поле                | pole      |           | q   | ʒ         | Z              | жур                  | tuk      |
|         | E   | ɛ   | "e    | телефон             | telefon   |           | x   | x         | x              | холод                | chlad    |
| Plozivý | p   | p   | p     | терен               | plot      | X         | xʲ  | x'        | хімія          | chemie               |          |
|         | P   | pʲ  | p'    | пір'я               | pero      | r         | r   | r         | рука           | ruka                 |          |
|         | b   | b   | b     | биква               | bukva     | R         | rʲ  | r'        | річка          | říčka                |          |
|         | B   | bʲ  | b'    | більшуй             | větší     | j         | j   | j         | дбайте         | šetřit               |          |
|         | t   | t   | t     | тема                | téma      | l         | l   | l         | Лелека         | čáp                  |          |
|         | T   | tʲ  | t'    | тіло                | tělo      | L         | lʲ  | l'        | любов          | laská                |          |
|         | d   | d   | d     | духату              | dýchat    | Q         | ʒʲ  | -<br>(zz) | прудорі<br>жжя | úpatí                |          |
|         | D   | dʲ  | d'    | дім                 | dům       | H         | hʲ  | h'        | гілка          | větev                |          |
|         | k   | k   | k     | кора                | kůra      | h         | h   | h\        | голос          | hlas                 |          |
|         | K   | kʲ  | k'    | кімната             | pokoj     | ɟ         | ɟ   | d-z       | дзвін          | zvonění              |          |
| Nazály  | g   | g   | g     | ґава                | vrána     | dq        | dʒ  | d-w       | бджола         | včela                |          |
|         | m   | m   | m     | молоко              | mléko     | c         | ts  | ts        | цнота          | bezúhonnost          |          |
|         | M   | mʲ  | m'    | міст                | most      | G         | tsʲ | t'S'      | цього          | tohoto               |          |
|         | n   | n   | n     | носуту              | nosit     | C         | ʃʲ  | tS'       | чоботу         | boty                 |          |
| N       | nʲ  | n'  | ніс   | nos                 | dZ        | dʒʲ       | dz' | дзьоб     | dзьоб          |                      |          |

Tab. 1: Fonetická abeceda ukrajinštiny. IPA a SAMPA jsou převzaty z [10].

### 3. Úvod do fonetiky ukrajinštiny

Fonetika je věda, která zkoumá zvuky řeči a zvukové složení jazyka. Nejmenší jednotkou řeči je foném. V ukrajinštině existuje 38 fonémů. Rozlišujeme souhlásky (konsonanty) a samohlásky (vokály), které se liší podle fonetických charakteristik. Obecně se klasifikace provádí podle artikulačních vlastností hlásek.

Předmětem studia artikulace jsou artikulační orgány, způsob tvoření jednotlivých hlásek i větších celků (např.: slabiky, slova, věty). Artikulační charakteristiky samohlásek a souhlásek se navzájem liší. V procesu tvoření souhlásek musí vydechovaný vzduch překonávat překážky. Svalové napětí během produkce souhlásek je mnohem vyšší než v případě samohlásek. Kromě toho se při tvoření samohlásek účastní jiné svaly než při tvorbě souhlásek: v prvním případě ty, které se používají k otevírání úst, a v druhém případě svaly, které slouží k zavírání úst.

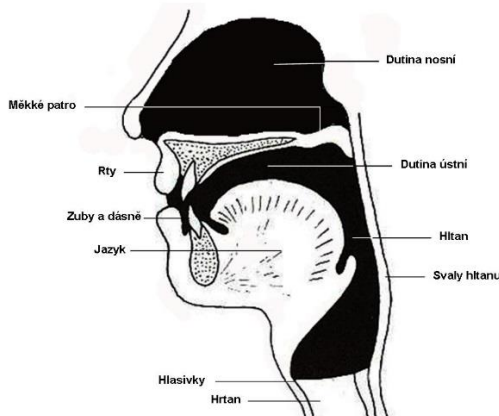
Funkčně nejsou samohlásky a souhlásky stejné. Osobitostí samohlásek je, že tvoří slabiky. Tato funkce však není absolutní. V některých jazycích mohou být základem slabiky i sonory. Obecně lze říci, že samohlásky mají větší význam pro vytváření slabik.

Proces artikulace zvuku obsahuje tři fáze:

- příprava řečového aparátu pro výslovnost zvuku;
- výslovnost zvuku určitou polohou orgánů;
- ukončení artikulačního procesu, který předpokládá změnu své polohy pro tvoření dalšího zvuku, nebo uklidňují se.

V procese tvoření hlásek se účastní artikulační orgány člověka nebo orgány vokálního traktu. Rozmístění jednotlivých jeho částí je vidět na obrázku 5. Na základě určitého místa jejich produkce se provádí fonetická klasifikace hlásek. Hlavním řečovým ústrojím je dýchací ústrojí (plíce, hrudní koš) – vytváří proud vzduchu, jeho výdechová část je základem mluvené řeči.

- fonační ústrojí (hrtan, hlasivky) - hlasivky rozkmitají proud vzduchu, v hrtanu se artikulují některé hlásky.
- dutina hltanová (stěna jazyka, kořen jazyka) - spojení hltanu s hlasivkami.
- dutina ústní (jazyk, rty, dásně, zuby, tvrdé patro, měkké patro) - tvoření všech hlásek s výjimkou hrdelních.
- dutina nosní - artikulace nosovek.



Obr. 5: Orgány vokálního traktu člověka. Převzato z [7].

Klasifikace artikulačních orgánů je podle jejich účasti na artikulačním procesu:

Aktivní orgány:

- dolní ret - labium
- špička jazyka - apex lingue [apex ligvé]
- hřbet jazyka - dorsum lingue
- hlasivková štěrbina - glottis

Pasivní orgány:

- hrtan - larynx
- tvrdé patro - palatum
- měkké patro - velum
- horní ret - labium
- horní zuby - dentes
- dásně - alveolae
- čípek – uvula

### 3.1. Samohlásky

Celkem v ukrajinštině existuje šest samohlásek [a, o, u, i, y, e]. Při psání tyto zvuky jsou reprezentované grafémy a, o, u, i, y, e, я, ю, є, і. V ukrajinské abecedě písmena mají obecný název bukvy. Bukvy я, ю, є, і mají název dvojhlásky. Bukva і vždy se označuje dvěma fóny: [ji]. Bukvy я, ю, є označují dva fóny [ja] [ju] [je] resp. v následujících případech:

- Na začátku slova. (яблуко [jAbluko]);
- Po samohlásce. (заява [zajAva]);
- Po měkkém znaku. (Мольєр [moLjEr]);
- Po apostrofu. (здоров'я [zdorOvja]);

V ostatních případech я, ю, є označují jeden fón hlásku. Například: лялька [LALka].

Samohlásky mohou být přízvučné a nepřízvučné. Přízvuk v rámci slova je zdůraznění slabiky ve slově pomocí hlasu. V ukrajinštině přízvuk je volný, tzn., že v různých slovech může být na různých slabikách. Kromě toho je pohyblivý a mění svou pozici například ve stejných slovech s různým významem. Je důležité, že stejná slova mohou mít jiný význam v závislosti na přízvučné slabice.

|            |            |
|------------|------------|
| байдУже    | бАйдуже    |
| доповідАч  | доповідАч  |
| договОру   | договОру   |
| маБуть     | маБУть     |
| первісний  | первісний  |
| помилка    | помИлка    |
| такОж      | тАкож      |
| весняний   | веснЯний   |
| завжди     | завждИ     |
| жалібний   | жалібний   |
| м'язовий   | м'язОвий   |
| перестАрок | перЕстарок |
| простий    | прОстий    |

Tab. 2: Dvojitý přízvuk některých ukrajinských slov.

Kvůli tomu často aby pochopit význam slov potřebujeme kontext, ve kterém vyskytují. Některá slova mají dvojité přízvuk (viz tabulce 3), jev kdy je možné zdůraznění jedné ze dvou slabik ve slově a význam se nemění. V mezinárodní abecedě IPA přízvuk označuje se symbolem (´), který se píše před přízvučným vokálem [ˈa, ˈe, ˈo]. V ukrajinské jazykovědě přízvuk se označuje znakem (˘), který se píše nad přízvučným vokálem [á, é, ó].

V následující tabulce jsou uvedena slova se stejným významem ale různými přezůvkami, kde velké písmeno označuje přízvuk.

Charakteristikou samohlásek je pozice ve slově. Rozlišuje se silná a slabá pozice. Silná pozice je v případě pokud samohláska ve slově je přízvučná nebo je vedle tvrdé souhlásky [pArta, dubOvyj]. Slabá pozice pro samohlásky je nepřízvučnost a existence vedle nich měkkých souhlásek.

### 3.1.1. Artikulační vlastnosti samohlásek

Vokály se vytváří pomocí hlasivek. Hlasivky člověka kmitají a vzduchový proud volně prochází z hrtanu, hrdelní a ústní dutinou a nakonec je vyzařován rty do okolí [1]. Stupeň posunutí jazyka dopředu a stupeň jeho zdvihu při jejich tvorbě určuje artikulační skupiny samohlásek. Při pohybu jazyka horizontálním směrem klasifikujeme samohlásky ke skupinám předních (palatálních), středních (centrálních) a zadních (velárních) samohlásek. Podle pravidel ukrajinské fonetiky žádná samohláska nepatří ke skupině středních samohlásek. Vertikální posun určuje výšku samohlásek, která může být vysoká, téměř vysoká, středová a nízká. Klasifikace ukrajinských samohlásek je představená v tabulce 4.

|                        | Přední | Střední | Zadní |
|------------------------|--------|---------|-------|
| Vysoká (zavřená)       | [i]    |         | [y]   |
| Téměř vysoká (zavřená) | [y]    |         |       |
| Středová               | [e]    |         | [o]   |
| Nízká (Otevřená)       |        |         | [a]   |

Tab. 3: Artikulační vlastností samohlásek.

Klasifikujeme také samohlásky podle postavení rtů, které mohou být zaokrouhlené (labiální) nebo nezaokrouhlené (nelabiální). Při tvorbě labiálních ([o],[u]) samohlásek rty jsou zaokrouhlené a protáhnuté dopředu. Při tvorbě nelabiálních ([a],[e],[y],[i]) samohlásek rty nevykonávají aktivní činnost.

#### 3.1.1.1. Artikulační vlastnosti předních samohlásek.

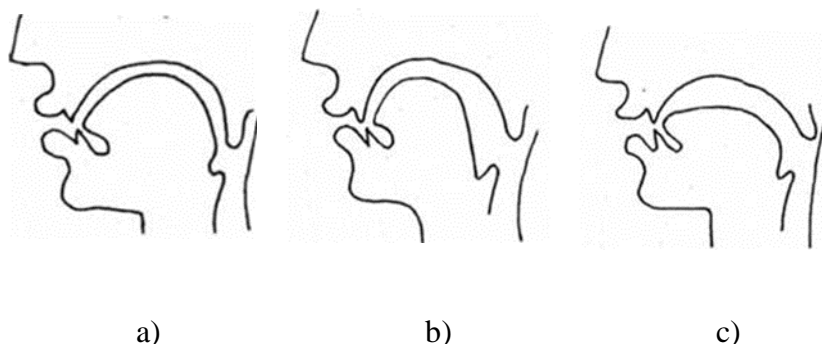
Hlavní charakteristiky fonu [e]: nelabiální, přední samohláska, středního zdvihu, otevřená. Při jeho tvoření dutina ústní je otevřená více než při tvoření všech ostatních předních samohlásek (viz obrázek 6a.). Rty nezúčastňují se aktivně při artikulaci [e] a ve srovnání s [i] a [y] méně vytažené, ale víc otevřené.

Zadní strana jazyka je vytažena dopředu, blíže ke středu patra. Poloha jazyka během artikulaci je střední, ale výrazně nižší než pro ostatní přední samohlásky. Rezonátory jsou částečně přední a zadní části ústní dutiny.

Hlavní charakteristiky fonu [y]: nelabiální, přední samohláska, vysoko-středního zdvihu, otevřená nebo také zavřená. Při artikulaci samohlásky [y] jako základního fonu /y/ dutina ústní je zavřená více než při artikulaci samohlásky [e] (viz obrázek 6b.).

Přední část jazyka je konvexní a jeho hlavní hmota vytažena dopředu. Stupeň posunu je střední ve srovnání s artikulací [e] a [i]. Poloha jazyku je vyšší, než při výslovnosti [e], ale nižší než při výslovnosti [i]. Rezonátorem je zadní část ústní dutiny.

Hlavní charakteristiky fonu [i]: nelabiální, přední samohláska, vysokého zdvihu, otevřená nebo také zavřená. Při artikulaci samohlásky [i] jako základního fonu [i] dutina ústní je otevřená nejmíň (viz obrázek 6c.). Kvůli tomu [i] patří mezi nejvíc zavřené přední samohlásky.

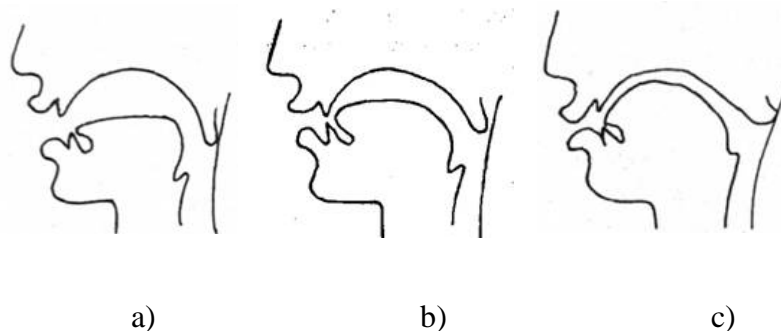


Obr. 6: Stav vokálního ustrojí při artikulaci ukrajinských předních samohlásek.

Celá hmota jazyka (přední a střední části) je vytažena dopředu k přední části patra nebo dolních zubů a alveolů. Jazyk nabývá konvexního tvaru. Míra zdvihu zadní části jazyka ke tvrdému patru při artikulaci [i] je nejvyšší ve srovnání se všemi ostatními předními samohláskami. Rty, které tvoří úzkou štěrbinu nataženou do stran mnohem více, než při artikulaci [y] a [e]. Měkké patro uzavírá vstup do nosní dutiny. K zadní části rezonátoru v ústní dutině dodatečně je připojený (jako výslovnosti přední samohlásky [a] a [e]), rezonátor generovaný mezi přední částí jazyka a zuby přední části ústné dutiny.

### 3.1.1.2. Artikulační vlastnosti zadních samohlásek.

Základní variantou fonému [a] je nelabializovaná zadní samohláska s nízkým zdvihem a otevřená. Samohláska [a] jako základní alofon fonému [a] je nejotevřenější ze všech samohlásek (viz obrázek 7a.). Při její výslovnosti jsou ústa otevřena na maximální šířku. Výslovnost se provádí s nejnižším zdvihem jazyka, který se oddaluje zpět ke stěně hrtanu a ponechává otvor pro vydechovaný vzduch. Během tohoto procesu se v přední části úst formuje akustický rezonátor. Rty neplní aktivní účast při artikulaci samohlásky [a]. Měkké patro a zadní stěna hltanu jsou přitlačeny, a tím se uzavírá vstup vedoucí do nosní dutiny.



Obr. 7: Stav vokálního ustrojí při artikulaci ukrajinských zadních samohlásek.

Pro foném [o] v roli hlavního alofonu je samohláska [o], která je otevřená, labiální, zadní řady a středního zdvihu. Při artikulaci alofonu fonémy [o] dutina ústní je otevřena méně než pro [a], ale více než pro [u] (viz obrázek 7b.). Proto je samohláska [o] považována za uzavřenější při artikulaci ve srovnání s [a].

Zadní část stěny jazyka se posouvá zcela zpět. Jazyk má o něco vyšší polohu než při výslovnosti samohlásky [a], má tedy střední polohu. Rty jsou během artikulaci [o] jsou vytaženy dopředu a vzniká mezi nimi kruhový otvor. K rezonátoru v ústní dutině se dodatečně připojuje rezonátor vytvořený mezi

rty a zuby v důsledku procesu labializace, přičemž se současně snižuje rezonanční tón. Vstup do nosní dutiny je uzavřený.

Pro foném [u] je v roli hlavního alofonu samohláska [u], která je otevřená nebo také zavřená, labiální, zadní řady a vysokého zdvihu.

V procesu výslovnosti samohlásky [y] jako hlavního alofonu fonémy [u] je ústní dutina otevřena méně než při výslovnosti samohlásek zadní řady (viz obrázek 7c.). Jazyk je ve velmi vysoké poloze. Jazyk směřuje vzhůru k měkké nebo dokonce k zadní části tvrdého patra.

Poloha nejvyššího bodu zvednutí horní části jazyka je posunutá dopředu nejvíce pro samohlásku [o]. Ze všech zadních samohlásek [y] je nejvíce přemístěný dopředu. Při artikulaci samohlásky [y] rty jsou velmi napjaté a protáhlé dopředu, více než při výslovnosti labiální samohlásky [o], a uzavírá se vstup do nosní dutiny. K rezonátoru v ústní dutině dodatečně připojuje se rezonátor vytvořený mezi rty protáhnuté dopředu, více než při artikulaci [o], kvůli tomu samohláska [y] je příliš labiální.

### 3.1.2. Akustické vlastnosti samohlásek

Zvuk je fyzikální jev, který má svoje fyzikální vlastnosti (viz kapitolu 2). Když mluvíme o zvuku, při popisu artikulaci zvuků je kladen důraz na fyziologické problémy, které ovlivňují produkování zvuku.

Mezi akustickými parametry zvuků relevantní charakteristikou slouží frekvenční spektrum. Takže, při klasifikaci zvuků, je nutně se spoléhat na jejich spektrální strukturu. Spektrální struktura závisí na formantech – oblastech s vyšší koncentrací akustické energie při tvoření zvuků. Formanty jsou nejvýznamnější akustické charakteristiky samohlásek. Frekvenční rozsah jednotlivých formantů pro každou ukrajinskou samohlásku je reprezentován v tabulce 4.

| Formanty<br>Samohlásky | $F_1[Hz]$ | $F_2[Hz]$ | $F_3[Hz]$ | $F_4[Hz]$ |
|------------------------|-----------|-----------|-----------|-----------|
| [a]                    | 750-780   | 1200-1230 | 2100-2220 | 3500-3700 |
| [e]                    | 520-540   | 1630-1690 | 2200-2300 | 3660-3860 |
| [y]                    | 350-380   | 2100-2150 | 2450-2550 | 3640-3790 |
| [i]                    | 280-300   | 2270-2310 | 2950-3150 | 3580-3670 |
| [o]                    | 450-480   | 750-790   | 2420-2620 | 3460-3710 |
| [u]                    | 350-400   | 600-670   | 2470-2590 | 3630-3830 |

Tab. 4: Hodnoty pásem formantů ukrajinských samohlásek.

Zvuk [a] je určen třemi rezonančními pásmy. První z nich je nejširší v rozsahu od 400 Hz do 1400 Hz. Formanty  $F_1$ ,  $F_2$ , jsou velmi blízko vedle sebe a jsou dva vrcholy jedné rezonanční pásma. Třetí formant je v rozsahu 1800-2300 Hz, čtvrtý je v rozmezí 3200-4000 Hz.

Spektrum zvuku [e] je charakterizován třemi rezonančními pásma. První rezonance má rozsah až 700 Hz, tady je umístěn formant  $F_1$ . Druhá rezonance je široká (od 1200 do 2500 Hz), a má dva vrcholy jsou to druhý a třetí formanty  $F_2$ ,  $F_3$ . Třetí rezonanční pásmo je nejslabší z hlediska tónů, a má rozsah od 3400 do 4500 Hz. V této sekci je také čtvrtý formant  $F_4$ .

Spektrum samohlásky [i] má tři významné rezonanční pásma. První rezonance je ve velmi úzkém pásmu - 150 až 600 Hz. V první rezonanční části je umístěn první formant  $F_1$ . Druhý rezonanční pás je 2000-2500 Hz a reprezentuje druhý formant  $F_2$ , třetí rezonance je v rozsahu 2800 - 4000 Hz a tady jsou umístěné formanty  $F_3$ ,  $F_4$ .



Spektrum samohlásky [y] má tři významné rezonanční pásma. První z nich je nejintenzivnější, zabírá nejužší šířku pásma do 500 Hz. Druhá rezonance je poměrně rozsáhlá asi 1700-3000 Hz. F2 a F3 formanty jsou blízko u sebe. Čtvrtý formant, který je rozmístěný ve třetím rezonančním pásmu v rozsahu od 3400 do 4500 Hz.

Hlasitý [o] má tři rezonanční pásma ve svém spektru. První rezonanční pásmo je od 200 do 1000 Hz, a tady jsou rozmístěny F1 a F2. Druhá rezonanční pásmo je v rozsahu od 2000 do 2600 Hz a špičkou je F3, třetí je ve frekvenci 3000 až 3800 Hz, s vrcholem F4.

Spektrum zvuku [u], je charakterizován třemi rezonančními oblastmi. První rezonance je ve frekvenčním rozsahu od 200 Hz do 1000 Hz, a tady jsou první dva formanty F1, F2. Třetí a čtvrtý formanty F4, F3 jsou v rozsahu 2100-2600 Hz, 3400-3800 Hz resp.

### 3.1 Souhlásky

Souhlásky jsou zvuky vytvořené pomocí hlasivek a šumu nebo jen šumu. Celkem v ukrajinském jazyce existuje 32 souhlásky. Klasifikace souhlásek je provedená na základě obecných principů jejich tvoření a podle poměru šumu a tónu, tvrdosti, palatalizaci. Klasifikujeme souhlásky podle akustických a artikulačních charakteristik. Vlastnosti souhlásek a hlavní rozdíl od samohlásek:

- Přítomnost šumu nebo konsonantnost;
- neschopnost tvořit slabiku;
- nepřízvučnost.

#### 3.2.1. Artikulační vlastnosti souhlásek

Artikulace souhlásek je uskutečněná jiným stylem, než při artikulaci samohlásek. To je kvůli tomu, že samohlásky jsou akusticky produkovány pomocí rezonančních vlastností vokálního traktu, souhlásky potřebují plynoucí vzduch přes různorodé zúžení podél vokálního traktu. Vokální trakt musí být při tvorbě samohlásek průchodný po celé jeho délce a otevřený na straně ústního otvoru. U souhlásek může být zúžení vokálního traktu nebo jeho blokování umístěno v jakémkoliv místě. Souhlásky se liší od samohlásek aktivní částí vokálního traktu. U samohlásek je využíván veškerý vokální trakt, zatímco u souhlásek se uplatňuje variabilní část vokálního traktu. Typy vytvářených zvuků jsou rezonanční frekvence u samohlásek a periodické zvuky a neperiodické šумы u souhlásek.

Podle místa artikulace všechny souhlásky jsou rozděleny na retné souhlásky, dásňové souhlásky, patrové souhlásky, hrtanové souhlásky (viz tabulku 8).

Retné souhlásky nebo labiály se vytvářejí pomocí sevření spodního okraje horního rtu a dolního rtu nebo dolního rtu a horních zubů (labiálo-dentály). Labiály se dělí na bilabiály ([b], [p], [v], [m]) a labiálo-dentály ([f]). Dásňové souhlásky nebo alveoláry jsou takové souhlásky, které se artikulují pomocí předkové nebo středové části jazyka a alveoly. V ukrajinštině k této skupině patří například [d], [D], [t], [T], [l], [L], [n], [N], [dz], [dZ], [dq], [z], [Z], [s], [S], [c], [G], [q], [C], [w], [r], [R].

| Místo artikulace / Způsob artikulace | Labiály | Labiálo-Dentály | Dentálo-alveolární | Alveoláry | Post alveoláry | Palatály | Veláry | Glotála |
|--------------------------------------|---------|-----------------|--------------------|-----------|----------------|----------|--------|---------|
| Ploziva                              | p b     |                 | t d                |           |                |          | k g    |         |
| Afrikáta                             |         |                 | ʦ ʣ                |           | ʧ ʤ            |          |        |         |
| Nazály                               | m       |                 | n                  |           |                |          |        |         |
| Vibranta                             |         |                 |                    | r         |                |          |        |         |
| Frikativa                            |         | f               | s z                |           | ʃ ʒ            |          | x      | h       |
| Aproximanta                          |         | v               |                    |           |                | j        |        |         |
| Laterální aproximanta                |         |                 | l                  |           |                |          |        |         |

Tab. 5: Klasifikace základních souhláskových fonémů ukrajinštiny.

Patrové souhlásky je skupina souhlásek, k nimž patří palatály [j], [D], [T], [Z], [S], [G], [dZ], [L], [N], [R] a veláry [g], [k], [x].

Další klasifikace souhlásek je podle způsobu artikulace. Mezi závěrové souhlásky patří nazály, plozivy a afrikáty. Nazály ([m], [n]) vytvářené při sklonění měkkého patra, při čemž část vydechaného vzduchu prochází přes nosní dutinu. Afrikáty jsou souhlásky, které se vytváří při otevírání semknutí artikulačních orgánů. K nim patří [b], [p], [d], [m], [g], [k]. Plozivy nebo výbuchové souhlásky ([b], [p], [d], [t], [k], [g], [D], [T]) jsou zvuky, které při řeči vznikají uvolněním závěru (okluze) v mluvidlech. Glotály jsou souhlásky vytvořené pomocí hlasové štěrbiny například [h], [x]. Frikativy ([v], [z], [q], [h], [f], [s], [w], [x]) jsou souhlásky, které vznikají částečným semknutím artikulačních orgánů při tomto vzniká mezera, a vzduch tře o stěny mezery. Vibranty (např.: [r]) jsou souhlásky, které se vznikají vibrací artikulátorů, který uzavírá nebo otevírá cestu pro vzduch. Laterální aproximanty [l], vyznačující se tím, že proud vzduchu prochází hrany ústní dutiny. V tabulce souhlásek vyznačuje se jakým způsobem a v jakém místě vokálního traktu souhláska se formují.

### 3.2.2. Akustické vlastnosti souhlásek

Všechny konsonanty ukrajinštiny jsou plicní, tj. vytvořené díky překážce v ústech nebo krku a následujícího nebo současného vytlačení vzduchu z plic. Souhlásky se vytvářejí pomocí šumu a tónu. Základní akustické charakteristiky souhlásek závisí na podílu složek šumu a tónu v jejím spektru. Souhlásky podle akustických charakteristik se dělí na sonory a šumové souhlásky.

Sonory (lat. Sonore - zvučný, hlasitý), se nazývají souhlásky, skládající se z tónu a šumu, kde tónová složka převládá. V ukrajinštině je 9 sonorních souhlásek. Mezi sonorní souhlásky ukrajinštiny patří [m], [n], [N], [v], [l], [L], [r], [R], [j].

Šumové souhlásky se skládají z tónu (nebo hlasu) a šumu, kde dominuje šum nebo jen ze samostatného šumu. Šumové souhlásky jsou pak rozdělené na znělé a neznělé. Znělými se nazývají souhlásky, skládající se z tónu a šumu, kde dominuje šum. Neznělé souhlásky jsou takové souhlásky, které se skládají pouze ze samostatného šumu.

V moderní spisovné ukrajinštině sonory netvoří korelační páry a znělé šumové mají korespondující neznělé šumové. Neznělá souhláska [f] nemá korespondující znělou souhlásku. Níže uvedená tabulka uvádí znělostní páry

|                |   |   |   |   |   |    |   |    |   |   |    |
|----------------|---|---|---|---|---|----|---|----|---|---|----|
| <b>Znělé</b>   | b | h | g | d | q | dq | z | dz | D | Z | dZ |
| <b>Neznělé</b> | p | x | k | t | w | C  | s | c  | T | s | G  |

Tab. 6: Znělostní páry.

Souhlásky na určitých pozicích ve slovech mohou být měkké nebo tvrdé. Všem měkkým souhláskám odpovídají tvrdé ekvivalenty. Výjimkou je jenom [j], který je vždy měkký. Páry souhlásek podle tvrdosti jsou reprezentované v tabulce 6.

|              |   |   |   |   |   |   |   |   |    |
|--------------|---|---|---|---|---|---|---|---|----|
| <b>Měkké</b> | D | T | Z | S | G | L | N | R | dZ |
| <b>Tvrdé</b> | d | t | z | s | c | l | n | r | dz |

Tab. 7 Měkké a tvrdé souhlásky ukrajinštiny.

Konsonanty [d, t, z, s, c, l, n, r] stávají měkkými, pokud na následující pozici ve slově jsou písmena я, ю, е, і, ъ. Například тiтка, [Títka].

Ostatní souhlásky mohou být změkčené, pokud jsou na pozici před я, ю, е, і, а nikdy nestojí před měkkým znakem. Skupiny takových souhlásek rozlišujeme podle způsobů tvoření:

- Labiály [b], [p], [v], [m], [f];
- Hrdelní [h];
- Frikativy [q], [w], [dq'], [C];
- Veláry [x], [g], [k];

Asimilace znělosti je pojem, který označuje změny znělosti fónu s cílem zjednodušit výslovnost. Znělé konsonanty vždy zůstávají znělými kromě slov легко, воґко, кігті, нігті, дьогтю, kde znělý r [h] se mění na neznělý [x]. Konsonant [z] uvnitř předpony з nebo ve tvaru předložky před neznělými fóny transformují se na [s]. Například зцілити [sGilyty], з кавою [s kAvoyu].

Neznělé konsonanty přebírají znělostní jakosti sousedních znělých souhlásek. Příklad: této změny slovo боротьба – [boroDbA], kde měkký konsonant т [T] mění se na odpovídající tvrdý д [D]. Předchozí souhláska změkčuje se pod vlivem následující. Toto pravidlo platí jen pro konsonanty [d], [t], [z], [s], [c], [l], [n], [dz], [r]. Například кузня [kUzNa].

Kromě toho existuje artikulační spodoba znělosti pro zvuky blízké podle způsobu tvoření. Je to pro zjednodušení výslovnosti během mluvení. Artikulační spodoba vždy platí pro následující skupiny písmen:

|        |             |
|--------|-------------|
| шся    | робишся     |
| ться   | молиться    |
| жся    | намажся     |
| чся    | не скалічся |
| шці    | комашці     |
| жці    | мережці     |
| зж     | розжився    |
| зч     | зчистиш     |
| здж    | з джурою    |
| зш, сш | заріши      |
| тч, тш | братчик     |

Tab. 8: Přehled souhláskových skupin blízkých podle způsobu tvoření.

## 4. Fonetická transkripce ukrajinštiny

Velmi často se v různých světových jazycích mluvená řeč se liší od psaného textu, a proto je nutné, aby existoval způsob její symbolické reprezentace. Fonetická transkripce je zápis fónů pomocí fonetické abecedy. Pro každý jazyk existuje minimální počet fonémů a jejich kombinací využívaných pro fonetickou transkripci, a v našem případě jsme použili fonetickou abecedu z kap. 2. Fonetická transkripce se využívá pro automatické zpracování řeči (například pro syntézu a rozpoznávání řeči). Během těchto procesů je nezbytné, aby fonetická transkripce přesně odpovídala zvolenému textu, což má vliv na úspěšnost těchto procesů. Přepis grafické reprezentace na fonetickou je vždy podporován pravidly fonetiky určitého jazyka. Je zřejmé, že vzhledem k tomu, že v ní využitá pravidla musejí platit pro větší počet případů, není automatická fonetická transkripce schopna vyřešit všechny možné řečové události, a proto se vybírají jen jejich obecné varianty.

Změny fonémů jsou popsány jeho levým a pravým kontextem, přičemž obecné produkční pravidlo je definováno ve tvaru [1] :

„JESTLIŽE řetězci znaků A bezprostředně předchází řetězec znaků C a je bezprostředně následován řetězcem PAK znaků D, se A přepíše na řetězec znaků B. Zjednodušená varianta zápisu:

$$A \rightarrow B/C\_D$$

A, C, D představují posloupnost písmen nebo posloupnosti symbolů použité fonetické abecedy vzniklé provedením fonetické transkripce. Řetězec B vždy reprezentuje posloupnost symbolů použité fonetické abecedy.“

Pro fonetickou transkripce potřebujeme vědět slovní přízvuk a předpokládáme na vstupu, že máme přízvuk označen. Problematiku přízvuku jsme neřešili a ručně označovali jsme přízvuk v každém slově.

Fonetická pravidla použité v procesu návrhu korpusu:

### **Symboly používané v pravidlech fonetické transkripce:**

Pro účely fonetické transkripce byl připraven základní přepis skupin jednoduchých písmen na fóny. K fonetické transkripci budeme pro zápis výslovnosti používat zjednodušenou fonetickou abecedu (viz kap. 2) Tabulka 1. ukazuje rozsah fonetického popisu a základní přepis písmen na fóny. Níže jsou uvedené skupiny písmen, které se mění vždy bez použití lingvistických pravidel:

Pravidlo, ve kterém přízvučné samohlásky se mění na odpovídající fóny:

**VLSTR - VPSTR**

Pravidlo, ve kterém nepřízvučné samohlásky se mění na odpovídající fóny:

**VLUST - VPUST**

Pravidlo přepisu souhláskového grafému na foném.

**CLT - CPT**

Pravidlo, ve kterém se přepisují ostatní souhláskové grafémy na měkké souhláskové fonémy.

**CLM - CPM**

| Zkratka skupiny | Popis   | Příklad:   |
|-----------------|---|--|
|                 | Hranice mezi slovy  |  |
| VPSTR           | Přízvukné samohlásky ve tvaru fónů                            | A, E, Y, O, U, I   |
| VLSTR           | Přízvukné samohlásky  | A, E, И, O, Y, I   |
| VLUST           | Nepřízvukné samohlásky  | a, e, и, o, y, i   |
| VPUST           | Nepřízvukné samohlásky ve tvaru fónů                          | a, e, y, o, u, i   |
| CPV             | Znělé párové souhlásky  | b, B, d, D, z, Z, q, g, h, H   |
| CPU             | Neznělé párové souhlásky                                      | p, P, t, T, s, S, w, k, x, X   |
| CLT             | Tvrdé souhlásky   | б, в, г, д, ж, з, к, л, м, н, п, р, с, т, ф, х, ц, ш, щ, р   |
| CPT             | Tvrdé souhlásky   | b, v, h, d, q, z, k, l, m, n, p, r, s, t, f, x, c, w, g  |
| CLM             | Měkké souhlásky   | б, в, г, д, з, к, л, м, н, п, р, с, т, ф, х, й, ч, щ, ц  |
| CPM             | Měkké souhlásky ve tvaru fónů                                 | B, V, H, D, Z, K, L, M, N, P, R, S, T, F, X, j, C, W, G  |
| SON             | Sonory  | m, M, n, N, j, r, R, l, L, v, V  |
| DPJ             | Dvojhásky způsobující změkčování                              | jI, jU, jA, ji, ju, ja, jE, je   |
| VPM             | Samohlásky způsobující změkčování - fóny                      | I, i   |
| CLAP            | Dvojhásky s apostrofem  | 'Ī, 'Ю, 'Я, 'ї, 'ю, 'я, 'Є, 'e   |
| DL              | Dvojhásky způsobující změkčování (grafém)                     | Ī, Ю, Я, ĩ, ю, я, Є, e   |
| DP              | Dvojhásky způsobující změkčování (foném)                      | I, U, A, i, u, a, E, e   |
| LABPS/LABPH     | Měkké labiály/Tvrdé labiály                                   | P, B, F, M/p, b, f, m  |
| VELPS/VELPH     | Měkké velary/Tvrdé velary                                     | K, g, X/k, x   |
| DENPS/DENPH     | Měkké dentály/Tvrdé dentály                                   | T, D, S, Z, N/t, d, s, z, n  |
| CORPH           | Koronární souhlásky   | t, d, n, s, z, r, l, c, h, C, H, w, q  |
| CORPH0L         | Koronární souhlásky bez likvidu                               | t, d, n, s, z, c, h, C, H, w, q  |
| LIQPH/LIQPS     | Tvrdé likvidy/Měkké likvidy                                   | r, l/R, L  |
| DBLP            | Zdvojené souhlásky  | kk, KK, gg, GG, mm, MM, nn, NN, pp, PP, rr, RR, ss, SS, tt, TT, ff, FF, cc, ll, LLdd, DD, ww, zz   |
| CPT2/CPM2       | Tvrdé/Měkké párové souhlásky                                  | b, v, h, d, z, k, l, m, n, p, r, s, t, f, x, w/B, V, H, D, Z, K, L, M, N, P, R, S, T, F, X, W  |
| CLMZ/ CPMZ      | Souhlásky s měkkým znakem/ Změkčené souhlásky po měkkém znaku | бь, вь, гь, дь, зь, кь, ль, мь, нь, пь, рь, сь, ть, фь, хь, йь, чь, шь, жь, щь, ць/B, V, H, D, Z, K, L, M, N, P, R, S, T, F, X, j, C, w, q, W, G |
| DBLL/ SGLL      | Zdvojená písmena  | гг, кк, лл, мм, нн, пп, рр, сс, тт, фф, хх, цц, чч, дд, шш, зз/ г, к, л, м, н, п, р, с, т, ф, х, ц, ч, д, щ, з                                   |

Tab. 9: Skupiny fónů využité ve fonetické transkripci.

## **Spojení souhlásky a samohlásky**

Spojení tohoto typu se vyskytuje jak uvnitř jedné slabiky, tak přes hranice slabik. Při výslovnosti souhlásky a samohlásky v rámci jedné slabiky většinou nedochází k žádným změnám. Výjimkou je změkčování souhlásek ve spojení se samohláskou [i].

Příklad:

xІмія [XIMija], япОнія [jarONija], тІнь [TiN]

K fonetické transkripci budeme pro zápis výslovnosti používat zjednodušenou fonetickou abecedu (viz tabulku) Tato tabulka ukazuje rozsah fonetického popisu.

### **Palatalizace souhlásek**

Palatalizace nebo změkčování souhlásek v ukrajinštině se provádí pomocí dvojhlásek, я, ю, є, ї, [ja], [ju], [je], [ji], ve spojení se samohláskou „i“ [i] a měkkého znaku „Ь“.

Pro základní pravidlo změkčování souhlásek bylo použito dva pravidla:

#### **CPT2 - CPM2 / \_ VPM, DL**

První pravidlo způsobuje přepis tvrdých souhlásek na měkké, pokud písmeno je na pozici před dvojhláskou, která způsobuje změkčování nebo před samohláskou „i“.

Příklad:: біблії [BIBLiji], віза [VIza], заporіжжя [zapORIqa], зІлля [ZILa]

#### **CLMZ - CPMZ / \_**

#### **DL - DP / CP \_**

Druhé pravidlo je základní a platí pro jakékoliv spojení souhlásky a měkkého znaku „Ь“. Pro toto spojení platí, že tvrdá souhláska se přepíše na měkkou ve fonetické transkripci.

Příklad:

рАдсья [rADSa], більшуй [BILwuj], лежҮть [leqYT]

### **Výslovnost dvojhlásek s apostrofem**

Podle pravidel apostrof se používá pro oddělení souhlásky a samohlásky, která způsobuje změkčování. Souhláska s apostrofem na jakékoliv pozici vždy se přepíše na tvrdou a apostrof zruší její případné změkčování.

#### **CLAP - DPJ / \_**

Příklad:

пІр'я [PIrja], м'Яч [mjAC]

Vždy ve fonetické transkripci dvojhláska se přepíše z [j] na první pozici.

### **Asimilace znělosti**

Podle pravidel ukrajinštiny během mluvení může docházet, že znělou souhlásku přečteme jako neznělou a naopak. Toto je nutně pro zjednodušení výlovy. Ačkoliv první situace je vzácný případ a pro ni existuje jen několik pravidel:

Znělá souhláska [h] vždy se přepíše na neznělou [x] pokud [h] je na pozici před neznělou souhláskou, souhláskou [k] a sonory

## **h - x / \_ CPU, k, K, |CPU, |k, |K, |SON, |VPSTR, |VPUST**

Пříklad:

лЕгко [lEhko], вОгко[vOhko], кІгті [Klxti]

## **z - s / ro, be, Cere \_ CPU**

Пříklad:

рОзповідь [rospoViD], рОзгун [rOstyn], безпАрнуї [bespArnyj].

Neznělá souhláska se přepíše na znělou častěji v případech, pokud následující souhláska je znělá. Hranice mezi slovy nemá vliv na pravidlo pro všechny uvedené písmena kromě [D] Pravidlo platí pro souhlásky níže:

## **D - T / \_ CPV**

**S - Z / \_ CPV, |CPV**

**C - q / \_ CPV, |CPV**

**w - q / \_ CPV, |CPV**

**k - g / \_ CPV, |CPV**

**x - h / \_ CPV, |CPV**

Пříklad:

незабУтне [nezabUTNe], Ось|дЕ [OZ|dE], нАш|брАт [nAq|brAt], вІснук|грозҮ [VIsnyg|hrozY].

## **Souhláskové skupiny**

Souhláskových skupiny mohou vyskytovat uvnitř jedné slabiky i přes hranice slabik. Přes hranice slabik mohou se vyskytovat uvnitř, na hranice slov pokud mezi nimi není pauza, na hranice předložky a slova nebo předpony a kmene. V této situaci dochází ke spodobě znělosti sousedních souhlásek co umožňuje jednodušší výslovnost celé skupiny. Pro ukrajinský jazyk existuje několik základních souhláskových skupin, pro které existuje pravidlo spodoby jejich znělosti. Pravidlo platí pro souhláskové skupiny na libovolné pozice kromě případů, pokud souhlásková skupina je rozdělena pauzou. Níže je vidět ukázku všech použitých pravidel pro základní souhláskové skupiny. Pravidlo ukazuje, jakým způsobem se přepíše jednotlivá skupina ukrajinských souhlásek a to včetně zjednodušení. Přepisuje se na fóny s použitím základní tabulky fónů (viz tabulku „Fonetická abeceda ukrajinštiny“).

|                      |                      |
|----------------------|----------------------|
| <b>шся - Sa / _</b>  | <b>щї - Gi / _</b>   |
| <b>ться - Ga / _</b> | <b>тці - Gi / _</b>  |
| <b>жся - zSa / _</b> | <b>зж - q / _</b>    |
| <b>чся - Gsa / _</b> | <b>зч - wC / _</b>   |
| <b>шці - Sgi / _</b> | <b>здж - qdq / _</b> |
| <b>жці - Zgi / _</b> | <b>ждж - qdq / _</b> |
| <b>чці - Gi / _</b>  | <b>зш - w / _</b>    |
| <b>сш - w / _</b>    | <b>тш - C / _</b>    |
|                      | <b>тч - C / _</b>    |

Пříklad:

розвАжся [rozvAZSa], пробувАється [probyvAjeGa], не|скалІчя [ne|skaLIGSa], квОчці [kvOGi], зжАту [qAty], з|чҮм [w|CYm], брАтчук [brACyк].

Kromě toho platí pravidlo přepisu zdvojených písmen tak že všechna zdvojená písmena se přepisují jako jedno písmeno ve fonologické transkripci.

#### **DBLL - SGLL / \_**

Пříklad:

спОсіб|вумовляння [spOSib|vumovLANa], дальнoБачення [daLnobACeNa], прудорІжжя [prydoRIQa], волОсся [volOSa], мІццю [Mİcu].

Podle pravidel ukrajinštiny existuje spodoba znělosti bukvý [z] v určité pozici před [q], [C] a před “дж” [dq]. Níže je ukázka všech použitých pravidel.

**з – q / \_|q**

**з – w / \_|C**

**з - q / \_|dq**

Пříklad:

з|жАру [q|qAru], з|джерелА [q|dqerelA]

Jev artikulační asimilaci souhlásek probíhá neuvědoměle a hlavní cíl je zjednodušení výslovnosti. Velice často se uplatňuje v rychlé mluvě, charakteristické pro běžný výslovnostní styl. Spodoby zúčastňují párové souhlásky i jedinečné souhlásky. Náhradou párové hlásky jejím znělostním protějškým nedochází k závažné změně artikulace, protože hlásky v rámci párových dvojic vznikají v podstatě na stejných místech a z hlediska artikulace se liší pouze znělostí. Do artikulační asimilace může docházet, ale to není nutné.

Souhlásky, které patří labiální skupině, se mění z tvrdé varianty na měkkou, před měkkými labiálami.

#### **LABPH – LABPS / \_ LABPS**

Пříklad:

вмісткҮй [VMistkYj]

Tvrde labiály se nemění před měkké dentály a měkkou “л” [L].

#### **LABPH - LABPH / \_ DENPS, L**

Пříklad:

плІвка [pLİvka],

Tvrde veláry se nemění před měkkýma B, V, G, D, Z, K, M, N, P, S, T, F, X, j, C, W.

#### **VELPH – VELPH / \_ B, V, G, D, Z, K, M, N, P, S, T, F, X, j, C, W**

Пříklad:

гнІй [hNij], глІОк [hLUk]



## 5. Příprava řečového korpusu ukrajinštiny

Textový korpus představuje v lingvistice souhrn textů sebraných a zpracovaných za použití určitých pravidel a používáný jako základní báze pro studium jazyka. Korpusově orientovaná syntéza řeči je v současné době nejpoužívanější. Pro využití této metody musí být texty foneticky popsány a roztrženy na segmenty. Tato fáze musí být provedena důkladně a pečlivě, protože má značný vliv na kvalitu generované umělé řeči. Formování korpusu musí kromě zachycení všech řečových segmentů splňovat následující požadavky:

- výsledný korpus musí být foneticky rozsáhlý, tzn., že ve fonetické transkripci textů jsou všechny řečové fonémy a jejich varianty. Je to nutné, aby výsledná umělá řeč obsahovala všechny jednotky řeči a jejich varianty, což zajišťuje její kvalitu;
- objem korpusu musí být minimalizovaný, pokud existuje taková možnost, a to pro usnadnění jeho zpracování pro účely řečové syntézy;
- korpus musí být foneticky vyvážen tzn., že řečové jednotky tohoto korpusu se vyskytují co nejčastěji a vyváženě, pokud existuje taková možnost.

### 5.1. Stažení a úprava textů

Hlavním úkolem předzpracování textů je převedení textu do jednotného formátu. Výstupem úpravy je textový soubor, který obsahuje jednu větu na každé řádce. Pro potřeby této práce bylo vybráno 10 ukrajinských zpravodajských portálů. Prostřednictvím jejich RSS kanálů poté speciální skript po dobu dvou týdnů automaticky stahoval nové články do textového souboru. Po stažení jsme získali rozsáhlý textový soubor, který obsahoval mnoho zbytečných znaků (například: #, &, \$, \*, @ atd.). Tyto znaky se vyskytovaly v emailových adresách, označeních měn, klíčových slovech atd. Pro nahrávání jsme potřebovali získat „čistý“ ukrajinský text, a proto bylo nutné provést úpravy shromážděného textového materiálu. Redigování jsme prováděli pomocí programovacího skriptu v jazyce Python. Jeho hlavním konceptem je porovnávání textu po řádce s regulárním výrazem a zápis každé věty na samostatné řádce. Během této fáze bylo hlavním cílem odfiltrout texty a získat „čistý text“, tzn. sadu ukrajinských oznamovacích a tázacích vět, které obsahují jen písmena ukrajinské abecedy, čárku, tečku a otazník. Zpracování textu spočívalo v následujících krocích:

1. Výskyt ruštiny a jiných jazyků. Během tohoto kroku byl největším problémem výskyt ruského textu způsobený multilingvismem některých textových zdrojů. Kvůli tomu, že se v některých ukrajinských větách současně vyskytovala ukrajinská a ruská slova, rozhodli jsme se vymazat celou větu, ve které bylo nalezeno ruské písmeno. Ukrajinština a ruština mají téměř stejnou abecedu, která se v obou případech označuje jako cyrilice, a proto stanovit, že slovo není z hlediska automatického zpracování ukrajinské, je poměrně náročný úkol. Zároveň bylo nutné ručně kontrolovat, zda v textovém souboru po zpracování nezůstala žádná ruská věta. Daný problém se podařilo úspěšně vyřešit pomocí regulárního výrazu v programovacím jazyce Python. V tomto výrazu jsme jako kritérium pro výmaz věty použili výskyt alespoň jednoho z ruských grafémů ы, э, ь. Další věcí, kterou jsme se zabývali, bylo vyhledání a výmaz všech jmen, názvů, zkratk atd. psaných jinak než pomocí ukrajinské abecedy (například Facebook, Skype, Google, Trump, EU, NATO apod.). Jako hlavní kritérium pro regulární výraz jsme použili celou anglickou abecedu, kterou je možné zapsat pomocí prvního a posledního písmena abecedy.

2. Zvláštní symboly a čísla. Pro zjednodušení přečtení věty je nutné, aby všechny speciální znaky (např. #, &, \$, \*, @ atd.), číselné výrazy a datumy byly ve výstupních větách vymazány. Zvláštní znaky jsme vymazali pomocí integrované knihovny String, ve které již existuje možnost odkázat na interpunkci (string.punctuation), což vrací seznam všech znaků a regulárního výrazu. Některé znaky jsme museli ponechat, a proto kontejnerový typ list, kterým byl reprezentován seznam zvláštních značek, byl vhodný pro snadné přidání a výmaz symbolů. Kromě toho byly odstraněny věty s citáty, zkratkami a přímou řečí,

kteře byly také detekovány podle specifických symbolů v textu. Čísla byla vymazána, aby se zjednodušilo čtení výsledných vět. K výmazu čísel byl rovněž použit regulární výraz.

3. Odstranění zkratk a názvů. V dalším kroku jsme měli za cíl odstranit všechny zkratky a případné názvy článků v textu. Zkratka je sada velkých písmen, což jsme snadno rozpoznali v textu pomocí regulárního výrazu. Názvy kapitol jsme rozpoznali tak, že na jejich konci není žádná interpunkce. Názvy nemají kontext, a proto mohou vést k výrazným artefaktům ve výsledném korpusu. Zkratky byly odstraněny, neboť mohou být různými řečníky přečteny různým způsobem.

4. Odstranění duplicitních vět a prázdných řádků. V tomto kroku jsme odstranili stejné věty, což bylo usnadněno předem provedenou „normalizací“ textu. Během procesu „normalizace“ jsme ověřili, zda se v textu nevyskytují nadbytečné nebo chybějící bílé znaky. Všechny prázdné řádky jsme vymazali.

Po úpravě textů jsme potřebovali vypsát každou větu na novém řádku. Kvůli tomu bylo nutné rozpoznat hranice vět. Pro tento účel existuje v jazyce Python balík `nlk.tokenize`, pomocí něhož je lze rozpoznat hranice vět v textu. Konec věty jen představen speciálními symboly, například `.`, `!`, `?` apod. Rozdělení textu na věty na základě těchto speciálních znaků se nazývá tokenizace. Kromě toho jsme pomocí tohoto balíku omezili délku výstupních vět. Filtrace všech vět podle jejich délky byla podle počtu slov, a to od pěti do deseti. Bylo to provedeno proto, aby byla zachována kvalita řeči během nahrávání a aby se hlas na začátku a na konci věty příliš nelišil.

Poslední fázi představovalo vypsání jednotlivých vět na samostatných řádcích. Po zpracování bylo získáno kolem 600 000 vět zapsaných v textovém souboru.

## 5.2. Výběr vět

Fáze výběru vět se provádí pro reprezentativní sadu vět, kterou později nahrajeme a na jejímž základě vytvoříme řečový korpus pro syntézu řeči. Důležitým požadavkem je minimalizace času, během něhož se budou věty nahrávat. Tento požadavek je nutný z důvodu, aby si řečník zachoval vlastnosti hlasu a styl mluvení. Větová sada musí dle možností obsahovat všechny fonetické a prozodické kontexty, které se mohou vyskytovat na vstupu řečového syntetizéru. Toto kritérium nelze pro obecný systém TTS v praxi uplatnit, protože by vedlo k příliš velkému počtu vět. Z tohoto důvodu musí být počet vět optimalizován. Při výběru vět často platí v praxi podmínka, aby obsahovaly zvolené řečové jednotky v různých kontextech.

Pro výběr vět byl využit tzv. nenasycený algoritmus (angl. greedy algorithm) nebo také add-on procedura. Pro zvýšení počtu výskytů uvažovaných fonémů alespoň  $t$ -krát (parametr „-mn“, viz tabulka 10.) jsme věty vybírali pomocí modifikovaného algoritmu, který je složen z následujících bodů [6]:

1. Ze vstupního souboru vět se vybere věta s největším počtem různých fonémů, které v dosud vybraných větách nejsou zastoupeny v požadovaném počtu  $t$ , a přesune se ze vstupního souboru do souboru dosud vybraných vět.
2. Jestliže lze ze souboru dosud vybraných vět nějakou větu odstranit tak, aby počet výskytů žádného fonému neklesl pod stanovený práh  $t$ , je tato věta z dosud vybraných vět vyloučena a přesunuta zpět do souboru vstupních vět.
3. Kroky 1 a 2 se opakují tak dlouho, dokud všechny fonémy nejsou ve vybraných větách zastoupeny v požadovaném počtu  $t$ .
4. Pro každou foneticky přeepsanou větu ze vstupního souboru se spočítá skóre  $S$ , v našem případě podle následujícího vztahu:

$$S = - \sum_{i=1}^I \frac{n_i+n'_i}{n} \log_2 \frac{n_i+n'_i}{n}, \text{ kde } n = \sum_{i=1}^I n_i + n'_i$$

a  $I$  je počet různých fonetických jednotek, které mají být obsaženy ve vybraných větách,  $n_i$  vyjadřuje

počet výskytů  $i$ -té fonetické jednotky v dosud vybraných větách,  $n_i'$  značí počet výskytů  $i$ -té fonetické jednotky v právě zkoumané větě a dále definujeme  $0 \log_2 \equiv 0$ , které vyjadřuje, jak právě zkoumaná věta přispívá k výslednému zastoupení difonů v souboru dosud vybraných vět.

5. Vybere se věta s maximální hodnotou skóre  $S$  a přesune se ze vstupního souboru do souboru dosud vybraných vět.

6. Provede se přepočítání skóre všech dosud vybraných vět a věta s nejhorsším skóre se přesune zpět do souboru vstupních vět. K přesunu však nedojde, pokud by vyřazovanou větou byla poslední přidaná věta (věta by se znovu vybrala v další iteraci) nebo pokud by vyřazením došlo k poklesu výskytu některého fonému pod daný práh  $t$ .

7. Kroky 4 až 6 se opakují, dokud není vybrán požadovaný počet vět nebo dokud není dosaženo požadované hodnoty skóre.“

Algoritmus byl realizován pomocí programu, který se nazývá „Výběr vět“. Tento program bylo nutné spouštět prostřednictvím příkazové řádky a operačního systému Linux. Pro zajištění potřebného mechanismu výběru vět je nutné spouštět tento program pomocí speciálně zadaných parametrů představených v tabulce 10.

K větám, jež byly automaticky vybrány pomocí výše popsaného algoritmu, jsme přidali ručně vybranou sadu tázacích vět. K ručnímu přidání tázacích vět bylo přistoupeno z toho důvodu, že jsme potřebovali jen takový typ tázacích vět, kde je logický přízvuk (zdůraznění slova ve větě) při čtení na začátku věty, neboť takový přízvuk je v ukrajinštině nejběžnější. Celkový počet vybraných vět na konci fáze výběru pro vytváření řečového korpusu ukrajinštiny je 2332 vět.

| Název parametru  | Typ    | Popis  |
|--|--------|--|
| -c   | file   | Konfigurační soubor. Nastavení lze přepsat pomocí příkazu z příkazového řádku.   |
| -iw  | file   | Vstupní soubor s větami v písemné podobě.  |
| -ip  | file   | Vstupní soubor s větami ve fonetickém tvaru.   |
| -o   | file   | Výstupní soubor s vybranými větami.  |
| -sn  | number | Počet vět vybraných pro každého řečníka je v rozsahu od 1 do 32000   |
| -sp  | number | Počet řečníků, pro které jsou vybrány věty, rozsah je od 1 do 1000   |
| <b>Volitelné parametry</b>                             |        |  |
| -sr  | number | Maximální počet řečníků, kteří mohou mít stejnou větu. Výchozí hodnota je 1.   |
| -id  | file   | Vstupní soubor s větami, které nesmí být vybrány (v písemné podobě).   |
| -if  | file   | Vstupní soubor s větami, které musí být vybrány (v písemné formě).   |
| -lf  | file   | Soubor se záznamem skóre, pokud je vybrán, skóre bude zaznamenáno, jinak ne.   |
| <b>Přepínače pro požadovaný počet výskytů jednotek</b> |        |  |
| -m   |        | Věty budou vybrány tak, aby obsahovaly požadovaný počet jednotek. Následující přepínače musí být nastaveny:  |
| -mu  | type   | Jednotky pro výběr, typy jsou:<br>P pro fonémy<br>T pro trifony<br>D pro difony  |
| -mo  | type   | Věty pro každého řečníka (typ S) nebo pro všechny vybrané věty (typ C) budou obsahovat požadovaný počet řečových jednotek.   |
| -mn  | number | Minimální požadovaný počet výskytů jednotek.   |
| -me  |        | Pokud je tento přepínač nastaven, program skončí s chybou, není-li možné vybrat věty s požadovaným počtem jednotek, pokud není nastaven, upozornění na soubory protokolu budou přidány pouze.  |
| <b>Přepínače pro vyvažování:</b>                       |        |  |
| -b   |        | Věty budou vyváženy, musí být nastaveny spínače.   |
| -bu  | type   | Jednotky pro výběr, typy jsou:<br>P pro fonémy<br>T pro trifony<br>D pro difony  |
| -bo  | type   | Věty pro každý reproduktor (typ S) nebo pro všechny vybrané věty (typ C) budou vyrovnány.  |
| -bk  | type   | Druh vyvažování, uniforma pro typ U a přírodní pro typ N.  |
| -bf  | file   | Soubor s relativními četnostmi výskytů jednotek pro přirozené vyvažování (pokud je nastaven parametr -bk N), pokud tento soubor není zadán, budou relativní výskyty vypočteny z vět pro čtení. |
| <b>Další parametry</b>                                 |        |  |
| -uu  | list   | Seznam jednotek bez kontextu.  |
| -ui  | list   | Seznam jednotek, které budou ignorovány, ve stejném formátu jako -uu.  |
| -H   |        | Nápověda.  |

Tab. 10: Parametry programu «Výběr vět».

### 5.3. Nahrávání vět

Další věcí, kterou jsme se zabývali při přípravě řečového korpusu, je nahrávání vybraných vět. Vzhledem k tomu, že jsme zvolili konkatenční syntézu řeči, je tato fáze vysoce důležitá pro kvalitu umělé řeči. Syntetizovaná řeč se vytváří řetěžením segmentů nahraných vět, a tak je snadou mít nahrávky bez vedlejších šumů.

Pro nahrávání jsme využili profesionální software SoundForge, který zaručil kvalitní nahrávání a ukládání signálů. Nedostatkem tohoto programu je, že nezobrazuje věty, které se následně čtou. Toto

bylo vyřešeno pomocí grafické nadstavby, „dialogového manažera“, naprogramovaného pomocí rozhraní C#. Jeho přidáním bylo umožněno nejen přečíst větu, ale i manipulovat procesem nahrávání. Řečník má možnost vidět číslo nahrávky, vynechat nevhodnou větu a přejít na další nebo se vrátit k předchozí, několikrát zopakovat nahrávání stejné atd. Kromě toho dialogový manažer během nahrávání kontroloval hladinu intenzity signálů; pauzu na začátku a konci promluvy; přítomnost nízkofrekvenčních šumů vznikajících kvůli nesprávnému nastavení nahrávacího systému atd. V případě jakéhokoli z těchto případů se objeví nové okno s upozorněním a není možné pokračovat další větou.

Věty jsme se snažili nahrát během několika sezení, aby artikulační osobitosti řečníka zůstávaly stejné. Po nahrávání jsme spočetli celkový čas nahrávek řečového korpusu ukrajinštiny bez zahrnutí pauz na začátku a na konci nahrávky. Celkový čas činil přibližně 2 hodiny 45 minut a konečný počet nahrávek byl 2332 ukrajinských promluv.

#### **5.4. Ortografické a fonetické anotace**

Z důvodu lidského faktoru občas dochází k situaci, že i nejzkušenější řečník přečte větu jinak, než je zapsáno, a proto se provádí další fáze anotování promluv. Důvodem pro to, aby této fázi byla věnována velká pozornost, je možnost, že se nahraný signál bude lišit od skutečné lingvistické informace v textovém korpusu. Jakákoli nepřesnost může být příčinou segmentačních chyb umělé řeči a následně špatné kvality syntetické řeči.

Ortografická anotace se provádí, aby bylo možné na slovní úrovni přesně popsat obsah každé nahrané věty. Základem pro daný proces jsou vybrané věty namluvené řečníkem. Nejčastější případy, ke kterým docházelo během nahrávání ukrajinských vět:

- řečník se hlasitě nadechoval;
- řečník se přeekl a místo vyznačeného slova použil jiné;
- řečník slovo vůbec nepřečetl nebo zaměnil pořadí několika slov ve větě.

Anotace pro ukrajinštinu byla provedena pomocí programu Transcriber ve dvou fázích. V první fázi jsme vyslechli každou nahrávku a případně opravili věty podle toho, co bylo reálně řečeno.

K přepisu byla použita sada pravidel, a to pro přepis číslovek, zkratek, akronymů, interpunkce, nesprávně řečených nebo netypických slov, neřečových událostí (nádechy, mlasknutí) atd.[6]

Následně byla provedena automatická fonetická anotace ortografických anotací a akustických signálů nahrávek vět. Jako vstupní data jsme použili pravidla fonetické transkripce, kterými jsme se detailně zabývali v kapitole 4. V případech, kdy existovalo několik možností fonetického přepisu, jsme zvolili frekventovanější variantu.

Po anotaci řečového korpusu ukrajinštiny jsme spočetli řečové jednotky:

- celkový počet fonémů: 101653;
- celkový počet různých difonů: 1248;
- celkový počet difonů: 98103.

#### **5.5. Segmentace řečového korpusu.**

Fáze segmentace řeči je proces hledání hranic akustických realizací řečových jednotek v nahrávkách vybraných vět. Segmentace řečového korpusu je segmentace všech nahraných řečových promluv, které jsou uloženy v korpusu.

Dříve se v konkatenačních systémech řečové syntézy využívala ruční segmentace. Byla obvykle používána pro korpusy řečových jednotek, které měly pouze jednu realizaci nějaké lingvistické jednotky. Tento proces vyžaduje velmi mnoho času a lidské práce, což snížilo jeho popularitu. Ruční segmentace se nepoužívá také z toho důvodu, že tento proces je řečově náročný a nejednoznačný.

V současných řečových systémech jsou populární algoritmy automatické segmentace řeči. Vzhledem k tomu, že proces segmentace často probíhá na úrovni fonémů, bývá taková segmentace někdy

označována také termínem automatická fonetická segmentace (angl. automatic phonetic segmentation, zkr. APS).

V dnešní době se používají dva základní přístupy pro automatickou segmentaci řečových korpusů, a to: skryté Markovovy modely (HMM) (viz kap. 1) a porovnávání se syntetickou řečí s využitím techniky dynamického borerení času. Nemůžeme prohlásit, že některá z těchto metod je lepší, nebo horší než druhá, protože obě vedou k přibližně stejným výsledkům.

Metoda HMM provádí konzistentní segmentaci, ale její průměrná chyba má často větší hodnotu [11].

Při použití DTW občas dochází k výskytům velkých segmentačních chyb. Má se za to, že velké chyby v segmentaci jsou horší, proto se v současnosti více využívají techniky založené na metodě HMM. Ty jsme pro navržený korpus použili i v naší práci.

## 6. Závěr

Syntéza řeči je komplikovaný proces, který vyžaduje velkou pozornost k detailům. Člověk začal zkoumat možnosti umělého generování řeči ještě před více než třemi stoletími a dosud tento zájem se neztrácí. Ačkoliv, kvalita znění a prozodické jakosti umělé řeči ještě nedosáhli vysoké úrovně, je vidět značný progres od začátku vývoje a do dnešní doby. Řečové generování je aktuální kvůli tomu, že během procesu svého rozvoje našlo uplatnění v různých sférách běžného života.

Existuje několik základních přístupů řečové syntézy, které se používají v současné době: konkatenční a statistická parametrická syntéza. Konkatenční přístup se stává více populární od doby technologie druhé generace. Konkatenční metoda poskytuje více přirozenou a individuálně znějící řeč, ale kvalita některých souhlásek se může lišit.

Práce měla za cíl návrh řečového korpusu ukrajinského jazyku. Ačkoliv jsme kláli velký důraz na podporu konkatenčního přístupu řečové syntézy, navržený korpus je možné používat i pro ostatní přístupy generování umělé řeči. Proces korpusového návrhu zahrnoval následující sadu kroků:

V rámci tohoto kroku jsme vybrali sadu ukrajinských novin pomocí RSS kanálů webových stránek. Během tohoto procesu upravili jsme texty tak, aby obsahovali jen požadované symboly (viz kap. 5)

Během tohoto etapu vybírali jsme věty, aby obsahovaly zvolené řečové jednotky a kontext jednotek neměl velkou roli. Snažili jsme se, pokud je to možné, zajistit pokrytí fonetických kontextů. Tento krok byl proveden s využitím algoritmu z kapitoly č. 5.

Pak jsme vybrané věty postupně nahráli. Pomocí speciálního softwaru řídili jsme proces nahrávání. Během tohoto procesu snažili jsme se, aby mluvená řeč byla nejvíc "čistou", tj., bez šumů a specifických řečových událostí.(viz kapitolu č. 5).

Ortografické anotace jsme provedli, aby popsaly na slovní úrovni, obsah každé nahrávky. Základem pro daný proces byli vybrané věty, které namlouval řečník.

Fáze segmentaci řeči je proces hledání hranic akustických realizací řečových jednotek v nahrávkách vybraných vět. Pro přípravu našeho korpusu použili jsme automatickou metodu segmentace řečových skrytých Markovových modelů (HMM).(viz kapitolu č. 5).

V rámci bakalářské práce vytvořili jsme základní řečový korpus ukrajinského jazyku. Ve výsledku získali jsme databáze řečových jednotek. Návrhem pro další rozvoj výstupu této práce je, že korpus ukrajinštiny může být využitý pro syntézu umělé ukrajinské řeči. V rámci bakalářské práce vytvořili jsme základní řečový korpus ukrajinského jazyku. Ve výsledku získali jsme databáze řečových jednotek. Návrhem pro další rozvoj výstupu této práce je, že korpus ukrajinštiny může být využitý pro syntézu umělé ukrajinské řeči.

Další práce se nabízí věnovat pozornost fonetickému odhadování ukrajinského prizvuku ve vstupních textech.

## 7. Literatura

1. PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Academia, Praha, 2006
2. CRYSTAL, D. How language works. Clays Ltd, England 2006
3. ALLEN J., HUNNICUT S., KLATT D. From Text To Speech, The MITTALK System, Cambridge University Press, 1987, 213 pp.
4. BENELLO J., MACKIE A. W., ANDERSON J. A. Syntactic category disambiguation with neural networks, Computer Speech and Language, 1989
5. Encyclopedia of Multimedia Technology and Networking
6. MATOUŠEK, J. Habilitační práce Počítačová syntéza řeči, Plzeň, 2008.
7. E-learningová podpora mezioborové integrace výuky tématu vědomí na UP Olomouc: <http://pfyziol.fup.upol.cz>
8. ФАЩЕНКО М. М., ВОЛОСЕВИЧ З. І., МИКИТИН-ДРУЖИНЕЦЬ М. Л., СЕНИК Г. В., ЯКОВЕЦЬ С. Г. Сучасна українська мова Фонетика. Фонологія. Орфоепія. Графіка, Академія, Київ, 2010
9. ТЕРЕЩЕНКО В. М. Українська мова, УЛА, Харків, 2013
10. WELLS J. C. Computer-coding the IPA: a proposed extension of SAMPA. University College London, 1995
11. KOMINEK, J., BENETT, C., Black, A. Evaluating and Correcting Phoneme Segmentation for Unit Selection Synthesis, Eurospeech, Geneva, Switzerland, 2003