UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS HUMANAS E SOCIAIS

Natural Language Processing Approach for Macedonian-French and Macedonian-English Interpreting based on Oral Sociopolitical Corpora

Aneta Rafajlovska

Dissertation for obtaining the Master Degree in

International Masters

in Natural Language Processing & Human Language Technology

Dissertation under the supervision of:

Professor Doutor Henri Madec

Professor Doutor Jorge Manuel Baptista

Faro, 2013

UNIVERSITÉ DE FRANCHE-COMTÉ

CENTRE DE RECHERCHE EN LINGUISTIQUE "LUCIEN TESNIÈRE"

UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS HUMANAS E SOCIAIS

Natural Language Processing Approach for Macedonian-French and Macedonian-English Interpreting based on Oral Sociopolitical Corpora

Aneta Rafajlovska

Dissertation for obtaining the Master Degree in

International Masters

in Natural Language Processing & Human Language Technology

Dissertation under the supervision of:

Professor Doutor Henri Madec

Professor Doutor Jorge Manuel Baptista

Faro, 2013

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to both of my supervisors for all the help and support they have provided me throughout these two years. I would like to especially thank my first supervisor for helping me determine, adapt and shape this dissertation project during my first year, so that I can work on what I wanted from the beginning.

I would like to thank the Erasmus Mundus Consortium for giving me this unique chance and for all the wonderful moments I have experienced during this two years.

I would sincerely like to thank all the teachers and all the participants in the program for the knowledge they have provided me and for the friendships I have made.

Specially, I must thank the interpreters who were consulted for the purpose of this research for their contribution and their effort.

Lastly, I would like to thank my family and everybody who missed me back home!

Faro, 2013

Aneta Rafajlovska

ABSTRACT

The interpreting is the technique of orally translating an oral message on the spot or with a particular delay. It is widely used for international conferences, meetings, reunions, etc. or in any situation where the speakers express themselves orally in different languages.

In this dissertation project our focus is concentrated on presenting the computational aid available for the interpreting. The central point of this research is a specific type of interpreting which is performed immediately after the speech has been uttered, its length being inferior then 6 minutes. This type of interpreting is called Consecutive Interpreting.

In order to interpret longer segments of speech the technique of note-taking is used to help the memory of the interpreter. The note-taking represents a special method for rapidly writing down the main points of the thought uttered by the speaker.

This being a challenging task for the beginners in this profession and also the fact that perhaps it tends to be easily forgotten if not practiced enough, led to our conclusion that computational aid is needed to facilitate the task of the note-taking. Unfortunately, there is no computer assisted tool that deals with this issue.

In this project we have extracted information about the frequency of words from oral corpora and used that information to give a conceptual design proposal for a computer assisted tool for the transcription of symbols into text, for the task of note-taking in consecutive interpreting.

KEYWORDS

Consecutive Interpreting

Note-taking

Symbol

Speech Recognition

Transcription

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

- AAAF: Association Aéronautique et Astronautique de France
- AESA: Agence Européenne de la Sécurité Aérienne
- API: Application Programming Interface
- ARPA: Advanced Research Projects Agency
- ASR: Automatic Speech Recognition
- ATM: Automatic Teller Machine
- CAT: Computer-Assisted Translation or Computer-Aided Translation
- CSST: Commission de la Santé et de la Sécurité du Travail
- DSP: Digital Signal Processing
- EU: European Union
- GMM: Gaussian Mixture Models
- HMM(s): Hidden Markov Model(s)
- LVCSR: Large-Vocabulary Continuous Speech Recognition
- MTV: Makedonska Televizija
- NATO: North Atlantic Treaty Organization
- NLP: Natural Language Processing
- NN: Neural Network
- PC: Personal Computer
- POS: Part-of-Speech
- RM: Republic of Macedonia
- SAPI: Microsoft Speech Application Programming Interface
- SR: Speech Recognition
- SSR: Supplemented Speech Recognition
- STT: Speech to Text
- TM: Translation Memory
- TTS: Text-to-Speech
- TU: Translation Units

- UAlg: Universidade do Algarve
- UFC: Université de Franche-Comté
- WFSTs: Weighted Finite-State Transducers

Chapter 1

1. Overview of the Project

1.1. Introduction to Interpreting

The interpreting is an old profession, dating back from the first relations between people speaking different languages (Herbert, 1952). It has a noble objective because it helps people to bond and understand each other. While removing the language barrier that separates people speaking different languages, the interpreter converts their thoughts and views, from one language into another, orally. The interpreter plays a role of a mediator.

The essence of the communication is the comprehension. The comprehension makes the communication possible. Thus, the role of the interpreter is very important because it enables the comprehension and therefore it enables the communication between humans.

The objective of the interpreter is to help individuals or groups of people understand each other, get to know each other's culture, respect each other and if they wish reach an agreement (Herbert, 1952).

Nowadays, interpreting is considered to be respectable profession. It is practiced during conferences, plenary sittings, summits, seminars, congresses, reunions, meetings, etc. or in any situation where people speaking different languages orally exchange their thoughts and views.

Interpreting is very different from translation. The main difference between these two, in some ways similar, but also very distinct professions, is their source. The source of the translation is written text, unlike the interpreting, whose source is spoken language.

Each and every language pair has its particularities that appear both while translating and while interpreting from one language into another. Similar languages, like languages that belong to the same language family, but not necessarily, make the task of the interpreting much easier, thanks to the similarities in vocabulary and other grammatical resemblances.

This work will be focused on the interpreting from Macedonian into French and from Macedonian into English. Research will be conducted using oral corpora from the domain of the political sociology.

Macedonian is a **Slavonic** language belonging to the Indo-European language family. More precisely, it belongs to the Eastern group of the **South Slavic** branch of languages. French belongs to the **Italic** subfamily of the Indo-European language family. It belongs to the group of **Romance** languages derived from the Latin. English also belongs to the Indo-European language family; more precisely it belongs to the **Germanic** subfamily.[1] The fact that these three languages derive from the Indo-European language, underlies why they do not show extreme divergences. But, the fact that they come from a completely different subdivision, the **Balto-Slavic**, the **Italic** and the **Germanic** subfamily, explains why they have a lot of differences.

Macedonian is a language spoken by approximately 2 and a half to 3 million people. It is the official language in Republic of Macedonia and holds the status of official minority language in parts of Albania, Serbia, Kosovo and Romania. It is also spoken by the Macedonian diaspora in Australia, Canada, the United States, Italy, Germany, Bulgaria, Greece, Turkey and other countries.

The writing system of Macedonian is the Macedonian alphabet which is an adaptation of the Cyrillic script. The Macedonian alphabet consists of 31 graphemes. For each phoneme in Macedonian there is a grapheme.

1.2. PROJECT OBJECTIVES

The initial goal of this project was to analyze the linguistic aspects of the oral speech while interpreting from Macedonian into French and from Macedonian into English. The

[1] Indo-European language family http://www.vaughns-1-pagers.com/language/european-languages.htm

main idea was to collect parallel oral corpus consisting of the original speech in Macedonian and the interpretation in French or English. Due to the fact that the ideal material that was required in order to conduct such research was not available to the general public, we had to base our research only on recordings that represent original spontaneous speech in Macedonian broadcasted publicly.

In order to do the analysis of the oral spontaneous speech in Macedonian, oral corpora will be collected. The oral corpora will consists of recordings of debates, interviews, etc. that were broadcasted on some of the Macedonian TV-channels. All of the recordings shall exemplify the spontaneous oral expression in Macedonian on the subject of political sociology.

Furthermore, review on the existent computational aid available for interpreters will be conducted in order to examine the possibility of automation of the interpreting. The main focus will be on the problem of note-taking in consecutive interpreting. The main issue in this case is transferring the oral speech of the speaker into a written text that will help the memory of the interpreter and it would facilitate his task. Thus, a brief research of the current state of the art of the automatic speech recognition software will be conducted.

Finally, an attempt of semi-automation of the note-taking in consecutive interpreting will be explained. Conceptual design of a computer assisted tool for the consecutive interpreting will be elaborated as a possible solution for semi-automation of the note-taking in consecutive interpreting.

1.3. THESIS STRUCTURE

The rest of this document is structured in the following way:

- Chapter 2, State of the Art, consists of the review of the state of the art in computational aid available for interpreters. This chapter tackles upon the available Computer-Assisted Translation Tools for Interpreters and the development of the Speech Technologies with focus on the Automatic Speech Recognition.

- Chapter 3, Differences between Translation and Interpreting, gives an overview of the resemblances and the divergences of Translation and Interpreting. Further on, it describes in detail the most widely used types of interpreting such as the Simultaneous Interpreting and the Consecutive Interpreting with special emphasis on the process of Note-taking. Finally, the other types of interpreting, which are not used so often as the previous two types, are also briefly described: Liaison Interpreting, Relay Interpreting, Remote Interpreting and Telephone Interpreting.

- Chapter 4, Oral Corpora, describes the oral corpus that was used as the basis of this research. In this chapter the variables, the topics and the word count are described, as well as the difficulties that were encountered while collecting and processing the oral corpus. Lastly, view of the results obtained can be found in this chapter.

- Chapter 5, Computational Aid for Consecutive Interpreting, describes the Multilingual Platform that is being proposed in this dissertation project as a possible conceptual solution for semi-automation of the note-taking in consecutive interpreting. Some possibilities for the input of the symbols, as well as for the output of the system can be found under this chapter.

- Chapter 6, Conclusions, presents the conclusions.

CHAPTER 2

2. STATE OF THE ART

In this chapter, a research on the existing computational aid available for interpreters will be made. It will focus both on the simultaneous and on the consecutive interpreting.

The first part of this chapter is concentrated on the simultaneous interpreting, regarding the *Computer-Assisted Translation* and the *Computer-Assisted Translation Tools for Interpreters*.

The second part of this chapter is dedicated to the consecutive interpreting. Research of the state of the art of the *Automatic Speech Recognition* will be made in the aim of investigating the possibility of complete automation of the transcription of the original oral speech into text for the needs of the consecutive interpreting.

Furthermore, a research on the *Speech Technologies for Macedonian* and the *Speech Recognition Software* will be elaborated.

2.1. COMPUTER-ASSISTED TRANSLATION

Computer-Assisted Translation or *Computer-Aided Translation* (CAT)[2] represents the act of translating a text, by a human translator, with the aid of specific computer software, dedicated for this purpose.

CAT is frequently used by translators, for translating written text. These tools are very useful because they perform sentence segmentation.

[2] http://www.cattools.org/

Moreover, the translation and the original text are saved together and they are presented as translation units (TU).

Lastly, the CAT tools save the translation unites in a database called the translation memory (TM).

Over all, the CAT tools can make the translation faster and easier thanks to the convenient segmentation of the text.

Furthermore, the quality check of the translation is made much easier because of the fact that the original text and the translation are saved together in the translation unites.

Finally, the translation memory makes it possible to reuse the same translation further in the text or in any other text. The software evaluates the percentage of how much the new sentence or paragraph matches with a sentence or paragraph which was already translated and memorized in the TM. If it is a 100% match the translation is done automatically. If not, the translator can decide accordingly to the percentage whether the translation can be done automatically or not. Thus, the CAT tools save time and effort and help the translator to use consistent terminology.

Some of the most commonly used CAT tools[3] are: SDL TRADOS, Déjà Vu X, Wordfast, SYSTRAN, Lingo, Araya, LogiTerm, Fortis Revolution Translation Suite, MetaTexis, Open Language Tools, Across, Pootle, Similis, Globalsight, AnyMEN, Tr-aid, TranslateCAD, Felix, Multiterm, TransSearch, AppleTrans, MultiCorpora MultiTrans, WordFisher, Swordfish, WinCATS, KeyTerm, SDL Termbase, System Quirk, TermStar, Poedit and many more.

[3] http://www.translationdirectory.com/articles/article1629.php

2.1.1. COMPUTER-ASSISTED TRANSLATION TOOLS FOR INTERPRETERS

Certain software can help improve the interpreter's performance. Some of the existing CAT tools can also be useful for interpreters.

The interpreting is very different from the translation of a written text. The main difference between the translation and the interpreting is that the source of the interpreting is an oral message.

Therefore, the interpreting requires different type of computational aid. When it comes to simultaneous interpreting, the use of bilingual dictionaries, glossaries of the terminology and other documents can be of great help only if it is possible to search through them extremely rapidly.

As explained in the next chapter "Simultaneous Interpreting", the interpreter in the booth needs to be fully concentrated on the speech in order to interpret with the delay of only few seconds after the speaker. So, if the interpreter needs some help with a specific term, he can consult rapidly a dictionary or a glossary and move on with the interpreting.

Some CAT tools are constructed particularly for this purpose. Taking all of the above into consideration, they provide the possibility of creating glossaries and consulting them rapidly in the interpreting booth.

- ApSIC Xbench

The ApSIC Xbench[4] is an integrated reference tool. Its main advantage is that it provides a clear and structured view of the terminology. This software supports a large variety of input formats. This is very appropriate because the interpreter can use glossaries and translation memories from other CAT tools which he had created beforehand. Moreover, the ApSIC Xbench provides an easy search within the glossaries and convenient visualization of the bilingual information.

[4] http://www.apsic.com/en/products_xbench.html

- Interplex Glossary Software for Interpreters and Translators

The Interplex Glossary Software for Interpreters and Translators[5] is a search engine for terminology glossaries developed by Peter SAND. It is especially designed to make search as easy as possible in order to be used also in an interpreting booth. Some of the advantages of the Interplex are: the search function says on; it searches all the languages in the glossary if the search is not limited to a particular language; it ignores accents, umlauts and tildes; it includes a "multi-glossary search" etc. (Sand, AIIC, 2003) All of the features mentioned above make the search of the term easy and fast – which is crucial in the interpreting booth.

- LookUp

LookUp[6] is a terminology tool developed by conference interpreters and translators for use during simultaneous interpreting and while translating. LookUp offers full support for lexicographical and terminographical projects. Its main purpose is consulting rapidly the terminology while in the interpreting booth.

- InterpretBank

The InterpretBank[7] is another CAT tool very similar to the others mentioned above. It is a modular tool for terminology and knowledge management for simultaneous interpreters. It is developed at the Translation and Interpretation Faculty, University of Mainz, Germersheim, Germany. InterpretBank consists of ConferenceMode, TermMode and MemoryMode for easy consulting and searching the glossaries, creating the glossaries and memorizing the glossaries. In the ConferenceMode the interpreter can load up to 3 language pairs and he can switch easily the language pair according to the source language of the speaker.

---

[5] http://www.fourwillows.com/interplex.html

[6] http://www.lookup-web.de/

[7] http://www.interpretbank.de/

There are many other CAT tools that are using a similar approach and that are used by interpreters. In fact, any CAT tool that provides the possibility of searching bilingual material stored in the translation memory (TM) can be useful.

2.2. SPEECH RECOGNITION

Speech Recognition (SR) is also known as *Automatic Speech Recognition* (ASR), *Computer Speech Recognition*, *Speech to Text*, or just *STT*. Speech recognition is the technique of converting human speech into written text using computer software. In fact, the computer software converts the audio signal, which is transmitted by a microphone or a telephone, into text, which appears on the screen of the computer or the mobile phone.

There are many different types of Speech recognition systems (Zue, Cole, & Ward, 1996):

- *Isolated-Word Speech Recognition System* as opposed to the *Continuous Speech Recognition System*. The latter does not require making pauses between the words, the user can speak fluently. Oppositely, the isolated-word speech recognition system requires making brief pauses between each word that is being pronounced.
- *Speaker Dependent System* as opposed to *Speaker Independent System.* The latter does not require *speaker enrollment.* Speaker enrollment is the process of recording samples of the speaker's voice. The speaker dependent systems only recognize the vocal expression of one user, accordingly to the voice samples recorded prior use.
- The Speech recognition systems also vary in the sense of the vocabulary. There are SR systems with small vocabulary, even less than 20 words, and systems with large vocabulary with more than 20.000 words. However, the SR systems with large vocabulary have higher error rate. The accuracy diminishes as the size of the vocabulary grows.
- When it comes to producing sentences, the Speech recognition systems use artificial grammars to limit the word order. There are systems that use very

constricted word order. Each word has a specific order after which word it can follow. This is the case of the finite-state network. On the other hand there are the systems which as language model use the context-sensitive grammar. In this case the system can recognize more general language closer to the natural language.

- Robust speech recognition is the ability of the system to still recognize accurately the words pronounced by the speaker even if the audio sound is not clear, in means that there is also background noise, the quality of the recording is degraded, the environment changes etc. (Stern, 1996) There are some obstacles that can significantly affect the speech recognition. For example: environmental noises, deformation of the voice caused by the environment, the emotional state of the speaker (stress, happiness…), co-articulation, dialect, etc. Even systems that are speaker independent are much less accurate in the speech recognition when some of the obstacles mentioned above take place and the conditions change (Stern, 1996).

- The quality of the speech recognition is also very dependent on some external parameters. The quality of the microphone and its placement can affect largely the accuracy of the speech recognition (Zue, Cole, & Ward, 1996). Therefore, some speech recognition programs on the market are sold along with a special headset microphone.

Most of the automatic speech recognition systems are based on the statistical approach of Bayes decision rule (Uszkoreit, 1996). The implementation of Bayes decision rule for automatic speech recognition is based on two kinds of stochastic models: the acoustic model and the language model which together are the basis for the decision process itself - the search for the most probable sentence.

Modern speech recognition systems are based on Hidden Markov Models (HMM) (Brugnara & De Mori, 1996). Neural Networks have also been used to estimate the frame based scores; these scores are then integrated into HMM-based system architectures, in what has become known as *hybrid systems* (Zue, Cole, & Ward, 1996). Dynamic time warping is an approach that was historically used for speech recognition but has now largely been displaced by the more successful HMM-based approach.

During the last decade, Weighted Finite-State Transducers (WFSTs) have become popular in speech recognition. While their main field of application remains hidden Markov model (HMM) decoding, the WFST framework is used for finding solutions for many central problems in automatic speech recognition as for Large-Vocabulary Continuous Speech Recognition (LVCSR) (Hoffmeister, 2012).

When it comes to the process of performance evaluation of the speech recognition systems there are three levels of specificity (Hirschman & Thompson, 1996).

The first level takes into consideration the *criterion* which determines what would be evaluated according to the requirements of the research – the precision, the speed or the error rate.

The second level is the *measure* which determines the specific property of system performance that will be reported in an attempt to get at the chosen criterion - ratio of hits to hits plus misses, seconds to process, percent incorrect.

Finally the third level is the *method* which shows how the appropriate value for a given measure and a given system is determined. Typically this consists of some form of concurrent or post-analytic measurements of system behavior over some benchmark task.

For speech recognition the criterion of performance evaluation is *recognition accuracy*. One measure is the *word error rate*. The method used in the speech recognition evaluation made by the U.S. Advanced Research Projects Agency (ARPA) consists of comparing system transcription of the input speech to the *truth* (i.e., transcription by a human expert), using a mutually agreed upon dynamic programing algorithm to score agreement at the word level (Hirschman & Thompson, 1996).

Assessment methods developed for speech recognition consist of a combination of system-based approaches with performance-based techniques. System-based approaches either deal with the recognition system as a whole (*black box methods*) or provide access to individual modules within the complete recognizer (*glass box methods*) (Pallett & Fourcin, 1996).

The most frequently used methods (e.g., those used within the ARPA program) are application-oriented techniques based on the use of general databases, collected under

what might be regarded as representative conditions. Much of the data used for speech recognizer performance assessment consists of *read* speech, not spontaneous, goal-directed speech (Pallett & Fourcin, 1996).

2.3. SPEECH TECHNOLOGIES FOR MACEDONIAN

Speech technologies are language-dependent. They have to be developed separately for each language, taking into account its particularities. In this section, the research made and the development of speech technologies for Macedonian will be presented. Many studies have been conducted, mostly in the field of the text-to-speech, mainly for helping people with damaged sight, as well as in the domain of speech recognition of Macedonian.

A hybrid system for speech recognition of Macedonian has been developed at the Faculty of Electrical Engineering, Ss. Cyril and Methodius University in Republic of Macedonia (Kraljevski, Mihajlov, & Gorgjevik, 2000). It is Isolated-Word Speech Recognition System and it is also speaker dependent. At its current development, the system uses only a very small vocabulary limited to the digits from the decimal numeral system. The system is built on hybrid architecture combining the hidden Markov models (HMM) with Artificial Neural Networks.

The digitalized speech signal is transformed into parameters. Therefore, it obtains a sequence of acoustical vectors, which contain information about the spectral characteristics. Acoustical vectors are input of the neural network probability classifier. The most probable phonetic categories sequence is chosen with Dynamical Programming methods. Subsequently, a word from the vocabulary that matches the most the phonetic sequence is selected by the system using specific criteria.

This hybrid system for speech recognition of Macedonian was mainly developed for use in ATMs, security systems, mathematical applications using voice-input, remote financial transactions, military communications, etc.

The accuracy of this system is about 85% correctly recognized words. The approach based on hybrid architecture has been proven to be effective. In order to improve the

performance of this system, it is required to build an oral corpus in Macedonian. This corpus would have to be composed of a larger number of sentences, pronounced by a larger number of speakers in order to achieve greater generalization.

The creation of an appropriate oral corpus is crucial for the development of the speech technologies in Macedonian, for the speech recognition as well as for the speech synthesis.

Another study, (Kraljevski, Chungurski, Mihajlov, & Arsenovski, 2008) conducted at the Faculty of Electrical Engineering, Ss. Cyril and Methodius University and the Faculty of ICT, FON First Private University in Republic of Macedonia is connected to the development of an oral corpus in Macedonian. It is a study about the segmentation of the audio recordings.

The segmentation is a process in which the sound wave is divided by putting boundaries between the different spectral and auditory features. In fact, the sound wave is divided into distinctive segments that correspond to certain phonetic categories. The labeling is a similar process to the segmentation, but it may also consist of a wider description of acoustic-auditory features of the speech sequence. Those may include prosody, accent, pitch or phonetic units as phonemes or diaphones.

In spite of the fact that there are several ways to automatically segment and label a speech sequences, still the most precise and painstaking method is the manual segmentation and labeling by human experts. Unfortunately, the manual segmentation is at the same time the most time-consuming and laborious method. Therefore, it is not practical for building large speech corpora.

The human experts determine the boundaries by examining the spectrogram, the speech waveform as well as the audio reproduction of the sequence. After the segmentation, the different parts are labeled using textual code, in this case phonetic unites.

The automatic or semi-automatic segmentation and labeling is done by several different algorithms. In the end, some corrections have to be done by human experts. Furthermore, the accuracy is lower than the manual segmentation, but acceptable for some applications.

Systems for fully automatic segmentation are based on:

i. acoustic alignment to a synthetic reference utterance;

ii. trained systems for automatic speech recognition with hidden Markov models;

iii. combination of acoustic-articulatory features in phoneme categories (elitist approach).

For creating text-to-speech software, a large quantity of segmented and labeled speech sentences is required. Thus, it is necessary to have appropriate speech corpora in Macedonian and an effective way of performing automatic segmentation.

A subsystem for real-time text-to-speech (TTS) conversion for Macedonian (Josifovski, Mihajlov, Gorgevik, & Loskovska, 1996) has been developed as a joint collaboration between the Faculty of Electrical Engineering and the Department for Rehabilitation of Children and Youngsters with Damaged Sight “Dimitar Vlahov” in Skopje, R. Macedonia.

The system provides automatic reading of printed Macedonian Cyrillic text, its archival, conversion to speech by text-to-speech system and printing on a Braille printer.

For the syllabification in Macedonian, Neural Network (NN) based estimator of the probability that there is a syllable break after a given letter or phone is chosen.

Study for building speech synthesizer for Macedonian language, based on concatenation of speech segments (Chungurski, Kraljevski, Mihajlov, & Arsenovski, 2008) has been conducted at the Faculty of Electrical Engineering, Ss. Cyril and Methodius University and the Faculty of ICT, FON First Private University in Republic of Macedonia.

Concatenative Speech Synthesizers use oral corpus and produce audio speech by taking segments from the corpus and putting them together in order to create the word or the phrase that was typed as text. These systems are simple and they do not require deep knowledge of phonetic transitions and co-articulation effects. On the other hand, there are also speech synthesizers based on rules defined by linguists, which take the phonetic transitions and co-articulation effects into consideration.

Some attempts of developing quality concatenative speech synthesizer in Macedonian were not so productive because they were based on speech corpora from other Slavic languages. This results in unnatural intonation of the synthetic speech in Macedonian.

The general functional organization of Speech Synthesis System for Macedonian consists of two modules: Natural Language Processing (NLP) module and Digital Signal Processing (DSP) module.

The Natural Language Processing (NLP) module consists of three main parts: text analyzer, grapheme-to-phoneme unit and prosodic generator. Text analyzer is where input sentences are transformed into word arrays. This unit also identifies the numbers and abbreviations and transforms them into words.

The module for morphological analysis, which is part of the text analyzer, executes morphological analysis of the text in order to recognize the prefixes and the suffixes added to the canonic form of the word. This module also determines the correct accent of the words.

In Macedonian, unlike other languages, the plural of the words, the definite articles, the comparative and the superlative and also other forms are constructed by adding prefixes or suffixes. This can cause some complications in the morphological analysis of the text.

The word stress falls on the antepenultimate syllable almost regularly. This makes the procedure very simple. However, there are only few exceptions of the accentuation rule and for those cases exceptions' dictionary for the accentuation rule is implemented.

In Macedonian standard language there is no vowel reduction, meaning that the vowels sound very similar in their accented and unaccented form. Therefore, this significantly simplifies the entire speech synthesis process for Macedonian. The vowels have only one sound variant, and not multiple, as for example in French.

One of the main characteristics of Macedonian is that each grapheme is an equivalent of one phoneme. Every grapheme of a word is pronounced. Consequently, in the grapheme-to-phoneme unit, Macedonian has a very simple rule for grapheme to phoneme transcription. So, the phonetization task for Macedonian is reduced to trivial checks. The more complex solutions based on dictionaries and morphophonemic rules, like in English or French are not necessary for Macedonian. This unit also makes contextual checks of each phoneme. This is convenient because concatenation of longer speech segments, if available in the speech corpus, produces more natural intonation.

The prosodic generator makes the speech sound more natural. Speech with natural intonation is not only more pleasant for listening, but also it is easier to understand. For speech synthesizer the proper intonation and the prosodic elements are very important. This task does not seem very complicated in Macedonian thanks to the accentuation rules.

The Digital Signal Processing (DSP) module consists of two phases, speech processing and sound processing. It uses the symbolic information for phones and prosody from the NLP module and after certain processing of the input information, gives synthetic speech as output. Speech processing is one of the most important phases in TTS synthesis. In this phase, the symbolic information for the phones and the prosody for the input text are applied on the recorded speech corpus. Furthermore, the matching segments from the corpus and their concatenation order are also established in this phase. Then, the equalization of the segments is done, which gives the uninterrupted flow of the segments. In the next phase, the sound processing, the adjustment of prosodic elements is performed. Finally, the synthesized speech is produced.

A group of scientists from the Faculty of Engineering, University of Novi Sad, Serbia in cooperation with the company AlfaNum[8] from Novi Sad, Serbia, have created systems for automatic speech recognition and text-to-speech synthesis in Serbian, Croatian and Macedonian (Delić, Sečujski, Pekar, & Jakovljević, 2006).

The constructed system is phoneme-based automatic speech recognition over the telephone line. Its accuracy varies between 95% and 99% depending on the vocabulary size and the sound quality. AlfaNum automatic speech recognition (ASR) system is based on hidden Markov models (HMM). State emitting probabilities are modeled by Gaussian mixture models (GMM). The parameters of each Gaussian in GMM are estimated using the Quadratic Bayesian classifier.

The goal of decoding in the AlfaNum ASR systems (Delić, Sečujski, Jakovljević, & Janev, 2010) is to find the most probable word sequence corresponding to the input speech, as

[8] http://www.alfanum-global.com/

well as a confidence measure for each recognition. The search for the most probable word sequence is performed by Viterbi algorithm.

AlfaNum text-to-speech (TTS) and automatic speech recognition (ASR) engines (Pekar, Mišković, Knežević, & Sedlar, 2010) can be used on many different interfaces, all of them built upon basic TTS and ASR libraries written in C++. The Speech Application Programming Interface or SAPI is an application programming interface (*API*) developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. Both SAPI 4 and SAPI 5.x interfaces are implemented in these systems.

The first applications of automatic speech recognition in this region have been introduced at the public telephone network, with support by intelligent network functionalities.

The most widely used TTS-based system in Western Balkans is *anReader*[9]. It is developed by AlfaNum and it is used by almost one thousand visually impaired persons in Serbia, Bosnia and Herzegovina, Montenegro, Croatia and Macedonia. Before the release of anReader, the most widely used system was *WinTalkerVoice*, originally built for Czech.

The Macedonian version of *anReader* currently uses the Serbian speech database, which results in diminished quality of speech. High-level speech synthesis of Serbian and Croatian is performed using expert part-of-speech (POS) taggers. In the case of Macedonian full POS-tagging is not performed because it is unnecessary. This is due to the simplicity of accentuation in Macedonian, which is not the case of the other two languages.

Moreover, application for Android[10] is available for translating from English to Macedonian and vice versa with integrated voice recognition and text-to-speech software.

To conclude, activities for construction and development of oral corpora in Macedonian are currently undertaken. No functional automatic speech recognition software is available for Macedonian (Kraljevski, Chungurski, Mihajlov, & Arsenovski, 2008).

---

[9] http://anreader.alfanum.co.rs/

[10] http://www.androidpit.fr/fr/android/market/applications/application/com.alterme.translator.macedonian/English-Macedonian-Translator

2.4. SPEECH RECOGNITION SOFTWARE

Currently on the market there is a wide variety of companies working on speech recognition software.

- **Dragon NaturallySpeaking**

Dragon NaturallySpeaking[11] is speech recognition software by Nuance. It is a worldwide leader in speech recognition. Nuance has won many awards for its solutions. In 2003, IBM's speech recognition software called ViaVoice was sold to Nuance[12].

There is also a free application from Nuance for iPad™, iPhone™ and iPod touch™ called *Dragon Dictation*. This application allows the users to dictate SMS or e-mails on their mobile phones. The vocal message cannot exceed 30 seconds. Afterwards the audio signal is sent to the company's server and sent back as a written text[13].

Dragon NaturallySpeaking is available for PC and Mac. Many different versions are offered. *Dragon NaturallySpeaking* for PC is compatible with Windows Vista, XP, Windows 2003, Windows 2008, Windows 7 and *Dragon Dictate Mac* is compatible with Mac OS X.

Dragon is available in American English, Australian English, Asian English, Indian English, UK English, Dutch, French, German, Italian and Spanish.

After the installation of *Dragon NaturallySpeaking* for PC or *Dragon Dictate Mac* on a computer, this software offers the possibility to dictate to the computer, to give orders and to search the Web. It is a speaker dependent system, so voice enrolment is necessary

[11] http://www.nuance.fr/for-individuals/by-solution/speech-recognition/index.htm

[12] http://www-01.ibm.com/software/pervasive/viavoice.html

[13] http://chronicle.com/blogs/profhacker/5-easy-speech-to-text-solutions/23016

before use. It is sold with a headset microphone, so the quality of the microphone plays a big role in the accuracy of the speech recognition. It is said that the accuracy improves with time. The more you use it, the more accurate it gets.

- **Speech Recognition by Microsoft**

There is integrated speech recognition software in Windows 7[14]. It allows voice control of the computer, searching the Web and dictating to the computer in any program that allows text input. This software is user dependent so training the computer to recognize only one voice should be done before use. For the punctuation of the text the system translates the punctuation words pronounced by the user into punctuation symbols. The user has to pronounce each punctuation symbol he wants to add to the text. The user can also select a particular word or an entire sentence. There is the possibility of correcting a word, the system gives a list of probable choices and if the word is not on the list there is the possibility to spell the word.

The speeches of the voice dictation are sorted regarding the topic and the language style. Topics include e-mail speech, programming speech, or formal writing. Then, the information for each topic is stored on the hard drive[15].

The performance of this speech recognition software was tested for the purpose of this research. The short testing of the recognition of dictation showed poor results in our case. The transcription on the screen did not match the uttering in most of the cases. There is the possibility to correct the errors vocally, but this was time-consuming and quite unnatural. Patience is the key when working with speech recognition systems, especially for the dictation. Perhaps the poor results that were obtained were due to the fact that the speaker was not a native Anglophone. The quality of the microphone also pays a big role. It was concluded that under this circumstances the use of this speech recognition software for transcription of spontaneous natural language is not possible.

- **Voice and Speech Recognition**

---

[14] http://windows.microsoft.com/en-us/windows7/Set-up-Speech-Recognition

[15] http://searchcio-midmarket.techtarget.com/definition/Speech-Application-Program-Interface

Voice and Speech Recognition by e-Speaking[16] is compatible with Windows 2000, XP, Windows Vista and Windows 7. It gives the possibility of controlling the computer with voice commands, the dictation of e-mails or letters. Furthermore, it has integrated speech synthesis software, the computer responds vocally.

Separate text-to-speech software created by e-Speaking[17] is also available. Amongst a large number of languages it also offers Macedonian. eSpeak uses a "formant synthesis" method and it is written in C.

Voice and Speech Recognition has over hundred voice commands built-in and it is possible to add more commands. It uses the latest technologies form Microsoft. Precisely, it uses Microsoft's Speech Application Program Interface (SAPI) and Microsoft's .NET Framework. It uses scripted language for both the speech recognition and the speech synthesis.

- **Speech Recognition by Google**

The first application using speech recognition developed by Google is GOOG-411[18]. This is an application for smartphone. Thanks to this application a query to search for a location on Google Maps can be entered by using voice command. The image of the map with the location will be sent afterwards on the smartphone by the server.

Another application using speech recognition is the Googol Voice[19]. This is an application for Android. It allows the possibility to transcribe your voice mail into a text message.

There is also the Voice search by Google available for PC[20].

---

[16] http://www.e-speaking.com/

[17] http://espeak.sourceforge.net/

[18] http://googlesystem.blogspot.fr/2008/10/machine-translation-and-speech.html

[19] http://www.frandroid.com/actualites-generales/63147_google-voice-sera-disponible-en-france-fin-2011/

[20] http://www.google.com/insidesearch/features/voicesearch/index.html

The Speech recognition applications by Google are based on a statistical speech recognition engine that uses large oral corpus.

- **TalkingDesktop 7**

TalkingDesktop 7[21] is speech recognition software which is compatible with Windows XP, Windows Vista and Windows 7. It includes dictation, text-to-speech, run applications by voice name, voice browser etc.

- **Tazti 2.4**

Tazti 2.4[22] is speech recognition software. It offers the possibility of voice command, opening and closing programs, even playing PC games. It also gives the possibility to dictate to the computer.

- **TalkItTypeIt**

TalkItTypeIt[23] is speech recognition software which allows voice dictation, voice navigation and transcription of oral messages. Created by Christopher Carey can be used in Win 98, Millenium, Windows 2000 and XP.

- **Voice Finger 2.5.8**

Voice Finger 2.5.8[24] is a speech recognition tool. It was created by Robson Cozendey. It can be used with Windows Vista or Windows 7.

Voice Finger is different form the others speech recognition programs mentioned above. With this tool you can move your mouse using voice commands. The screen is divided in sections and each section it tagged by a number or a letter. By pronouncing the number or the letter of the section the mouse will move to that part of the screen and it will click on it. Whereas on the other speech recognition programs you would say ”open Microsoft

[21] http://www.talkingdesktop.com/index.htm

[22] http://www.tazti.com/index.php

[23] http://www.xpressionsmedia.com/index.php

[24] http://voicefinger.cozendey.com/

Word" and the system will execute the program. With this tool you would need to say "click double" and then the number of the square where the Microsoft Word icon is located on the desktop for example "46".

There is also the possibility to use the keyboard. By pronouncing the letter of the button you want to click, the software will write the given letter to the screen. There is also the possibility of dictating entire words, and phrases.

It is especially developed for people with disabilities or injuries. It enables the user to control the computer strictly by its voice.

- **Responding Heads** 3.5

Responding Heads 3.5 is speech recognition software created by Adsa Life Development[25]. It is compatible with Win 98, Millenium, Windows 2000, XP and Windows Vista. It uses the MS Speech recognition engine.

Responding Heads is mainly created for voice commands and it has the possibility of text-to-speech. But it does not offer the possibility of dictation.

- **Supplemented Speech Recognition (SSR)**

Supplemented Speech Recognition (SSR)[26] is an award winning product created by InvoTek. It is specially designed for people with difficult to understand speech (dysarthric speech). SSR is user dependent system, so firstly the user has to do the voice enrollment. Then the user has to type the first letter of the word and then pronounce the word. Thus, it uses a combination of typing and voice recognition. The voice recognition and the word prediction reduce the amount of typing up to 65% even if the speech is hard to understand. Over time, the performance of the system augments.

[25] http://www.adsalife.com/rh3/index.html

[26] http://www.invotek.org/products/speech-recognition/

- **DictaLink 4 SR**

DictaLink 4 SR is developed by Mysoft[27] and is compatible with Windows Vista, XP, 7 and Windows 2000. It gives the possibility to create and transcribe vocal messages or to create and to correct written and oral files. For the speech recognition, it requires the use of Dragon NaturallySpeaking.

2.5. CONCLUSION

To conclude, currently Dragon NaturallySpeaking is the leading voice recognition software on the market. Many reviews claim that it is the best and the most accurate speech recognition software with even up to 99% accuracy[28].

The website TopTenREVIEWS has rewarded this voice recognition software with the TopTenREVIEWS Gold Award. According to their research Dragon NaturallySpeaking performs with 93% accuracy[29]. Their research shows that it is slightly more accurate when the user is speaking at the normal speaking pace than when the user is speaking slowly. It has difficulty with the short words like “the”, “to”, “our”, “her” etc. At the beginning the user has to be patient and give the software enough training time to adjust to the user’s voice characteristics. Their conclusion is that with time it improves and is quite accurate and useful.

According to other reviews such as an older review from 2008 by PCMag.com's Michael Muchmore[30] and a 2011 review by consumersearch.com[31] Dragon NaturallySpeaking has 98% accuracy.

---

[27] http://www.mysoft.com/produit/dictalink.htm

[28] http://marcnorris.hubpages.com/hub/Dragon-Naturally-Speaking-Review

[29] http://voice-recognition-software-review.toptenreviews.com/dragon-naturally-speaking-review.html

[30] http://www.pcmag.com/article2/0,2817,2327354,00.asp

[31] http://www.consumersearch.com/voice-recognition-software/windows-speech-recognition

Following, with also high reviews, is the Windows Speech Recognition integrated in the Vista and Windows 7 operating systems with around 96% accuracy. It has been said that is nearly as accurate and easy to use as Dragon NaturallySpeaking.

The speech recognition engines use the technique of "language modeling" to differentiate the word that has been pronounced. To analyze the audio sound pronounced in the microphone the system uses statistical models and it interprets those sounds. The system does not understand the meaning of the words. In fact, the system takes into account how frequently words occur by themselves and in the context of other words. The program picks up the word that is most probable in the given context.

The reason why the speech recognition systems are more accurate when the user is speaking at the normal speaking pace is the fact that when the user pronounces an entire sentence there is more context and it is easier for the program to identify the words. The best pace to follow is the way newscasters read the news.

Some characteristics of the natural spontaneous speaking amongst humans, such as mumbling, leaving words out, hesitations, gaps etc. cannot be treated by the speech recognition systems. Moreover, in order to have a structured text appear on the screen each and every punctuation mark has to be pronounced.

In order to get accurate transcription of the dictation, the user has to be in a quiet room, avoiding additional noises. Speech recognition systems cannot discriminate the speech sounds from other sounds coming from the environment. The user has to speak clearly into the microphone. The microphone also plays an important role. It has to be of good quality and positioned in the right position[32].

The main disadvantage is that even the most developed speech recognition systems are completely user dependent. They cannot be used for transcribing interviews, meetings or conferences. With the "training", the user dependent systems learn how to recognize the

[32] http://www.nuance.fr/for-business/by-product/dragon/product-resources/frequently-asked-questions/index.htm

voice of a single user and they cannot understand the speech of more than one speaker. The software adjusts to the unique vocal characteristics of only one speaker[33].

So, overall normal spontaneous speech or a dynamic debate between two or more interlocutors would not be accurately transcribed by speech recognition systems in their current state of the art. Speech recognition systems still use scripted language.

Taking into account that automatic speech recognition systems are still not able to treat spontaneous speaker-independent continuous speech, no affiliation between any speech recognition system and the consecutive interpreting has been done.

Speech recognition systems have never been used as tools for facilitating the process of note-taking for the consecutive interpreting. Still, the process of taking notes for the consecutive is done by hand on a regular notepad. No computational aid is available for facilitating this complex and crucial step in the consecutive interpreting, which poses a lot of problems for beginners in this field.

[33] http://www.nuance.com/for-business/by-product/dragon/product-resources/frequently-asked-questions/index.htm

CHAPTER 3

3. DIFFERENCES BETWEEN TRANSLATION AND INTERPRETING

The interpreting between two languages is the act of orally transferring the massage from one language into another language. In both translation and interpreting, the language of the original massage is called source language and the language in which the massage should be translated into is called target language. The term ‘translation’ is used to describe the process of transferring the meaning from the source language into the target language while working on a text source and producing another text source. In fact the term ‘translation’ designates the text to text conversion of the message from one language into another. Whilst the term ‘interpreting’ means paraphrasing and conversion of the meaning of an oral massage in the source language and the production of its equivalent oral massage in the target language. Thus, the interpreting is speech to speech method.

To conclude, the source which is being translated makes the main difference between the translation and the interpreting. If the source is text the method concerned is translation and if the source is any form of oral message the method concerned is interpreting.

Even though these two techniques of enabling the communication between the speakers of different languages seem quite similar at first, they have many differences and require different skills and competences.

For doing translation of a text in a given source language, sometimes it is sufficient to have a passive knowledge of the source language. Many professional translators make translations of texts only in one direction usually the target language being their mother tongue. The use of dictionaries and other reference materials it is not only permitted, but it is also highly recommended to enable a quality translation. The domain is also very important. It is recommended that a translator specializes in a particular domain and usually he works in this particular domain. Therefore, the extensive knowledge of the

target language and of the subject treated in the text is crucial for making quality translation.

On the other hand, the interpreting demands a high level of proficiency in both the source language and the target language. The main reason for this is that most of the interpreting is bidirectional. Most often the interpreter is invited to interpret the oral communication between two parities speaking different languages. In this case, the interpreter is confronted with a sudden alternation of the source and the target language, the source language becomes the target language and vice versa. Thus, to provide a quality interpretation, the interpreter must have vast knowledge of the target language as well as the translator, but also superior language ability in the source language.

The vast knowledge of the domain is equally important in interpreting as in translation. The interpreter requires at last the basic knowledge of the matter that is being discussed so that he can pass the message correctly.

Besides the knowledge of the languages and of the domain, another very important part is the knowledge of the culture, the civilization and the country of both the source and the target language. In order to make a good translation or interpretation the translator or the interpreter should be familiar with the culture and the particularities of the country. It is highly recommended that he has spent some time abroad in the particular country or in another country where the given language is an official language. This is due to the fact that the language is in constant evolution, it is constantly changing. The things taught at schools and universities are more or less different than the language spoken in the country of interest. The experience of the real contact with the foreign language is something that cannot be taught in school. So, being a good interpreter and even translator demands having spent at least one year abroad in order to improve the language proficiency. The interpreter in particular should be fluent in the foreign language and should start feeling the foreign language, to a certain degree, as his native language.

3.1. TYPES OF INTERPRETING

The most commonly used methods of interpreting are the simultaneous interpreting and the consecutive interpreting. Sometimes these two types of interpreting are referred to as 'face-to-face interpreting' or 'on-site interpreting'. This designation comes from the fact that all parties of the discussion as well as the interpreter are in the same room face to face. It is preferable to use this type of interpreting rather than the remote interpreting while dealing with sensitive and more complex situations.

3.1.1. SIMULTANEOUS INTERPRETING

The simultaneous interpreting is the most common and the most widespread technique when it comes to transferring an oral message from one language – the source language into another language – the target language.

During the simultaneous interpreting, the interpreter is located in a soundproof booth. For the reason that while addressing the audience the speaker uses gestures which can be very helpful in understanding the message, it is very important that the booth is located in a way that the interpreter can see clearly the speaker during the interpreting. Since the interpreting requires a lot of mental effort usually the booths are designed for two interpreters so that they can change each 20 or 30 minutes.

The basic equipment of the booth consists of headphones and microphone. It is recommended that in the booth the interpreter would have the possibility to adjust the volume on the headphones. When doing a relay interpreting in larger venues, where there is more than one booth, the interpreter can adjust the headphones to listen either one of the boots in one ear and the original speaker in the other ear, or two different boots, or in some cases listen to the speaker in one ear and his own interpretation in the other ear.

Moreover, in the booth the interpreter should be able to turn on and off the microphone and pause the microphone to avoid disturbing the audience with unnecessary noises. In better equipped boots there is also a tape recorder so that the interpreter can record either

the original speech or his interpretation. It is recommended that the interpreter has a piece of paper and a pen, pencil is not recommended because it makes more noise. The interpreter writes down the numbers, the percentage, the names and similar facts that are harder to memorize.

Central issue while interpreting whether it comes to simultaneous or consecutive or any other type of interpreting is the fact that it is important to translate the message and not the words. The interpreting is to a great extent paraphrasing, shortening the unnecessary and focusing of the main points of the message.

In the case of simultaneous interpreting, the main focus of the interpreter after getting into the booth, adjusting the volume on the headphones and making sure that everything is working properly is concentrating on the message and not on the words! The interpreter listens and processes the meaning of the words. He cannot start interpreting until he has understood the general meaning of the first sentence. Only after he has grasped a meaningful thought, he can start composing the same thought in the target language. While he utters the interpretation of the first thought, at the same time the interpreter has to listen and process the next thought. Usually the interpreter takes 5 to 10 seconds or even more delay, to properly understand the message in the source language and to compose the proper interpretation in the target language. Thus, the most challenging thing while preforming simultaneous interpreting is listening and speaking at the same time. This requires a profound concentration, a good memory, rapidity when thinking and talking, decisiveness when choosing the words etc.

The ideal speaker for the simultaneous interpreting does not speak too fast because much of the original message would be lost. On the other hand, the ideal speaker does not speak too slowly because in that case it would be very difficult for the interpreter to grasp the meaning of the original message. The interpretation would also have too many pauses due to the fact that the interpreter has to wait long time to get a complete thought that can be interpreted.

It is obvious that an interpreter doing simultaneous interpreting has to be proficient in both the source language and the target language. The simultaneous interpreter has so many challenges to confront that it would be impossible to achieve a quality interpretation of the

message if he has to struggle with trying to think of the correct word or tense. Since everything is happening in a matter of seconds it is understandable that using dictionaries or another reference material is not so easy to do. The interpreter has to have an extensive vocabulary in both languages, and also to be able to paraphrase rapidly, to use synonyms in case he cannot think of the best solution at that moment.

One of the key skills in the simultaneous interpreting is being decisive. There is no time to think of every possible variant translation or the correct idiom. The interpreter must choose rapidly the best solution he can think of at the moment.

Furthermore, the best rule is to keep it simple. The sentences should be short, clear and easy to understand. Most often, it is recommended to do a simple sentence composed only of subject, verb and objet and to set aside the parts of the message which are not so important. The interpreter should not worry about rhetorical devices, figures of speech or about making perfect literary sentences. On the contrary, for a translator, when making a translation of a literary text, it is very important to preserve all the components of the original literary text including the rhetorical devices, the figures of speech etc. For the interpreter the goal is to translate the message in the simplest, shortest way possible. That way it is more probable that the interpreter would achieve to translate most of the speech rather than if he tries to make perfect sentences. Not only the interpreter would need more time, but also while concentrating on making perfect sentences, he also risks missing an important part of the original message.

In the case when the speaker is speaking faster, the best thing to do is to concentrate only on the essential parts and omitting the details. Most often the words that can be omitted without changing too much the meaning of the sentence are the adjectives, the adverbs, the personal pronouns in Macedonian, which is not the case in French, etc. But then again, in some cases even the omission of an adjective can be harmful for the meaning. Therefore, the interpreter should decide on the spot what is possible to omit and what absolutely cannot be omitted. The repeated parts of the sentence, the redundancies can also be omitted. In other words everything that does not bring a new meaning to the sentence is omitted. The paraphrasing is not only permitted but also recommended in some cases. The interpreter can even make one sentence in the target language out of two or more

sentences in the source language without changing the meaning. This way the goal is achieved and the interpreter has saved a lot of time to concentrate on the following sentences.

The main objective of the simultaneous interpreting is not to repeat everything that was said by the speaker. It is in fact to express the same meaning in the target language in a way that the native speaker of the target language would not have difficulties in understanding. Hence, what should be avoided at all costs is making literal translation of each and every word that follows in the source language. That way the main goal of the interpreting – the passing of the meaning of the message from one language into another language – would not be achieved.

Given the fact that most of the conferences on which simultaneous interpreting is used are technical it is obligatory that the interpreter knows the subject of the conference in advance. The quote of Benjamin Franklin "By failing to prepare, you are preparing to fail." is greatly relevant to the simultaneous interpreting. The interpreter should revise the technical vocabulary from the particular domain in both the source language and the target language. Most of the professional interpreters make glossaries before the conference. They make research in the given field to become more familiar with the subject. They also look up the previous works of the participants in the conference etc.

Furthermore, it is very important that the interpreter is well informed of the current situation in the world in general and particularly in the two countries (of the source language and the target language). Professional interpreters watch the news and read the newspapers of two or more countries daily. It is crucial for an interpreter to be up-to-date with the political, economic, social, cultural situation in the two or more countries depending on how many languages he interprets, in order to be able to interpret. It is quite often that some parts of the discussions on the conference are related with the global situation in the country. This aspect is as important as knowing the culture, the customs, the habits, the tradition, the lifestyle of the country etc.

3.1.2. Consecutive Interpreting

The consecutive interpreting is the second most common method of interpreting.

In the consecutive interpreting the speaker speaks without any interruptions for up to 6 minutes at most. During this time the interpreter listens carefully, memorizes what is being said and takes notes to help his memory. Only after the speaker has finished a complete segment which would be somewhat an equivalent to an entire paragraph in a written text, the interpreter starts interpreting.

The main difference between the simultaneous and the consecutive interpreting is the fact that while the speaker addresses the audience the interpreter only listens, he does not interpret. After the speaker has finished, he lets the interpreter to address the audience passing the message into the target language. Unlike the simultaneous interpreting, in the consecutive interpreting the speaker and the interpreter are not talking at the same time. In this case the audience cannot chose who they want to listen, the audience is listening consecutively the speaker and the interpreter. Therefore, the consecutive interpreting doubles the time of the conference or the meeting.

This type of interpreting does not require specific equipment unlike the simultaneous interpreting. The only thing the interpreter needs is a notepad and a pen.

The ideal speaker for the consecutive interpreting speaks slowly. Unlike the simultaneous interpreting where it is best if the speaker is neither too fast neither too slow, in this case it would be better if the speaker speaks a bit slower than for the simultaneous interpreting.

The consecutive interpreting demands greater capacity of memorizing than the simultaneous interpreting. In the simultaneous interpreting the interpreter should memorize just about 15 seconds to 1 minute ahead. On the other hand, in the consecutive interpreting the interpreter should memorize at once up to 6 minutes of the speech. Even though the interpreter takes notes, the notes are not at all a text that the interpreter should translate afterwards. The notes are just little triggers of the memory of the interpreter, they are usually bits and pieces that become an entire thought when they are connected with the

bits and pieces memorized by the interpreter. The key to quality consecutive interpreting is writing less, using more symbols than words, and memorizing as much as possible.

There is also another aspect of the consecutive interpreting which is not present in the simultaneous interpreting. It is the direct contact with the audience. During the simultaneous interpreting the interpreter is practically invisible. He is located in the booth, most often behind the audience. In this case the voice of the interpreter is the only thing that gets in contact with the audience. On the other hand, during the consecutive interpreting the interpreter plays an equal role with the speaker. It is required that the consecutive interpreter has oratory skills. He should stand straight in front of the audience with firm voice without any hesitations, without making big pauses, without reading too much from the notes. In this case is not just the voice that he has to pay attention to but also the entire posture, the eye contact with the audience etc.

The concentration plays essential role in the consecutive interpreting. The concentration should be very intense in order the interpreter to be able to pass the message. From the beginning of the speech the interpreter has to concentrate deeply because he is not in a soundproof booth. The consecutive interpreter should learn how to ignore all kind of disturbing noises and concentrate on the speech. He may also experience difficulties hearing the speaker if he speaks quietly etc. Moreover, while interpreting, the interpreter should not be disturbed by different noises or situations coming from the surroundings, the concentration should be very intense.

It is recommended that the interpretation of the interpreter is only two-thirds of the original segment. For a sequence of 6 minutes the interpreter should interpret for 4 minutes and 30 seconds at most. All the repetitions, redundancies, unnecessary details should be omitted. The audience is impatient to hear the essential part of the message. The consecutive interpreting, in general, doubles the time of the meeting or the conference. Consequently, if the interpreter does not respect this instruction he would be confronted with the impatience of the participants and of the audience. It is very important for the consecutive interpreter to be efficient, fast, and brief.

NOTE-TAKING

Having developed and effective note-taking system is crucial for the quality of the consecutive interpreting.

The technique of note-taking is completely different from the normal hand writing. Moreover, the notes taken by the interpreters are different then the shorthand used in stenography. It is essential that the notes are neat and easy to discriminate at only one glace. This would be impossible if the interpreter used shorthand, because even the experienced stenographers cannot read one or two lines at only one glance (Herbert, 1952). Using shorthand, it would not be possible to correct or add the missing parts fast enough while the speaker takes a small break.

So, the interpreters use signs and symbols in their notes. The biggest advantage of the signs and the symbols is that they do not belong to a particular language (Herbert, 1952). They do not represent the words, but the ideas and the objects. So, they are universal and can be used regardless the language, similar to the use of the Arabic numbers in Indo-European languages. The signs and the symbols are also convenient because if the signs and symbols are well chosen, not too complicated, it is faster to write them down than to write the entire word or an abbreviation. Furthermore, by using signs and symbols the notes are much more visual and easy to read at only one glance.

The signs and symbols used by the interpreters are individual and subjectively chosen by the interpreter himself. There is no given list or a dictionary containing the signs and symbols used by the interpreters. Every interpreter invents its own system. There are only few examples of the signs and symbols which can be found in literature concerning the note-taking in consecutive interpreting. Then again, those are only examples not a constraints, the interpreter can chose whether he would implement these symbols or invent other signs and symbols.

The interpreting notes play key role in the consecutive interpreting. It must be stressed once again that the technique that the interpreters use to take notes is strictly individual. Each and every interpreter uses different symbols and different abbreviations. Nevertheless, there are some general patterns for the method of taking notes which are

followed by most of the professional consecutive interpreters and which can be very helpful for beginners.

It is advised to use A5 notepad with top spiral. The top spiral is important because it makes it easier to turn pages. The interpreter can fold down the notes from the previous segment of the speech and leave unfolded the notes from the part that he should interpret next. So, when the speaker finishes the segment, the interpreter should just turn to the first unfolded page and immediately start interpreting from the beginning of that segment. It is very unsightly to have the entire audience waiting for the interpreter to finish writing, then start searching and hesitating for minutes which part was from the previous segment and which is from the new segment. The interpreter should start interpreting within seconds after the speaker has finished. Up to half a minute delay is tolerable, more than that would be inappropriate.

On the notepad the interpreter makes a vertical margin where he writes down the logical connectives or logical operators. It is very important that they are written down in order to know the connection between the different thoughts or sentences. Since the speaker is not dictating to the interpreter, but he is speaking fluently as he would speak normally while making a speech, it is impossible for the interpreter to write down everything.

One of the few rules in note-taking for the consecutive interpreting is to write diagonally on the notepad. The subject is written on the left, the verb is written in the center and slightly lower than the subject and on the right is written the object slightly lower than the verb. After each finished thought or sentence the interpreter draws a horizontal line from end to end of the notepad or just a little line on the margin to know where a sentence starts and where it ends.

The language of the notes is also important. Generally it is considered that it would be easier for the interpreter if he takes the notes directly in the target language instead of in the source language. But in practice, for beginners, I myself being one, in my opinion it is easier to take the notes in the source language and then interpret on the spot. So, it purely depends on the choice of the interpreter.

The ideal situation would be having symbols for every state, action or thing. The less the interpreter uses words the faster and the better would be to interpret the segment. The

advantage of the symbols is that they are not language related and the interpreter can use the same symbols regardless of the source language or the target language. Just as an example some interpreters use the following symbols:

□ country

□ → export

→ □ import

◯ the world

The symbols should really have a particular meaning and remind the interpreter of the given thing. That is why the symbols are individual. For example the symbol:

♡ can represent: heart,

so  would mean heart failure.

The same symbol for another interpreter can represent: love, couple, dedication, compassion, peace etc.

The symbol  can represent: church, religion or have a completely different meaning like: death, victim etc.

It is very important to choose well the symbols and to use a given symbol only in one meaning. The symbols are easy to read and make the notes much more visual. The interpreter does not have much time to read the notes. He can glance from time to time in the notes but he should make sure that is not overly noticeable.

Consequently, the symbols should be omnipresent in the notes of the interpreter but, it is obvious that it is impossible to have a symbol for everything. Quite frequently the interpreters also use abbreviations combined with symbols that describe some morphological features. For example:

Indemnisation = indemni^{t}

Indemniser = indemni$^{®}$

Indemnisable = indemni^{a}

Symbols for some common prefixes and suffixes are usually used by the interpreters. So for each and every noun ending in –tion the interpreter would use a small ‘t’ at the end, if it is a verb he will put a different sign, for example something like a small ‘R’, or if it is an adjective ending in –able he would use a small ‘a’ or any other symbol that reminds the interpreter of the given type of word.

The acronyms of the institutions and the organizations are used always in one language regardless in what direction the interpreter is interpreting. For exemple:

AAAF: Association Aéronautique et Astronautique de France

AESA: Agence Européenne de la Sécurité Aérienne

CSST: Commission de la Santé et de la Sécurité du Travail

When writing the notes the interpreter writes much bigger letters than his normal handwriting. He or she leaves a lot of space in between the words and usually writes diagonally, not in a straight line. On one page of the note pad, the interpreter can fit about up to tree sentences. Thus, taking notes is significantly different from writing. The reason for all of this is the fact that the notes should be very neat and organized, well separated and legible at only one glance.

Besides the importance of having developed, unambiguous system of signs, symbols, abbreviations and acronyms, another very important factor when taking notes is to know what to write down. The key is not to rush to write down everything because that is impossible and the interpreter risks making chaotic notes which would not help him at all during the interpreting. The notes can be of help only if they are well organized, legible and unambiguous. If not, they will make the situation even worse.

Furthermore, it is very important that the notes are complete. Even if the interpreter writes down the subject, the object and all the other parts of the sentence but does not note the verb, that sentence could easily become meaningless. If he does not remember the verb he would not be able to interpret correctly that sentence.

Then again, as it was the case in the simultaneous interpreting it is very important to be decisive and to make a quick judgment of what is essential and what is not so important in the message. In some cases only one word is enough to remind the interpreter of an entire sentence. Taking notes is very individual and depends on the type of speech. Some speeches are easier to memorize than others. In some cases taking a very few notes is sufficient but, in speeches filled with facts and numbers, all this facts should be noted.

Another general rule followed by most of the professional consecutive interpreters is that the facts like: numbers, percentages and proper nouns should always be written down. Anyhow, in the end the technique of taking notes is developed and molded in time. With practice and experience the interpreter discovers the best solution for him.

In Table 1: Examples of Symbols, few examples of symbols used in note-taking in consecutive interpreting taken from the electronic source Interpreter Training Resources[34] can be observed.

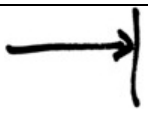
TABLE 1: EXAMPLES OF SYMBOLS

industry	
problem	
say	
meeting	

[34] http://interpreters.free.fr/consecnotes/symbolexamples.htm

http://interpreters.free.fr/consecnotes/symbolexamples2.htm

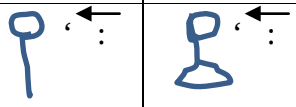
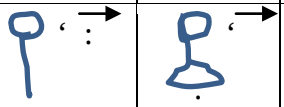
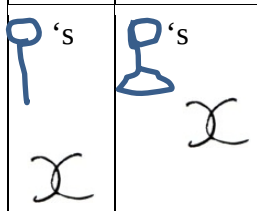
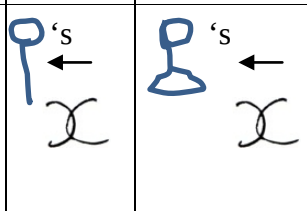
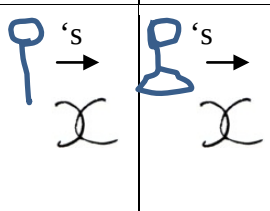
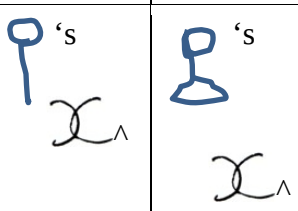
http://interpreters.free.fr/consecnotes/symbolexamples3.htm

end	
continue	
country	
decide	
any	
agriculture	
impact	
relations	
consequences	
until	

CONJUGATION

In the note-taking it is very important to always write down the verbs. Even though, it is advised to significantly shorten the phrases and avoid taking too many notes, the verbs should always be written down because they usually carry the meaning of the sentence. So, in general the interpreter should always write down the subject, the verb and the object. Few examples of conjugated verbs can be observed in Table 2: Conjugation.

TABLE 2: CONJUGATION

Present		**Past**		**Future**		**Conditional**	
Je parle		J'ai parlé		Je vais parler		Je parlerais	
I "		I "		I "		I "^	
Tu parles		Tu as parlé		Tu vas parler		Tu parlerais	
II "		II "		II "		II "^	
Il pense	Elle pense	Il a pensé	Elle a pensé	Il va penser	Elle va penser	Il penserait	Elle penserait
							
Nous voulons		Nous avons voulu		Nous allons vouloir		Nous voudrions	
							
Vous avez		Vous avez eu		Vous allez avoir		Vous auriez	
Y θ		Y θ		Y θ		Y θ^	
Ils lient	Elles lient	Ils ont lié	Elles ont lié	Ils vont lier	Elles vont lier	Ils lieraient	Elles lieraient
							

For every personal pronoun there is a particular symbol, which should be always added before the verb in order to avoid confusion. To express the past tense or the future tens the interpreter adds a little symbol above the verb and the conditional is expressed also by another symbol. Table 2: Conjugation

LOGICAL CONNECTIVES

In the process of note-taking, the interpreter always writes down the logical connectives or grammatical conjunctions, usually in the margin. They are really important because they connect the sentences and they show the interpreter in what direction the speech is going.

Here are some examples of some grammatical conjunctions. For all the synonyms the interpreter uses only one symbol.

- Addition
    - et **&** ; ensuite, voire, d'ailleurs, encore, de plus, quant à, non seulement… mais encore, de surcroît, en outre... **D +**
- Alternative
    - ou, soit ... soit, tantôt ... tantôt, ou ... ou, ou bien, seulement ... mais encore, l'un ... l'autre, d'un côté ... de l'autre, d'une part... d'autre part **V**
- Opposition
    - mais, or, cependant, pourtant, toutefois, néanmoins, en revanche, au contraire, malgré tout, certes... **HO**
- Cause
    - car, parce que, puisque, étant donné que, comme, vu que, sous prétexte que.. **Ċ** ⟹

- Consequence
    - donc, aussi, finalement, ainsi, voilà pourquoi, c'est pourquoi, par conséquent, tout compte fait... Ċ ⟹
- Hypotheses
    - si, au cas où, en admettant que, pourvu que, à condition que... Ħ
- Reason

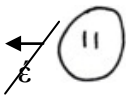
    pour que, de peur que, de crainte que, afin que... Ц

NEGATION

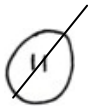
It is quite easy to describe the negation. The interpreter only crosses out the verb or the symbol which is negated.

For example in the sentence: ‘Le Président n’était pas à la réunion.’

The negation will be expressed in the following way:

Pdt 

Or it could be even more simplified:

Pdt 

MATHEMATICAL SYMBOLS

A lot of mathematical symbols are being used in the note-taking. The symbols like +, -, *, <, >, = are omnipresent in the interpreter’s notes.

For the large numbers, a single line over the number can represent thousand:

43.000 43

or respectively two lines would represent millions:

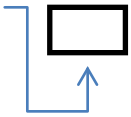
317.000.000 317

COMPLETE SENTENCES

Here are few examples of complete sentences, showing how the symbols connect to each other to create a whole sentence.

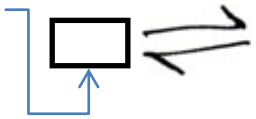
Je voudrais voir la Macédoine dans l'Union européenne en 2015.

I MK EU '15

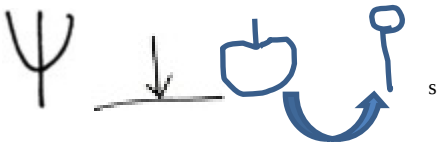


D'ailleurs, le peuple macédonien pense qu'il faut des changements dans les relations entre la Macédoine et la Grèce.

D + MK : MK – GR



L'agriculture assure l'alimentation des humains.



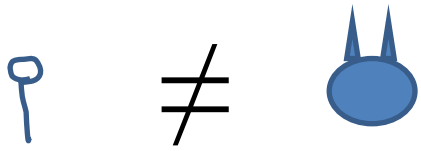
L'impact du developpement sur l'environement est enorme.

This passive sentence is converted into an active sentence to simplify it.

(Le developpement a un impact enorme sur l'environement)

L’énergie nucléaire entraine des répercussions.

L'homme se distingue de la bête.



3.1.3. LIAISON INTERPRETING

The liaison interpreting is used for smaller groups or meetings. It is frequently used for business meeting, visits to a foreign country, one-on-one interviews etc.

The participants of the conversation speak two or more different languages. During liaison interpreting the interpreter interprets in both directions. In this situation there is a sudden alternation of the source language and the target language. As an example we can imagine a business meeting between two executives. One of them is from Macedonia and the other is from France. The French executive starts talking in French, and then the interpreter interprets into Macedonian. Then the Macedonian executive answers in Macedonian and the interpreter interprets into French.

This technique is similar to the consecutive interpreting but is much less formal. The segments are much shorter so the interpreter does not require taking notes. The communication is spontaneous and flexible. The crucial requirement in this type of interpreting is being fluent in both languages and having good memory.

3.1.4. RELAY INTERPRETING

The relay interpreting is a subcategory of the simultaneous interpreting. It is used for large multilingual meetings like for example the meetings of the EU Institutions. In this type of conferences there are several target languages. Respectively there are several booths. One interpreter interprets from the source language into a pivot language which is a language known by all the interpreters. Usually the pivot language is English because it is most commonly used language by the interpreters. On the other hand, French is used as a pivot language quite often, because of its good structure and the richness of its vocabulary. All the other interpreters listen to the interpretation in the pivot language and interpret into their respective languages. For example Macedonian source language would be translated by one of the booths into French, and the other booths would not translate from the source language – Macedonian in this case, but from the pivot language – French into their respective languages Japanese, Spanish etc.

The task of the pivot simultaneous interpreter is even more challenging than the task of the simultaneous interpreter. This is due to the fact that the pivot interpreter has to provide almost perfect interpretation because his eventual omissions and inaccuracies would be multiplied in all the other languages. The pivot interpreter stays closer to the original text in terms of leaving less delay (time gap), and avoiding the free paraphrasing, sticking closer to the exact meaning of the message, using the exact words and not synonyms etc.

The technique of the relay interpreting can also be used with the consecutive interpreting. In this case the pivot consecutive interpreter interprets into the pivot language and the other interpreters take notes during the interpretation of the pivot interpreter, afterwards they interpret into their languages. The use of the relay interpreting with the consecutive interpreting is much less common than with the simultaneous interpreting. This is due to the fact that with the consecutive interpreting the duration of the meeting will double, triple or quadruple according to the number of target languages.

3.1.5. Remote Interpreting

The remote interpreting (in French ‘interprétation à distance’) is a form of simultaneous interpreting in which the interpreter is not located in a booth in the conference room. He is on another location and follows the conference through video-conferencing set up. Thus, in remote interpreting all the participants in the conference or the meeting are located in one venue, while the team of interpreters is located at another location and interprets via video conferencing. The main difference between the remote interpreting and the video conference interpreting is the fact that the latter describes a type of interpreting where the team of interpreters is located at the same location as a part of the conference participants and possibly part of the audience.

In the remote interpreting, the only connection between the interpreter on one side and the participants in the conference along with the audience on the other side is made possible by large screens or monitors. Usually, on the monitors in the background is shown a large frame of the entire conference, then few smaller frames on top. There is a smaller frame of the speaker in the upper left corner of the screen or in the lower right corner of the screen. Furthermore a smaller frame of the chairman is located at the same parallel as the frame of the speaker.

The interpreter uses the usual equipment to perform the interpretation – the headphones and the microphone regardless the fact that he is not in the same venue as all the rest of the participants in the conference.

3.1.6. Telephone Interpreting

The telephone interpreting is also referred to as "over-the-phone interpreting", "telephonic interpreting" and "tele-interpreting". It is a type of interpreting which is done via telephone. There are few different possibilities whereas the telephone interpreting is concerned. Either the interpreting is three directional – the three parties, the interpreter and

the two participants in the conversation are all at separate locations or the two participants are face to face and only the interpreter participates over the phone.

The telephone interpreting, the remote interpreting and the video conference interpreting cost less because the company that hires the interpreter does not have to pay for travel expenses. Also, another positive side of these three types of interpreting is the fact that since at a larger meetings as for example the plenary sittings of the European Parliament where there are over 500 possible language combinations (23 x 22 languages) there is not enough space in the conference rooms to fit all the interpreting boots. Therefore, interpreting from another location is not only cost-effective but also saves a lot of space in the conference room.

Anyhow, these immerging new types of interpreting, made possible by the development of the new technologies, are disputed by interpreters. Interpreters consider them to be more artificial and that the difficulties of the interpreting are greater than when dealing with a live situation. But still, it is possible that the remote interpreting will completely replace the simultaneous interpreting in the future.

This was also the case when the simultaneous interpreting first appeared. Many interpreters were against it because they considered that it treats the interpreters as machines or objects. The consecutive interpreter which was an active participant in the conference was pushed aside, in the background and made invisible in the interpreting booth. Nowadays, not only that the interpreters accept the simultaneous interpreting but, some even prefer the simultaneous over the consecutive interpreting.

Then again, in the case of the remote interpreting the interpreter is not only put in a little corner of the room as a machine but it is also pushed out of the room into another place as if he does not even participates at all in the communication. The interpreter is not even an object in this case but just an immaterial voce that is transmitted from somewhere.

Therefore, if the simultaneous interpreting in time replaced largely the consecutive interpreting despite the discontent of interprets, perhaps one day the remote interpreting will replace largely the simultaneous interpreting. This raises the question what will be the new technology that will replace the remote interpreting? How much can the new

technology change the way of interpreting? To which extent can the way of interpreting change? Can the human interpreter also be replaced?

CHAPTER 4

4. ORAL CORPORA

Oral corpora are excellent source of real data for the research into the spontaneous oral expression. As mentioned above in “Differences between Translation and Interpreting”, the interpreting consists of transmitting an oral message into another language. For this reason using oral corpora will provide the possibility to undertake pragma-linguistic research into the domain of interpreting.

The oral corpora will be used to analyze the spontaneous oral expression in Macedonian language. It is composed of audio recordings from diverse debates and interviews where the participants in the discussion express themselves in a spontaneous manner. All of the recordings, that are being used, have been broadcasted by some of the Macedonian TV channels.

The main subject treated in the recordings is the domain of the political sociology. In particular, the recordings treat some aspects of the contemporary political situation in Macedonia; the international relations between Macedonia and the European Union (EU); the integration of Macedonia in the EU and the North Atlantic Treaty Organization (NATO); the relations between Macedonia and Greece; the liberalization of visa regime with EU; the democracy and the freedom of speech in Macedonia etc.

The political aspect and the content of the recordings would not be of any interest for this research. This work aims to analyze the frequency of words in the spontaneous speech in Macedonian within the constricted domain of the political sociology. The recordings are randomly chosen taking into account the following factors:

- The source language is Macedonian;
- The recording is audible;

- The speakers express themselves spontaneously. Prewritten and edited speeches do not comply;
- The subject of the discussion is the sociopolitical situation in Macedonia;
- The content and the positions taken by the speakers are irrelevant.

For the transcription of the oral corpus we used the software Sonal[35]. It gives the possibility to transcribe manually the recordings in format *.wav. Different parts of the speech can be marked with different colors depending on the subject that is being discussed and afterwards an organized and clear view of the subject of interest can be extracted. It gives the possibility to make a synthesis of only the parts of interest and listen or read the transcription of only these parts, to do a research only on this selection and also export the text of the transcription of the parts of interest in Word or in *.rtf.

There is a clear view of the transcribed parts that have already been finished and the remaining parts that are yet to be done. There is the possibility to add up to 99 variables for each recording in order to determine the interlocutors, classify them according to their age, sex, profession etc., determine the type of the recording, whether it is a debate or an interview etc. and many other possibilities that are advantageous when making research.

Furthermore, the most important reason why we chose this particular software (Sonal) to treat the oral corpus is the possibility to determine the lexis. It gives the possibility to calculate the occurrence of the words in a particular topic or in the entire corpus. It calculates the number of different words and how many times they appear. It is also possible to see the full list of the recordings where a particular word appears and the exact context in which the given word appears.

This would be very advantageous for determining the particular terminology that is used in our oral corpus. Moreover, a clear view of the most frequently used words will help to determine for which of them we will create symbols for the note-taking in the consecutive interpreting.

[35] http://www.sonal-info.com/

Thus, as the source of the interpreting is an oral message this research is based on oral corpora. The oral corpora used for this research consists of 44 recordings, all of which were broadcasted on some of the following Macedonian TV-channels:

- MTV1,
- Sitel,
- A1,
- Kanal 5,
- Alfa,
- Nasa TV,
- AB,
- Moris
- Radio Slobodna Evropa

4.1. VARIABLES

Three variables were introduced in the corpus to classify the recordings more easily. There is information about what TV-channel they were retrieved from. Then, there is information about what is the nature of the recording and lastly, who is the main speaker in the recording. In order to see the frequency of this variables in the corpus see Figure 1: Frequencies

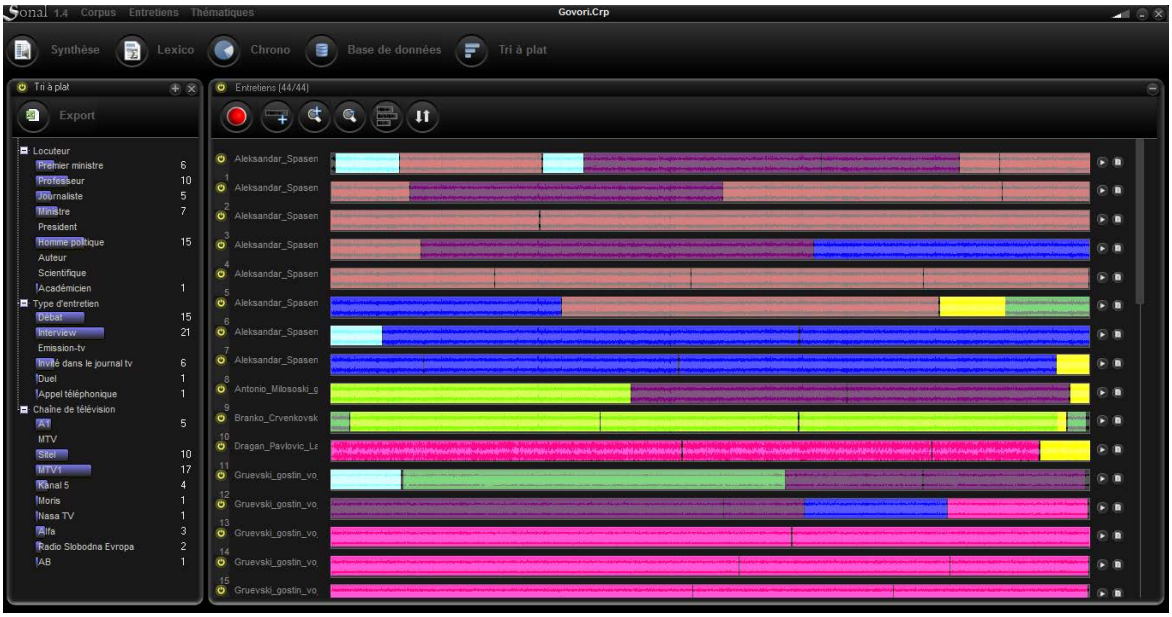


FIGURE 1: FREQUENCIES

In Figure 2: Speaker we can observe the frequency of the variable Speaker. The numbers and the dashes represent the number of recordings in which the main speaker has the given profession. In most of the recordings the main speaker is a politician.

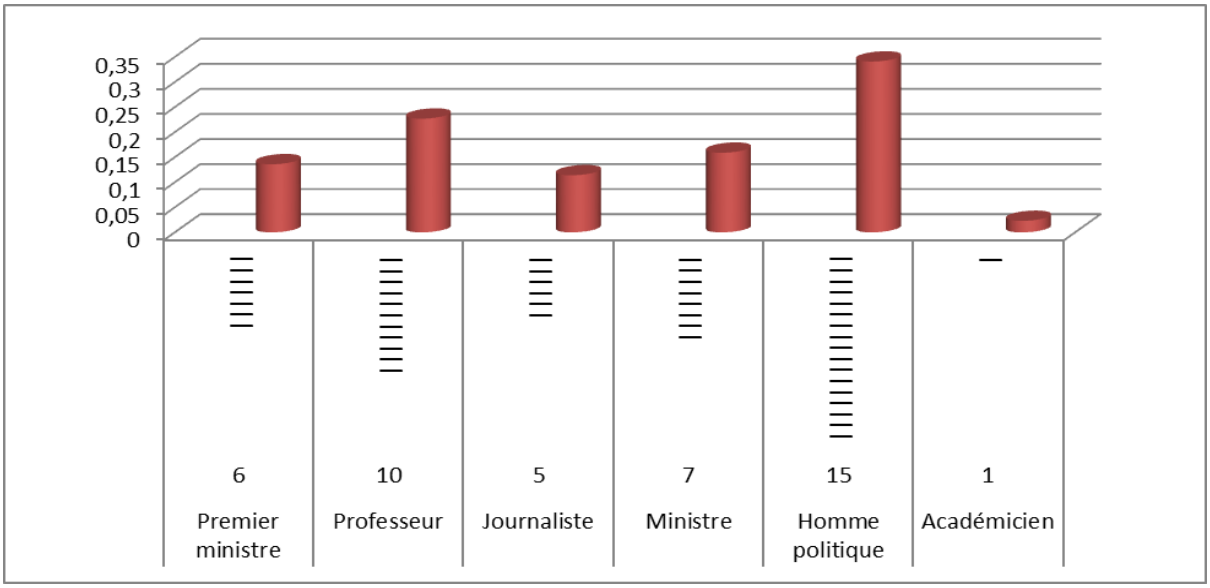


FIGURE 2: SPEAKER

In Figure 3: Nature of the Recording we can observe that most of the recordings are interviews.

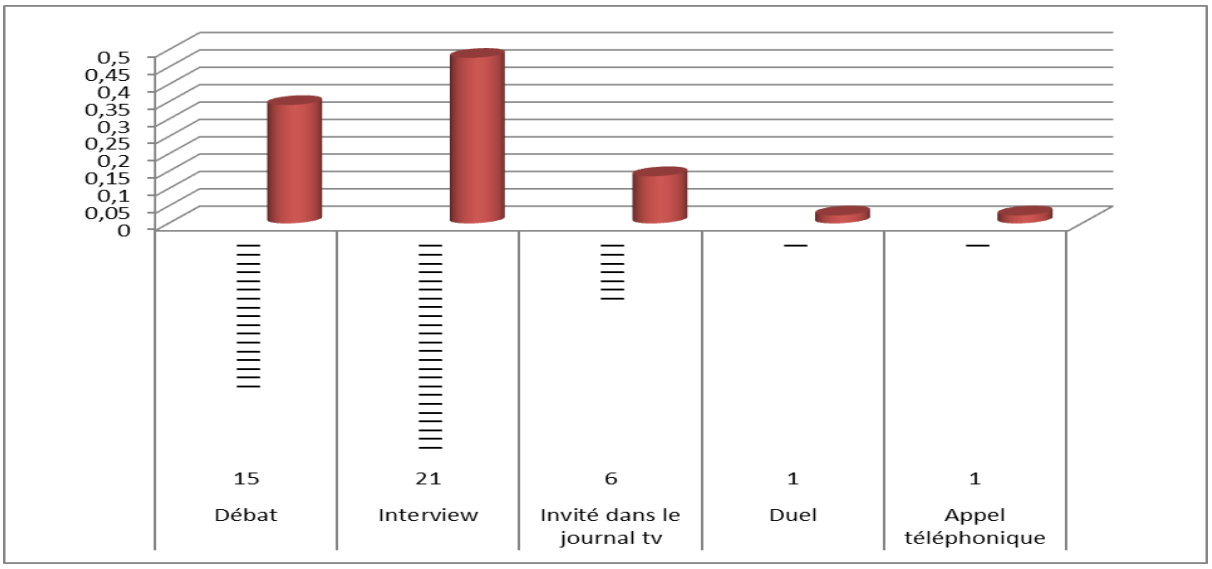


FIGURE 3: NATURE OF THE RECORDING

In Figure 4: TV-Channel we can note that most of the recordings are taken from the public national TV-channel Makedonska Televizija 1.

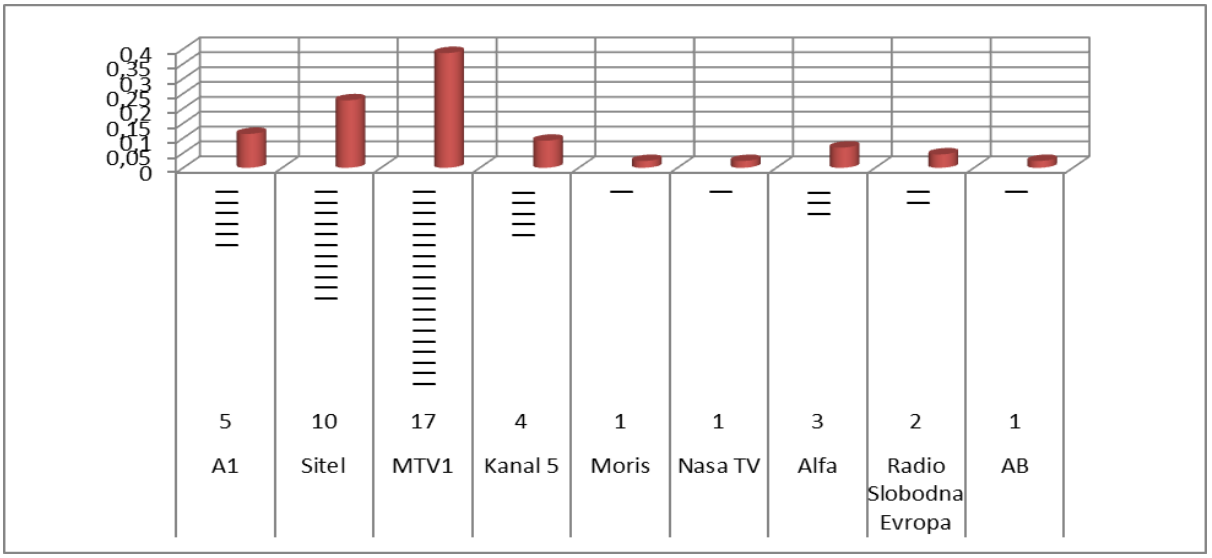


FIGURE 4: TV-CHANNEL

Sample of the database showing the variables for some of the recordings in the oral corpus can be observed in Table 3: Database. Detailed view of all the variables for all of the 44 recordings from the oral corpus can be found under Appendix C: Database.

Recording	Date added	Length	Speaker	Type of recording	TV channel
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del2.wav	27/04/2012 02:12	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del3.wav	27/04/2012 02:13	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del5.wav	27/04/2012 02:14	10:00	Politician	Interview	MTV1
Dragan_Pavlovic_Latas_vs__Borjan_Jovanovski.wav	27/04/2012 21:10	09:13	Journalist	Debate	MTV1
Gruevski_gostin_vo_„NIE__25_04_2010__del1.wav	27/04/2012 22:01	10:10	Prime Minister	Interview	Sitel
Makedonija_megu_osudata_na_komunismot__5.wav	27/04/2012 21:57	09:42	Professor	Debate	A1
Makedonskata_Istorija_vo_EU_-_Tret_del.wav	27/04/2012 21:09	13:49	Academician	Telephone call	Moris
Milososki__intervju_za_Kanal_5__30_07_09.wav	26/04/2012 22:01	08:43	Minister	Guest in the News	Kanal 5
Milososki_po_povod_viznata_liberalizacija__16_07_09.wav	26/04/2012 22:00	08:57	Minister	Guest in the News	Sitel
Nikola_Todorov_za_Radio_Slobodna_Evropa__06_02_2010.wav	27/04/2012 02:01	10:39	Minister	Interview	Radio Slobodna Evropa
Stavreski__intervju_za_Kanal_5__27_08_09.wav	27/04/2012 02:03	08:26	Minister	Guest in the News	Kanal 5
TRN_VO_OKO_SO_MILENKO_I_CHOMOVSKI_28_Sep_2011_1_del_mpg.wav	26/04/2012 21:53	11:57	Journalist	Duel	AB
11Novinarstvoto_vo_Makedonija_.wav	10/11/2012 18:27	13:20	Journalist	Debate	MTV1
1„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	11:40	Professor	Debate	MTV1

TABLE 3: DATABASE

4.2. TOPICS

The recordings form the oral corpora were all segmented and a topic was attributed according to the content discussed in the segment. The entire corpus consists of 9 topics Figure 5: Topics. Out of those 9 topics 3 topics were introduced to segment the parts of the recordings that are not so important. Those are:

- Introduction;
- Farewells;
- Music.

The other 6 topics were predetermined and the recordings were chosen to take part in the corpus under the condition that they threat some of these topics.

- Contemporary political situation in Macedonia;
- International relations between Macedonia and the European Union (EU);
- Integration of Macedonia in the North Atlantic Treaty Organization (NATO);
- Relations between Macedonia and Greece;
- Liberalization of visa regime with EU;
- Council of Europe;

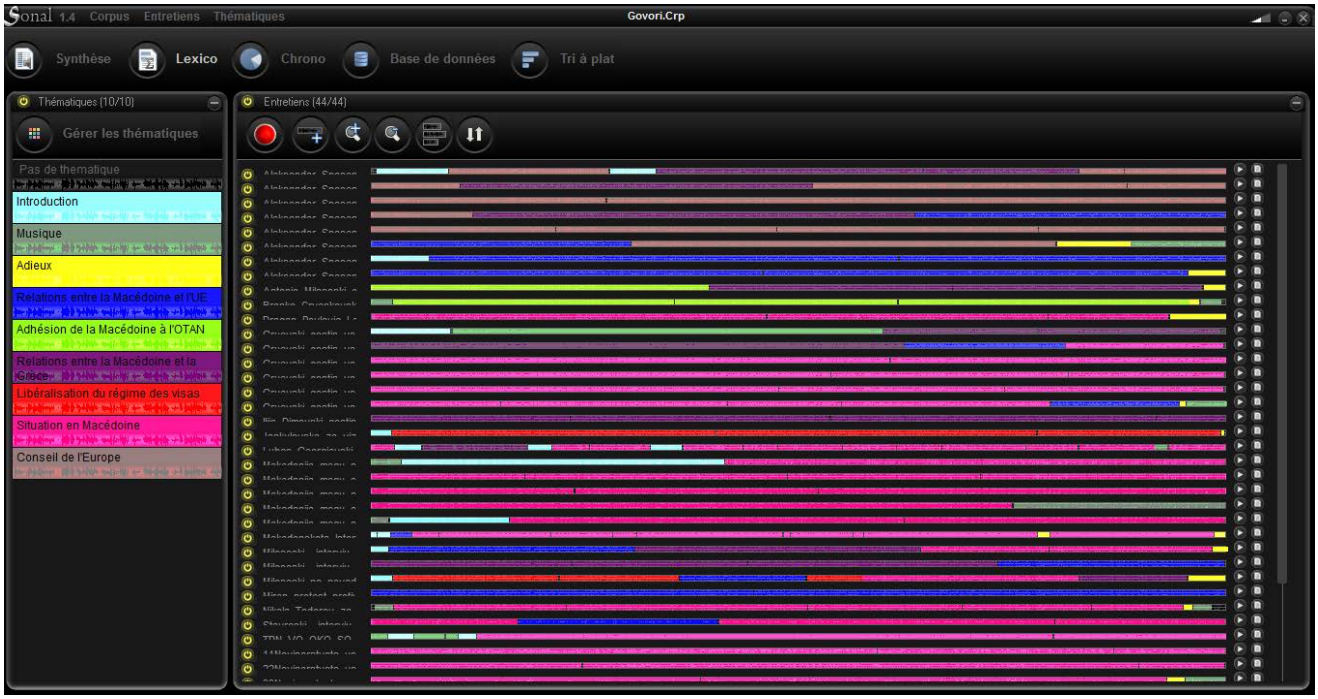


FIGURE 5: TOPICS

In Figure 6: Chrono we can note that almost half of the corpus 47% focuses on the matter of the *Contemporary political situation in Macedonia*. Next, 20% of the corpus is on the subject of the *International relations between Macedonia and the European Union (EU).* Then, the topic the *Relations between Macedonia and Greece* is represented with 10.9% and *Council of Europe* with 7.5%. The total length of the entire oral corpus is 7hours 8minutes and 41seconds.

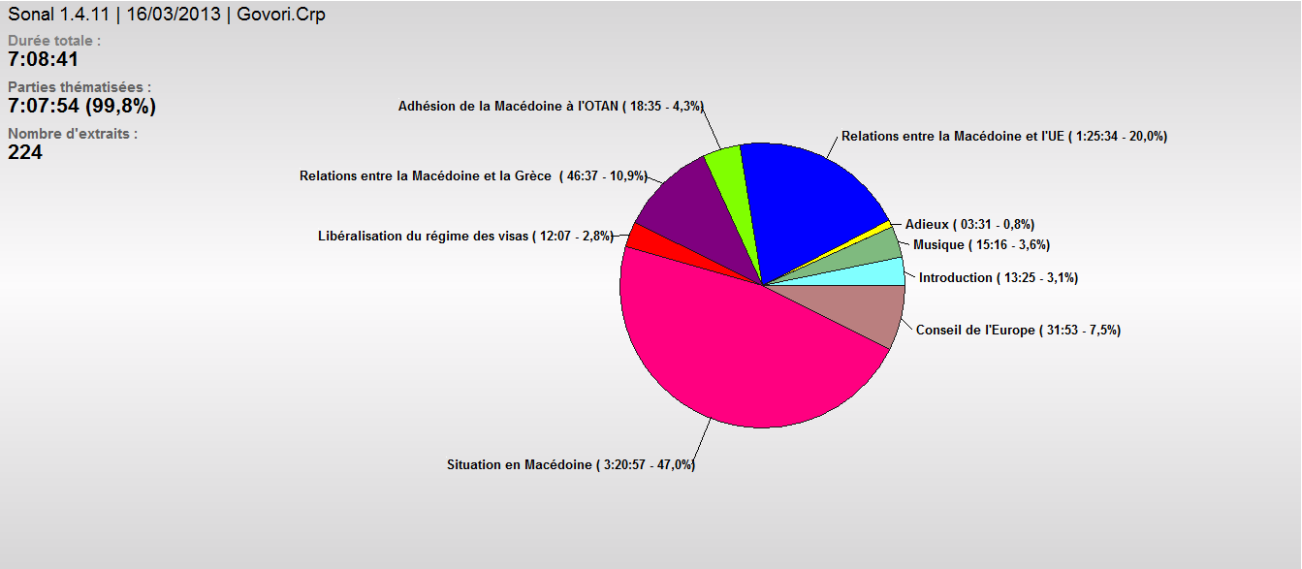


FIGURE 6: CHRONO

4.3. WORD COUNT

In order to process the oral corpora it was indispensable to transfer the oral speech into written text. For that reason all the recordings were transcribed manually using the software Sonal. The main focus of this research is the frequency of the words. We are interested in knowing which are the most frequent words that appear in this particular domain. The software Sonal was chosen for this task amongst other transcription software particularly because it has integrated word count system. Figure 7: Word Count.

The corpus consists of 64.155 words, out of which 8.930 are different words.

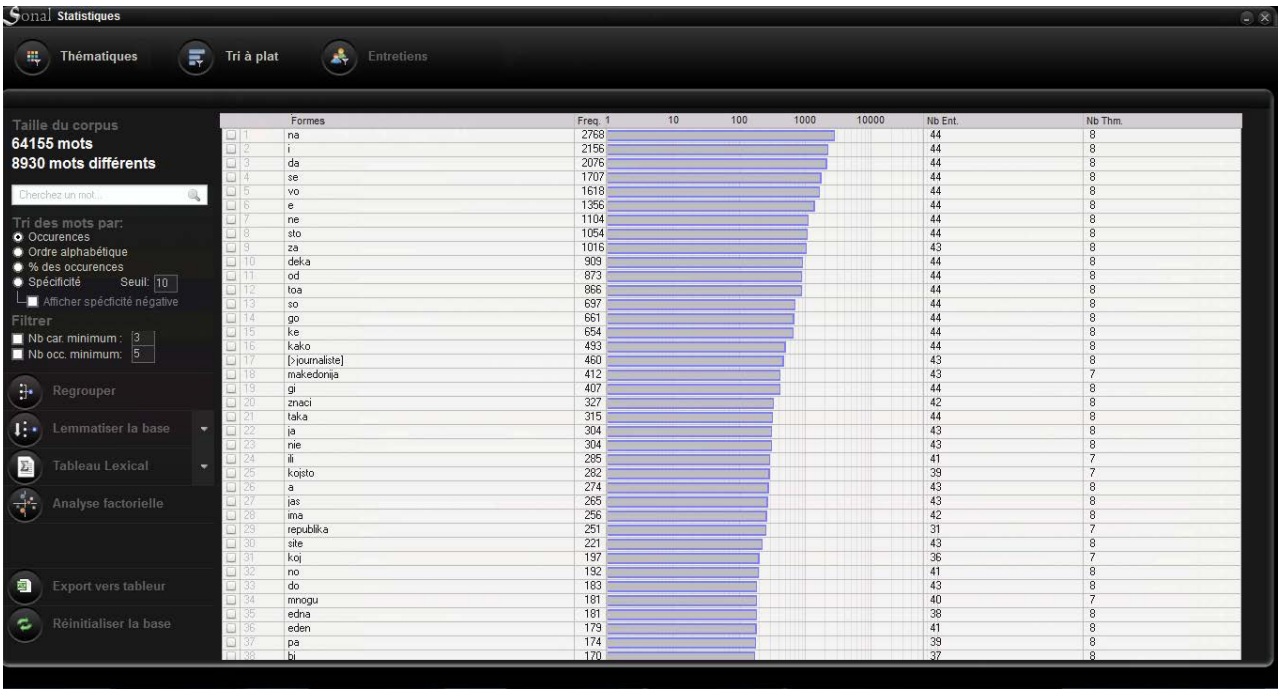


FIGURE 7: WORD COUNT

4.4. DIFFICULTIES

CYRILLIC SCRIPT

While collecting the oral corpora some difficulties arose. First of all, the most disturbing inconvenience was that the software Sonal did not support the Cyrillic script. Macedonian is written exclusively with the Macedonian alphabet, which is an adaptation of the Cyrillic script.

Taking into consideration all the advantages that this software offers and its compatibility with the research in question, made us decide to use this software regardless this significant inconvenience which is to use the Latin script for the transcription of Macedonian.

The Macedonian alphabet consists of 31 graphemes representing the 31 phonemes of the Macedonian standard language. For the transliteration of the Cyrillic script into Latin

script we adopted transliteration similar to the *Scientific transliteration*[36] of Cyrillic script to the Latin script (Romanization).

Moreover, we have simplified the transliteration further on to only the Latin characters present on a regular French or English keyboard in order to gain time in typing.

We avoided as much as possible using double characters for the very same reason of saving time in typing. Every phoneme is represented by a single grapheme from the Latin script available on the keyboard, with exception to:

i. Љ љ, /ʎ/ - Lj lj

ii. Њ њ, /ɲ/ - Nj nj

iii. Џ џ, /dʒ/ - Dz dz

Therefore, some of the Latin graphemes are used for more than one phoneme in Macedonian. There is ambiguity in the case of:

i. **Ж ж /ʒ/**; **З з /z/** and **S s /dz/** the three transliterated by **Z z**

ii. **Г г /g/** and **Ѓ ѓ /ɟ/** both transliterated by **G g**

iii. **С с /s/** and **Ш ш /ʃ/** both transliterated by **S s**

iv. **К к /k/** and **Ќ ќ /c/** both transliterated by **K k**

v. **Ц ц /ts/ and Ч ч /tʃ/** both transliterated by **C c**

[36] http://en.wikipedia.org/wiki/Cyrillic_transliteration

http://portal.unesco.org/culture/en/files/32319/11625494823macedonian_en.pdf/macedonian_en.pdf

http://en.wikipedia.org/wiki/Romanization_of_Macedonian

In Table 4: Romanization of Macedonian, only the ambiguous Latin characters that were used for more than one phoneme and the doubled graphemes are represented.

TABLE 4: ROMANIZATION OF MACEDONIAN

Cyrillic	IPA	Latin
Г г	/ɡ/	G g
Ѓ ѓ	/ɟ/	G g
Ж ж	/ʒ/	Z z
З з	/z/	Z z
S s	/dz/	Z z
Љ љ	/ʎ/	Lj lj
Њ њ	/ɲ/	Nj nj
С с	/s/	S s
Ш ш	/ʃ/	S s
К к	/k/	K k
Ќ ќ	/c/	K k
Ц ц	/ts/	C c
Ч ч	/tʃ/	C c
Џ џ	/dʒ/	Dz dz

Full table of the transliteration code used for the transcription of the oral corpus can be found in Appendix E: Romanization of Macedonian.

Thus, in the software all the recordings are transcribed with the Latin script and there is no possibility to change to Cyrillic script. But, the software Sonal gives the possibility to extract all the transcriptions of the recordings in a Word document format. So, there is the possibility to change the font of the transcriptions into the MAC C Times or to use some transliteration software[37]. The only problem then would be to change manually all the characters that are not present in the Latin script like љ, њ, s, ч, ќ, ж, ѓ, ш, џ, etc.

[37] http://tools.forret.com/translit/macedonian.php

http://www.microsoft.com/en-us/download/details.aspx?id=1438

Unfortunately, this process cannot be done automatically, because the transliteration from Latin to Cyrillic depends on the context for the ambiguous graphemes.

FORMAT

After collecting the recordings we encountered some difficulties in uploading them into the software. The problem was that all the recordings were in .MP3 format and the system offers more possibilities if the recordings are in .WAV format.

For that purpose we used of the software for phonetic analysis Praat[38] to convert them into .WAV file.

LENGTH

After uploading all the recordings into the corpus we noticed that the length of the recordings varied significantly. For this reason, the recordings were incomparable and it was much more difficult to process the long sound files that were about 1 hour long.

In order to deal with this problem and to have recordings with comparable length, we used the software Praat once again. We uploaded the long sound files into Praat and we extracted parts from them. In fact, we did not lose any part of the recordings, we just segmented them into shorter in length recordings. Because of the fact that most of the recordings we had in the beginning were about 10 minutes long we decided to take that length as default length and we shortened all the recordings approximately to this length. For example, the recordings that were 1 hour long were segmented into 6 recordings about 10 minutes long.

The shortest recording in length in our oral corpus is 03min 21sec and the longest recording in length is 13min 49sec. Another advantage of this software Sonal is that any

[38] http://www.fon.hum.uva.nl/praat/

recording can be turned off if it should not be taken into consideration. If we decide to do some other research on these corpora regarding the lexis or the topics we could turn off the recordings that seem to be too short or too long and do the research only on the recordings that are 10 minutes long.

The real length of the recordings form the oral corpus can be observed in Figure 8: Length. The numbers above the recordings (00.00, 05.00 and 10.00) represent minutes.

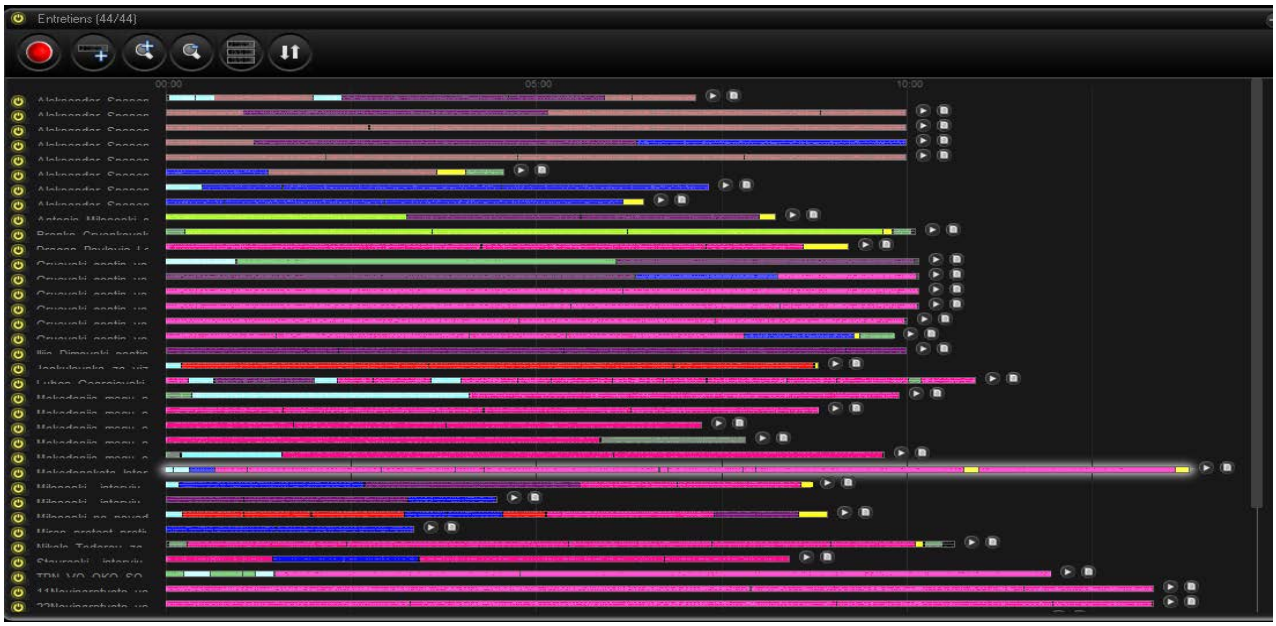


FIGURE 8: LENGTH

Anyhow, the length of the recordings is probably not the best criterion if we want to have comparable segments of speech. Perhaps it is better to take the number of words as the primary criterion because in the recording we can have long breaks or a speaker that speaks very slowly, etc. For that reason, after the manual transcription of the recordings was done, we created a table with the number of words in each recording, along with the length of the recording. The longest recording in words in our oral corpus has 2.749 words and it is 13min 20sec long.

The length and the number of words of the recordings in the oral corpus can be found in Table 5: Words & Length and in Figure 9: Words & Length

TABLE 5: WORDS & LENGTH

Recording	Words	Minuties:Seconds
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del1.wav	1277	07:09
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del2.wav	1749	10:00
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del3.wav	1791	10:00
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del4.wav	1698	10:00
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del5.wav	1648	10:00
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del6.wav	744	04:34
Aleksandar_Spasenovski_za_Evropski_klub__11_10_09_del1.wav	1217	07:20
Aleksandar_Spasenovski_za_Evropski_klub__11_10_09_del2.wav	1045	06:27
Antonio_Milososki_gostin_vo_vestite_na_Sitel__2mart_2011_.wav	1261	08:15
Branko_Crvenkovski_intervju_za_Radio_Slobodna_Evropa.wav	1434	10:08
Dragan_Pavlovic_Latas_vs__Borjan_Jovanovski.wav	1572	09:13
Gruevski_gostin_vo_„NIE__25_04_2010__del1.wav	865	10:10
Gruevski_gostin_vo_„NIE__25_04_2010__del2.wav	1846	10:10
Gruevski_gostin_vo_„NIE__25_04_2010__del3.wav	1791	10:10
Gruevski_gostin_vo_„NIE__25_04_2010__del4.wav	1800	10:10
Gruevski_gostin_vo_„NIE__25_04_2010__del5.wav	1837	10:01
Gruevski_gostin_vo_„NIE__25_04_2010__del6.wav	1779	09:50
Ilija_Dimovski_gostin_vo_Faktor5__28_04_2010__del3.wav	1660	10:00
Jankulovska_za_viznata_liberalizacija__15_07_09.wav	1408	08:49
Lubco_Georgievski_-_28_05_2010_Vo_centar_2_6.wav	1788	10:56
Makedonija_megu_osudata_na_komunismot__1.wav	1222	09:55
Makedonija_megu_osudata_na_komunismot__2.wav	1200	08:49
Makedonija_megu_osudata_na_komunismot__3.wav	1023	07:14
Makedonija_megu_osudata_na_komunismot__4.wav	801	07:50
Makedonija_megu_osudata_na_komunismot__5.wav	1268	09:42
Makedonskata_Istorija_vo_EU_-_Tret_del.wav	2000	13:49
Milososki__intervju_za_Kanal_5__30_07_09.wav	1546	08:43
Milososki__intervju_za_Sitel__07_10_09.wav	809	04:28
Milososki_po_povod_viznata_liberalizacija__16_07_09.wav	1708	08:57
Miren_protest_protiv_„evropskiot_genocid.wav	451	03:21
Nikola_Todorov_za_Radio_Slobodna_Evropa__06_02_2010.wav	1534	10:39
Stavreski__intervju_za_Kanal_5__27_08_09.wav	1344	08:26
TRN_VO_OKO_SO_MILENKO_I_CHOMOVSKI_28_Sep_2011_1_del_mpg.wav	1662	11:57
11Novinarstvoto_vo_Makedonija_.wav	2432	13:20
22Novinarstvoto_vo_Makedonija_.wav	2749	13:20
33Novinarstvoto_vo_Makedonija_.wav	2315	11:50
1„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	1516	11:40
2„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	1511	11:40
3„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	1226	11:40
4„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	1706	11:40

55„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	1982	13:36
1Pretsedatelot_Crvenkovski_vo_„Presing.wav	1547	10:50
2Pretsedatelot_Crvenkovski_vo_„Presing.wav	1628	10:50
3Pretsedatelot_Crvenkovski_vo_„Presing.wav	1573	11:02
TOTAL	66963	25720 sec = 7,144 hours
AVERAGE	1521,89	09:44
STANDARD DEVIATION	439,45	135,8
PEARSON		0,8010

Figure 9: Words & Length shows that the length of the recording is not so relevant if we are interested in grasping equal segments of oral speech. It can be observed that even though the length is similar the number of words varies from one recording into another. Some of the recording have the same length but show some difference in the number of words. The Pearson product-moment correlation coefficient between the number of words and the length is 0,8010.

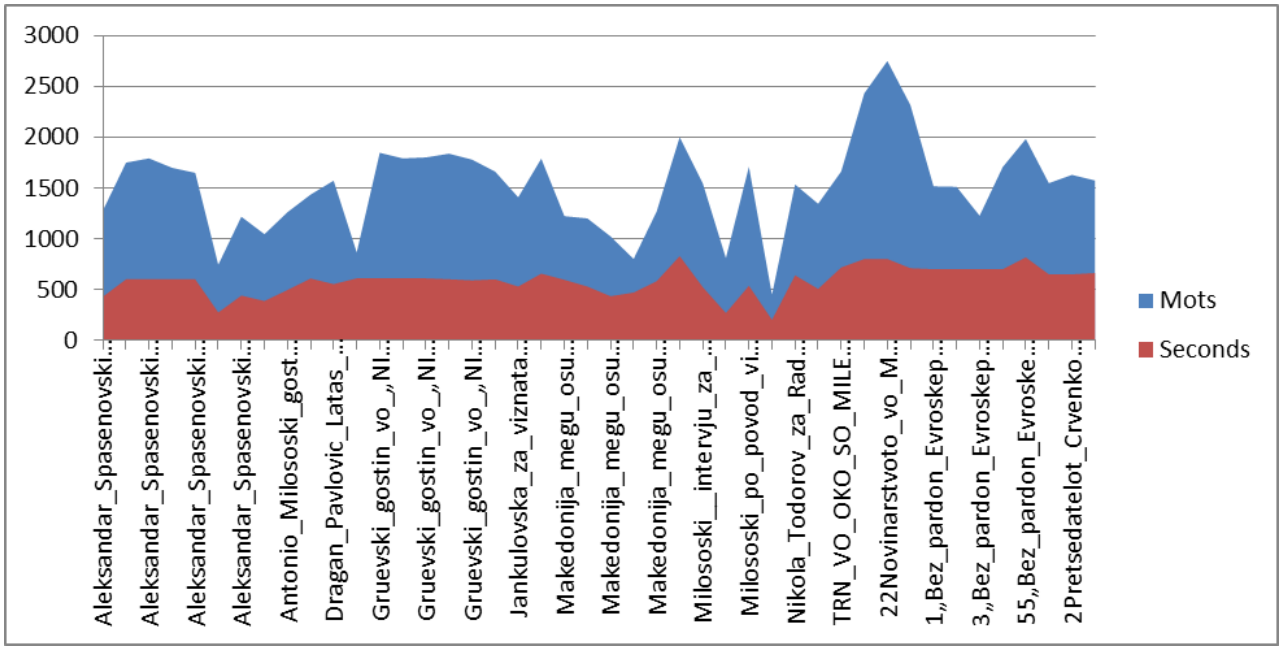


FIGURE 9: WORDS & LENGTH

4.5. Results

The oral corpora was used as means of extracting the information required for discovering the data that we needed in order to have an overview of how to create significant amount of symbols so that we can arrange them together in order to build a computer assisted tool for the note-taking in consecutive interpreting.

After the collection of the recordings that met the required criteria, we uploaded them into the software Sonal. Furthermore, we proceeded with their manual segmentation with regards of the particular topic that was discussed on the recording. The topics were predetermined and constituted one of the criteria for selection as mentioned above.

One of the most important steps, if not the most important one, was the transcription of the oral recordings into text. The transcription was done by hand with the help of the software Sonal. The reason why this step is so important is because we could not treat the corpus in its original state, being oral recordings. In order to collect the data that we needed we required tangible text source. The manual transcription of the recordings, besides being a painstaking task, provided us with the information we needed. The transcription of 10 minutes of recording equals to about 1.700 words and it took about 1 hour to do the transcription.

Once the transcription of the entire corpus was done we extracted the list of most frequent words that appear in the oral corpus. The list was examined and the misspellings were corrected into the system. There were some bizarre words that are not in the literary standard Macedonian, but since they were uttered in the oral corpus they were left as such. Then after all the transcriptions were corrected, the final correct list of most frequent words was extracted. Besides that, all the transcriptions of the recordings were extracted into Word document files (.docx format). All together there are about 185 pages of transcriptions.

Some segments of the transcriptions can be found in Appendix A: Transcriptions of the Oral Recordings.

Discovering the most frequent words in the oral expression in Macedonian from the political domain was the main goal of this research. Once the final correct list of most frequent words was obtained, the task of lemmatization of these words took place. The lemmas of the words were extracted because symbols are attributed to the lemma of the word, the word form that can be found in a dictionary entry.

The list of the most frequent words was processed and reduced only to the words that had a particular meaning, or words that are noted by the interpreters in the traditional note-taking. All the prepositions and conjunctions were put in the stop list. For the time being, the adjectives and the adverbs were also left aside. As it is also the common practice in the traditional note-taking, only constructing the simplest possible sentences is of interest.

After all the words that belong to the stop list were cleared from the list, all the inflected forms of the nouns and the verbs were reduced to their lemma. The verbs were grouped together according to some similarity in meaning. Some of the most frequent nouns that did not require symbols were not added. For example, in the corpus the noun 'Republic' appeared 251 times and the noun 'Union' appeared 132. But, in fact in context in the corpus almost always the noun 'Republic' preceded a name of a country (Република Македонија), so to note this the interpreter can only write 'RM' or just the name of the country 'MK' as he usually notes it, so that a special symbol for 'Republic' is not necessary. The same principle applies also in the case of 'Union', in the context in the oral corpus it was almost always used as 'European Union' (Европската Унија), so most likely the interpreter would note this as 'EU'. The list of lemmas of the nouns and the verbs that was obtained represented the words for which we required symbols for the note-taking.

SURVEY

Five professional interpreters, with some international experience in the consecutive interpreting, were consulted on the matter of choosing the symbols.

The subjects of the survey are all conference interpreters. One of the subjects has 19 years of experience as conference and business interpreter (German <-> English) and is currently undertaking PhD Studies in the field of the note-taking.

Another subject is an interpreter and translator (French <-> Macedonian and English <-> Macedonian) with 10 years of experience and also 8 years of experience as French teacher and Professional collaborator at the Faculty of Philology “Blaze Koneski” in Skopje, Macedonia.

Another subject has two years of experience with a Bachelor degree in Language Interpretation and Translation (English <-> Macedonian and French <-> Macedonian) and is currently doing a Masters in Conference Interpreting in Skopje, Macedonia.

The other subject has about 5 years of experience as freelance interpreter and translator (French <-> English). The last subject also works as an interpreter (Russian <-> English).

The survey had a Macedonian and an English version. It consisted of three parts. The first part contained some of the most frequent nouns of the oral corpus. In the survey there were 19 nouns: дел del ‘part’, држава drzava ‘country’, влада vlada ‘government’, однос odnos ‘with regard to = in relation to’, јазик jazik ‘language’, време vreme ‘time’, страна strana ‘side’, начин nacin ‘manner’, работа rabota ‘work’, народ narod ‘people’, име ime ‘name’, Собрание Sobranie ‘Assembly’, ден den ‘day’, избори izbori ‘elections’, период period ‘period’, новинарство novinarstvo ‘journalism’, политика politika ‘politics’, проблем problem ‘problem’, and ситуација situacija ‘situation’.

In the Macedonian version of the survey, an example of a sentence extracted directly from the corpus, with the noun used in the most frequent meaning, was also provided. This was done in order to give the interpreters some additional information so that they can see how they would connect the symbol of the noun in the given context. It was specified that it is not necessary to stick strictly to the given example. The examples from the corpus were not translated in the English version of the survey because the meaning of the translated nouns in English seemed straight-forward.

The second part of the survey consisted of some of the most frequent verbs in the corpus. The given verbs were grouped by some similarity in meaning. Synonyms, partial

synonyms and antonyms were grouped in a same section in order to make the task of the interpreters easier. It was indicated that if they consider that the verbs from the same section can be represented by the same symbol to draw only one symbol for a section. There were 37 verbs grouped in 12 sections:

1. demand ≠ give; бара bara ≠ дава dava
2. come =~ enter ≠ go; доаѓа doaga = влезе vleze ≠ оди odi
3. watch = see; гледа gleda = види vidi
4. have ≠ lack; има ima ≠ нема nema
5. say = speak = tell = state =~ discuss =~ propose; каже kaze = рече rece = вели veli = вика vika = зборува zboruva =~ разговара razgovara =~ дискутира diskutira =~ предложи predlozi
6. think = reckon =~ decide =~ understand =~ believe =~ know; мисли misli = смета smeta =~ реши resi =~ разбира razbira =~ верува veruva =~ знае znae
7. must =~ should =~ can; мора mora =~ треба treba =~ може moze
8. hope =~ expect =~ want; надева nadeva =~ очекува ocekuva =~ сака saka
9. bring =~ return ≠ take; носи nosi =~ врати vrati ≠ однесува odnesuva
10. improve =~ succeed =~ get; подобри podobri =~ успее uspee =~ добие dobie
11. work =~ lead; работи raboti =~ води vodi
12. happen =~ be =~ become =~ exist =~ represent =~ mean; случи sluci =~ сум sum =~ станува stanuva =~ постои postoi =~ претставува pretstavuva =~ значи znaci

The third part of the survey contained a list of 14 symbols retrieved from an electronic source, Interpreter Training Resources[39]. The interpreters were asked to write the word for which they use the symbol from the list. If they are not using the given symbols, they were asked to enter the words that would be compatible for the given symbols according to

[39] http://interpreters.free.fr/consecnotes/notes.htm

them (choosing the words mentioned in the survey or other words). The answers provided by the interpreters for the third part of the survey can be observed in Table 6: Survey symbols.

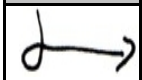
TABLE 6: SURVEY SYMBOLS

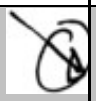
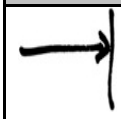
Symbol	Interpreter Training Resources [40]	Interpreter 1	Interpreter 2	Interpreter 3	Interpreter 4	Interpreter 5	Matches
	industry	/	industry	/	фабрика fabrika 'factory'	industry	2/5
	problem	/	tensions	fluctuation, crisis	планина planina 'mountain'	/	0/5
	say	/	/	think	во однос на vo odnos na 'in relation	say, speak, talk	1/5

[40] http://interpreters.free.fr/consecnotes/symbolexamples.htm

http://interpreters.free.fr/consecnotes/symbolexamples2.htm

http://interpreters.free.fr/consecnotes/symbolexamples3.htm

					to’		
	meeting	/	/	propose	човек covek ‘man’	discussion round	1/5
	end	/	/	end, to sum up, to conclude, all in all	пречка precka ‘obstacle’	/	1/5
	continue	/	/	looking forward to	пресврт presvrt ‘change, turn’	/	0/5
	country	country	country	country, state	држава, земја drzava, zemja ‘country, state’	city	4/5
	decide	/	/	no, negation	нула, ништо nula, nisto ‘zero, nothing’	average	0/5
	any	/	/	answer	има ima ‘have’	missing	0/5
	agriculture	/	/	agriculture	дрво drvo	agriculture	2/5

					‘tree’		
	impact	objective	/	influence, impact	се фокусира на se fokusira na ‘to focus on’	target	1/5
	relations	/	/	compare	дилема dilema ‘dilemma’	either… or	0/5
	consequences	/	/	consequence	круг (на луѓе) krug (na luge) ‘circle (of people)’	/	1/5
	until	/	before, up until now	until	завршува zavrsuva ‘to end’	obstacle	2/5
Matches :	14/14	1/14	3/14	6/14	1/14	4/14	
Average :	2,996/14 = 1.071/5 = 0.214						

In average, the 5 interpreters agree with only about 3 meanings for the given symbols out of 14 taken from the Interpreter Training Resources. For each symbol, in average, only 1 interpreter out of 5 agrees with the choice of the word for the given symbol.

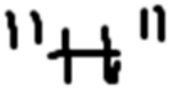
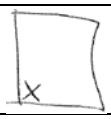
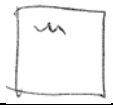
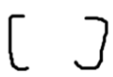
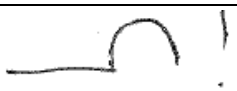
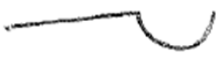
To conclude, the 5 interpreters were asked to provide the symbols for 56 words and to attribute words for 14 symbols.

The interpreters provided the symbols they use for the traditional note-taking in consecutive interpreting for the words of our list. Based on their choices, the symbols that were used by more than one interpreter were chosen to be the default symbols in this project.

A segment of the dictionary consisting of the Macedonian words obtained from the oral corpus and their translation into French and English, the part-of-speech tags and the symbols for the words, is shown in Table 7: Dictionary. The full symbol dictionary can be found in Appendix D: Symbol Dictionary.

TABLE 7: DICTIONARY

MK	EN	FR	PoS	SY
дел	part	part	n.	
држава	country	état	n.	
влада	government	gouvernement	n.	
однос	with regard to = in relation to	en ce qui concerne = par rapport à	n.	
јазик	language	langue	n.	
време	time	temps	n.	
страна	side	côté = parti	n.	

начин	manner	manière	n.	
работа	work	travail	n.	
народ	people	peuple	n.	
име	name	nom	n.	
Собрание	Assembley	Assemblée	n.	
ден	day	jour	n.	
избори	elections	élection	n.	
период	period	période	n.	
новинарство	journalism	journalisme	n.	
политика	politics	politique	n.	
проблем	problem	problème	n.	
ситуација	situation	situation	n.	
бара	demand	demander	v.	
дава	give	donner	v.	
доаѓа	come	venir	v.	
влезе	enter	entrer	v.	

оди	go	aller	v. 
гледа = види	watch = see	regarder = voir	v. 

As it can be observed in Table 7: Dictionary, the same symbol is used for the three languages in the dictionary. In case another language is added to the dictionary, regardless of the specificities of that language, the symbol would remain the same. The reason for this is that the symbols shown in the table above do not represent phonetic nor morphological nor syllabic writing. They are not based on the form of the word in a given language or on the pronunciation of the word. This type of symbols used for the note-taking can be treated as pictographic writing of the ideas behind the words that we use to represent those ideas. Still, the symbols shown in the dictionary above are not concrete or 'natural', they are to a great extent abstract and conventional in some way.

As there is no perfectly precise limit for depicting the level to which a script belongs to, whether it is the lexemic level, the morphemic level, the syllabic level or the phonemic/phonetic level (Coulmas, 1989), for any of the nown scripts in the world, the proto writing script used for the note-taking is no exception. In any script, there can be situations in which the graphemes used in the script depict word writing, morhpeme writing, syllable writing and phonetic writing at the same time.

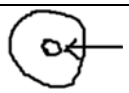
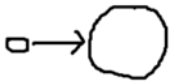
In the traditional note-taking there are symbols or signs that represent morphological features or syllabic features. For example the 's' symbol in the symbol for people  can also be found in many other symbols, represents a morphological feature, describing the plural form. In the case of abbreviations some symbols that represent syllable writing can be found quite often like in the example: Indemniser = indemni®.

If needed the infinitive form of the verbs is indicated by the symbol ® right after the symbol of the verb, just as it is in the example above. On the other hand, if the verb is in an inflected form, before noting the symbol of the verb, a symbol of the personal pronoun

describing the person (1st person, 2nd person or 3th person), the number (singular or plural) and the gender (masculine, feminine or neuter) of the inflected form of the verb, must be noted first.

List of the personal pronouns for the Macedonian, the French and the English along with the symbols for noting these personal pronouns can be found in Table 8: Personal Pronouns.

TABLE 8: PERSONAL PRONOUNS

Personal Pronouns	MK	EN	FR	PoS	SY
1s	јас	I	je	pp.	
2s	ти	you	tu	pp.	
3sm	тој	he	il	pp.	
3sf	таа	she	elle	pp.	
3sn	тоа	it	/	pp.	
1p	ние	we	nous	pp.	
2p	вие	you	vous	pp.	
3pm (=3pf*)	тие*	they*	ils	pp.	
3pf	/	/	elles	pp.	

First of all the interpreter notes the personal pronoun and then the symbol of the verb. If the verb is in the Present tense the interpreter does not need to specify that with a specific symbol. The symbols of the verbs without the specification of the infinitive form or any

other additional symbols are considered to be by default in the Present tense. On the contrary, if the verb is in the Past, the Future or the Conditional tense, its tense must be specified by an additional symbol placed after the symbol of the verb.

The noting of the verbs described above is done in view of the digitalization of the process of note-taking. This rule for indicating gender and tense exists also in the traditional note-taking. Rozan in his masterpiece "Note-taking in Consecutive Interpreting" amongst the seven principles of note-taking proposes this rule:

"Rule: To indicate gender and number we add ^{e} or ^{s} to the symbol or abbreviation. To indicate tense we add ^{ll} for the future and ^{d} for the past**.** (Rozan, 1956)"

As it is describet in his book, these symbols are language related. The 'e' represents the French feminine ending, the ^{ll} comes from the English, etc. According to Rozan any letter can be used and this will depend on the languages involved.

Therefore, for this conceptual design we chose universal symbols for specifying the tense, which can be observed in Table 9: Tense.

TABLE 9: TENSE

Present	Past	Future	Conditional
no need for extra symbol besides the symbol of the verb	Symbol + ←	Symbol + →	Symbol + 

Chapter 5

5. Computational Aid for Consecutive Interpreting

5.1. Introduction

In this chapter an overview of a conceptual approach for the semi-automation of the consecutive interpreting will be elaborated. The main focus is to find a possibility to automatize the process of note-taking in consecutive interpreting, in the view of making the task of the interpreter easier.

As we saw in the previous chapters the consecutive interpreting is quite a challenging task even for humans. Seeking a possible way of facilitating this task we made an extended research on the existing computer assisted translation tools and the computer assisted translation tools for interpreters. As it was revealed that some computer assisted translation tools are used for the simultaneous interpreting, none was available for the consecutive interpreting.

Taking into consideration that the consecutive interpreting, being a serious and demanding technique, is preferable to be used in some circumstances over the simultaneous interpreting, we thought it worthwhile to try and conceive a conceptual design of a computer assisted tool for facilitating the task of the consecutive interpreting. The key technique for doing the consecutive interpreting, taking into account that the speech that is to be interpreted can be up to 6 minutes long, is the note-taking. Without a well-developed technique of note-taking the consecutive interpreting is most probably not going to be done in a correct way.

For that reason, we focused our attention on the possibility of facilitating the process of note-taking itself.

Unfortunately, we discovered that no computer assisted aid is available for this process. So, we made an overview of the state of the art in some similar domains, software that is not conceived for the interpreting in particular but that is based on the oral expression and

therefore similar in its approach. We wanted to see if it is possible to implement some computational aid originally conceived for dictating and for voice control[41]. Therefore, an overview of the state of the art in the domain of the speech technologies, the speech recognition and the speech recognition software was shown. Moreover, our attention was particularly concentrated on the development of speech technologies for Macedonian because speech technologies are language-dependent and this research is made for the interpreting from Macedonian into French and Macedonian into English.

As described in chapter 'State of the Art', these types of speech recognition software were never used for the consecutive interpreting in order to make automatic transcription of the oral message of the speaker into written text that could be used by the interpreter to help his memory. In their current state of the art, they are still not very accurate while dealing with spontaneous natural language.

Therefore, our research was shifted towards conceiving a possible tool or platform for semi-automation of the process of note-taking in consecutive interpreting, given that complete automation cannot be performed so far.

5.2. MULTILINGUAL PLATFORM

The traditional note-taking in consecutive interpreting is made on a regular note-pad preferably with a pen rather than with pencil. Discovering the advantages of using symbols and signs for the traditional note-taking, this advantages being universality, rapidity and representation of ideas rather than words, it was considered that these advantages could significantly contribute if symbols and sings are also used in a digitalized way in view of a possible semi-automation of the note-taking.

[41] http://www.nuance.com/

Given that the symbols and the sings in most cases represent either usually an equivalent of one word or same morphological feature, we can draw the conclusion that it would be faster to write down one symbol than to write the entire word.

Therefore, our conceptual proposal for semi-automation of the process of note-taking in consecutive interpreting consists of input of set of symbols used by interpreters into computer software and generation of text by the software based on the symbols that were entered.

5.2.1. INPUT

There is a wide variety of symbols used by interprets. Some of them can be found on a regular keyboard, but some of them are very particular because they are traditionally written or drawn by hand. Thus, in order to input the symbols into a computer a regular keyboard would not suffice. For the input of the symbols and the signs into a computer we propose some conceptual possibilities described further on in this chapter. It is very important to clarify that all the possible solutions that will be mentioned here, are just conceptual, none of them was implemented nor tested due to limited time and resources.

DIGITAL WRITING DEVICE

Possible solution for the input of the symbols into the system could be the use of note-taking software. On the market there is a wide variety of note-taking software with different characteristics[42]. Some of them include handwriting recognition, drawing images

[42] http://en.wikipedia.org/wiki/Comparison_of_notetaking_software

and synchronization with audio recording[43]. Furthermore, there are applications that allow conversion of handwritten notes uploaded from any digital writing device[44].

For example *Lifetrons note writer*[45], it is a tool that allows the user to take notes on a regular piece of paper with a specially designed pen. Afterwards, the software detects the hand-writing and transforms it to digitalized text, as if it was typed on a computer. This tool also allows drawing images on paper by hand, which can be used on a computer later on.

The advantage of this method is that if the notes of the interpreter mainly consist of hand-written words they would be transcribed automatically into digitalized text. Still, special software for the transcription of the hand-written symbols into digitalized text is to be designed, otherwise the use of this tool would not be possible.

GRAPHICS TABLET

Graphics tablets also known as *pen tablets* or *digitizers*[46] can also provide a possible input solution for the input of the symbols. Graphics tablets are input devices that are used for drawing and hand writing in a similar way as doing it on paper with a pen. The users can hand-draw images and graphics directly on the graphics tablets. The pen tablets consist of a flat surface for the input and a stylus for writing.

---

[43] http://luminantsoftware.com/iphone/audionote.html

[44] http://www.visionobjects.com/en/myscript/note-taking-and-forms-applications/

[45] http://www.lifetrons.ch/?lid=2#!36

[46] http://www.wacom.eu/index2.asp?lang=fr

TABLET COMPUTER

Possible solution for the input of the symbols into the system is having a specially made *symbol keyboard*. This keyboard should be used together with the regular keyboard, it does not exclude the regular keyboard. The reason for this is that we cannot only use symbols to generate text. Regular keyboard would be necessary for the input of numbers, proper names, toponyms, very short words consisting of one or two letters, etc.

This keyboard could be designed for a *touchscreen tablet computer* as a virtual keyboard. The virtual keyboard should consist of symbols for the consecutive interpreting. The virtual symbol keyboard should appear on the screen when needed and the input will be done by typing the symbols on the touchscreen keyboard of the tablet by the interpreter.

While listening to the original speech, the interpreter instead of writing and drawing the symbols by hand could type the symbols on the special symbol keyboard. The probable advantages of this method would be that it is probably faster to type one key on a keyboard than to draw a symbol. Secondly, the symbols tend to be forgotten if not used regularly. Thus, the clear view of all the symbols on the keyboard could help the interpreter's memory rather than having to know all the symbols by heart. On the other hand, there could be difficulties in finding the symbols on the keyboard. The interpreter would have to undertake training in order to learn how to use the symbol keyboard and to learn the position of the symbols on the keyboard.

The digital writing devices and graphics tablets offer a different way of input of the symbols if compared with the symbol keyboard. On one hand, the method of using note-taking software is much closer to the traditional note-taking because the symbols are hand-written. It could be that interpreters would need much less time in adjusting to this new technique of input because it is very similar to the technique they have used so far. On the other hand, probably there would not be any difference in terms of speed. It is more or less the same for the interpreter whether he would use the traditional note-taking or this tool when it comes to the input. Still, this platform would probably offer different output as opposed to the traditional note-taking so, this could be interesting to tackle upon further on in this paper.

CONCLUSION

As we mentioned before, there are several possibilities for the input of the symbols into computer. In this chapter we named some of the most likely possible solutions. Of course, there can be a lot of other modifications of the methods mentioned above for the use on different gadgets like smartphones, tablets, notebooks etc. The main idea is to have an input into computer software that can be processed by the computer and then possibly transcribed into plain text.

We must point out that all of the possible solutions of input mentioned above cannot be used as they are in their current state. Specially designed software for the transcription of the interpreting symbols into text must be developed if they were to be used for the specific purpose of the note-taking in consecutive interpreting.

Furthermore, we must indicate once again, that to our knowledge none of the methods mentioned above was implemented or tested in a real situation of note-taking for the consecutive interpreting.

Similar problem has been treated for the pictogram-based Augmentative and Alternative Communication (AAC) (Pahisa-Solé, 2012). Compansion system that expands telegraphic language into natural language sentences has been built for Catalan. The evaluation of the system has shown very high results: 98,7% of the generated sentences were considered correct.

5.2.2. OUTPUT

Having in mind that with today's technology we are not able to transcribe interpreting symbols into meaningful text it is obvious that software must be constructed in order to complete the computational aid for consecutive interpreting. In theory, the software that needs to be constructed should consist of:

- Symbol database
- Linguistic rules
- Linguistic inference engine

The symbol database consists of parallel dictionary containing the translation of the words in natural language, their part-of-speech tags and the symbol that describes the idea behind those words as described in Table 7: Dictionary.

Once the construction of the symbol database is finished, the linguistic rules need to be formulated. Consecutive interpreter's expert experience is instrumental in the process of determining the linguistic rules. Using the expert knowledge and taking into account the particularities of the natural language a set of linguistic rules should be constructed.

Finally, the linguistic inference engine needs to be built. This is a logical computer code that reads the input, compares it to the symbol database and using the linguistic rules determines the output text.

In short the flow of the program would be as follows:

After the input of the symbols in the system, the symbols are saved in an array of variables. These variables are processed possibly by a linguistic inference engine operating on a set of linguistic rules. The set of linguistic rules is made using the expert knowledge of the language in question. Using the linguistic rules the inference engine takes into account the arrays of variables and transforms them into meaningful text. Finally, the generated text appears on the screen so the user can see it. The graphic user interface of the program then allows the user to save the displayed text into a .doc or .pdf file by a simple click of a button. Figure 10: Program Flow Diagram

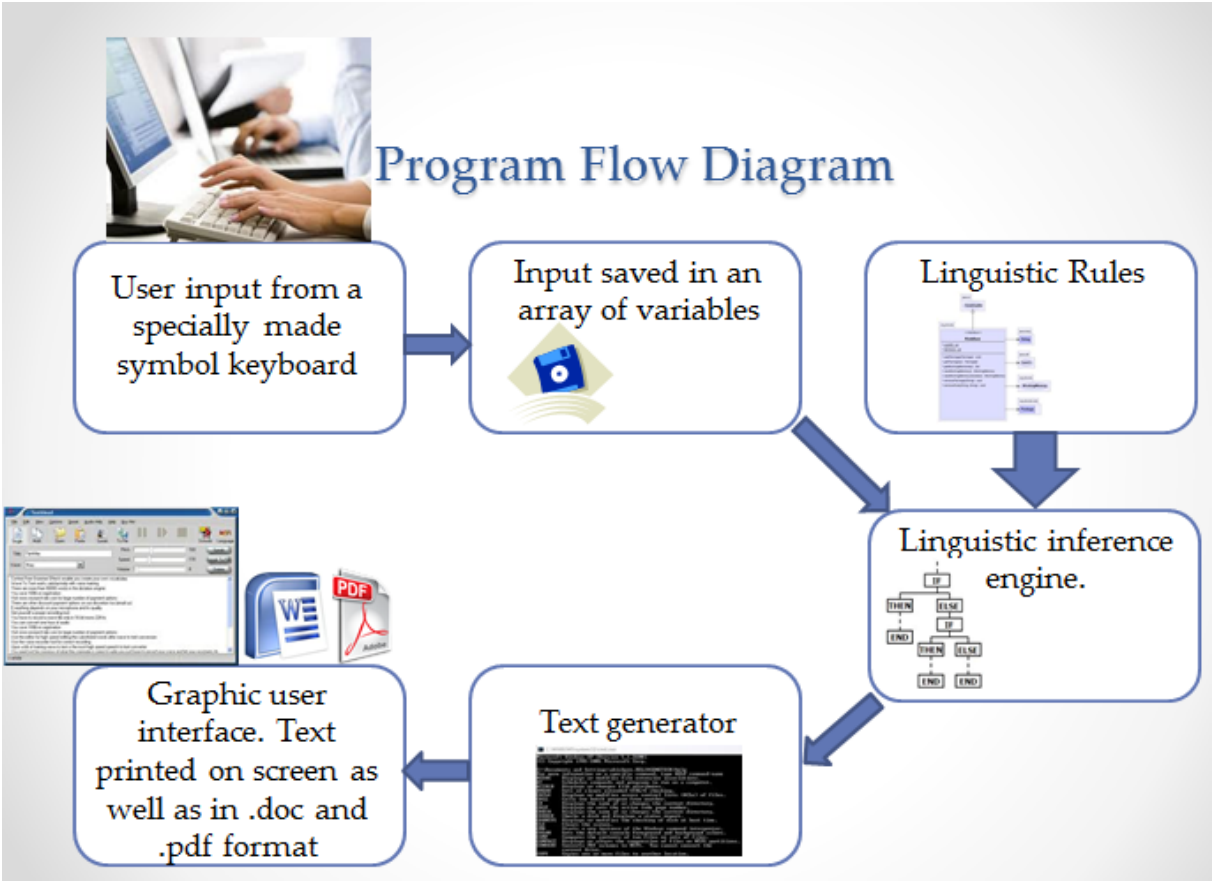


FIGURE 10: PROGRAM FLOW DIAGRAM

Probably the main advantage of this platform would be the fact that the notes of the interpreter that consist of many symbols and signs and some text would be transcribed completely into text. The output of the system would be plain text.

This is useful because one of the main problems of interpreters in the beginning of their career and for some also further on, is the problem of deciphering or “reading” the notes at the speed appropriate for a well-trained consecutive interpreter, which means only glancing once over the sentence and not losing too much time trying to deal with the notes. Thus, if the interpreter is confronted with a clearly intelligible typed text it would be probably easier for him to grasp the thought rather than if he is confronted with his notes scrabbled in haste.

Another possibility, in particular situations where appropriate, the output of the system could be transmitted on a screen visible to the audience. The translation of the original

speech could appear as a type of subtitle of what is being uttered by the speaker. Of course, the delay would be significant, given the fact that the consecutive interpreting is done after the speaker has finished the speech. In this case, the translation on the screen can appear only after the entire sentence, from the beginning until the end, has been uttered.

Chapter 6

6. CONCLUSIONS

As we have seen so far, interpreting is a profession that demands promptness and accuracy in the shortest delay possible. Furthermore, the consecutive interpreting as a specific type of interpreting demands an exquisite faculty of memorizing long speech in the correct order. It is very challenging for the beginners in this profession to start grasping the technique and to perform complete quality interpretation.

The overview of the State of the Art in this field has shown that some computer assisted translation tools are available for the simultaneous interpreting. Unfortunately, not much is available for the consecutive interpreting.

Moreover, the overview of the State of the Art in Speech Technologies and Automatic Speech Recognition has shown that none of the available Speech Recognition Software would be able to tackle the task of automatically transcribing spontaneous oral speech. Their performance, when it comes to dictation, under specific conditions and voice enrollment of one user, is remarkable, with up to 99%[47] accuracy. Significant progress is done in this domain, but, still when it comes to Large-Vocabulary Continuous Speech Recognition (LVCSR) this systems are not able to treat the natural language.

Therefore, in our knowledge, no affiliation between the Automatic Speech Recognition and the consecutive interpreting has been done so far.

When it comes to the development of the Speech Technologies for Macedonian some progress in this domain has been done. Activities for construction and development of oral corpora in Macedonian are currently undertaken. Studies about the segmentation of the

[47] http://marcnorris.hubpages.com/hub/Dragon-Naturally-Speaking-Review

audio recordings have been done. Subsystem for Real-Time Text-to-Speech Conversion (Josifovski, Mihajlov, Gorgevik, & Loskovska, 1996) and a Hybrid Isolated-Word Speech Recognition System for Macedonian have been developed in Republic of Macedonia (Kraljevski, Mihajlov, & Gorgjevik, 2000). But still, no functional ASR software is available for Macedonian (Kraljevski, Chungurski, Mihajlov, & Arsenovski, 2008).

Basing this research on extracting information from oral corpus has been fruitful for the task determined in the beginning of this research. This task was discovering the most frequent words in the spontaneous oral speech from the sociopolitical domain in Macedonian.

Oral corpus consisting of 44 recordings, with approximately comparable length of about 10 min was built. These recordings were selected amongst the available sources to the general public, according to the topic discussed, the main domain being the political sociology.

The total length of the corpus being 7hour 8minutes and 41seconds, it consists of 64 155 words, out of which there are 8 930 different words. The recordings were manually transcribed with the help of the transcription software Sonal. Once the information about the frequency of different words in the entire corpus was obtained, we proceeded to the task of lemmatization.

Brief bilingual survey (in Macedonian and in English) consisting of 56 lemmas of the most frequent words and 14 symbols taken form an electronic source available for interpreters, was conducted. Five professional interpreters with experience in consecutive interpreting provided the symbols that they use for the given words.

Having obtained this data, proceeding with the construction of parallel dictionaries along with interpreting symbols was made possible. These dictionaries could be used for creating computer assisted tool for semi-automation of the note-taking in consecutive interpreting. Conceptual design proposal for multilingual platform that deals with this task was elaborated in Chapter 5 Computational Aid for Consecutive Interpreting.

BIBLIOGRAPHY

(n.d.). Retrieved January 2012, from Cattools.org: http://www.cattools.org/

Adda-Decker, M. (2007, January). Corpus pour la transcription automatique de l'oral . *Revue française de linguistique appliquée (Vol. XII)*, pp. 71-84.

Brugnara, F., & De Mori, R. (1996). HMM Methods in Speech Recognition. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology*, (pp. 20-28).

Cardey-Greenfield, S. (1987). *Traitement algorithmique de la grammaire normative du français pour une utilisation automatique et didactique.* Besancon: Thèse d'État de Lettres.

Chungurski, S., Kraljevski, I., Mihajlov, D., & Arsenovski, S. (2008). Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language. *ITI 2008 30th Int. Conf. on Information Technology Interfaces.* Cavtat, Croatia.

Coulmas, F. (1989). *The Writing Systems of the World.* Oxford UK & Cambridge USA: Blackwell.

Delić, V., Sečujski, M., Jakovljević, N., & Janev, M. (2010). Speech Technologies for Serbian and Kindred South Slavic Languages. In N. Shabtai (Ed.), *Advances in Speech Recognition.* Sciyo.

Delić, V., Sečujski, M., Pekar, D., & Jakovljević, N. (2006). A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian. *IS–LTC*, (pp. 257–260). Ljubljana, Slovenia.

Herbert, J. (1952). *Manuel de l'interprète, comment on devient interprète de conférences.* Genève: Librarie de l'Université.

Hirschman, L., & Thompson, H. (1996). Overview of Evaluation in Speech and Natural Language Processing. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology.*

Hoffmeister, B. (2012). WFST Enabled Solutions to ASR Problems: Beyond HMM Decoding. 551 - 564 .

Josifovski, L., Mihajlov, D., Gorgevik, D., & Loskovska, S. (1996). Text-to-speech conversion for Macedonian as part of a system for support of humans with damaged sight. *18th Int. Conf. Information Technology Interfaces ITI'96.* Pula, Croatia.

Kraljevski, I., Chungurski, S., Mihajlov, D., & Arsenovski, S. (2008). Semi-Automatic Speech Segmentation for Macedonian Language. *ITI 2008, 30th International Conference on Information Technology Interfaces.* Dubrovnik, Croatia.

Kraljevski, I., Mihajlov, D., & Gorgjevik, D. (2000). Hybrid HMM/ANN system for speech recognition of Macedonian language. *Fifth National Conference With International Participation ETAI'2000,* (pp. I1-1 - I1-6). Ohrid, Republic of MACEDONIA.

Lederer, M. (1980). *La traduction simultanée, fondaments theoriques.* Lille: Service de reproduction des theses, Université de Lille III.

Luzzati, D. (1983). *Recherches sur la structure du discours oral spontané.* Paris: 3' cycle, Paris III, sous la direction de J. Pinchon.

Pahisa-Solé, J. (2012). *Compansion System for a Pictogram-based AAC application in Catalan.*

Pallett, D. S., & Fourcin, A. (1996). Speech Input: Assessment and Evaluation. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology,* (pp. 400-404).

Pekar, D., Mišković, D., Knežević, D., & Sedlar, N. V. (2010). Applications of Speech Technologies in Western Balkan Countries. In N. Shabtai (Ed.), *Advances in Speech Recognition.*

Rozan, J.-F. (1956). *La prise de notes en interprétation consécutive.* Genève : Georg.

Sand. (2003). Retrieved 2012, from AIIC: http://aiic.net/ViewPage.cfm/article1028.htm

Seleskovitch, D. (1968). *L'interprète dans les conférences internationales.* Paris: Lettres modernes.

Seleskovitch, D. (1975). *Langage, langues et mémoire étude de la prise des notes en interprétation consécutive.* Paris: Lettres modernes.

Stern, R. M. (1996). Robust Speech Recognition. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology*, (pp. 14-19).

Uszkoreit, H. (1996). Mathematical Methods. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology*, (pp. 315-332).

Zue, V., Cole, R., & Ward, W. (1996). Speech Recognition. In R. A. Cole (Ed.), *Survey of the State of the Art in Human Language Technology*, (pp. 3-9).

ANNEXES

APPENDIX A: TRANSCRIPTIONS OF THE ORAL RECORDINGS

Here is a small sample of the transcriptions of the oral recordings.

==========

Recording exported by Sonal (v.1.4) on the 16/11/2012 at 01:23:08

==========

Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del6.wav

4 extract(s)

****

Characteristics:

Speaker: Politician

Type of recording: Interview

TV channel: MTV1

****

Observations:

****

00:00 > 00:00 [Relations between Macedonia and EU]

[>Aleksandar_Spasenovski]: 0,1] Republika Makedonija so zgolemuvanjeto na diplomatskata i konzularnata mreza, so profesionalizacijata na diplomatite vo Republika Makedonija i so voveduvanjeto na noviot kadar mene mi dava za pravo da zaklucam deka toa ke dovede i do odredeni kvalitativni uspesi, no sepak treba da bideme, a, da bideme iskreni, da kazeme deka Republika Makedonija sepak e mala drzava. Deka postojat, deka sproti sebe ima drzavi so pogolema drzavotvorna tradicija i so mnogu porazviena diplomatska mreza, so pogolemi kapaciteti, pobogati drzavi i taka natamu no, sepak nie imame pak druga komparativna prednost a toa e vistinata kojasto vo delot osobeno na preku razgovorite okolu razlikite okolu imeto Republika Makedonija na nasa strana i toa pak e od ogromna korist bidejki mnogu lesno mozete lugeto da gi ubedite vo odnos na prasanje koesto e prilicno prirodno, a toa e deka sekoj ima pravo da se imenuva taka kako sto smeta deka treba da se imenuva i taka kako sto toj samiot saka da se imenuva. I vo taa smisla veruvam deka Republika Makedonija uvazuvajki gi realnostite na megunarodnata politika, imajki go predvid faktot deka e drzava kojasto saka cas poskoro da se integrira vo Evropskata Unija, no isto taka, imajki gi predvid svoite legitimni interesi vo odnos na sopstvenoto ustavno ime deka ke se izbori da najde eden soodveten nacin kojsto ke znaci zadovoluvanje na site prioriteti, a toa e zacuvuvanje na imeto no istovremeno integracija vo Evropskata Unija.

01:24 > 01:24 [Council of Europe]

[>Journaliste]: Ona sto e za pofalba e sto Makedonija vo izminatite barem godini uspeva da gi iskoristi site formi, toa e i makedonskoto pretsedavanje so Sovetot na Evropa na pozitiven nacin da se pretstavi sebesi a, ne da se nametne kako zemja kojasto samo pravi konflikti nekade, tuku eve se obiduvame i se nametnuvame so dobro kade sto mozeme ostavajki gi nastrana problemite kojsto gi imame bilateralni, kojsto sepak neli bolat, megutoa toa e na politikata da go resava. Dojdovme do krajot na emisijata, moram licno da ti postavam edno prasanje, posle sest meseci sto ocekuvas da e najgolemata pridobivka od makedonskoto pretsedatelstvo?

[>Aleksandar_Spasenovski]: Jas iskreno veruvam deka Republika Makedonija vo ramkite na ovie sest meseci dodeka pretsedava so Sovetot na Evropa ke uspee da ostvari dve celi. Prvata cel e na dostoinstven nacin da go vodi Sovetot na Evropa, da rakovodi so ovaa institucija i vtoro da uspee da se nametne preku takvata profesionalnost i preku takviot pristap i preku pokazuvanjeto na vistinskite kapaciteti kojsto gi ima ovaa drzava deka ke uspee isto taka da gi, vo pogolema mera da gi osvoi simpatiite na evropskata javnost, no isto taka i na evropskite politicari kako drzava kojasto navistina uspeva soodvetno i ramo do ramo da se nosi so evropskite problemi, so tezinata na evropskite zadaci kako i golemite evropski drzavi kojsto vo minatoto pretsedavaa so Sovetot na Evropa. Nie treba da bideme svesni deka Republika Makedonija vakva sansa ke dobie posle dvaeset i cetiri godini, najrano posle dvaeset i cetiri godini, imajki predvid faktot deka pretsedatelstvuvanjata traat po sest meseci, a Sovetot na Evropa ima cetirieset i sedum drzavi clenki tie se rotiracki, jas veruvam deka, veruvam vo serioznosta na pristapot na instituciite na Republika Makedonija. Veruvam vo podgotvenosta i vo zelbata na makedonskite politicki faktori kojsto neposredno ke rakovodat so ovoj proces na pretsedatelstvuvanje. Veruvam vo kapacitetite na makedonskoto Ministerstvo za nadvoresni raboti i na krajot na kraistata ke vidime i vremeto toa ke go pokaze no jas iskreno veruvam deka Republika Makedonija ke uspee da ostvari dve celi a toa e da se prikaze kako dostoinstven partner i seriozna drzava na megunarodnata politicka scena a takvata pretstava kojasto ke ja osvoime ke ni ovozmozi da ostvarime drugi nadvoresno politicki celi, a toa e pobrzata integracija vo Evropskata Unija.

03:40 > 03:40 [Farewells]

[>Journaliste]: Ti balgodaram mnogu za gostuvanjeto vecerva vo emisijata "Bez pardon"

[>Aleksandar_Spasenovski]: Blagodaram.|

[>Journaliste]: Ke bideme seste meseci prisutni i nie ke sledime kako ke rabotite. Da se nadevame deka ke gi ostvarime ovie dve celi i mnogu poveke imame sansa treba da ja iskoristime.Vam dragi gledaci vi blagodram sto bevte eden cas so emisijata "Bez pardon". Bidete povtorno na ovaa frekfencija i idniot vtornik ke imame povtorno interesna i aktuelna tema. Prijatno.

04:04 > 04:04 [Music]

APPENDIX B: WORD COUNT

This is a sample of the most frequent words in the oral corpus.

Word count edited by Sonal1.4.11				
Date :11/03/2013				
Statistics taken from the corpus : Govori.Crp				
Word	Total occurrences	Occurrences in the sub-population	Percentage of occurrences	Probability (Lafon)
na	2768	0	0	0,00%
i	2156	0	0	0,00%
da	2076	0	0	0,00%
se	1707	0	0	0,00%
vo	1618	0	0	0,00%
e	1356	0	0	0,00%
ne	1104	0	0	0,00%
sto	1054	0	0	0,00%
za	1016	0	0	0,00%
deka	909	0	0	0,00%
od	873	0	0	0,00%
toa	866	0	0	0,00%
so	697	0	0	0,00%
go	661	0	0	0,00%
ke	654	0	0	0,00%
kako	493	0	0	0,00%
[>journaliste]	460	0	0	0,00%
makedonija	412	0	0	0,00%
gi	407	0	0	0,00%
znaci	327	0	0	0,00%
taka	315	0	0	0,00%
ja	304	0	0	0,00%
nie	304	0	0	0,00%
ili	285	0	0	0,00%
kojsto	282	0	0	0,00%
a	274	0	0	0,00%
jas	265	0	0	0,00%
ima	256	0	0	0,00%
republika	251	0	0	0,00%

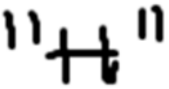
site	221	0	0	0,00%
koj	197	0	0	0,00%
no	192	0	0	0,00%
do	183	0	0	0,00%
mnogu	181	0	0	0,00%
edna	181	0	0	0,00%
eden	179	0	0	0,00%
pa	174	0	0	0,00%
bi	170	0	0	0,00%
treba	170	0	0	0,00%
dali	170	0	0	0,00%
ako	167	0	0	0,00%
toj	161	0	0	0,00%
evropskata	160	0	0	0,00%
samo	159	0	0	0,00%
moze	156	0	0	0,00%
tie	154	0	0	0,00%
bese	154	0	0	0,00%
ova	152	0	0	0,00%
imame	150	0	0	0,00%
sega	143	0	0	0,00%
mislam	142	0	0	0,00%
ona	142	0	0	0,00%
sme	136	0	0	0,00%
taa	136	0	0	0,00%
[>katerina_canevska-arsovska]	133	0	0	0,00%
unija	132	0	0	0,00%
nema	128	0	0	0,00%
po	122	0	0	0,00%
megutoa	120	0	0	0,00%
koga	119	0	0	0,00%
bide	116	0	0	0,00%
del	116	0	0	0,00%
dobro	114	0	0	0,00%
ni	113	0	0	0,00%
evropa	112	0	0	0,00%
godina	109	0	0	0,00%
ovoj	108	0	0	0,00%
uste	108	0	0	0,00%
kojasto	108	0	0	0,00%

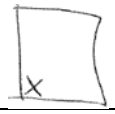
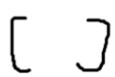
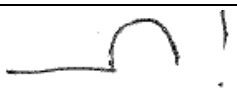
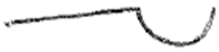
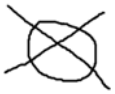
APPENDIX C: DATABASE

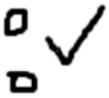
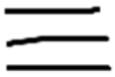
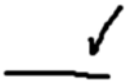
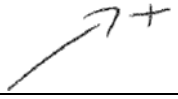
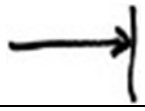
Recording	Date added	Length	Speaker	Type of recording	TV channel
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del1.wav	27/04/2012 02:11	07:09	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del2.wav	27/04/2012 02:12	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del3.wav	27/04/2012 02:13	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del4.wav	27/04/2012 02:13	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del5.wav	27/04/2012 02:14	10:00	Politician	Interview	MTV1
Aleksandar_Spasenovski_vo_„Bez_pardon__26_04_2010_del6.wav	27/04/2012 02:15	04:34	Politician	Interview	MTV1
Aleksandar_Spasenovski_za_Evropski_klub__11_10_09_del1.wav	27/04/2012 02:16	07:20	Politician	Interview	MTV1
Aleksandar_Spasenovski_za_Evropski_klub__11_10_09_del2.wav	27/04/2012 02:17	06:27	Politician	Interview	MTV1
Antonio_Milososki_gostin_vo_vestite_na_Sitel__2mart_2011_.wav	26/04/2012 21:54	08:15	Minister	Guest in the News	Sitel
Branko_Crvenkovski_intervju_za_Radio_Slobodna_Evropa.wav	26/04/2012 21:55	10:08	Politician	Interview	Radio Slobodna Evropa
Dragan_Pavlovic_Latas_vs__Borjan_Jovanovski.wav	27/04/2012 21:10	09:13	Journalist	Debate	MTV1
Gruevski_gostin_vo_„NIE__25_04_2010__del1.wav	27/04/2012 22:01	10:10	Prime Minister	Interview	Sitel
Gruevski_gostin_vo_„NIE__25_04_2010__del2.wav	27/04/2012 22:01	10:10	Prime Minister	Interview	Sitel
Gruevski_gostin_vo_„NIE__25_04_2010__del3.wav	27/04/2012 22:01	10:10	Prime Minister	Interview	Sitel
Gruevski_gostin_vo_„NIE__25_04_2010__del4.wav	27/04/2012 22:01	10:10	Prime Minister	Interview	Sitel
Gruevski_gostin_vo_„NIE__25_04_2010__del5.wav	27/04/2012 22:02	10:01	Prime Minister	Interview	Sitel
Gruevski_gostin_vo_„NIE__25_04_2010__del6.wav	27/04/2012 22:02	09:50	Prime Minister	Interview	Sitel
Ilija_Dimovski_gostin_vo_Faktor5__28_04_2010__del3.wav	26/04/2012 21:58	10:00	Politician	Debate	Kanal 5
Jankulovska_za_viznata_liberalizacija__15_07_09.wav	26/04/2012 21:59	08:49	Minister	Guest in the News	Sitel
Lubco_Georgievski_-_28_05_2010_Vo_centar_2_6.wav	27/04/2012 21:10	10:56	Politician	Interview	Kanal 5
Makedonija_megu_osudata_na_komunismot__1.wav	27/04/2012 21:55	09:55	Professor	Debate	A1
Makedonija_megu_osudata_na_komunismot__2.wav	27/04/2012 21:56	08:49	Professor	Debate	A1
Makedonija_megu_osudata_na_komunismot__3.wav	27/04/2012 21:56	07:14	Professor	Debate	A1

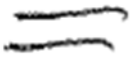
Makedonija_megu_osudata_na_komunismot__4.wav	27/04/2012 21:57	07:50	Professor	Debate	A1
Makedonija_megu_osudata_na_komunismot__5.wav	27/04/2012 21:57	09:42	Professor	Debate	A1
Makedonskata_Istorija_vo_EU_-_Tret_del.wav	27/04/2012 21:09	13:49	Academician	Telephone call	Moris
Milososki__intervju_za_Kanal_5__30_07_09.wav	26/04/2012 22:01	08:43	Minister	Guest in the News	Kanal 5
Milososki__intervju_za_Sitel__07_10_09.wav	26/04/2012 22:03	04:28	Minister	Guest in the News	Sitel
Milososki_po_povod_viznata_liberalizacija__16_07_09.wav	26/04/2012 22:00	08:57	Minister	Guest in the News	Sitel
Miren_protest_protiv_„evropskiot_genocid.wav	26/04/2012 22:04	03:21	Politician	Interview	Nasa TV
Nikola_Todorov_za_Radio_Slobodna_Evropa__06_02_2010.wav	27/04/2012 02:01	10:39	Minister	Interview	Radio Slobodna Evropa
Stavreski__intervju_za_Kanal_5__27_08_09.wav	27/04/2012 02:03	08:26	Minister	Guest in the News	Kanal 5
TRN_VO_OKO_SO_MILENKO_I_CHOMOVSKI_28_Sep_2011_1_del_mpg.wav	26/04/2012 21:53	11:57	Journalist	Duel	AB
11Novinarstvoto_vo_Makedonija_.wav	10/11/2012 18:27	13:20	Journalist	Debate	MTV1
22Novinarstvoto_vo_Makedonija_.wav	10/11/2012 18:27	13:20	Journalist	Debate	MTV1
33Novinarstvoto_vo_Makedonija_.wav	10/11/2012 18:27	11:50	Journalist	Debate	MTV1
1„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	11:40	Professor	Debate	MTV1
2„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	11:40	Professor	Debate	MTV1
3„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	11:40	Professor	Debate	MTV1
4„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	11:40	Professor	Debate	MTV1
55„Bez_pardon_Evroskepticizam_25_10_2011_MTV.wav	10/11/2012 18:49	13:36	Professor	Debate	MTV1
1Pretsedatelot_Crvenkovski_vo_„Presing.wav	10/11/2012 20:09	10:50	Politician	Interview	Alfa
2Pretsedatelot_Crvenkovski_vo_„Presing.wav	10/11/2012 20:09	10:50	Politician	Interview	Alfa
3Pretsedatelot_Crvenkovski_vo_„Presing.wav	10/11/2012 20:09	11:02	Politician	Interview	Alfa

APPENDIX D: SYMBOL DICTIONARY

MK	EN	FR	PoS	SY
дел	part	part	n.	
држава	country	état	n.	
влада	government	gouvernement	n.	
однос	with regard to = in relation to	en ce qui concerne = par rapport à	n.	
јазик	language	langue	n.	
време	time	temps	n.	
страна	side	côté = parti	n.	
начин	manner	manière	n.	
работа	work	travail	n.	
народ	people	peuple	n.	
име	name	nom	n.	
Собрание	Assembly	Assemblée	n.	
ден	day	jour	n.	

избори	elections	élection	n.	
период	period	période	n.	
новинарство	journalism	journalisme	n.	
политика	politics	politique	n.	
проблем	problem	problème	n.	
ситуација	situation	situation	n.	
бара	demand	demander	v.	
дава	give	donner	v.	
доаѓа	come	venir	v.	
влезе	enter	entrer	v.	
оди	go	aller	v.	
гледа = види	watch = see	regarder = voir	v.	
има	have	avoir	v.	
нема	lack	manquer	v.	
каже = рече = вели = вика = зборува =~ разговара =~ предложи	say = speak = tell = state =~ discuss =~ propose	dire = parler = déclarer = affirmer =~ discuter = traiter =~ proposer	v.	

мисли = смета =~ реши	think = reckon =~ decide	penser = estimer =~décider	v. 
разбира =~ верува	understand =~ believe	comprendre =~ croire	v. 
знае	know	savoir	v. 
мора	must	falloir	v. 
треба	should	devoir	v.
може	can	pouvoir	v. 
надева =~ очекува	hope =~ expect	espérer =~ s'attendre à	v. 
сака	want	vouloir	v. 
носи	bring	apporter	v. 
однесува	take	emporter = emmener	v. 
врати	return	retourner = revenir	v. 
подобри	improve	améliorer	v. 
успее	succeed	réussir	v. 
добие	get	obtenir	v. 
работи =~ води	work =~ lead	travailler = diriger	v. 

случи	happen	arriver = se passer	v. 
сум =~ станува =~ постои =~ претставува	be =~ become =~ exist =~ represent	être = devenir = exister = représenter	v. 
значи	mean	vouloir dire = signifier	v. 

APPENDIX E: ROMANIZATION OF MACEDONIAN

This is the transliteration code used in the Oral Corpora, to transcribe the Macedonian oral recordings into Latin script.

Comparative table of the Romanization of the Macedonian letters							
Cyrillic	**IPA**	**ISO 9 (1995)**	**BGN/PCGN/UN**	**ISO 9 (R:1968) + National Academy**	**ISO 9 (R:1968, b)**	**In passports**	**Latin (used in the corpus)**
А а	/a/	A a					A a
Б б	/b/	B b					B b
В в	/v/	V v					V v
Г г	/ɡ/	G g					G g
Д д	/d/	D d					D d
Ѓ ѓ	/ʝ/	Ǵ ǵ	G/Đ g/đ	Ǵ ǵ	Ǵ ǵ	Gj gj	G g
Е е	/ɛ/	E e					E e
Ж ж	/ʒ/	Ž ž	Ž ž	Ž ž	Zh zh	Zh zh	Z z
З з	/z/	Z z					Z z
S s	/dz/	Ẑ ẑ	Dz dz	Dz dz	Dz dz	Dz dz	Z z
И и	/i/	I i					I i
J j	/j/	J̌ ǰ	J j	J j	J j	J j	J j
К к	/k/	K k					K k
Л л	/l/	L l					L l
Љ љ	/ʎ/	L̂ l̂	Lj lj	Lj lj	Lj lj	Lj lj	Lj lj
М м	/m/	M m					M m
Н н	/n/	N n					N n
Њ њ	/ɲ/	N̂ n̂	Nj nj	Nj nj	Nj nj	Nj nj	Nj nj
О о	/ɔ/	O o					O o
П п	/p/	P p					P p
Р р	/r/	R r					R r
С с	/s/	S s					S s
Т т	/t/	T t					T t
Ќ ќ	/c/	Ḱ ḱ	K/Ć k/ć	Ḱ ḱ	Ḱ ḱ	Kj kj	K k
У у	/u/	U u					U u
Ф ф	/f/	F f					F f
Х х	/h/	H h	H h	H h	Kh kh	H h	H h
Ц ц	/ts/	C c	C c	C c	Ts ts	C c	C c
Ч ч	/ʧ/	Č č	Č č	Č č	Ch ch	Ch ch	C c
Џ џ	/dʒ/	D̂ d̂	Dž dž	Dž dž	Dž dž	Dj dj	Dz dz
Ш ш	/ʃ/	Š š	Š š	Š š	Sh sh	Sh sh	S s

APPENDIX F: SURVEY

Here is an extract of the survey in Macedonian and in English

ПРАШАЛНИК ЗА БЕЛЕЖЕЊЕТО ПРИ КОНСЕКУТИВНОТО ТОЛКУВАЊЕ

Ќе Ве молиме да ги внесете симболите што го користите за наведените зборови кога бележите при консекутивно толкување. Со двоен клик на белите полиња непосредно под зборовите ќе се отвори нов прозорец во Paint кој овозмжува да го нацртате симболот со помош на глувчето на Вашиот компјутер. По внесувањето на симболот само затворете го прозорецот (не е потребно да го зачувате) и Вашиот симбол ќе се појави на Word документот. Доколку не сте задоволни можете со двоен клик повторно да се вратите на Paint прозорецот.

Втора опција за да ја пополните оваа анкета би била да го испечатите овој документ и своерачно да ги нацртате симболите. Потоа, да го скенирате и да го прикачите на повратен мејл.

По пополнувањето на овој прашалник, срдечно ве молиме пополнетиот прашалник да го испратите најдоцна до 25ти март 2013 год. на следната е-адреса:

r.aneta@yahoo.com

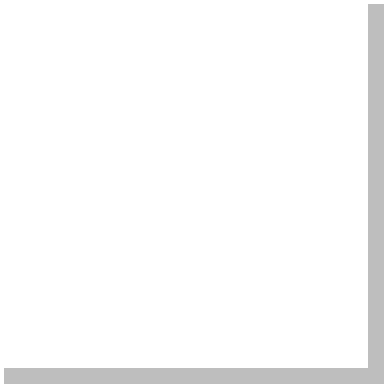
I Под дадените зборови наведени се и примери во кои зборот е употребен во контекст. Не е задолжително да се држите строго до дадениот контекст.

дел

„И мислам дека тоа е еден апсолутен израз на еден значаен дел на демократските капацитети на македонското новинарство.“

држава

„Трошењето на државата да се прилагоди на состојбите во реалниот сектор.“

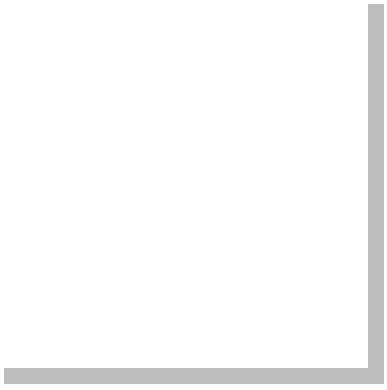


влада

„Она што максимум ние можеме да го направиме е да не ѝ отежнуваме на Владата.“

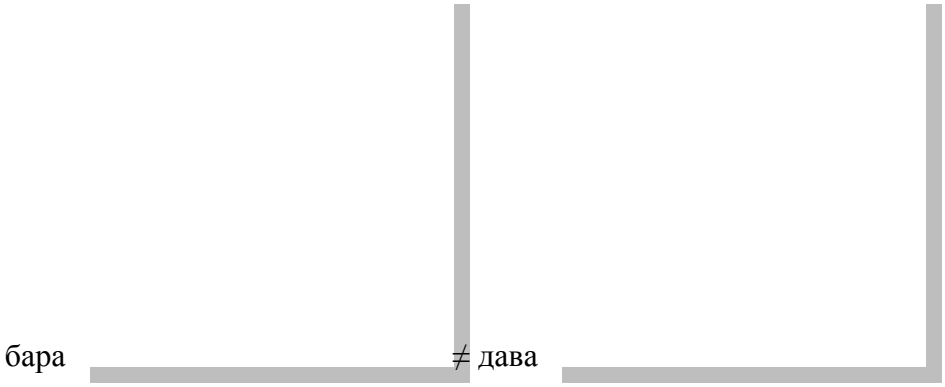
однос

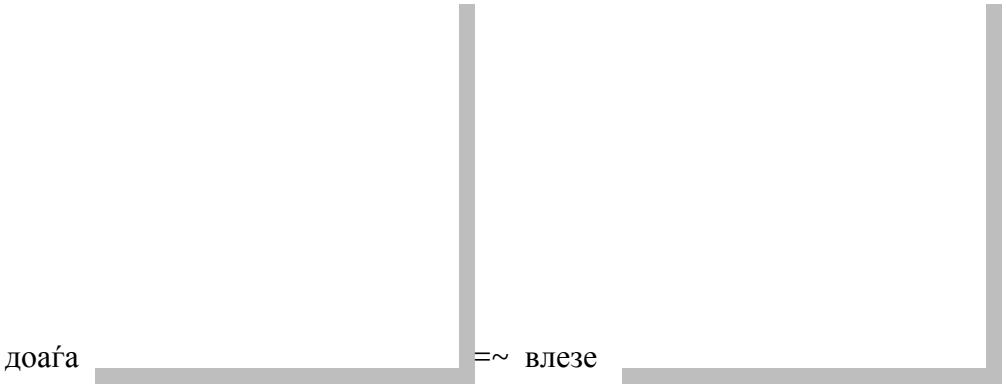
„Верувам дека тие компаративни предности во однос на приоритетите и целите на Советот на Европа можат многу да дојдат во израз во однос на претседателството на Република Македонија.“



II Дадените глаголи се групирани по одредена сличност. Доколку сметате дека глаголите од иста група може да се претстават со истиот симбол нацртајте го само

симболот што би го користеле (за еден или повеќе глаголи од таа група) и оставете ги другите полиња празни.





SURVEY REGARDING THE NOTE TAKING IN CONSECUTIVE INTERPRETING

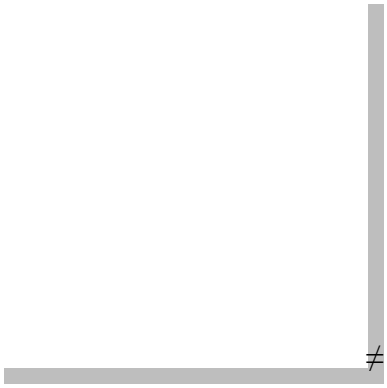
Please enter the symbols you are using for the given words while taking notes for the consecutive interpreting. Double click on the white fields and you will be redirected to a Paint window where you can draw the symbols using the mouse. Once you have entered the symbol just close the Paint window (no need to save it) and the symbol should appear in the Word document. You can reopen the Paint window, if you wish to change the symbol, by double-clicking the field again.

Another possibility to fill in this survey would be to print this document and draw the symbols by hand. Then, scan the document and send it by mail.

After filling in this survey, please send it by mail at the latest by the 26th of March 2013 on the following e-mail address:

r.aneta@yahoo.com

II The given verbs are grouped by some similarity in meaning. If you consider that the verbs from the same section can be represented by the same symbol draw only one symbol (for one or more verbs from the same section).

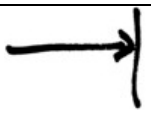
demand  ≠ give 

think = reckon _____ =~ decide _____

=~ understand _____ =~ believe _____

III If you are using some of the symbols from the list of symbols bellow please enter the words you are using them for. If not, please enter the words that would be compatible for the given symbols according to you (you can use the words mentioned in this survey above or other words).

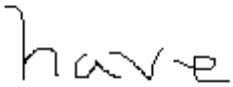
	
	
	
	
	
	
	
	
	
	

APPENDIX G: NOTES FROM THE INTERPRETERS

Here are some segments of the answers on the survey provided by the interpreters.

гледа = види

gleda ‘watch’ vidi ‘see’



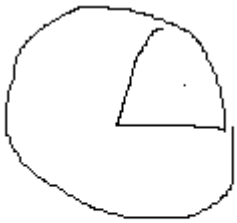
има ≠ нема

ima ‘have’ nema ‘lack’

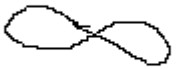
каже = рече = вели = вика = зборува

kaze ‘say’ = rece ‘say’ = veli ‘say’= vika ’say’ = zboruva ‘speak’

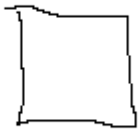
part



in relation to



country



work

government

people

APPENDIX H: TOPICS AND VARIABLES

This table shows the relation between the three variables: Speaker, Type of recording and TV-channel, and the topics in the oral corpus. Calculations are shown in seconds and in percentage.

Variable :	(1)Speaker										
Sec.	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
Prime Minister	0	57	335,9	4,65	203,51	0	621,21	0	2400,34	0	3622,61
Professor	0	489,49	358,14	15	3206,49	0	0	0	2151,43	0	6220,55
Journalist	0	36,78	84,28	52,08	0	0	0	0	3403,67	0	3576,81
Minister	0	37,66	29,44	56,07	419,88	195,75	725,17	726,83	1291,82	0	3482,62
President	0	0	0	0	0	0	0	0	0	0	
Politician	0	164,06	108,4	59,51	1283,43	919,46	1450,28	0	2044,28	1913,18	7942,6
Author	0	0	0	0	0	0	0	0	0	0	
Scientist	0	0	0	0	0	0	0	0	0	0	
Academician	0	19,63	0	23,24	20,81	0	0	0	765,08	0	828,76
Total	0	804,6199	916,16	210,55	5134,12	1115,21	2796,66	726,83	12056,62	1913,18	25673,95
% lig	Without topic	Introduction	Music	Farewells	Relations between	Integration of	Relations between	Liberalization of visa regime	Political situation in	Council of Europe	

					Macedonia and EU	Macedonia in NATO	Macedonia and Greece	with EU	Macedonia		
Prime Minister	0%	1,60%	9,30%	0,10%	5,60%	0%	17,10%	0%	66,30%	0%	100%
Professor	0%	7,90%	5,80%	0,20%	51,50%	0%	0%	0%	34,60%	0%	100%
Journalist	0%	1,00%	2,40%	1,50%	0%	0%	0%	0%	95,20%	0%	100%
Minister	0%	1,10%	0,80%	1,60%	12,10%	5,60%	20,80%	20,90%	37,10%	0%	100%
President	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Politician	0%	2,10%	1,40%	0,70%	16,20%	11,60%	18,30%	0%	25,70%	24,10%	100%
Author	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Scientist	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Academician	0%	2,40%	0%	2,80%	2,50%	0%	0%	0%	92,30%	0%	100%
Total	0,00%	3,10%	3,60%	0,80%	20,00%	4,30%	10,90%	2,80%	47,00%	7,50%	100%
% col	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
Prime Minister	0%	7,10%	36,70%	2,20%	4,00%	0%	22,20%	0%	19,90%	0%	14,10%
Professor	0%	60,80%	39,10%	7,10%	62,50%	0%	0%	0%	17,80%	0%	24,20%
Journalist	0%	4,60%	9,20%	24,70%	0%	0%	0%	0%	28,20%	0%	13,90%
Minister	0%	4,70%	3,20%	26,60%	8,20%	17,60%	25,90%	100,00%	10,70%	0%	13,60%
President	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
Politician	0%	20,40%	11,80%	28,30%	25,00%	82,40%	51,90%	0%	17,00%	100,00%	30,90%
Author	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
Scientist	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%

Academician	0%	2,40%	0%	11,00%	0,40%	0%	0%	0%	6,30%	0%	3,20%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Variable :	(2)Type of recording										
Sec.	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
Debate	0	489,49	391,53	67,08	3206,49	0	600,01	0	4926,03	0	9680,63
Interview	0	221,06	473,74	71,17	1486,94	919,46	1471,48	0	5034,91	1913,18	11591,94
TV-show	0	0	0	0	0	0	0	0	0	0	
Guest in the News	0	37,66	0	49,06	419,88	195,75	725,17	726,83	701,53	0	2855,88
Duel	0	36,78	50,89	0	0	0	0	0	629,07	0	716,74
Telephone call	0	19,63	0	23,24	20,81	0	0	0	765,08	0	828,76
Total	0	804,6199	916,16	210,55	5134,12	1115,21	2796,66	726,83	12056,62	1913,18	25673,95
% lig	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
Debate	0%	5,10%	4,00%	0,70%	33,10%	0%	6,20%	0%	50,90%	0%	100%
Interview	0%	1,90%	4,10%	0,60%	12,80%	7,90%	12,70%	0%	43,40%	16,50%	100%

TV-show	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Guest in the News	0%	1,30%	0%	1,70%	14,70%	6,90%	25,40%	25,50%	24,60%	0%	100%
Duel	0%	5,10%	7,10%	0%	0%	0%	0%	0%	87,80%	0%	100%
Telephone call	0%	2,40%	0%	2,80%	2,50%	0%	0%	0%	92,30%	0%	100%
Total	0,00%	3,10%	3,60%	0,80%	20,00%	4,30%	10,90%	2,80%	47,00%	7,50%	100%
% col	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
Debate	0%	60,80%	42,70%	31,90%	62,50%	0%	21,50%	0%	40,90%	0%	37,70%
Interview	0%	27,50%	51,70%	33,80%	29,00%	82,40%	52,60%	0%	41,80%	100,00%	45,20%
TV-show	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
Guest in the News	0%	4,70%	0%	23,30%	8,20%	17,60%	25,90%	100,00%	5,80%	0%	11,10%
Duel	0%	4,60%	5,60%	0%	0%	0%	0%	0%	5,20%	0%	2,80%
Telephone call	0%	2,40%	0%	11,00%	0,40%	0%	0%	0%	6,30%	0%	3,20%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Variable :	(3)TV channel										
Sec.	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	

A1	0	306,47	150,76	0	0	0	0	0	2151,43	0	2608,66
MTV	0	0	0	0	0	0	0	0	0	0	
Sitel	0	83,66	335,9	43,71	354,8	195,75	1171,62	726,83	2536,44	0	5448,71
MTV1	0	272,96	271,34	107,81	4288,69	0	769,62	0	2774,6	1913,18	10398,2
Kanal 5	0	75,12	10,63	10	268,59	0	855,42	0	1064,26	0	2284,02
Moris	0	19,63	0	23,24	20,81	0	0	0	765,08	0	828,76
Nasa TV	0	0	0	0	201,23	0	0	0	0	0	201,23
Alfa	0	10	36,71	11,21	0	353,55	0	0	1545,45	0	1956,92
Radio Slobodna Evropa	0	0	59,93	14,58	0	565,91	0	0	590,29	0	1230,71
AB	0	36,78	50,89	0	0	0	0	0	629,07	0	716,74
Total	0	804,6199	916,16	210,55	5134,12	1115,21	2796,66	726,83	12056,62	1913,18	25673,95
% lig	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
A1	0%	11,70%	5,80%	0%	0%	0%	0%	0%	82,50%	0%	100%
MTV	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Sitel	0%	1,50%	6,20%	0,80%	6,50%	3,60%	21,50%	13,30%	46,60%	0%	100%
MTV1	0%	2,60%	2,60%	1,00%	41,20%	0%	7,40%	0%	26,70%	18,40%	100%
Kanal 5	0%	3,30%	0,50%	0,40%	11,80%	0%	37,50%	0%	46,60%	0%	100%
Moris	0%	2,40%	0%	2,80%	2,50%	0%	0%	0%	92,30%	0%	100%
Nasa TV	0%	0%	0%	0%	100,00%	0%	0%	0%	0%	0%	100%
Alfa	0%	0,50%	1,90%	0,60%	0%	18,10%	0%	0%	79,00%	0%	100%

Radio Slobodna Evropa	0%	0%	4,90%	1,20%	0%	46,00%	0%	0%	48,00%	0%	100%
AB	0%	5,10%	7,10%	0%	0%	0%	0%	0%	87,80%	0%	100%
Total	0,00%	3,10%	3,60%	0,80%	20,00%	4,30%	10,90%	2,80%	47,00%	7,50%	100%
% col	Without topic	Introduction	Music	Farewells	Relations between Macedonia and EU	Integration of Macedonia in NATO	Relations between Macedonia and Greece	Liberalization of visa regime with EU	Political situation in Macedonia	Council of Europe	
A1	0%	38,10%	16,50%	0%	0%	0%	0%	0%	17,80%	0%	10,20%
MTV	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0,00%
Sitel	0%	10,40%	36,70%	20,80%	6,90%	17,60%	41,90%	100,00%	21,00%	0%	21,20%
MTV1	0%	33,90%	29,60%	51,20%	83,50%	0%	27,50%	0%	23,00%	100,00%	40,50%
Kanal 5	0%	9,30%	1,20%	4,70%	5,20%	0%	30,60%	0%	8,80%	0%	8,90%
Moris	0%	2,40%	0%	11,00%	0,40%	0%	0%	0%	6,30%	0%	3,20%
Nasa TV	0%	0%	0%	0%	3,90%	0%	0%	0%	0%	0%	0,80%
Alfa	0%	1,20%	4,00%	5,30%	0%	31,70%	0%	0%	12,80%	0%	7,60%
Radio Slobodna Evropa	0%	0%	6,50%	6,90%	0%	50,70%	0%	0%	4,90%	0%	4,80%
AB	0%	4,60%	5,60%	0%	0%	0%	0%	0%	5,20%	0%	2,80%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%