



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

## Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in  
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale  
Classe LM-38

Tesi di Laurea

# *Towards an automatic recognition of hybrid languages: The case of Surzhyk*

Relatore  
Prof. Giorgio Maria Di Nunzio

Laureando  
Nataliya Sira  
n° matr.1137410 / LMLCC

Anno Accademico 2017 / 2018

## **Abstract**

**Introduction:** Surzhyk is a hybrid Ukrainian-Russian language that emerged after the lasting coexistence of Ukrainian and Russian languages in Ukraine. It presents characteristics of mixed language and is also considered a “folk-spoken language”, a variety of vernacular language. (Serdeha and Saharovskiy, 2011: 41). Surzhyk takes its roots at the end of the 17th century when due to the policies of russification Ukrainian speaking people started to adopt or were forced to adopt a more prestigious language, Russian. The current status of Surzhyk is characterised as a non prestigious spoken language that is commonly used in the whole Ukraine with the largest number of its speakers in the central, southern and eastern parts of Ukraine.

**Purpose of research:** Together with russification, code-mixed elements began to be part of people’s every day vocabulary; consequently, these hybrid words slowly started to shift onto the pages of blogs, social media, Internet discussions and other types of written language. Surzhyk-speaking people are not always aware of using Surzhyk. In this research the purpose is to find out how the elements of Surzhyk could be identified automatically in the written source text, without human help. The current research aims to identify the non-standard verbs of the first person plural in Present tense.

**Materials and methods:** The research involves different fields of studies and thus is interdisciplinary. Starting from transcribing the records collected by Del Gaudio in different geographical areas of Ukraine (2010), the text samples were analysed on the presence of Surzhyk. In the created text documents we identified lexical elements that did not belong to the standard Ukrainian language, or were defined as incorrect Ukrainian by the automatic corrector of Google Drive Documents. After the manual identification we created Excel tables aimed to represent all non-standard terms. Consequently, every term was compared to Ukrainian and to Russian. The code implementation in R allowed to produce example-based rules on identification of the first person plural verbs of Surzhyk in Present tense. The research is based on the online corpus of Ukrainian and Russian and on the vocabularies of the both languages, as well as on my personal knowledge of Ukrainian, Russian and familiarity with Surzhyk.

**Key words:** Hybrid languages, Surzhyk, Text mining, NLP, R.

## **Acknowledgements**

I would like to express my sincere gratitude to my family for supporting me throughout my studies and my life in general. Special gratitude is to my sister Anastasiya, to my best friends and to Marco.

## **Table of contents**

### Introduction

### First part

#### 1. 1. Language situation in Ukraine

##### 1. 1. 1. The group of Slavic languages

##### 1. 1. 2. The Old Church Slavonic language

##### 1. 1. 3. Ukrainian language development

##### 1. 1. 4. Constitution of Ukraine and the regulations on language

##### 1. 1. 5. The history of coexistence between Ukrainian and Russian language

##### 1. 1. 6. The presence of dialects in Ukraine

##### 1. 1. 7. The question of Surzhyk

#### 1. 2. Hybrid languages, from language interaction to language mixing

##### 1. 2. 1. Bilingualism and language interference

##### 1. 2. 2. Language alternation: from codeswitching through language mixing to fused lect

###### 1. 2. 2. 1. Codeswitching

###### 1. 2. 2. 2. Language mixing

###### 1. 2. 2. 3. Fused lect

##### 1. 2. 2. Surzhyk classification

###### 1. 2. 2. 1. Previous classifications of Surzhyk

###### 1. 2. 2. 2. Surzhyk as a hybrid language with the characteristics of mixed language

#### 1. 3. Natural Language Processing

##### 1. 3. 1. NLP as a field of studies

##### 1. 3. 2. The meeting point of different research areas

##### 1. 3. 3. NLP main areas and problems

##### 1. 3. 4. NLP application in the current research

###### 1. 3. 4. 1. A variety of subfields and tools

###### 1. 3. 4. 2. R as an environment for testing

## Second part

### 2. 1. Preparatory phase

#### 2. 1. 1. Material collection and terms classification

#### 2. 1. 2. The instruments of R

#### 2. 1. 3. Identification of the elements to be analysed

##### 2. 1. 3. 1. Background information

##### 2. 1. 3. 2. Aspectual distinction

##### 2. 1. 3. 3. Russian verbal system

##### 2. 1. 3. 4. Ukrainian verbal system

##### 2. 1. 3. 5. Present tense endings of the first person plural

##### 2. 1. 3. 6. The particles analysed in the practical part of the research

### 2. 2. Project development

#### 2. 2. 1. First phase

##### 2. 2. 1. 1. Creation of the application in R

##### 2. 2. 1. 2. Checking the output in application user mode

##### 2. 2. 1. 3. Rules on identification

###### 2. 2. 1. 3. 1. General rules

###### 2. 2. 1. 3. 2. Specific rules

###### 2. 2. 1. 3. 3. General comments on rules

#### 2. 2. 2. Second phase

##### 2. 2. 2. 1. Summary of the rules

##### 2. 2. 2. 2. Testing of the second group of specific rules in user mode

##### 2. 2. 2. 3. General rules

##### 2. 2. 2. 4. Testing of the second group of general rules (with a generalised final pattern “-M”) in user mode

##### 2. 2. 2. 5. General considerations on the second phase

#### 2. 2. 3. Third phase

##### 2. 2. 3. 1. Additional rules

- 2. 2. 3. 2. Testing of the rules.
- 2. 2. 3. 2. 1. Specific rules testing in creator mode
- 2. 2. 3. 2. 2. General rules testing in creator mode
- 2. 2. 3. 3. Analysis of the results and comments
- 2. 2. 4. Fourth phase. Testing on the Russian texts

Conclusion

References

Summary in Italian

Summary in Russian

Appendices

## **Introduction**

Language interference is a common argument in today's society where, due to globalisation, many languages are concentrated on the same area. Multilingualism became one of the main characteristics of the end of 20th and beginning of 21st century. In the European Union moving from country to country for educational, employment or personal reasons became a routine and together with these transfers, the number of multilingual families continues to increase. At the higher level, in the areas of official or unofficial multilingualism, it is inevitable that the languages involved in the communication come in contact and somehow influence each other. The presence of two or more languages in discourse may lead to language alternation phenomena expressed by Auer (1999: 309) as continuum from codeswitching through language mixing to fused lects. In the situations of language contact a completely new language may emerge (Appel, Muysken, 1993: 175) that would present elements of both languages that were involved in its creation. The hybridity of the mixed languages consists in the fact that they take their lexicon from one source and grammar from another. Their classification becomes difficult as on the basis of the lexicon they could appertain to one language family and on the basis of the morphology, syntax and general grammatical characteristics they may belong to another language family. (Bakker, Mous, 1994: 5). The hybrid languages phenomena is discussed in the first part of this work together with the case of the hybrid Ukrainian-Russian language Surzhyk that was created as a result of Ukrainian and Russian languages coexistence. The study of the formation and development of Ukrainian literary language, its current status and relationship with Russian are to be considered mandatory for further approach with Surzhyk.

The research of this Master's thesis studies the language situation in Ukraine which is generally characterised by the prevalence of two Slavonic languages: Ukrainian and Russian. It focuses specifically on the identification of the hybrid elements, known as

Surzhyk and common to the spoken language, that do not appertain neither to Ukrainian nor to Russian standard languages. Manual identification of these hybrid elements in the spoken language will then lead to their automatic identification. This part of the work is based on the field of study that focuses on the interactions between human language and computers, Natural Language Processing, commonly known for topic segmentation, speech recognition, automatic summarisation and used in machine translation, grammar correcting and automated question answering. The main purpose of this research is to allow automatic recognition of the Ukrainian-Russian hybrid elements without human intervention and to provide the basis for the further works on the automatic recognition of mixed languages.

The research method adopted in this thesis is an example based one. Since Surzhyk is more characteristic as a spoken language rather than a written one, the collection of Surzhyk samples was an important task. During the first stage of the research, we analysed the interviews registered by Del Gaudio in different areas of Ukraine and published as a CD together with his PhD dissertation *On the nature of Surżyk* (2010). The interviews of Del Gaudio involved people from Kyiv, Chernihiv and Kharkiv areas, but in the current research only the first group of conversations regarding the area of Kyiv were elaborated in detail. In order to have digital data, it was necessary to transcribe manually all the registrations since no efficient application for Speech-to-Text automatic recognition was found neither for Ukrainian nor for a non-standard spoken language Surzhyk. The first stage of the analysis consisted in, first of all, creating the Surzhyk corpus. In order to have the material in the proper format for computation and further elaboration, all the registrations collected by Del Gaudio were transcribed manually and saved as text documents. The process was guided by the rule of writing what was heard and Ukrainian writing rules were taken as basis. It is important to underline that not always audio registrations were noiseless and clear, for these reasons there may be some omissions in the texts. In addition, there may be some small typing errors. Secondly, in the previously created text documents we identified lexical elements that did not belong to the standard Ukrainian language, or also words



that were defined as not correct Ukrainian by the automatic corrector of Google Drive Documents. This analysis was limited and mainly consisted in identifying of the non-standard Ukrainian lexical elements in the text documents and creating their representation in Excel documents. Consequently, they were compared to Ukrainian terms and to Russian terms. While for the Ukrainian terms there is only the corresponding written version, for the Russian terms we both present the written version that respects the rules of Russian and the phonetic pronunciation of the term according to Ukrainian phonology. This becomes necessary as soon as “Russian content morphemes introduced to the Ukrainian morphosyntax are subject to Ukrainian phonology and are often pronounced in accordance with Ukrainian phonological rules” (Kent, 2010: 45). Finally, we created and filled the terminological records that in total present more than 1000 non-standard Ukrainian terms. Table 1 is an example of the structure of the Simplified terminological tables.

Table 1.

Term	Ukrainian written	Russian written	Russian pronounced	Additional comments if needed	Context	English written
A	B	C	D	E	F	G

In the first column A we have the term to analyse, written in Cyrillic according to the Ukrainian rules when the whole dialogue is in Ukrainian. When the whole dialogue is in Russian, the term may be written in Cyrillic according to the Russian rules.

In the second column B we have the Ukrainian version of that term written in Cyrillic according to Ukrainian rules.

In the third column C we have the Russian version of the term written in Cyrillic according to the Russian rules and in the fourth column D there is the Russian version of the term written in Cyrillic according to the pronunciation of the word in Russian but adapting Ukrainian alphabet.

The column E contains additional comments in English where needed.

The column F may present the context in which the term was used in the text if it is a specific term.

The last column G contains English translation of the term.

Some may argue that a term that corresponds perfectly to the Russian pronunciation should be classified as a Russian one, but we always have to consider the following statements: first of all, the identified elements are characteristic and widespread especially as a non-standard spoken language; secondly, in this type of analysis we elaborate texts that were registered and collected in Ukraine and the main language of the conversation was Ukrainian. Previous works on Surzhyk, for example Kent (2010), proposed to adopt the Matrix Language Frame model (MLF) elaborated by Myers-Scotton (2002, 2003) for analysing a mixed language variety. According to this model only one language is the source of the abstract morphosyntactic frame in a bilingual clause and that is the Matrix Language. The other participating language is the Embedded Language and it must agree with structural requirements stipulated by the Matrix Language (Myers-Scotton 2002, 2003). If we consider that in the analysed texts we have Ukrainian as a ML and Russian as an EL, our samples should contain a great amount of Russian lexical items and mostly Ukrainian morphosyntax. Moreover, Surzhyk is actually created by introducing Russian content morphemes to the Ukrainian morphosyntax, and Russian lexical items are often pronounced in accordance with Ukrainian phonological rules. (Kent, 2010: 45). So if the main language of the conversation was Russian, we would surely affirm that terms that correspond to the Russian pronunciation are Russian, but in this specific analysis we are trying to define the spoken language, that even some speakers define as “not a clean language” (Del Gaudio, 2010: registration DW\_A0031\_Kyjiv\_BasicSurzyk\_F), a some kind of “mixed Ukrainian” (Kent, 2010: 43), “this type of Ukrainian” (Kent, 2010: 43).

The preparatory phase required a significant amount of time and for organisational reasons it was decided to identify a particular characteristic of Surzhyk verbs. Due to

the recurring repetition of the ending “-м” in the first person plural of the Present tense in the processed texts that was noticed during the process of manual transcription of the registrations collected by Del Gaudio, it was decided to analyse the presence of this particular particle more in detail. We provided a list of rules based on the examples available in transcriptions. To carry out this analysis, it was necessary to compare the Russian and Ukrainian verbal systems, and, specifically, the endings of the first person plural in the Present tense. Together with the facts on historic development of the first person plural endings in Present tense in Slavonic languages, the present differences of the endings in both Russian and Ukrainian languages were taken into consideration.

The research involves different fields of studies and is to be considered interdisciplinary. The first part of the thesis is theoretical and it is composed of three chapters. The first chapter studies the language situation in Ukraine, the main dialects diffused on the Ukrainian territory and the regulations on the language policies of the Constitution of Ukraine; it defines the relationship between Ukrainian and Russian languages and explains the origins of the hybrid Ukrainian-Russian language Surzhyk. The second chapter presents different phenomena that characterise situations of bilingualism or multilingualism. Together with the definition of hybrid languages we discussed the current classification of Surzhyk as a hybrid language that presents mixed language characteristics. The third chapter presents the basis of Natural Language Processing, an area of Computer Science that deals with research on how computer systems can analyse, understand or produce natural languages.

The second part of the thesis is practical and it is divided into two chapters. The first chapter presents a preparatory phase that consisted in the material collection and its manual elaboration; familiarising with the programming language and environment R by studying its main functions; and the description of the elements that have to be analysed during the phases of testing. The second chapter of the second part regards the project development and is divided into four phases of testing. The first three phases deal with Surzhyk samples regarding the area of Kyiv that were published by Del

Gaudio as a CD together with his PhD dissertation *On the nature of Suržyk* (2010). The final phase presents an alternative testing of the created rules on Russian language texts. To conclude there is a summary of the results obtained during this research, general comments on the importance of this topic in a multilingual society and a discussion on its possible developments.

As a final part of this thesis it was decided to provide appendices that concern the creation of the terminological tables for the spoken language Surzhyk. In the appendices we included the general tables with all the hybrid elements present in the analysed texts (Simplified terminological table of Surzhyk) and a model of the deeper description of the hybrid language terms (Standard terminological table of Surzhyk). The Simplified terminological tables of Surzhyk were designed in order to represent all non-standard Ukrainian terms found in the Kyiv audio records collected by Del Gaudio. The structure of the Standard terminological table of Surzhyk model was drafted with Federica Vezzani, a PhD student in Linguistic, Philological and Literary Science of University of Padua, the 33th cycle of the Department of Linguistic and Literary Studies DISLL, in order to define important aspects of Surzhyk terms. Although creating terminological tables for Surzhyk was only a part of the work in the preparatory phase and not the main task of this thesis, we consider that it may be useful for further studies on Surzhyk.

Transliteration of the Ukrainian alphabet was performed following the rules approved by the resolution of the Cabinet of Ministers of Ukraine № 55 dated 27 Jan. 2010. Table 2 is its representation.<sup>1</sup>

Table 2.

А	А	
Б	В	
В	У	

---

<sup>1</sup> Based on the official Table of transliteration of the Ukrainian alphabet in Latin script (Таблиця транслітерації українського алфавіту латиницею).

Г	H	
Г	G	
Д	D	
Е	E	
Є	Ye ie	At the beginning of the word In other positions
Ж	Zh	
З	Z	
И	Y	
І	I	
Ї	Yi i	At the beginning of the word In other positions
Й	Y i	At the beginning of the word In other positions
К	K	
Л	L	
М	M	
Н	N	
О	O	
П	P	
Р	R	
С	S	
Т	T	
У	U	
Ф	F	
Х	Kh	
Ц	Ts	
Ч	ch	

Ш	Sh	
Щ	Shch	
Ю	Yu iu	At the beginning of the word In other positions
Я	Ya ia	At the beginning of the word In other positions

It should be noticed that, according to the rules, Ukrainian soft sign and apostrophe are not reproduced when transliterated. In certain circumstances some Ukrainian letters were transliterated differently if compared to the presented table; this occurred when works that adopted different rules on transliteration were quoted. For example “ž” was maintained when being part of the title in English, or part of quotations.

## **1. 1. Language situation in Ukraine**

As an introductory part to the topic of this research it is important to understand the current linguistic situation in Ukraine and the development of Ukrainian language through different historical periods. This chapter presents, firstly, information on the Slavic group of languages that is useful to define the provenience of Ukrainian. Secondly, there is a brief summary of the Old Church Slavonic — a language that characterised the period of Christianity adoption and consequently the diffusion of writing. Then, follows an undertopic on the formation and development of Ukrainian literary language. Consequently, there are quotations and analysis of the two articles regarding the regulations on the language policies of the Constitution of Ukraine. Then this chapter defines the relationship between Ukrainian and Russian languages and the main dialects diffused on the Ukrainian territory. Finally, the chapter ends with the description of the hybrid Ukrainian-Russian language Surzhyk, created as a result of Ukrainian and Russian languages coexistence.

### **1. 1. 1. The group of Slavic languages<sup>2</sup>**

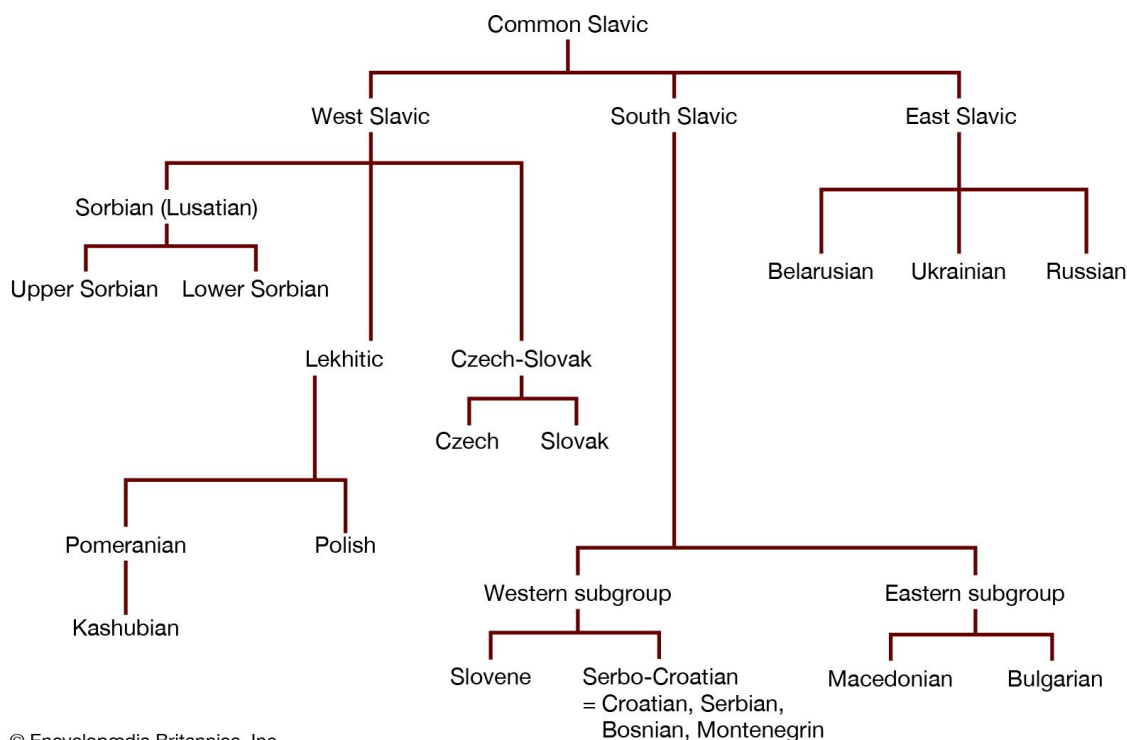
Indo-European Group of languages to which appertain the Germanic and Romance languages includes also the Slavic, or Slavonic, languages. (Gardiner, 1984: 1). The Slavic languages are divided into three groups: West Slavic, East Slavic and South Slavic. The following figure (ILLUSTRATION 1) illustrates the Slavic languages' family tree.

---

<sup>2</sup> Based on Gardiner, 1984: 1-2.

## ILLUSTRATION 1

The Slavic languages' family tree.



© Encyclopædia Britannica, Inc.

*Encyclopædia Britannica Online*, [www.britannica.com/topic/Slavic-languages?oasmId=61368](http://www.britannica.com/topic/Slavic-languages?oasmId=61368). Accessed 12 Jun. 2018.

The West Slavic branch includes Polish, Czech, Slovak, Upper and Lower Sorbian and Kashubian. The South Slavonic group includes Slovene, Croatian, Serbian, Bosnian, Montenegrin, Macedonian and Bulgarian. East Slavic branch consists of Belarusian, Ukrainian, and Russian. According to Gardiner, South Slavonic dialect of Bulgarian and Macedonian type was the first Slavic language used for writing and because of its function is called Old Church Slavonic. Old Church Slavonic language was regenerated on the basis of the 10th-11th centuries copies of the lost original texts produced by Cyril, Methodius and their disciples in Moravia, Bulgaria or Macedonia. These texts consisted of Gospels, liturgies, lives of saints, homilies and were aimed to spread Christianity, a policy that took place starting from 9th century. (Gardiner, 1984: 1-2).



## 1. 1. 2. The Old Church Slavonic language

The Old Church Slavonic language was introduced into the liturgy in Great Moravia, the first Slavic national state founded in the 9th century that united different groups speaking West Slavic dialects. (Ivanov, Browne). In 862 Prince Rostislav of Moravia demanded the Byzantine Emperor Michael III to send to Moravia Slavonic-speaking missionaries that could spread Christianity. For such an important task were chosen brothers Constantine and Methodius, two natives of the inhabited by Slavonic-speaking population city Salonica. Constantine had already proved himself as a scholar and missionary, while Methodius was an experienced administrator. Their mission in Moravia was approved by the pope and in 869 two brothers travelled to Rome, where Constantine had taken monastic vows and assumed the name of Cyril and then died. Afterwards Methodius returned to Moravia and Prince Rostislav died, leaving power in the hands of his successor Svatopluk. Under Svatopluk the German clergy were in the ascendant. In 885 Methodius died and his disciples had to escape, some of them succeeded to establish themselves in the Bulgarian Empire at the court of Tsar Simeon and in Macedonia. (Gardiner, 1984: 1). In 865 Bulgaria adopted Christianity and in the end of 9th century and in the 10th century it became the center of Slavic writing. This explains that it was from the Bulgaria that at the end of the 10th century the first ministers of the church were invited to Kyivan Rus. Moreover, the church books, written in the language, used for translation by Cyril and Methodius, were imported into Kyivan Rus also from the Bulgaria.<sup>3</sup> (Tokar, 1993: 19). In the first half of the 11th century, after the Byzantium pushed down the independence of Bulgaria, development of the Slavonic culture in the Balkans declined. (Gardiner, 1984: 1-2).

---

<sup>3</sup> “Болгарія прийняла християнство у 865 р., наприкінці IX - у X ст. вона стала центром слов'янської писемності. Саме з Болгарії було запрошено у кінці X ст. до Київської Русі перших служителів церкви і завезено церковні книги, писані мовою, якою перекладали Кирило і Мефодій. У науковому обігу цю мову називають старослов'янською.”

### 1. 1. 3. Ukrainian language development

The written documents of the 12th - 13th centuries contained phonetic, grammatical and lexical traits of Proto-Ukrainian. Already in the 14th century Ukrainian language was prospering as a separate lingual code. “From the Carpathians to the upper reaches of the Don River, from the Polissya marshes to the Black Sea, was heard one language, which, however, retained traces of the ancient tribal language division.”<sup>4</sup> (Rusanivskyi in Tokar, 1993: 23). The development of every language is directly connected to the historical events that can contribute to its expansion or to its decline and the formation and evolution of the Ukrainian literary language constantly depended on concrete events that took place on the Ukrainian territories. In the second half of the 17th century, for example, after Ukraine was attached to Russia in accordance with the Pereiaslav agreement of 1654, the country gradually lost its Cossack liberties and fall into political and economic dependence from Russia. This period and the whole 18th century was characterised by the gradual development of capitalist relations in Ukraine accompanied by the emergence of the common for Ukraine and Russia leading, actually Russian, state cultural and economic center. Due to the conditions of such unequal coexistence and the dominant role of Russia in the former Russian Empire, Russian language was privileged the first place, becoming the only state language. Nevertheless, even under these unfavorable conditions for the formation of the Ukrainian nation and the limited possibilities for the development of its culture, Ukrainian language gradually formed as a literary language. (Tokar, 1993: 24).<sup>5</sup>

---

<sup>4</sup> “Протоукраїнські риси — фонетичні, граматичні і лексичні — засвідчують писемні пам’ятки XII - XIII ст. На XIV ст. українська мова існувала вже як окремий лінгвальний організм, що розвивався за своїми законами. Від Карпат до верхів’їв річки Дону, від поліських боліт до Чорного моря чулася одна мова, яка, однак зберігала сліди давнього, ще племінного мовного поділу.”

<sup>5</sup> “Формування та розвиток української літературної мови постійно залежали від конкретних подій життя нашого народу. Наприклад, у другій половині XVII ст. Україну за Переяславською угодою 1654 р. приєднано до Росії. Поступово Україна втрачає свої козацькі вольності, потрапляючи в політичну і економічну залежність від Росії. На цей же час і на все XVIII ст. припадає поступовий розвиток капіталістичних відносин в Україні. Цей процес супроводжувався, зокрема, появою єдиного для України та Росії провідного, насправді російського державного й культурно-економічного центру. За умов такої нерівності і панівної ролі Росії в колишній Російській імперії російська мова висувалась на перше місце, все більше перетворюючись на єдину державну мову. Однак українська мова навіть і за цих несприятливих умов формування української нації та обмеженого розвитку її культури поступово формувалась саме як літературна мова...”

An important contribute to the formation of new Ukrainian literary language was done by Kotliarevskiy. The author of first major literature works in Ukrainian language, such as “Aeneid”, “Natalka Poltavka”, “Moskal-Charivnyk”, played a significant role in creating the literature that presented real language spoken by folks. 1798, the year when the first three parts of the famous poem “Aeneid” came out in print, is conventionally considered the year of beginning of the new Ukrainian literary language. In the first half of the 19th century, were published first collections of Ukrainian songs gathered from people’s memories, literature works of various genres created by Ukrainian writers such as Kvitka-Osnovianenko, Hulak-Artemovskiy, Hrebinka, Borovykovskiy, and the the first Western Ukrainian almanac “Rusalka Dnistrovaia” prepared by Holovatskiy, Shashkevych and Vagylevych. The process of formation of the Ukrainian literary language on the folk speech basis was completed by Taras Shevchenko. One of the greatest Ukrainian poets, writer and artist, a fighter against the oppression of the working people, he became the founder of the Ukrainian literary language. (Tokar, 1993: 25).<sup>6</sup>

After 1917 in the Ukrainian ethnic territory were formed the Ukrainian National Republic and the Western Ukrainian National Republic. The years of official recognition of the Ukrainian state existence and Ukrainian people's fight for the liberation permitted Ukrainian language to augment its functions of usage. Consequently, when the Bolsheviks came to power in Ukraine, was proclaimed Ukrainian Soviet Socialist Republic. With the beginning of the Soviet period Russian language took the function of the language of international communication.<sup>7</sup> At the turn

---

<sup>6</sup> “Значну роль у формуванні нової української літературної мови на живій народній основі відіграв І. П. Котляревський - автор перших великих художніх творів українською мовою (“Енеїда”, “Наталка Полтавка”, “Москаль-чарівник”). Початком нової української літературної мови умовно вважається 1798 р., коли вийшли друком три перші частини славнозвісної поеми “Енеїда” [...]. У першій половині ХІХ ст. вийшли перші збірки українських пісень, записаних з народних уст, різноманітні за жанром і формою художні твори українських письменників (Г. Ф. Квітки-Основ’яненка, П. П. Гулака-Артемівського, Є. П. Гребінки, Л. І. Боровиковського, “Русалка Дністровая” Я. Головацького, М. Шашкевича, І. Вагілевича). Процес становлення української літературної мови на живій народній основі завершив великий поет, мислитель і непримиренний борець проти гноблення трудящого народу Тарас Григорович Шевченко. Він став основоположником української літературної мови.”

<sup>7</sup> “Після 1917 р. в межах української етнічної території сформувались дві держави — Українська Народна Республіка і Західноукраїнська Народна Республіка.” “Існування протягом кількох років

of the 20's and 30's, with the assertion of Stalin cult of personality and the policies on forced Russification, associated with his name, the development of the modern Ukrainian literary language has considerably slowed down. After the collapse of the USSR in late 1991, on the ruins of the last empire were created the sovereign independent states. Ukraine became independent on 24 August 1991 and the historic fate of Ukrainian people radically changed. (Tokar, 1993: 26-27).<sup>8</sup>

#### **1. 1. 4. Constitution of Ukraine and the regulations on language**

The Constitution of Ukraine is the Fundamental Law of Ukraine. It was adopted at the Fifth Session of Ukrainian parliament Verkhovna Rada, on 28 June 1996, guided by the Act of Declaration of the Independence of Ukraine of 24 August 1991.<sup>9</sup> The Fundamental Law of Ukraine is composed of fifteen titles (chapters) containing articles. The articles that regard language policies are Article 10 of the Title I and Article 53 of the Title II.

Article 10 is located in the Title I of Ukrainian Constitution that contains also Article 1 on Ukraine as a “sovereign and independent, democratic, social, law-based state” (Constitution of Ukraine, Title I. Article 1)<sup>10</sup> and regulates general principle of Ukraine.

---

(1917-1920) Української держави, визвольні змагання українського народу зумовили розширення функцій української мови [...]. Згодом, після приходу до влади в Україні більшовиків, було проголошено Українську Радянську Соціалістичну Республіку. Почався радянський період [...].”

<sup>8</sup> “Розвиток сучасної української літературної мови суттєво уповільнився вже на зламі 20 — 30-х рр., з утвердженням у державному житті культу особи Сталіна і пов'язаної з його іменем політики русифікації... Історична доля українського народу докорінно змінилася тоді, коли на руїнах останньої імперії — колишнього СРСР — утворилися суверенні незалежні держави, в тому числі й незалежна Україна.”

<sup>9</sup> A full version of the official English translation of the Constitution of Ukraine used in this chapter is available at the Official website of Ukrainian Government Portal, available as document to be downloaded:

[https://www.kmu.gov.ua/storage/app/imported\\_content/document/110977042/Constitution\\_eng.doc](https://www.kmu.gov.ua/storage/app/imported_content/document/110977042/Constitution_eng.doc).

Accessed 14 Jun. 2018.

All quotations referred to the Constitution of Ukraine were consulted in Ukrainian language at the Official web portal of Verkhovna Rada of Ukraine. A full text is available at: <http://zakon0.rada.gov.ua/laws/show/254%D0%BA/96-%D0%B2%D1%80>. Accessed 14 Jun. 2018.

<sup>10</sup> “Україна є суверенна і незалежна, демократична, соціальна, правова держава.” (Конституція України, Розділ I. Стаття 1).

“The State language of Ukraine shall be the Ukrainian language.

The State shall ensure comprehensive development and functioning of the Ukrainian language in all spheres of social life throughout the entire territory of Ukraine.

Free development, use, and protection of Russian and other languages of national minorities of Ukraine shall be guaranteed in Ukraine.

The State shall promote the learning of languages of international communication.

The use of languages in Ukraine shall be guaranteed by the Constitution of Ukraine and shall be determined by law.” (Constitution of Ukraine, Title I. Article 10).<sup>11</sup>

Article 53 is located in the Title II of the Constitution of Ukraine that regulates human and citizen rights, freedoms, and duties. The article regarding the right to education, guarantees national minorities the right to have education in their native language or to have the possibility to study their native language.

“Everyone shall have the right to education.

Complete general secondary education shall be compulsory.

The State shall ensure accessible and free pre-school, complete general secondary, vocational and higher education at the state and communal educational establishments; the development of pre-school, complete general secondary, extra-curricular, vocational, higher and post-graduate education, various forms of study; the provision of state scholarships and privileges to pupils and students.

Citizens shall have the right to obtain free higher education at the state and communal educational establishments on a competitive basis.

Citizens belonging to national minorities shall be guaranteed, in accordance with law, the right to education in their native language, or to study their native language at the state and communal educational establishments or through national cultural societies.” (Constitution of Ukraine, Title II. Article 53).<sup>12</sup>

---

<sup>11</sup> “Державною мовою в Україні є українська мова.

Держава забезпечує всебічний розвиток і функціонування української мови в усіх сферах суспільного життя на всій території України.

В Україні гарантується вільний розвиток, використання і захист російської, інших мов національних меншин України.

Держава сприяє вивченню мов міжнародного спілкування.

Застосування мов в Україні гарантується Конституцією України та визначається законом.” (Конституція України, Розділ I. Стаття 10).

<sup>12</sup> “Кожен має право на освіту.

Повна загальна середня освіта є обов'язковою.

Держава забезпечує доступність і безоплатність дошкільної, повної загальної середньої,

### 1. 1. 5. The history of coexistence between Ukrainian and Russian language

Throughout its history Ukrainian language had to fight against Russian language in order to become an official and nationally recognised language. “The formation of the standard Modern Ukrainian language dates back to the 18th century when Ukrainian lands were partitioned between the Russian and Austrian (later Austro-Hungarian) Empires.” While the Austro-Hungarian Empire’s language policies aided in the development of the standard Ukrainian language, culture and distinct national identity, the Russian Empire marginalized Ukrainian language out of urban centres by providing the extensive Russification. (Kent, 2012: 3). In 1720 Peter the Great ordered to adapt all the Ukrainian language scriptures to the Russian language before publishing.<sup>13</sup> Consequently, the printing of the books in the press of Kyiv and Chernihiv was in Russian. In 1863 and in 1876, the use of Ukrainian in schools, press, musical compositions and theatrical performances was prohibited, as well as the import of Ukrainian-language printed materials.<sup>14</sup> Ukrainian language was often persecuted and its linguists suffered from repressions. Relocations and extermination of the intelligentsia were typical phenomena together with the widespread accusation of the national separatism dissemination. The 1932-33 famine provoked by the Soviets costed Ukrainians 3-6 million lives, 6-7 million, according to other estimates. (Besters-Dilger,

---

професійно-технічної, вищої освіти в державних і комунальних навчальних закладах; розвиток дошкільної, повної загальної середньої, позашкільної, професійно-технічної, вищої і післядипломної освіти, різних форм навчання; надання державних стипендій та пільг учням і студентам.

Громадяни мають право безоплатно здобути вищу освіту в державних і комунальних навчальних закладах на конкурсній основі.

Громадянам, які належать до національних меншин, відповідно до закону гарантується право на навчання рідною мовою чи на вивчення рідної мови у державних і комунальних навчальних закладах або через національні культурні товариства.” (Конституція України, Розділ II. Стаття 53).

<sup>13</sup> Referred to: “для совершенного согласия с великороссийским, с такими ж церковными книгами справлявать прежде печати с теми великороссийскими печатями, дабы никакой розни и особаго наречия во оных не было” (*Polnoe sobranie zakonov Rossijskoj imperii T.6, 1830: 245*) in Besters-Dilger J., 1996: 129: “Im Jahr 1720 anordnete Peter der Grosse, dass “alle kleinrussischen (d.h. ukrainischen) Schriften vor dem Druck an das Großrussische angepasst werden müssten.”

<sup>14</sup> “1863 und 1876 wurde der Gebrauch des Ukrainischen in der Schule, im Druck von Büchern und Liedern und bei Theateraufführungen untersagt, ebenso auch der Import ukrainischsprachiger Druckerzeugnisse.”

1996: 129, 130, 130-131).<sup>15</sup> However, Ukrainian was adopted as an official state language in 1989, “the uneven development caused the current dichotomy in language use and language attitudes in Ukraine with a Ukrainian speaking west and a predominantly Russian speaking east.” (Bilaniuk, Taranenko in Kent, 2012: 4).

Nowadays in Ukraine there are no particular problems with communication in everyday life, according to Olszański: “spoken colloquial Ukrainian and Russian are mutually intelligible.” The standpoint that people should adopt the language that give them impression of being more comfortable in usage for them personally is widespread in sociological research. This could prove that the difficulties in oral communication are absent. (Olszański, 2012: 23). Even though Ukraine is considered a bilingual country with a common use in communication of both Ukrainian and Russian, especially in the eastern regions, it should be underlined that Russian spoken in Ukraine presents both phonetic and lexical differences when compared to Russian spoken in the Russian Federation. Some important particularities of the difference between the Russian language spoken in Russia and Russian used in Ukraine was underlined in 2014, during the war in eastern Ukraine. “Although Russia denied sending troops and argued that the separatist movement was run by local residents, observers frequently commented on the markedly non-local pronunciation of the militants.” Lexicon served as a proof of the Russian combatants involvement in the war. (Bilaniuk, 2017: 357-358). The analysis of the lexical aspect of the speaker could be useful in the process of identification of the speaker provenience.

### **1. 1. 6. The presence of dialects in Ukraine**

Ukrainian territory extends from north to south for 893 kilometers and from east to west

---

<sup>15</sup> “Typische Erscheinungen waren Umsiedlungen und Vernichtung der Intelligenz mit Begründung, sie betriebe nationalen Separatismus; in der Ukraine kam die von sowjetischer Seite provozierte Hungerkatastrophe von 1932/33 hinzu, die 3-6 Millionen (nach anderen Schätzungen 6-7 Millionen) Ukrainer das Leben kostete.”

for 1316 kilometers. (“Україна та її географічні дослідження”).<sup>16</sup> With such an extended territory, the country has developed regional nonstandard dialects that became a common cultural phenomenon with its serious and humorous aspects. (Bilaniuk, 2017: 355).

Dialect (*narichchia*) is a set of specific linguistic characteristics that do not appertain to the literary language. These features can concern phonetic, morphological, syntactic and lexical aspects of the language and are present in the oral speech of population living in a certain locality.<sup>17</sup> Ukrainian dialects are divided into sub-dialects, so called *pidnarichchia*, within which are defined different types of *hovir* and *hovirka*. The latter is the smallest dialectal unit. In Ukraine are present three main dialects, or *narichchia* (dialect groups) that are: Northern (or Polissian), Southwestern and Southeastern.<sup>18</sup> (Tokar, 1993: 30).

---

<sup>16</sup> “Максимальна протяжність нашої країни із заходу на схід (1 316 км) значно більша, ніж з півночі на південь (893 км).”

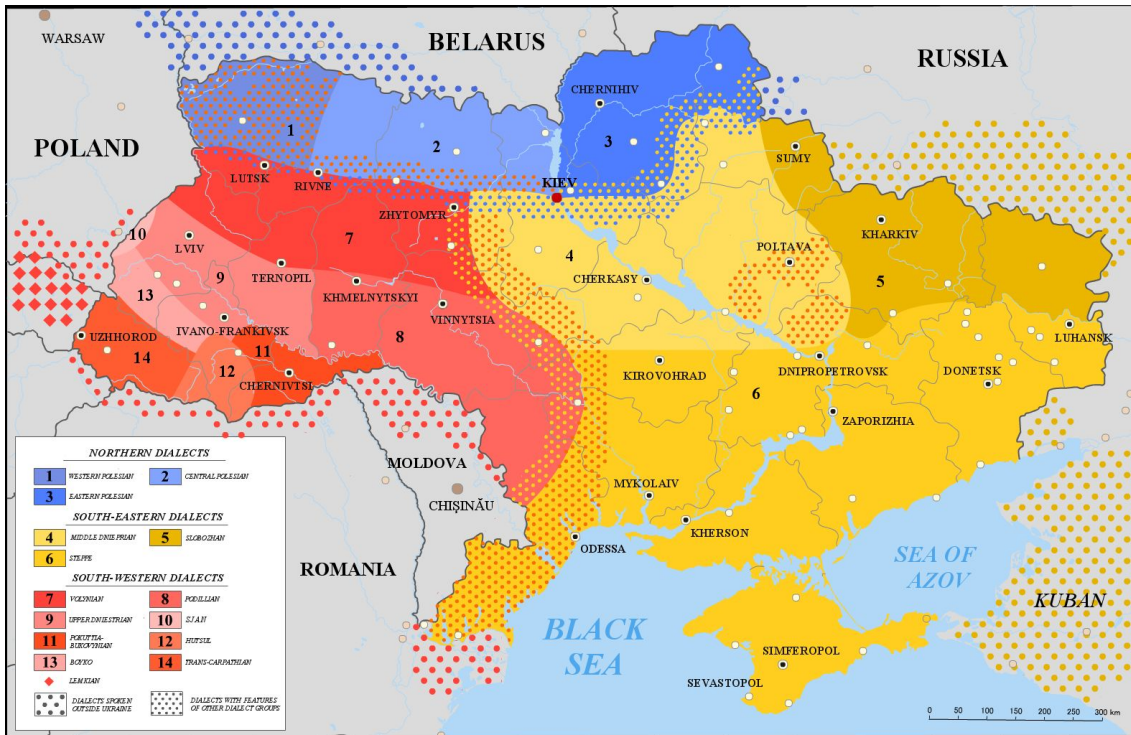
<sup>17</sup> “Діалектом (наріччям) називається сукупність специфічних мовних особливостей — фонетичних, морфологічних, синтаксичних, лексичних, що не властиві літературній мові, але характерні для усного мовлення населення, яке проживає у певній місцевості... Діалект членується на піддіалекти (піднаріччя), в межах яких виділяють говори і говірки. Говірка є найдрібнішою діалектною одиницею...”

<sup>18</sup> “Розрізняють три основні діалекти, або наріччя (діалектні групи), української мови: північний (або поліський), південно-західний і південно-східний.”



ILLUSTRATION 2.

Map of Ukrainian dialects.



Wikimedia, [commons.wikimedia.org/wiki/File:Map\\_of\\_Ukrainian\\_dialects\\_en.png](https://commons.wikimedia.org/wiki/File:Map_of_Ukrainian_dialects_en.png). Accessed 14 Jun. 2018.

The first group of dialects is Northern, also called Polissian, that is spread in the territory of Chernihiv, northern part of Sumy, Kyiv, Zhytomyr, Rivne and Volyn regions. Its southern boundary can be approximately defined as follows: Volodymyr-Volynskiy — Lutsk — Rivne — Novograd-Volynskiy — Zhytomyr — Fastiv — Pereiaslav-Khmelnytskyi — Pryluky and further on east following the River Seim. In the Northern *narichchia* are distinguished Left-bank Polissian (Eastern Polissian), Right-bank Polissian (Central Polissian) and West Polissian (Volyn-Podilian) groups of *hovir*.<sup>19</sup> (Tokar, 1993: 30).

The second group of dialects is the Southwestern dialect and it is present in the regions

<sup>19</sup> “Північний (або поліський) діалект поширений на території Чернігівської, північної частини Сумської, Київської, Житомирської, Рівненської, Волинської областей. Південна межа проходить приблизно так: Володимир-Волинський — Луцьк — Рівне — Новоград-Волинський — Житомир — Фастів — Переяслав-Хмельницький — Прилуки і далі на схід по р. Сейм. У межах північного (поліського) наріччя виділяють такі групи говорів: лівобережнополіські (східнополіські), правобережнополіські (середньополіські) і західнополіські (волинсько-подільські).

of Vinnytsia, Khmelnytskyi, Ternopil, Chernivtsi, Ivano-Frankivsk, Lviv, and Transcarpathian regions, as well as the northern part of the Mykolaiv region, northern and western parts of Odesa region, southern Volyn, Rivne and Zhytomyr regions.<sup>20</sup> (Tokar, 1993: 30). In the Southwestern group of dialects, also *narichchia*, “there are three groups of *hovir*: Podillia-Volynian (with the *hovir* of Podillia and Volyn); Galicia-Bukovynian (with the *hovir* of Upper Dnistran, or Galician; Upper Sannian; Pokuttia-Bukovynian; East Carpathian, or Hutsul); Carpathian (with the *hovir* of: Boiko, or North Carpathian; South Carpathian, or Transcarpathian; Lemko, or West Carpathian).”<sup>21</sup> The Southwestern group of dialects can be approximately delimited from another group of dialects, Southeastern, the crossing a line from north to south as follows: Fastiv — Bila Tserkva — the lower reaches of the Dnister River. (Tokar, 1993: 32).

The third group of dialects is the Southeastern dialect that, if compared to other dialects, encompasses the largest linguistic array. Its distribution is delimited in the southern part of the Sumy and Kyiv regions, on the territory of Kharkiv, Poltava, Dnipropetrovsk, Luhansk, Donetsk, Kirovograd, Zaporizhzhia, Cherkasy, Kherson, in the southern part of the Odesa and Mykolaiv regions. This Southeastern dialect is divided into the Middle Dniprian, Slobozhan and Steppe typologies of *hovirka*.<sup>22</sup> (Tokar, 1993: 33).

These various local dialects could be often combined with Russian language features. This phenomena could be repeatedly observed in the current speech practices of people

---

<sup>20</sup> “Південно-західний діалект охоплює території Вінницької, Хмельницької, Тернопільської, Чернівецької, Івано-Франківської, Львівської, Закарпатської областей, а також північну частину Миколаївської, північну і західну частину частини Одеської, південну Волинської, Рівненської Житомирської областей.”

<sup>21</sup> “Лінія розмежування південно-західного і південно-східного діалектів проходить приблизно так (з півночі на південь): Фастів — Біла Церква — нижня течія Дністра. Останнім часом у південно-західному наріччі (діалекті) виділяють три групи говорів: подільсько-волинські (з говорами подільськими і волинськими); галицько-буковинські (з говорами: наддністрянськими, або галицькими; надсянськими; покутсько-буковинськими; східнокарпатськими, або гуцульськими); карпатські (з говорами: бойківськими, або північно-карпатськими; південнокарпатськими, або закарпатськими; лемківськими, або західнокарпатськими).”

<sup>22</sup> “Південно-східний діалект охоплює найбільший мовний масив порівняно з іншими діалектами. Він поширений у південній частині Сумської і Київської областей, на території Харківської, Полтавської, Дніпропетровської, Луганської, Донецької, Кіровоградської, Запорізької, Черкаської, Херсонської, у південній частині Одеської та Миколаївської областей [...]. Цей діалект поділяється на три групи говірок: середньонаддніпрянська, слобожанська, степова.”

in different regions of Ukraine. Many would call it Surzhyk. (Bilaniuk, 2017: 356).

### 1. 1. 7. The question of Surzhyk

According to Serdeha and Saharovskiyi (2011: 41), the oral and written language of all Ukrainians, regardless of their education, profession or residence, is universal and it is manifested in two varieties or forms, called literary and dialectal languages. Between these two opposing varieties of the Ukrainian national language are verified other phenomena. These intermediate formations could denoted by the terms “folk-spoken language”, “*prostorichchia*” (vernacular language). As stated by Serdeha and Saharovskiyi, one of the signs of the Ukrainian *prostorichchia* is Surzhyk. “*Prostorichchia* is one of the varieties of a universal language, which, not being limited by territorial or narrow social terms, together with dialects and jargon is opposed to the literary language, its colloquial style. The basic form of the functioning of the *prostorichchia* is the supradialectal speech of urban residents (rarely - rural) who are poorly literate because of the lack of sufficient education.”<sup>23</sup> (Serdeha, Saharovskiyi, 2011: 41).

In the Hrinchenko’s dictionary *Slovar ukrainskoi movy* (1958: 231) a term “surzhyk” was defined as a mixed grain bread or mixed flour of rye and wheat, rye and barley or oats and barley and so on; as a second meaning the term could define a man of mixed

---

<sup>23</sup> “Усна і писемна мова всіх українців, незалежно від їх освіти, професії місця проживання, називається загальнонародною. Українська загальнонародна мова, як і багато інших мов, виявляється в двох різновидах або формах, які виступають під назвами літературна і діалектна мова. Між цими двома діалектично протилежними різновидами української загальнонародної мови є ще проміжні утворення, що позначають термінами «народно-розмовна мова», «просторіччя». Просторіччя – один із різновидів загальнонародної мови, який, не будучи обмеженим територіальними або вузькосоціальними рамками, разом із діалектами та жаргонами протиставлений літературній мові, її розмовному стилю. Основна форма функціонування просторіччя – наддіалектне мовлення міських жителів (рідше – сільських), які погано володіють літературними нормами через брак достатньої освіти. Наприклад, суржик є однією з ознак українського просторіччя.” Official website of V. N. Karazin Kharkiv National University, Department of Philology: [http://www-philology.univer.kharkov.ua/katedras/prof\\_sites/serdega/serdega\\_saharovski\\_dialect.pdf](http://www-philology.univer.kharkov.ua/katedras/prof_sites/serdega/serdega_saharovski_dialect.pdf). Accessed 17 Jun. 2018.

race.<sup>24</sup> In 1978 it was already possible to speak about a transferred meaning of “surzhyk”, its colloquial meaning delineated elements of two or more languages combined artificially, without observing the norms of written language, it was determined as an unclean language.<sup>25</sup> (*Slovník ukraïnskoï movy*, 1978: 854). A very detailed description of the term was given officially in the *Entsyklopediia Ukraïnska mova*:

“Surzhyk (literally — a mixture of rye, wheat, barley, oats, etc., and also the flour from this grain) — the language in which the elements of different languages are united artificially, without adhering to literary norms. Used predominantly in relation to the Ukrainian *prostorichchia*, littered with unmotivated borrowings (due to Ukrainian-Russian interference) of Russian elements. Surzhyk is deprived of national colors, beauty and expressiveness. Mostly used in everyday language, could be also found on the pages of newspapers and magazines, books and brochures. A fight against Surzhyk is one of the most important tasks in order to improve the culture of the Ukrainian language. In literary works, its usage is considered a stylistic device that characterises a character individually or that provides the comical, ironic effect.” (*Entsyklopediia Ukraïnska mova*, 2000).<sup>26</sup>

Nowadays Surzhyk is very often characterised as a destructive phenomenon, because of its negative impact on the Ukrainian language.

In general it can be distinguished between two types of Surzhyk: Ukrainian-Russian Surzhyk and Russian-Ukrainian Surzhyk. The Ukrainian-Russian Surzhyk is defined as the language used by people with Ukrainian as L1<sup>27</sup> that were trying or were forced to

---

<sup>24</sup> “Суржик, -ка, м. Смѣшанный зерновой хлѣбъ или мука изъ него, напр. пшеница съ рожью, рожь съ ячменемъ, ячень съ овсомъ и пр. 2) Человѣкъ смѣшанной расы.”

<sup>25</sup> “1.) Суміш зерна пшениці й жита, жита й ячменю, ячменю й вівса і т. ін.; борошно з такої суміші. 2.) перен., розм. Елементи двох або кількох мов, об’єднані штучно, без додержання норм літературної мови; нечиста мова.”

<sup>26</sup> “СУРЖИК (букв. — суміш жита з пшеницею, ячменю з вівсом і т. ін., а також борошно з такого зерна) — мова, в якій штучно об’єднані без дотримання літ. норм елементи різних мов. Ужив. переважно щодо укр. просторіччя, засміченого невмотивовано запозиченими (внаслідок укр.-рос. інтерференції) рос. елементами. С. — це збіднена мова, позбавлена нац. колориту, краси й виразності. Найпоширеніший у побут. мовленні, звідки проникає на сторінки газет і журналів, книжок і брошур. Боротьба з С. — одне з гол. завдань у галузі підвищення культури укр. мови. У худож. мові С. — стиліст. засіб типізації та індивідуалізації персонажів, створення коміч., іроніч. ефекту.” Available at <http://litopys.org.ua/ukrmova/um107.htm>. Accessed 18 Jun. 2018.

<sup>27</sup> L1 - first language, mother tongue.

speak Russian due to the Russification policies in Ukraine. This period is limited between the end of the 17th century and the beginning of the 18th century. The introduction of the Russian language in administration and in the educational system led to the creation of figures of scribe, deacon, seminarian and teachers with a specific spoken language. In this spoken language, the elements of Old Church Slavonic, the plates and forms of Russian chancery and Russian literary expressions were combined with phonetic basis and lexical and morphological elements of the Ukrainian language. (Masenko, 2008: 5).<sup>28</sup>

The Russian-Ukrainian Surzhyk defines the language spoken by people with Russian as L1 and who had tried to speak Ukrainian because of Ukrainisation. (Masenko, 2008: 3). In the 1920s the prestige of the Ukrainian language raised. The process of Ukrainisation that regarded administrative/managerial sphere, education, and the mass media increased the status of Ukrainian. Moreover, the new policy slowed down the proceedings of ethnic and social marginalisation of the provincials, that were used to be the main social group that contributed to the distribution of Surzhyk before. People that were Ukrainian native speakers started to firmly adhered to their language and, on the contrary to the past periods, did not adapt to the Russian-speaking environment.<sup>29</sup> (Masenko, 2008: 11). Consequently, the number of speakers that did not adopt Russian language in the big Ukrainian cities increased.

The next wave of the Surzhyk development was registered between 1930 and 1970 due to the increasing of Russification policies. Although in the the early 1990s the Russian language had more prestige, after the law on the state language was adopted and after the independence was achieved, the situation in Ukraine has changed. Ukrainians, who

---

<sup>28</sup> “Впровадження російської мови в адміністративно-управлінську сферу й освіту спричинило формування в середовищі місцевого чиновництва — канцеляристів, писарів, дяків, а також бурсаків і шкільних учителів специфічного макронічного усного мовлення, в якому елементи старослов'янщини, кліше і штампи російського канцеляриту й літературні звороти химерно сполучались з фонетичною основою й лексичними та морфологічними елементами української мови.”

<sup>29</sup> “Українізація в 1920-х роках адміністративно-управлінської сфери, освіти, в тім числі вищих навчальних закладів, засобів масової інформації підняло престиж української мови і загальмувало процеси етнічної й соціальної маргіналізації провінціалів, які перед тим становили основну соціальну базу поширення суржикового мовлення. В містах з'являються носії української мови, які твердо дотримуються своєї мови не пристосовуючись до російськомовного оточення.”

earlier spoke in Russian, tried to express themselves in Ukrainian, creating in this way a Russian-Ukrainian Surzhyk. As a very clear example of Russian-Ukrainian Surzhyk, one can analyse the speech of Ukrainian politician Viktor Yanukovich grown up in a Russian-speaking environment in Donbas. Ex president of Ukraine tried to speak Ukrainian, but he failed repeatedly: the language he used to express himself in public presented marked elements of Surzhyk.

Statistical studies show the decrease in the diffusion of Surzhyk, the results of survey demonstrate that the number of Surzhyk-speaking people decreased from 18.2% of the overall interviewees in 1996-1999 to 14.7% in 2000-2003. (Khmelko, 2004 in Masenko, 2008: 25).<sup>30</sup> Nevertheless the problem consists in the fact that, as seems, Surzhyk-speaking people are not always aware of using Surzhyk. Moreover, it is often difficult that during the sociological research Surzhyk-speakers admit to be such. (Olszański, 2012: 21). This is exactly the reason why more studies on Surzhyk should be done. However, Surzhyk prestige is remarkably low both from the point of view of experts and of that of common public (Masenko, 2008: 25),<sup>31</sup> this spoken language is commonly used in the whole Ukraine. The territories of the country where was registered the largest number of Surzhyk speakers were eastern, southern and central parts of Ukraine. (Vakhtin et al., 2003 in Kent, 2010: 43).

According to Olszański, it could be admitted that Surzhyk is a Ukraine-specific language phenomenon,<sup>32</sup> enabled by the co-existence of two mutually intelligible languages in the same territory. (Olszański, 2012: 12). But, relying on the Lewis' classification of languages' mutual intelligibility, Russian and Ukrainian result to be mutually unintelligible languages. (Kent, 2012: 11). The current situation of mass

---

<sup>30</sup>“Згідно з узагальненими по всій Україні даними соціологічного дослідження лінгвістичної структури України, “упродовж 1996-1999 років частка населення, що вживає суміш української і російської мов, становила 18,2 %, а впродовж 2000-2003 років — 14,7%.”

<sup>31</sup>“В перспективі, за умови впровадження послідовного мовного планування, зокрема українізації системи освіти, яка вже здійснена в більшості регіонів, мішана субмова може звужити ареал побутування, поступившись місцем українській літературній мові та координативному білінгвізмові. За наявності нормативних стандартів обох контактних мов мішаний субкод не має шансів на стабілізацію і більш-менш тривале існування. Його престиж надзвичайно низький не тільки в експертній оцінці, а і в масовій свідомості.”

<sup>32</sup> Trasianka present in Belarus has a similar nature as Surzhyk.

media and book publishing in Ukraine demonstrates that Ukrainian and Russian, followed by English, are the languages in prevalence in Ukraine, they are both requested, that also could mean that the languages in question are not mutually intelligible. Recently, on the official website of State Statistics Service of Ukraine was displayed statistical data on mass media and book publishing in Ukraine regarding 2017.<sup>33</sup> According to the State Scientific Institution Book Chamber of Ukraine named after Ivan Fedorov that provided statistical data, the number of printed publications and printed titles (books, brochures) in Ukrainian is 15,789 and in Russian is 3,886. When it comes to the circulation of publications, the number of copies in Ukrainian is 32,859.7 thsd. whereas in Russian 9,439 thsd.<sup>34</sup> It should be underlined that data exclude the temporarily occupied territory of the Autonomous Republic of Crimea, the city of Sevastopol and a part of temporarily occupied territories in the Donetsk and Luhansk regions where the Russian-speaking population is in prevalence. The voice on the distribution of newspapers by type of publication language reveals that the annual circulation in thousand of copies is 945,569.7 in Ukrainian and 1,116,394.6 in Russian.<sup>35</sup> The presence of both Ukrainian and Russian languages in the press, television and radio should guarantee a good level of command in Ukrainian and Russian but what happens in practice is that these two languages interact and interference each other.

---

<sup>33</sup> Available on the official website of State Statistics Service of Ukraine <http://www.ukrstat.gov.ua/> under the voice “Засоби масової інформації та книговидавництва в Україні у 2017 році” or as a download with a direct link:

[http://www.ukrstat.gov.ua/operativ/operativ2012/cltr\\_rik/zas\\_mas\\_inf\\_knyg/zas\\_mas\\_inf\\_knyg\\_17.xls](http://www.ukrstat.gov.ua/operativ/operativ2012/cltr_rik/zas_mas_inf_knyg/zas_mas_inf_knyg_17.xls). Accessed 19 Jun. 2018.

<sup>34</sup> Data available in the previous document under the voice *Distribution of books and brochures by language in 2017*, 10.

<sup>35</sup> Data available in the previous document under the voice *Distribution of newspapers by type of publication language in 2017*, 8.

## **1. 2. Hybrid languages, from language interaction to language mixing**

Language interference is not a new argument in today's society where many languages are concentrated on the same area due to globalisation. Multilingualism became one of the main characteristics of the end of 20th and beginning of 21st century. In the European Union moving from country to country for educational, employment or personal reasons became a routine and together with these transfers, the number of multilingual families continues to increase. At the higher level, in the areas of official or unofficial multilingualism, it is inevitable that the languages involved in the communication come in contact and somehow influence each other. Moreover, "not only can one language take over elements from another one, but an entirely new language can emerge in situations of language contact." (Appel, Muysken, 1993: 175). In this chapter, there is firstly a description of the range of phenomena that characterise the contact of two languages and the way they are juxtaposed in discourse or within a sentence. Lastly, possible classifications of Surzhyk that were already expressed by the previous studies on Surzhyk are presented.

### **1. 2. 1. Bilingualism and language interference**

It is common that language contact leads to bilingualism, generally distinguished between societal and individual type. The first type of bilingualism characterises a society where two or more languages are spoken,<sup>36</sup> the second one is individual, even though determining whether a person is bilingual or not is not simple. The two extreme definitions of bilingualism are those of Bloomfield and Macnamara. (Appel, Muysken, 1993: 1). According to Bloomfield, a bilingual individual should possess a control of two or more languages that is native-like. (1933: 56 in Appel, Muysken, 1993: 2). Macnamara proposed to identify speakers as bilingual when, in addition to the first-language skills, they possess second-language skills in one of the four modalities that can be speaking, listening, writing and reading. (1969, in Appel, Muysken, 1993: 2). Another criteria for bilingualism distinction proposes the definition of monolingual

---

<sup>36</sup> In Appel and Muysken "the terms bilingual and bilingualism also apply to situations where more than two languages are involved." (1993: 3).



and bilingual mode. Grosjean distinguishes between the monolingual mode and the bilingual mode. The first one takes place when bilinguals adopt one language with monolingual speakers of that language, while the second mode occurs when bilinguals are together and have the possibility for codeswitching. What happens in the monolingual mode is that bilinguals may occasionally mix languages, this mixing may be often provoked by the dominant language that influences the less dominant. This type of influence was defined interference, the term transfer is also used and sometimes preferred. (Grosjean, 1992 in Malmkjaer, 2004: 69).<sup>37</sup>

Within the concept of interference could be delimited two main typologies, Grosjean, for example distinguishes between static and dynamic interference. From one side, “static interference describes the relatively permanent influence of one of the bilingual’s languages on the other”, while “dynamic interference recognises the features from one language are transferred temporarily into the other.” In the static interference may be present accent, intonation and the pronunciation of the individual sounds whereas in the dynamic interference the transfer can concern syntax, phonology or vocabulary in both written and spoken language. However, the term “interference” is seen negatively by most bilinguals as it reveals a monolingual, “fractional” perspective, switching between languages may be often necessary. Recently, the codeswitching between two languages has become the norm in many bilingual situations. (Grosjean 1992, in Malmkjaer, 2004: 69).

### **1. 2. 2. Language alternation: from codeswitching through language mixing to fused lect**

The way two languages are juxtaposed in discourse or / and within a sentence may be explained through the language alternation phenomena expressed as continuum.

---

<sup>37</sup> Partially available at:

[books.google.it/books?id=agODAgAAQBAJ&pg=PA69&lpg=PA69&dq=Grosjean+1992+language+interference&source=bl&ots=R6zaLYK0x2&sig=iRzYc8zDINTB9QxzZoAqxMVoKL0&hl=uk&sa=X&ved=0ahUKEwjiasLqT8eHbAhVCPBOKHeluAxxkO6AEIKDAA#v=onepage&q=Grosjean%201992%20language%20interference&f=true](https://books.google.it/books?id=agODAgAAQBAJ&pg=PA69&lpg=PA69&dq=Grosjean+1992+language+interference&source=bl&ots=R6zaLYK0x2&sig=iRzYc8zDINTB9QxzZoAqxMVoKL0&hl=uk&sa=X&ved=0ahUKEwjiasLqT8eHbAhVCPBOKHeluAxxkO6AEIKDAA#v=onepage&q=Grosjean%201992%20language%20interference&f=true)  
[/https://books.google.it/books?id=hDqCAgAAQBAJ&printsec=frontcover&hl=uk&source=gbs\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.it/books?id=hDqCAgAAQBAJ&printsec=frontcover&hl=uk&source=gbs_summary_r&cad=0#v=onepage&q&f=false). Accessed 19 Jun. 2018.

According to Auer (1999: 309), the language alternation continuum can be presented as a range between three cases: codeswitching, language mixing, and fused lects. Codeswitching and fused lects represent the polar extremes of continuum and language mixing is located between them. The process of development proceeds from codeswitching through language mixing to fused lect.

#### **1. 2. 2. 1. Codeswitching**

Codeswitching defines the pragmatic pole of the Auer's continuum. The contrast between one code and the other that are involved in the process of codeswitching is meaningful. These two codes may be languages, as well as dialects or other varieties and the action of codeswitching does not presuppose deep knowledge of the language from which the insertion originates. (Auer, 1999: 310, 313, 314).

The reason of the codeswitching usage can vary in base of the situation. Speaker can decide to adopt it with the purpose of emphasizing a particular word or phrase, substituting a not known word or phrase, expressing a concept without equivalent, reinforcing the request with switching to a majority language in a situation of majority / minority language situation, to clarify a point, to change the attitude of relationship or to create a distance, etc. The last aims of the codeswitching usage are common for example for people who meet and, initially, use the common majority language but, as the conversation proceeds and personal information is revealed (for example: status and ethnic identity), may switch to a regional language. This codeswitching may signal the crossing of boundaries and lessening of social distance. On the contrary, a change from minority language or dialect to majority language may demonstrate the speaker's wish to create a distance from the listener, or establish a more formal conversation or relationship. (Malmkjaer, 2004: 70-71).

#### **1. 2. 2. 2. Language mixing**

Besides codeswitching there are many other cases of language contact and juxtaposition of two languages, one of them is mixing (of languages, dialects, etc.). (Auer, 1999:

314). However the definition of the “mixed language” concept is not a simple one, Bakker and Mous proposed “the term “language intertwining” for the process forming mixed languages showing a combination of the grammatical system (phonology, morphology, syntax) of one language with the lexicon of another language.” (1994: 4-5). Relying on the discussions that took place during the workshop on mixed languages in Leiden,<sup>38</sup> Bakker asserts that the most common type of the mixed language is the one that takes the lexicon of one language and combines it with the grammatical system of another language. A language could be defined as a mixed one if it could be ascribed to two different language families with equal rationalisation. For example if the mixed language in question is a mixture of two languages, A and B, the speakers of both languages can assert that this mixture is their language. (Bakker, 1994: 26-27). According to Golovko, most mixed languages are composed by some already formed parts of grammar systems of the two participating languages unified to the lexicon of the other language. The grammar system may adopt a simplified grammatical structure of one of the two languages, which grammatical elements may be further replaced by the ones of another language involved in mixing. (Golovko, 1994: 119). A mixed language takes its lexicon from one source and grammar from another, “on the basis of the lexicon one would classify such languages as belonging to one language family and on the basis of the morphology, syntax and general grammatical characteristics one would classify them as belonging to another language family.” (Bakker, Mous, 1994: 5). Some of the languages that were classified as mixed could be: Media Lengua that combines Quechua grammar to Spanish lexicon, Callahuaya that unifies Quechua grammar to Puquina lexicon, Michif that mixes Cree (Algonquian, Amerindian) verbs to the French nouns. In the presented cases the source languages belong to different language families or branches which allows to distinguish the provenience of each individual morpheme. (Bakker, 1996: 10-13).

Researchers who confronted the language mixing with the codeswitching claimed that

---

<sup>38</sup> The workshop in Netherlands was organised by Peter Bakker (University of Amsterdam) and Maarten Mous (Leiden University) on the 24th of May, 1993. The linguists working on mixed languages were invited to Leiden and contributed to the creation of the volume *Mixed languages: 15 Case Studies in Language Intertwining*, edited by Bakker P. and Mous M.

between these two phenomena could be delineated a clear structural difference. (Bakker, Mous, 1994: 9). Moreover, if compared to codeswitching, language mixing seems to require a higher bilingual competence and there is a hypothesis that there is a tendency to move from codeswitching to language mixing, but not vice versa. It has to be noticed that Auer does not claim that moving from codeswitching to language mixing is the only way in which a mixed speaking style may emerge, but only that moving from language mixing to codeswitching is excluded. According to Auer, a possible counterexample could be a bilingual speech that is subject to conscious effort of purification. In the transition phase from codeswitching to language mixing that may be quite prolonged, the older codeswitching pattern and the newer language mixing pattern coexist. (Auer, 1999: 318, 319).

In general people's attitudes towards language mixing is negative. Whereas the behaviour of bilinguals and monolinguals regarding mixing may vary from strong positive or (more often) negative reaction, among the linguists, mixing can provoke strong reactions, usually of the negative kind. (Auer, 1999: 318).

### **1. 2. 2. 3. Fused lect**

The final point of the language alternation continuum delineated by Auer is expressed by the phenomena of fused lects. This term is taken by Auer from Matras (1996). There is mainly a grammatical difference between mixing and fused lects. Apparently, a fused lect may resemble language mixing and the difference may be noticed only when analysed at a deeper grammatical level. The language mixing phenomena allows variations in juxtaposition, this means that two codes may be juxtaposed, but they need not be. Contrarily, in fused lects the use of one code or the other for certain constituents is obligatory, it results to be part of their grammar, and speakers have no choice for a variation. The fused lects grammar results to be simplified if compared to the two languages that are involved, language A and language B. Additionally, the structures provenient from language A and B that are more or less equivalent in monolingual use, when transported to the fused lects AB, may develop specialised uses. Moreover, fused

lects may have to adapt their structure to the combination of elements from language A and B. This adaptation is executed through the development of the new structures that are identical neither to those of A nor of B language. According to Auer many of the European Romani dialects may be classified as fused lects and one of them is Gadschkene dialect<sup>39</sup> spoken in Germany that emerged from Romani and German bilingualism. (Auer, 1999: 321, 322). The extract from Holzinger (1993: 324 in Auer, 1999: 322) provided by Auer as example of this fused lect, point out the presence of German nominal, verbal, and adjectival stems partly integrated into an older Romani morphology. Besides, there are also diverse German particles, conjunctions, and adverbials that are inserted into Romani and have no correspondence in this variety of Romani language. The case demonstrates that the German elements are obligatory in Gadschkene, they are part of its grammar and lexicon. This characteristic is also valid for others Sinti dialects. (Auer, 1999: 323).

Generally, speakers of a fused lect may, but need not to be proficient speakers of the involved languages A and / or B. Mixed languages are good candidates for fused lects.<sup>40</sup> However fused lects may certainly have other origins in addition to language mixing, the case of fused lects that could be traced back to language mixing is common. It is rather impossible that fused lect can develop towards language mixing. The transition between language mixing and fused lect could be investigated through the analysis of their origin or by the definition of the mechanisms that solicit this transition. (Auer, 1999: 321, 324). During the research on Copper Island Aleut that is a fused variety of Russian and Aleut, Golovko (1994: 118 in Auer, 1999: 324) admits that in the emergence of a mixed language, codemixing is necessary but not sufficient.

---

<sup>39</sup> The varieties of Romani languages, in particular the Sinte Romani family that includes Gadschkene dialect could be visualised in Glottolog 3.2. elaborated by Hammarström, Harald & Bank, Sebastian & Forkel, Robert & Haspelmath, Martin available at: [glottolog.org/resource/languoid/id/gads1256](http://glottolog.org/resource/languoid/id/gads1256). Accessed 17 Jul. 2018.

<sup>40</sup> According to Auer (1999: 322) one of the best examples is the language of the Métis Buffalo hunters in Canada - Michif. It was researched and classified by Bakker (1994) as the Cree-French mixed language. If we suppose that Michif may develop from mixed language to fused lect, then it may be possible that Surzhyk could also follow this continuum. The idea of Surzhyk as a mixed lect was already proposed by Kent (2012). At the current stage of Surzhyk development I would say that it could be premature to classify its hybridity as fused lect but I cannot exclude that one day it may be possible.

## 1. 2. 2. Surzhyk classification

### 1. 2. 2. 1. Previous classifications of Surzhyk

Previous studies on Surzhyk expressed its possible classifications. In this part of the chapter diverse interpretations of Ukrainian-Russian spoken language are present.

- **Taranenko: hybrid language / mixed sociolect / component of Ukrainian *prostorichchia***<sup>41</sup>

According to Taranenko, Surzhyk is a codemixing and not codeswitching.<sup>42</sup> On the Auer's continuum that extends from codeswitching through codemixing to fused lects, Surzhyk is located in the middle of continuum but not as a mixed language but rather as a mixed sociolect.<sup>43</sup> Ukrainian-Russian Surzhyk could be seen as a sociolect, as a component of the Ukrainian *prostorichchia* with admixture of russianisms. Its emergence is linked to the prolonged Ukrainian-Russian bilingualism in its asymmetric, diglossic form.<sup>44</sup> Further mass distribution of this hybrid language may have a destructive character.<sup>45</sup> (Taranenko, 2008: 20, 21, 16, 27).

- **Del Gaudio: bilingual mixed language / partial mixed language**

According to Del Gaudio, Surzhyk is a mixed language, if considered in relation to contemporary Russian and Ukrainian. Similarly to any other bilingual mixed language, Surzhyk is the result of widespread bilingualism, and it basically involves Russian and

---

<sup>41</sup> As mentioned in the previous chapter "*prostorichchia* is one of the varieties of a universal language, which, not being limited by territorial or narrow social terms, together with dialects and jargon is opposed to the literary language, its colloquial style. The basic form of the functioning of the *prostorichchia* is the supradialectal speech of urban residents (rarely - rural) who are poorly literate because of the lack of sufficient education." (Serdeha, Saharovskiy, 2011: 41).

<sup>42</sup> "Суржик - змішування мовних кодів, а не їх перемикання."

<sup>43</sup> "За поділом процесів взаємодії мов на етапи від перемикання кодів через змішування мов до їх "сплавлення" суржик, таким чином, перебуває посередині цього континууму, але не як змішана мова, а як змішаний її соціолект."

<sup>44</sup> "УРС як соціолект - як певний компонент українського просторіччя з досить помітною домішкою русизмів, що давно вже став розмовним мовленням і фактично основним засобом спілкування більшості україномовного населення країни як у містах, так і в сільській місцевості внаслідок тривалої масової українсько-російської двомовності в її асиметричній (диглосній) формі."

<sup>45</sup> "Масове поширення цього мовного гібриду є деструктивним як для подальшої долі української мови, так і для її носіїв."

Ukrainian and their dialectal varieties. It could be assumed that the residue of the 18th century language poly-codedness has been maintained at an oral level: similar Russian forms, Church Slavonic, archaic and dialectal Ukrainian elements - all together have contributed to the creation of a hybrid language. Del Gaudio defines Surzhyk a hybrid language in relation to the standard, normative Ukrainian or Russian languages. The author proposes a theory on the prototype Surzhyk nucleus that is composed of oral dialectal, and archaic forms of Ukrainian with a certain percentage of Russian admixture mainly in vocabulary. If this theory could be accepted, then it would be justified to consider Surzhyk a partially mixed language. (Del Gaudio, 2010: 234, 209, 250).

Additionally, Del Gaudio premises that the hypothesis of triglossic<sup>46</sup> traits which characterise part of Ukrainian sociolinguistic habits may be promoted. In this situation the extremely formal language, also a very high variety, would be Ukrainian, Russian would be considered as a high variety due to its ordinary communicative function at a semi-formal level, whereas Surzhyk could have a role of the very low variety as characterising informal communicative situations. In this hypothesis the various degrees of interference and marked Russian or Ukrainian varieties could be considered as the low variety. (Del Gaudio, 2010: 259).

- **Kent: Russian-Ukrainian mixed lect**

Although in 2010 (51) Kent assumes that from the morphosyntactic perspective Surzhyk could be defined as composite codeswitching, in 2012 she claims that Russian-Ukrainian codeswitching is not Surzhyk.<sup>47</sup> The main reason is that in codeswitching a passage from one code to another could not be predicted, whereas in

---

<sup>46</sup> In linguistics it is common to speak about diglossia as a phenomena that characterises the usage of two or more varieties of the same language in the same speech community. In the presence of diglossia there are a High variety, which has more prestige and is regarded as superior, and a Low variety, said to be an inferior variety. Diglossia phenomena is described in Ferguson, 1959: 325-340.

<sup>47</sup> Kent adopts the term “Russian-Ukrainian Surzhyk” in reference to the Ukrainian-based Surzhyk with admixture of Russian elements. In the samples she analysed she did not find Russian-based Surzhyk. (Kent, 2012: 30). Though, in the present research I use the term “Ukrainian-Russian Surzhyk” in reference to the Ukrainian-based Surzhyk and “Russian-Ukrainian Surzhyk” to denote the Russian-based Surzhyk.

Surzhyk it is possible to observe an internal structure and to predict which elements may be supplied by Russian and which by Ukrainian. Kent concluded that in base of its structural features, Surzhyk could be best classified as a mixed lect. (Kent, 2012:7). In Auer's continuum mixed lect "is a speech of bilinguals that contains elements of two or more languages, has internal structure but is not stable enough to be considered a mixed language" and its speakers are "reportedly near-balanced bilinguals." (Auer, 1999 in Kent, 2012: 7, 144). The older population of Kent's research participants who spoke Surzhyk were not active bilinguals in Ukrainian and Russian, even though they understood both languages. It can be assumed that from a structural point of view Surzhyk presents some vocabulary substitution, it behaves as a single code, and it demonstrates systematic inter- and intra-clausal switches. According to Kent, "even though sociolinguistic evidence does not support the analysis of Surzhyk being a mixed lect, grammatical evidence is more important than sociolinguistic" (Kent, 2012: 144, 146).

In the earlier work of Kent (2010) Surzhyk is subject to the Matrix Language Frame model (MLF) testing. The MLF model was elaborated by Myers-Scotton (2002, 2003) for analysing a mixed language variety. In the MLF the matrix language is the only source of the abstract morphosyntactic frame in a bilingual clause. The other participating language is the embedded language and it must agree with structural requirements stipulated by the matrix language (Myers-Scotton 2002, 2003 in Kent, 2010: 41). According to Flier, the structural properties of the Ukrainian-Russian Surzhyk are "lexical transfers, lexical extensions, lexical calques, syntactic calques, morphological selection, morphological attraction, phonological integration" (Flier, 2008: 45) and following the MLF, in the case of Ukrainian-Russian Surzhyk, Ukrainian results to be a matrix language while Russian an embedded language.

- **Zaporozhets: Surzhyk as a variety of idiom**

According to Zaporozhets, it is inappropriate to consider Surzhyk as a dialect, *prostorichchia*, jargon or slang as all these phenomena are stratification markers of the



same language. The concept of Surzhyk consists in the correlation of two closely related languages, it is a bilingual phenomena. Surzhyk could be seen as a variety of idiom, a two-component idiom that has a national specific feature based on the contact of Ukrainian and Russian and is characterised by the numerous violations of the codified norms of both languages. Zaporozhets admits that “idiom” is not only a general term used for the denomination of different language formations, such as language, dialect, *hovir*, literary language and its variations, but also a term that could be adopted in the conditions where, besides the state language, there are other languages and such situation causes difficulties in the definition of the exact linguistic status of the phenomena in question. The latter meaning of the “idiom” concept is quoted by Zaporozhets from the Encyclopedic Dictionary of Linguistics of 1990.<sup>48</sup> (Zaporozhets, 2011: 149, 150).

It could be noted that studies on Surzhyk produce different opinions not only on its status but also on its classification. Most of them analyse diverse aspects of this phenomena and present convincing argumentations. In the following subchapter is described the way Surzhyk is denominated in this research.

#### **1. 2. 2. 2. Surzhyk as a hybrid language with the characteristics of mixed language**

Since the definition of Surzhyk classification is not the purpose of this research, we will consider Surzhyk as a hybrid language that presents mixed language characteristics. Its hybridity may be explained by the involvement not only of Ukrainian and Russian languages but also of the different dialects present in the Ukrainian territory. Moreover, it can vary depending on the social affiliation of the speakers, their age and language experience.<sup>49</sup> According to the criterion aimed to define mixed language nature provided by Bakker in 1992, “a language is mixed if it can with equal justification be assigned to two different language families.” Following the argument of Bakker on the case of

---

<sup>48</sup> ЛІС, 171.

<sup>49</sup> For language experience I certainly mean language proficiency of the speakers, as well as the dominant language in their family and professional circles. Here we could also consider the language in which the speaker normally reads, as it could be common for Ukrainian speakers to opt for a book in Russian if it is available for free instead of buying a Ukrainian version. Not to be underestimated also the language of music that speaker is used to listen.

Michif, the Cree-French mixed language of the Métis, where “both French and Cree speakers may say it is their language when they recognise roughly half of what is said, but they have to admit that much of it remains unintelligible”, one can see some similarities with the phenomena of Surzhyk. The criterion proposed by Bakker in 1992 may be applied to the cases where “the lexicon is from one language and the grammatical system (phonology, morphology and syntax) from another” (Bakker 1992 in Bakker, 1994: 26) and Surzhyk seems to be the case, especially if we consider that the process of “intertwining” mixes lexicon and grammar of different codes and leads to the creation of the new language that typically presents stems from one language and affixes from another. (Bakker, 1996: 18). In the current research the idea of the Matrix Language Frame model (MLF) elaborated by Myers-Scotton (2002, 2003) and proposed by Kent for Surzhyk analysis (2010) is taken into consideration. The nature of Surzhyk is complex: in Ukrainian-Russian Surzhyk structure, Ukrainian is a matrix language whereas Russian is an embedded language. In addition it has a huge variety of possible mutations in base of the geographical area and speaker’s characteristics (age, dominant language, type of education). It may be supposed that in the nature of Ukrainian-Russian Surzhyk there are two, or even three types of mixing. The primary mixing could regard the mixing of Ukrainian and Russian languages and define the basic Surzhyk structure and could be considered more or less fixed. Then the secondary mixing may involve Ukrainian dialects. Considering the number of dialects present in Ukraine, it could be assumed that the secondary mixing adds further mutations to the already existing primarily mixed language. However this second phase of mutations may vary in base of geographical area; if analysed together with the dialects main characteristics, a general list of rules could be identified. The third mixing passage may regard the speaker. This third phase could be called personalisation, as the spoken language, that has already undergone previous mutations, is subject to the personal style of the speaker’s speech. Personalisation of the spoken language does not only depend on the age of speakers and their social affiliation but also on the impression they aim to give when speaking the language in question. These mutations, or additional mixing, may consist in further lexical insertions or phonological deviations. In order to establish

if these three-level mixing could be confirmed or not, one may try to stratify the samples of spoken Ukrainian-Russian Surzhyk and sort the elements in base of their affiliation to Ukrainian or Russian standard language, dialects (if so, establish also which one), style markers. Such type of systematic corpus-based analysis would require depth description of every element and, if carried on long term, could enable to demonstrate the percentage of every category of elements present in the samples. Being Russian and Ukrainian languages of the same branch of language family (East-Slavic), it may happen that in several cases it could be difficult to establish the affiliation of some archaic forms, however these forms could be grouped and analysed afterwards.

### **1. 3. Natural Language Processing**

#### **1. 3. 1. NLP as a field of studies**

Natural language processing (NLP) is an area of computer science that deals with research on how computer systems can analyse, understand or produce natural languages. Natural languages are all possible human languages and they can be available as text, spoken language, or keyboard input. (Allen, 2003: 1218). NLP was founded on the basis of a number of different research areas such as computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and psychology. (Chowdhury, 2003: 51). Natural language processing can test whether a computer system seems to understand the natural language or not. This test could be verified when analysed whether computer system can successfully perform its tasks that may vary between translation from one language to another, summarisation of the text and others. The classical model of such testing was proposed by the British mathematician Alan Turing (Allen, 2003: 1218) whose 1936 model of algorithmic computation is regarded by many as the foundation of modern computer science. (Jurafsky, Martin, 2000: 10).

#### **1. 3. 2. The meeting point of different research areas**

Throughout its history, the behaviour towards speech and language processing varied according to the field of study that could be computer science, electrical engineering, linguistics, and psychology. The mechanism of speech and language processing requires the involvement of different overlapping fields of studies in the following areas: computational linguistics in linguistics, natural language processing in computer science, speech recognition in electrical engineering, computational psycholinguistics in psychology. (Jurafsky, Martin, 2000: 6, 10). Depending on the researcher's formation, for example, one could distinguish between natural language processing and computational linguistics, that deal the same problem but from different points of view. The belief that computers could process language as humans do is based on the fact that the use of language is linked to the people's general cognitive abilities. The

computational implications of this connection were firstly considered, among others, by Turing.

## **2. 1. The Turing test**

In 1950 Turing proposed to consider the idea of question on whether machines could think.<sup>50</sup> A possible reply to the research question could be given through the process of a game called “imitation game” that would involve three players. Firstly, it was proposed to consider three human players: the first one a man (A), the second one a woman (B), and the third one an interrogator (C) that could be of either sex. During a game the interrogator is located in a room apart front the other two players and he does not know which one of the players is woman and which one is man. His task is to determine this by making questions, obviously he could also ask other two players if they are man or woman, but it was supposed that one of the players (A) had the task to cause the interrogator to make a wrong identification while another player’s (B) charge was to help the interrogator. In such conditions both of the players could answer that they are woman by adopting their game strategy (A lies whereas B says truth) confusing in such way the player C. In order to avoid that the voices of man or woman help the interrogator in his final decision, it was proposed to present written or typewritten answers. It was expressed that the ideal arrangement would be to have a teleprinter that could communicate between the two rooms, in alternative, the figure of intermediary could be adopted in order to repeat the question and answers. The subsequent question of the research was about what could happen if the part of A would be taken by a machine acting as a man. (Turing, 1950: 433-435). In this game, called Turing’s game, the interrogator, that is a human player, has to determine which of the other players is the machine. The task of the machine is to make interrogator think that it is a person, while the human player’s role is to convince the interrogator that the other participant is machine while she is human. (Jurafsky, Martin, 2000: 7). It was concluded that if the interrogator could determine a machine to be a human player, then it would be a demonstration that a machine can think. The modern version of the Turing test is

---

<sup>50</sup> Full text of the paper Turing A. M. “Computing Machinery and Intelligence” available at: [academic.oup.com/mind/article/LIX/236/433/986238](http://academic.oup.com/mind/article/LIX/236/433/986238). Accessed 2 Jul. 2018.

characterised by more than one human judge that can question and chat with both subjects and the test is considered as successful if after five minutes of interaction, more than 30% of interrogators are convinced that the machine is human. (Investopedia).<sup>51</sup>

Notice that, originally, the Turing test was a mere thought experiment, designed to test a very specific aspect of machine intelligence. The choice of defining intelligence as the ability to disguise as a human being represented a narrow portion of the human intelligence spectrum, and thus was arbitrary: it introduced a clear, although necessary, bias in the test. For example, some machines have been instructed to cheat on the test by mimicking unintelligent human behavior, such as typing errors (“Artificial Stupidity”). Nevertheless, the Turing test has become a staple in NLP research and even more in Artificial General Intelligence. The test is so iconic that it has even made its way into pop culture.<sup>52</sup>

### **1. 3. 3. NLP main areas and problems**

Nowadays the progress in the NLP enables the creation of different applications useful in a variety of fields of study, including machine translation, natural language text processing and summarisation, user interfaces, multilingual and cross-language information retrieval, artificial intelligence and speech recognition. (Chowdhury, 2003: 51).

An important area of research in NLP is Natural Language Text Processing. It enables the elaboration and structuring the large bodies of textual information with the purpose to retrieve particular information. Through the manipulation of texts natural language text processing systems allow data extraction, automatic indexing and abstracting. The process may begin with morphological analyses, stemming of terms, and evolve to the lexical and syntactic processing that involves the words characteristics definition, their part of speech recognition, words and phrases determination and sentence parsing. The

---

<sup>51</sup> [www.investopedia.com/terms/t/turing-test.asp](http://www.investopedia.com/terms/t/turing-test.asp). Accessed 2 Jul. 2018.

<sup>52</sup> For example written and directed by Alex Garland science fiction psychological thriller film “Ex Machina” (2014), [https://en.wikipedia.org/wiki/Ex\\_Machina\\_\(film\)](https://en.wikipedia.org/wiki/Ex_Machina_(film)). Accessed 19 Jul. 2018.

systems of automatic abstracting and text summarisation are able to select part of texts for the subsequent abstract or summary production relying on linguistic and statistic criteria. (Chowdhury, 2003: 59-60).

Another active area of NLP deals with the research on the natural language systems that elaborate not written but spoken language. This sub-field is denominated Speech Recognition (also Automatic Speech Recognition, Computer Speech Recognition). In such type of NLP, in addition to the previously defined disambiguation problems, are introduced both negative and positive aspects. On the one hand, the interpretation could be slowed down by the new uncertainty and ambiguity, since the input words are not always clear or well-defined. On the other hand, the processing could be aided by additional information that is common to the spoken language, such as intonation and prosody. (Allen, 2003: 1222).

The process of creation and development of the computer systems able to understand natural languages is characterised by three main problems: thought process, linguistic input representation and meaning, and world knowledge. Natural languages processing systems may start from the word level and determine the morphological structure and nature of the word, for example part-of-speech identification and single meaning. The next level could be the sentence analysis where the word order, grammar and meaning of the whole sentence are determined. The final level of NLP may be the context and the domain. In different context or domain the same word, or sentence, may have different meaning or connotation. The main difficulty in NLP is the ambiguity that is characteristic of all natural languages and that distinguishes them from the artificial languages. There are different forms of ambiguity such as: simple lexical ambiguity, structural or syntactic ambiguity, semantic ambiguity, pragmatic ambiguity, referential ambiguity, etc. that may interact and produce an extremely complex input for the process of interpretation. (Chowdhury, 2003: 55, 51- 52).

In the language processing applications knowledge of language and its use is fundamental; this is actually what differs them from other data processing systems. The analysis of complex language behavior necessitates different acquaintances that are phonetics and phonology, morphology, syntax, semantics, pragmatics and discourse conventions knowledge. (Jurafsky, Martin, 2000: 2-3).

### **1. 3. 4. NLP application in the current research**

#### **1. 3. 4. 1. A variety of subfields and tools**

NLP has a variety of subfields. Consequently it is important to introduce a list of functions adopted in the second part of this research, in the process of testing. Additionally, I present some other terms that may not directly regard the current stage of research but are involved partially or could deal more closely with the purpose of our research in its further development.

- **Information extraction**

The process aimed to extract useful textual information from texts written in natural language is called information extraction (IE). (Gaizauskas, Wilks, 1998 in Chowdhury, 2003: 61). There are different techniques that enable information extraction and help to prepare a summary, extract databases, fill in slots in frames, and identify keywords for information retrieval. (Chowdhury, 2003: 61). The extracted information may vary according to the final purpose, for example one may aim to extract temporal events, locations, names, etc. One of the subtasks of the IE is named entity recognition (NER).

- **Named Entity Recognition**

NER extracts information about named entities. (Adriaans, Katrenko, 2007: 88). Under “named entities” are understood entities that carry information on location, person, organisation. “NER is typically framed as a sequence labelling task using the BIO scheme (Ramshaw and Marcus, 1995; Baldwin, 2009) i.e., given an input sentence, the goal is to assign a label to each token.” In order to obtain high performance in NER it is necessary to possess a well-developed annotated corpus of the language in question; in



terms of annotated sentences it should be tens of thousands. However, for most languages this quantity results to be unfeasible. (Cotterell, Duh, 2017: 91-92).

- Part-of-Speech-Tagging

Another subfield of NLP is the part-of-speech (POS) tagging of sentences that allows to classify the single word entities in base of their grammatical properties. There are already different POS tagging packages for widespread languages. With the help of the recent package RDRPOSTagger,<sup>53</sup> the part-of-speech tagging can be performed on more than forty languages. This package is available for a computational language and environment R.

- Text mining

Text mining is an interdisciplinary field of activity that includes different theoretical approaches and methods that deal with the same type of input information - text. For computers texts are mainly unstructured collections of words, if compared to data frames or tables. Text mining analysis includes the import of the text into the environment, for example in R; the organisation of the text structure; text tidying up that often consists in its whitespace removal, stopword elimination, stemming process, ecc.; the creation of the term-document matrix that is a common format used for text representation in computation. A text mining framework offers a range of functionalities aimed to manage text documents and to allow efficient work with them. Moreover, it provides filtering functions that may extract patterns of interest from text collections. (Feinerer, Hornik, Meyer, 2008: 1-5). Tidy text mining transforms a text in the tidy text format, that could be defined as a table with one-token-per-row. A token may be defined as “meaningful unit of text” that can be a single word, an n-gram, a sentence, a paragraph and in tidy text every token has its row. (Silge, Robinson, 2018).

---

<sup>53</sup> Available at: [www.bnosac.be/index.php/blog/65-natural-language-processing-on-40-languages-with-the-ripple-down-rules-based-part-of-speech-tagger](http://www.bnosac.be/index.php/blog/65-natural-language-processing-on-40-languages-with-the-ripple-down-rules-based-part-of-speech-tagger). Accessed 3 Jul. 2018.

#### **1. 3. 4. 2. R as an environment for testing**

The computational language and environment R “is an integrated suite of software facilities for data manipulation, calculation and graphical display.” Among different functions it includes data handling and storage facility, operators for calculations on arrays, in particular matrices, large collection of tools for data analysis, graphical facilities and a simple and effective programming language. It is preferred to call R “environment” due to its fully planned and coherent system, that differs it from some other data analysis software that often develop incrementally specific and inflexible tools. (“What is R?”). R works with numerous data types and the most basic are for example numerics, integers, logical data with the Boolean values “TRUE” or “FALSE” and characters that have text or string values. (“Introduction to R”). It contains a set of functions that are useful in pattern matches finding, where for pattern is intended a “certain amount of text” that in this research can vary between a simple letter, a combination of letters, word or combination of words. This “certain amount of text”, a pattern that describes a set of strings is called regular expression and mostly all scripting languages like Perl, Python, Java and others present different tools for their elaboration. (Sanchez, 2013: 68, 55). Another R function that deals with this research is the tool for finding the most common words present in the source text that may be useful for statistical representation of the elements repetition and usage.

The strategies of NLP become more complex when have to elaborate multilingual texts. Being Surzhyk a mixed language of Ukrainian and Russian languages, the automatic language identification, which would allow then Surzhyk processing, will surely require efforts. According to Barman, Das, Wagner, Foster (2014: 13) in the situations where speakers switch between languages or mix them, the automatic language identification process is increasingly important as it facilitates the further language processing. In the multilingual language use the phenomena of code mixing became normal and it could be considered as a natural product of multilingualism. In social media communication, for example, multilingual speakers often switch between languages (Barman, Das, Wagner, Foster, 2014: 13-14), and the former is just one of the proofs that confirms a

rapid and irreversible development of mixed languages and as follows a problem of their identification.

## **Part 2**

### **2. 1. Preparatory Phase**

The preparatory phase of the practical part of this research is divided into three main stages. The first one consists in the collection of the material and its manual elaboration. The second one regards familiarising with the programming language and environment R and studying its main functions. The last section deals with the choice of elements to be analysed and their background information.

#### **2. 1. 1. Material collection and terms classification**

During the first stage of the preparatory phase the interviews collected by Del Gaudio in different areas of Ukraine published as a CD together with his PhD dissertation *On the nature of Surżyk* (2010) were analysed. The interviews of Del Gaudio involved people from Kyiv, Chernihiv and Kharkiv areas, but in the current master dissertation research only the first group of conversations regarding the area of Kyiv were elaborated in detail. In order to have digital data it was necessary to transcribe manually all the registrations since no efficient program for Speech-to-Text automatic recognition was found neither for Ukrainian nor for a non-standard spoken language Surzhyk. The first stage of the analysis consisted in, firstly, transcribing all the registrations collected by Del Gaudio, secondly, identifying lexical elements in previously created Word document that did not belong to the standard Ukrainian language, and, finally, filling the simplified terminological tables.

During the process of manual transcription of the registrations we used Ukrainian or Russian rules according to the main language of the conversation. In order to prepare the material for the analysis and to take into consideration only “useful data”, it was decided to eliminate parts of the speech that belonged to Del Gaudio or to his first rate informants, also “A informant(s)”, and to concentrate on the speech of B informant(s) that according to Del Gaudio “are to be understood as just ordinary informants, who were questioned only once, and who in the majority of cases were in all circumstances

Surzhyk speakers.” It is important to stress that for “useful data” I mean real and spontaneous speech, therefore parts of the speech that belong to Del Gaudio were cut off from the analysis considering that he is an Italian native speaker, while words of A informant(s) may be classified only as an instrument adopted with the aim to get “good data”. Actually the function of the first rate informants was to help Del Gaudio to “get in touch with other people and thus increase the number of respondents” and it was common practice for them to adopt intentionally special techniques that “had the purpose of testing the informant’s linguistic habits and at the same time confusing them”, as “code-switching, with the introduction of a few conscious mistakes” (Del Gaudio, 2010: 49, 54, 51). However one of Del Gaudio informants states that Surzhyk is her native speech because she is “originally from Dnipropetrovs’k” and “had been living in Kyjiv for the past thirty years” (Del Gaudio, 2010: 51, 51), it was decided not to analyse her speech as after the first two registrations “DW\_A0031\_Kyjiv\_BasicSurzyk\_F.docx” and “DW\_A0037\_Kyjiv\_BasicSurzyk\_F.docx”, when the parts of speech of the informant A were put into the general term table, it was noticed that the informant was changing the speech language from Ukrainian to Russian not in a very natural way. Therefore, in all the following texts prepared for testing we will only have the speech of the interviewed people and eventually also of people that interact with the interviewed person due to work, e.g. some interviews took place while the interviewed were working as concierges so they had a lot of time but sometimes they could be disturbed.

When the documents regarding the areas of Kyiv were created it was possible then to identify the non-standard elements present in texts with the help of the automatic corrector of Google Drive Documents that underlined non-standard terms as not correct with Ukrainian settled as the language of input. However, some underlined terms were simply uncommon geographic names in Ukrainian unknown to the system, it was decided to take them into consideration. Additionally, I selected also the terms that seemed to me not Ukrainian. This strategy was adopted with the idea that in the

following phase of analysis the meaning and correctness of these terms will be checked with the help of online dictionaries and corpus of Russian and Ukrainian.

Consequently, for every registration of Kyiv area were created spreadsheet documents (we used Microsoft Excel). These were based on the previous analysis and were designed in order to represent the non-standard terms, or also elements identified by the automatic corrector of Google Drive Documents as not correct Ukrainian. All the terms were found in the analysed text and were consequently compared to Ukrainian and to Russian corresponding terms. While for Ukrainian terms there is only a written version, for Russian terms we present both written version that respects the rules of Russian and a phonetic pronunciation of the term according to Ukrainian phonology. It becomes necessary as soon as “Russian content morphemes introduced to the Ukrainian morphosyntax are subject to Ukrainian phonology and are often pronounced in accordance with Ukrainian phonological rules” (Kent, 2010: 45). It was decided to present English translation of the terms that in most cases was provided by myself, as the meaning of the mixed elements often was based on both Ukrainian and Russian meaning and there are no online dictionaries for Surzhyk. The spreadsheet documents were organised in the following way:

Table 3.

Term	Ukrainian written	Russian written	Russian pronounced	Additional comments if needed	Context	English written
A	B	C	D	E	F	G

In the first column A there is the term to analyse, written in Cyrillic according to the Ukrainian rules when the whole dialogue is in Ukrainian. When the whole dialogue is in Russian, it is written in Cyrillic according to the Russian rules. The second column B contains the Ukrainian version of the term A written in Cyrillic according to Ukrainian

rules. In the third column C is included the Russian version of the term written in Cyrillic according to the Russian rules and in the fourth column D there is the Russian version of the term written in Cyrillic according to the pronunciation of the word in Russian but adapting Ukrainian alphabet. The column E presents additional comments in English where needed. The column F may contain the context in which the term was used in the text if it is a specific term. In the last column G English translation of the term is provided.

Some may argue that when the analysed term perfectly corresponds to the Russian pronunciation it should be classified as a Russian one, but the following statements should be taken into consideration. First of all, the identified elements are characteristic and widespread especially as a non-standard spoken language. Secondly, this type of analysis involves texts that were recorded and collected in Ukraine and the Matrix Language of the conversation was Ukrainian. So if the main language of the conversation was Russian, it could be affirmed that terms that correspond to the Russian pronunciation are Russian, but in this specific analysis the purpose is to define the spoken language, that even some speakers define as “not a clean language” (Del Gaudio, 2010: registration DW\_A0031\_Kyjiv\_BasicSurzyk\_F), a some kind of “mixed Ukrainian” (Kent, 2010: 43), “this type of Ukrainian” (Kent, 2010: 43).

This simplified version of the terminological table presented before was adopted in the classification of the terms that were identified in the registrations regarding the area of Kyiv. Table 4 illustrates the amount of terms identified for every file.

Table 4.

File	Terms found
DW_A0027_Kyjiv_BasicSurzyk_F	231
DW_A0031_Kyjiv_BasicSurzyk	80
DW_A0037_Kyjiv_BasicSurzyk_F	111

DW_A0038_Kyjiv_BasicSurzyk_M	140
DW_A0039_Kyjiv_BasicSurzyk_M	344
DW_A0044_Kyjiv_BasicSurzyk_F	252
DW_A0129_Kyjiv_DialectalSurzyk_M	250

For seven registrations of Kyiv area collected by Del Gaudio were created seven tables for a total amount of 1408 terms. The final part of this dissertation includes seven simplified terminological table that cover Del Gaudio registrations of the Kyiv area (Simplified terminological table of Surzhyk). Additionally, a model useful for further studies on Surzhyk terminology will be provided. The proposal model (Standard terminological table of Surzhyk) includes just few terms from Del Gaudio registrations of the Kyiv area and has the purpose to demonstrate the structure of the table and the criterion that are important for the study of this specific non-standard language.

### **2. 1. 2. The instruments of R**

The second stage of the preparatory phase consisted in completing of the “Introduction to R” course and familiarising with the main functions that may be used in R for processing characters data type. In the current research R was adopted as an environment for testing. There are different tools present in almost all scripting languages that could be used for their elaboration. (Sanchez, 2013: 68, 55). In this research an integrated development environment RStudio was adopted as it allows to interact with R more readily. (Roecker, Yoast, Wills, D’Avello, 2018). Fundamental for the second phase of the dissertation are the word and pattern extraction functions of R. Hereafter there is a summarising list of main commands that were used in RStudio:<sup>54</sup>

- load ( ) reads and writes an R data file;

---

<sup>54</sup> Based on “The R Base Package”; Sanchez 2013; Surles 2017.



- library ( ) loads the package into the session and prepares all its functions available to use;
- str ( ) shows the structure of the data set;
- getwd ( ) prints the current working directory;
- ls ( ) lists the objects in the current workspace;
- filter ( ) permits to filter data;
- data.frame ( ) creates two dimensional objects that can hold numeric, character or logical values;
- c ( ) joins elements into a vector;
- word ( ) extracts words from a sentence;
- grepl ( ) finds regular expression matches in x. To notice that x is a character vector where matches are attempted to find and pattern is a character string with a regular expression that corresponds to the fixed logical value to be found in the character vector.;
- inner\_join ( ) joins two data frames into one by retaining rows in both datasets. The pipe operator (%>%) strings together these joins.;
- print ( ) prints its arguments.

Another R function that deals with this research is the tool for finding the most common words present in the source text that may be useful for statistical representation of the elements repetition and usage.

### **2. 1. 3. Identification of the elements to be analysed**

Due to the recurring repetition of the ending “-M” in the first person plural of the Present tense in the processed texts, it was decided to analyse the presence of this particular particle more in detail.

#### **2. 1. 3. 1. Background information**

To carry out this analysis, it was necessary to compare the Russian and Ukrainian verbal systems, and, specifically, the endings of the first plural person in Present tense.

Together with the facts on historic development of the first person plural endings in Present tense in Slavonic languages, the present differences of the endings in both Russian and Ukrainian languages were taken into consideration. The following part contains a brief description of the main characteristics of Ukrainian and Russian verbal systems and of the endings in the Present tense in both languages. A particular analysis is dedicated to the flexions of the first person plural in Present tense, that are actually a purpose of the second part of our study.

### **2. 1. 3. 2. Aspectual distinction**

Russian and Ukrainian verbal systems are characterised by aspectual distinction that can be perfective or imperfective. Verb aspect carries the meaning of the action expressed by the verb and almost every verb has its aspectual pair. In Ukrainian “perfective verbs mean an action completed in the past or which will surely happen in the future”<sup>55</sup> while “imperfective verbs determine an act in progress, limited in time and they are used in the Present, Past and Future tense.” (Tokar in Burlaka et al. 1993, 326).<sup>56</sup> In Russian, similarly to Ukrainian, the categorical meaning of imperfective verbs is to define an action that does not have signs of being limited or integral. Hence the ability of the imperfective verbs to express an action in the course of its development, in particular an action striving to reach the limit.”<sup>57</sup> “In perfective verbs the type of the limit reached by the action is often interpreted as a certain critical point, after the achievement of which the action, having exhausted itself, ceases.” ( Akademiya Nauk SSSR 1980, 583).<sup>58</sup>

### **2. 1. 3. 3. Russian verbal system**

---

<sup>55</sup> “Дієслова доконаного виду означають дію, завершену в минулому або яка обов'язково відбувається в майбутньому.”

<sup>56</sup> “Дієслова недоконаного виду означають дію незавершену, обмежену в часі. Вживаються вони у теперішньому, минулому і майбутньому часі.”

<sup>57</sup> “Глаголы несов. вида не обладают признаком ограниченности действия пределом, признаком целостности действия. В этом заключается категориальное значение несов. вида. Отсюда вытекает способность глаголов несов. вида выражать действие в процессе его протекания, в частности действие, стремящееся к достижению предела.”

<sup>58</sup> “В глаголах сов. вида достигаемый действием предел чаще всего осмысливается как некая критическая точка, по достижении которой действие, исчерпав себя, прекращается.”

Russian verb is conjugated in base of person, time, inclination, number, and in the Past tense according to the genre. Depending on the system of flexions in personal forms of singular and plural in the Present tense and Simple Future tense all the verbs are divided into two conjugations (except for isolated verbs: *зигдигтьсяс, зыбигть(сяс), мяукать, хотеть, бежать, чтить, есть, надоесть, дать, создать, быть* (and also *забыть, добыть, сбыть* and other verbs with components *-есть, -дать, -быть*), *идти, ехать*).<sup>59</sup> (The part on the Russian verb from now on in this chapter is based on personal translation from *Russkaja Grammatika Vol. I, Akademija Nauk SSSR Institut russkogo jazyka, Izdatelstvo Nauka, Moskva, 1980*).

In Russian there are two verb conjugations: Conjugation I and Conjugation II and the affiliation of the verb to the first or second conjugation is determined by the differences in the flexions of the person (first, second, third) and number (singular and plural), with the exception of the flexion of first person singular.<sup>60</sup> There are also several verbs that have different endings of personal forms. (Akademija Nauk SSSR 1980, 636).<sup>61</sup> The verbs belonging to the Conjugation I are characterised by the following endings in Present tense: **-ю (-у), -ешь (ёшь), -ет (-ёт), -ем (-ём), -ете (-ёте), -ют (-ут)**. They have the ending with vowel **-e-** when it is atonic, and **-ë-** when it is toned and in base of their affiliation to one of the ten verbs classes they can also take **-ю** or **-у** in the first person singular and **-ют** or **-ут** in third person plural.

The verbs of the Conjugation II are characterised by the following endings in Present tense: **-ю (-у), -ишь, -ит, -им, -ите, -ят (-ат)**. The endings of the first person singular

---

<sup>59</sup> “Спряжение - это изменение глагола по лицам, временам, наклонениям, числам и - в прош. вр. и сослагат. накл. - по родам. В зависимости от системы флексий в личных формах ед. и мн. ч. наст. и буд. простого вр. все глаголы (за небольшим исключением изолированных глаголов: *зигдигтьсяс, зыбигть(сяс), мяукать, хотеть, бежать, чтить, есть, надоесть, дать, создать, быть* (а также *забыть, добыть, сбыть* и другие глаголы с компонентами *-есть, -дать, -быть*), *идти, ехать*) разделяются на два спряжения: первое и второе.” (Akademija Nauk SSSR 1980, 647, 663).

<sup>60</sup> “Различия во флексиях, за исключением флексии I л. ед. ч., определяются принадлежностью глагола к первому или второму спряжению: флексии *-ёшь, -ешь; -ёт, -ет; -ём, -ем, -ёте, -ете, -ут, -ют* характеризуют глаголы I спряжения, флексии *-ишь, -ит, -им, -ите, -ат, -ят* глаголы II спр.”

<sup>61</sup> “Несколько глаголов имеют иные системы окончаний личных форм.”

and the third person plural, **-ю** and **-ят** correspondingly, can become **-у** and **-ат** in the presence of spelling incompatibilities.

Two conjugations can be confronted and resumed in the Table 5.

Table 5.

	Conjugation I	Conjugation II
1 pers. sing.	-ю (-у)	-ю (-у)
2 pers. sing.	-єшь (ёшь)	-ишь
3 pers. sing.	-ет (-ёт)	-ит
1 pers. plur.	-ем (-ём)	-им
2 pers. plur.	-ете (-ёте)	-ите
3 pers. plur.	-ют (-ут)	-ят (-ат)

#### 2. 1. 3. 4. Ukrainian verbal system

Ukrainian “verbs are divided into two conjugations according to the nature of the personal endings of the Present tense (or Future tense for perfective): the first and the second. Only four verbs (*дати, їсти, вісти, бути*) and their derivatives constitute a separate group.” (From now on the part on the Ukrainian verb is mainly based on *Ukrajinskyj pravopys*, Kyjiv, 2015: 108-117<sup>62</sup>).<sup>63</sup>

The endings of the Conjugation I in Present tense are: **-у (-ю), -єш (-єш), -є (-є), -ємо (-ємо), -ете (-єте), -уть (-ють)** (*Ukrajinskyj pravopys* 2015, 108).<sup>64</sup> Reflexive verbs that are marked by the presence of the affix **-ся (-сь)**, will take this particle added to the

<sup>62</sup> Available at [www.litopys.org.ua/pravopys/rozdil2.htm#par80](http://www.litopys.org.ua/pravopys/rozdil2.htm#par80). Accessed 25 Jul. 2018.

<sup>63</sup> “За характером особових закінчень теперішнього часу (або майбутнього дієслів доконаного виду) дієслова поділяються на дві дієвідміни: першу й другу. Лише чотири дієслова (дати, їсти, вісти, бути) становлять окрему групу.”

<sup>64</sup> “До першої дієвідміни належать дієслова з особовими закінченнями: **-у (-ю), -єш (-єш), -є (-є), -ємо (-ємо), -ете (-єте), -уть (-ють)**.”

regular endings listed before, moreover such verbs in the third person singular take additionally the ending **-ть** plus then the affix **-ся (-сь)** identical to the other verbs that have **-ся (-сь)** in their infinitive form. (Ukrajinskyj pravopys 1946 in Babych 1993, 321).<sup>65</sup>

Ukrainian verbs of the Conjugation II are characterised by the following endings in Present tense: **-у (-ю), -иш (-їш), -ить (-їть), -имо (-їмо), -ите (-їте), -ать (-ять)**.<sup>66</sup> It should be noted that “verbs that have **-и-, -і- (-ї-)** or **-а-** (after **ж, ч, ш**) in their infinitive basis, in the first person singular and in the third person plural of the Present tense (or of the Future tense if the aspect of the verb in question is perfective) lose these vowels; such verbs appertain to the Conjugation II (Ukrajinskyj pravopys 2015, 108).<sup>67</sup> If verbs are marked by the presence of the affix **-ся (-сь)**, they maintain it as adding to the regular endings listed before. (Ukrajinskyj pravopys 1946 in Babych 1993, 321).

The two Ukrainian conjugations can be confronted and recapitulated in the Table 6.

Table 6.

	<b>Conjugation I</b>	<b>Conjugation II</b>
1 pers. sing.	stem + <b>-у (-ю)</b>	<b>-у (-ю)</b>
2 pers. sing.	stem + <b>-еш (-єш)</b>	stem + <b>-иш (-їш)</b>
3 pers. sing.	stem+ <b>-е (-є)</b>	stem + <b>-ить (-їть)</b>
1 pers. plur.	stem+ <b>-емо (-ємо)</b>	stem+ <b>-имо (-їмо)</b>
2 pers. plur.	stem+ <b>-ете (-єте)</b>	stem+ <b>-ите (-їте)</b>

<sup>65</sup> Babych compares rules of Ukrainian spelling of different years, the quotation refers to the one of 1946 that was published under editorship of Academician L. Bulakhovskyi.

<sup>66</sup> “До другої дієвідміни належать дієслова з особовими закінченнями: **-у (-ю), -иш (-їш), -ить (-їть), -имо (-їмо), -ите (-їте), -ать (-ять)**.”

<sup>67</sup> “Дієслова, які мають в інфінітиві основу на **-и-, -і- (-ї-)** або на **-а-** (після **ж, ч, ш**) і в першій особі однини та в третій особі множини теперішнього часу (або майбутнього дієслів доконаного виду) ці голосні втрачають, належать до другої дієвідміни...”

3 pers. plur.	stem+ -уть (-ють)	stem+ -ать (-ять)
---------------	-------------------	-------------------

### 2. 1. 3. 5. Present tense endings of the first person plural

In the first person plural of the Present tense a vowel that we have considered as a part of the ending before, from another point of view can be seen as a thematic or connector vowel between a stem and an ending of a verb. Consequently we can admit that possible endings of the Present tense in first person plural can be **-mo** in standard Ukrainian or **-m** in standard Russian. The difference that nowadays seems to be clear in two standards, was not such definite before, especially for Ukrainian language. Ukrainian grammar of 1927 presents the ending **-m** as possible variant to the ending **-mo** in the first person plural. (Buzuk 1927 in Horbach 1985, 81).<sup>68</sup>

Slavic linguist Fedot Filin assumed that endings forms of the first person plural **-m** and **-mo**, “according to the comparative-historical grammar of the Indo-European languages, could go back to the Indo-European **-mos** (> Proto-Slavic with the fallout of the finite **-s** > **-mo**) and **-mus** (> **-mъ**) or **-mon** (> **-mo**) and **-mun** (> **-m**), and **-mo** could directly reflect the probable Indo-European **-mo**. If **-m** and **-mo** came from the same platform, then the conditions for their differentiation could be only guessed. (Filin 1972, 463).<sup>69</sup> Here the author adopts Latin-based symbols, for example **-m** and **-mo**, when describing Slavonic languages in general, that means not only the languages which alphabet is characterised by Cyrillics; by describing singularly a language Filin adopts the appropriate for every language alphabet, for example Latin-based for Polish and Cyrillic-based for Ukrainian. In this thesis **-m** and **-mo** correspond to **-м** and **-мо** and it was decided to leave them in Cyrillics, as they refer to the analysed texts with Ukrainian as matrix language and to the research input during the addestration process of the application on automatic recognition. According to linguist Petr Kuznetsov “the

<sup>68</sup> “В 1-й ос. множ. маємо варіанти **-мо** й **-м**, у лемків **-ме**.”

<sup>69</sup> “По данным сравнительно-исторической грамматики индоевропейских языков, могут восходить к индоевропейским **-mos** (> праслав. с отпадением конечного **-s** > **-mo**) и **-mus** (> **-mъ**) или **-mon** (> **-mo**) и **-mun** (> **-mъ**), а **-mo** может непосредственно отражать вероятное индоевропейское **-mo**. Если **-m** и **-mo** произошли из одной платформы, то об условиях их дифференциации можно только гадать.”

difference between the endings **-mo** and **-mь** can be explained by the fact that the latter form is the result of the development of the corresponding form before the pause, where the rules of the end of the word are particularly distinctive, while the form **-mo** originally appeared when the verb form was in close contact with the subsequent words, as a result it developed as a vowel of the inner syllable. Subsequently, as a result of the influence of analogical principles, the connection of the respective forms with the position before the pause was lost.” (Kuznetsov 1953, 206).<sup>70</sup> In order to reply to the question on which forms were used in the common Old Slavonic language and how they related to each other, one can give different answers. “If we suppose that **-m**, **-mo**, **-me**, and **-my** were the result of unequal development of the Indo-European forms, then we must admit that from the very beginning of its occurrence the common Old Slavonic language may be divided into separate dialectal zones, where different predominant forms were used. Such a proposal is possible, but it cannot be proved”.<sup>71</sup> The use of the forms of the first person plural were also complex in the Ukrainian territories: “The common endings are **-m**, **-mo**, and also **-мы** and **-ме**. The form **-mo** along with **-mь** could be found in Izbornik 1073, in The words of Gregory Nazianzen of 11th century and in The Uspensky izbornik of 12th century, etc. This could not be attributed to the influence of the Old Slavonic language, since in the Old Slavonic language the ending **-mь** was the norm.” (Filin 1972, 463, 467).<sup>72</sup> “The ending **-mo** has been preserved in Ukrainian till nowadays.” (Kuznetsov 1953, 206).<sup>73</sup> However the inflection **-mo** is used

---

<sup>70</sup> “Различие окончаний *-mo* и *-mь* может объясняться тем, что последняя форма представляет собой результат развития соответствующей формы перед паузой, где действуют особенно отчетливо законы конца слова, форма же *-mo* первоначально фигурировала в таких случаях, когда глагольная форма являлась в тесном сочетании с последующими словами, вследствие чего *o* развивалось как гласный внутреннего слога. Впоследствии в результате аналогических воздействий связь соответствующих форм с положением перед паузой была утрачена.”

<sup>71</sup> “Если *-m*, *-mo*, *-me*, и *-my* являются результатом развития неодинаковых индоевропейских форм, то мы должны будем признать, что древний общеславянский язык уже с самого начала его возникновения, по-видимому, распадался на отдельные диалектные зоны, в каждой из которых преимущественно употреблялась одна из форм. Такое предположение возможно, но доказать его нельзя.”

<sup>72</sup> “...Известны окончания *-m*, *-mo*, а также *-мы* и *-ме*. Форма *-mo* наряду с *-mь* встречается в памятниках древнерусской эпохи: Izbornik 1073, Слова Григория Назиянзина XI в., Успенский сборник XII в., и др. Это *-mo* мы не можем приписать влиянию старославянского языка, поскольку в старославянском языке нормой была *-mь*.”

<sup>73</sup> “Окончание *-mo* сохранилось до настоящего времени в украинском языке.”

more often, the ending **-м** can also appear, even in the text or in speech of the same source, where the form **-мо** is used; the first form is fixed as the leading one grammatical and lexicographical literature, while in the language of writers both from the Eastern and Western Ukraine, as well as in the language of folklore, both forms can coexist, for example in works of Kotliarevskyi. (Akademija Nauk USSR 1978, 318).<sup>74</sup>

### **2. 1. 3. 6. The particles analysed in the practical part of the research**

In the transcribed texts, it was noticed a repetition in the usage of Ukrainian verbs with endings of the first person plural in the Present tense similar to the Russian standard. For example, in the speech of the interviewed people were identified elements that in the Present tense took the endings **-ем**, **-єм**, so apparently part of the Conjugation I or **-їм**, **-нм**, when probably being part of the Conjugation II, differently from the endings of Ukrainian in Present tense. Such endings seem to be a reduction of Ukrainian endings, some of them correspond perfectly to the Russian ones. For these reasons, it was decided to concentrate on the identification of these first person plural endings in Present tense, as possible indicators of Surzhyk. The testing phase of the practical part of this research regards the use of the ending **-м** in the first person plural of the Present tense in Del Gaudio spoken language samples of Kyiv area.

## **2. 2. Project development.**

### **2. 2. 1. First phase**

#### **2. 2. 1. 1. Creation of the application in R**

At the preliminary stage of the project development basic functions in R were created with the aim of automatically finding the elements discussed in the preparatory phase of the research. In order to prepare text material for the elaboration we started with the creation of the preparatory project in RStudio that enables reading files and generating the data frames through the usage of the following libraries and their tools:

---

<sup>74</sup> “...Частіше виступає флексія **-мо**, хоч нерідко у мові того ж самого джерела може трапитися і флексія **-м**. Перша форма фіксується як провідна і в тодішній граматичній та лексикографічній літературі. У мові письменників як із східних так і з західних українських земель, а також у мові фольклору уживаються обидві форми.”



- `tibble` provides strict checking and improve formatting;
- `dplyr` enables working with data frame;
- `textreadr` allows reading text documents into R;
- `tidytext` provides text mining for processing of words.

Consequently we prepared a project on Surzhyk recognition where the previously created data frames of the interviews were required. In the first phase, the application in R allows to find affixes and endings within words of the corpus. The corpus is based on the speech of Kyiv speakers collected by Del Gaudio that was prepared in the previous phase. The application does not only allow to find the pattern of interest but also to define which text and which line contains the particle in question, since there are several texts regarding different speakers. The interface of the “Search Surzhyk Interviews” application is simple in usage and can be considered user-friendly.<sup>75</sup>

### 2. 2. 1. 2. Checking the output in application user mode

First of all it was decided to check how many times correct endings of the first person plural of Present tense is used in Ukrainian by putting into the search bar one by one the endings of the Conjugation I **-емо** and **-ємо** and then of the Conjugation II **-імо**, **-имо**.

For the first Ukrainian ending **-емо** there are four results that perfectly correspond to what was aimed to find, that means Ukrainian verbs in Present tense of the Conjugation I in first person singular as presented in the Table 7:<sup>76</sup>

Table 7.

File	Line	Word	Text
DW_A0027.docx	20	їдемо	Ну от дивіться. В мене дочка приїхала сюди, я кажу «Подожди», або кажу «Сахар». А вона каже: «Мамо, не сахар – цукор»... Я ще говорю «нащо», а вона каже «навіщо». А я кажу ти вже мене не переведеш на

<sup>75</sup> “Search Surzhyk Interviews”, [gmdn.shinyapps.io/Surzhyk/](http://gmdn.shinyapps.io/Surzhyk/). Accessed 30 Apr. 2018.

<sup>76</sup> It was decided to represent here all the output phrases without eliminating the extra sentences that do not concern the identified word in order to facilitate a future reader to understand better the context and usage of the word in question.

			свій лад. В мене ще таке слово є – «щас», а вона мені каже «зараз». Вона мені повторяє, а я кажу «я же забуду». <b>Їдемо</b> в автобусі, я вибачаюся, а я кажу «Де будем злазити?» Вона така в мене дуже мужня. Ето виходить. А вона каже до мене «Мамо, не кричи».
DW_A0031.docx	19	в'яжемо	Сама роблю: і сіємо, і <b>в'яжемо</b> .
DW_A0031.docx	22	в'яжемо	Самі сієм самі <b>в'яжемо</b> . Труд дуже важкий, коло його дуже багато роботи, пилява і все, але ще треба якось жити.
DW_A0044.docx	32	живемо	Надя. Батько Федір, Надежда Фьодоровна, ну Надя. Я завжди хочу бить молодістю. Надежа, Вера і Любов в нас є празник. І мати Софа. О, так я була спочатку швейною-мотористкою, ну як чотири роки провчилася. Ну в мене був ну як один парень мене подвів, ну як сказати, як завжди знаєте в молодості, трошки ж дур...Ну я родила сина, а він як я отак в селі, а я в районі, нас було семеро дівчат з нашого села.. Ну ...Бог з тобою, не нада так і всьо. Дак я ето самос...а тоді прийшов цей муж, ну мій чоловік щяс що <b>живемо</b> . А якраз було старшому сину Сірьожка звать год і сім місяців. Ну він приїхав, він хотів до другої піти, а так як ми гостеприймні, ну як в нас було. Я з сім'ї двенадцять чоловік дітей, я восьма на щоту. Після мене ще брат і сестра, і менший самий брат, виродок ми на нього кажем.

For the second possible ending of the first person plural in the Conjugation I **-ємо** are present four results and, similarly to the previous research, they perfectly correspond to what was aimed to get, Ukrainian verbs of the Conjugation II in Present tense, first person plural:

Table 8.

File	Line	Word	Text
DW_A0027.docx	32	вивчасмо	А ми визнаєм, я вже всіх знаю хто тут живе, ми ж <b>вивчасмо</b> жильців усіх по фамілії.

DW_A0027.docx	51	включаємо	Вхадная рiшотка. Ми <b>включаємо</b> , само по сiбе отключається i зачинається, автоматiчески. Автаматiка, а на сходи iдуть хто не по лiфту а на первий етаж нема лiфта, а вони пеша, пешка, пешком, тудою iдуть.
DW_A0027.docx	83	гуляємо	То ми <b>гуляємо</b> , а чоловiки наквасяться.
DW_A0031.docx	19	сiємо	Сама роблю: i <b>сiємо</b> , i в'яжемо.

Before proceeding to the analysis of the use of the Ukrainian endings of the second conjugation, it would be appropriate to firstly make a comparison with the use of the endings of the non-standard Conjugation I **-ем** and **-єм** (or it could be also said **-е-** (**-є-**) as a thematic vowel plus the ending **-м**).

Table 9.

File	Line	Word	Text
DW_A0027.docx	20	будем	Ну от дивiться. В мене дочка приiхала сюди, я кажу «Подожди», або кажу «Сахар». А вона каже: «Мамо, не сахар – цукор»... Я ще говорю «нащо», а вона каже «навiщо». А я кажу ти вже мене не переведеш на свiй лад. В мене ще таке слово є – «щас», а вона менi каже «зараз». Вона менi повтора, а я кажу «я же забуду». Iдемо в автобусi, я вибачаюся, а я кажу «Де <b>будем</b> злазить?» Вона така в мене дуже мужня. Ето виходiть. А вона каже до мене «Мамо, не кричи».
DW_A0027.docx	75	ножем	Лучше. Краще – це красивiше в нас получається, а лучше. В мене зять вiн отого не хоче. Вiн хоче, шоб я привезла з села молока, i шоб воно скисло. Воно ж тодi сильно добре, з корови, свiже, а тодi через сутки скисне, в нас кажуть – хоть ножем рiж. Воно дуже туге.
DW_A0031.docx	23	iдем	Електричкою. До вокзала <b>iдем</b> автобусом, а вже електричкою три часа нам iзди, ну так стоiмо тут.
DW_A0039.docx	28	будем	Там сiм. Ну тога у нас був як я ето працював у нас було шистiдневка, в суботу ми робили до двух часов, а так сiм часов работали. Зароботок там був непоганий, там я трошки i денг як то говориться скопив. А

			сюди приїхав збиріженія у нас пропали. Не знаю..може слухаш радіо, кажуть <b>будем</b> возвращать, а коли вернуть ніхто не знає. Короче обдерли народ. А знаєш що таке обдерли?
DW_A0044.docx	31	вопшем	Я раніше робила вопше як трудову діяльність в селі, так я була швея-мотористка, знаєте що таке? Вопшем підодіяльниці шили, труси шили мужські, я в смену шила по трицять пар трусів, представляєте? В нас електрическа біла машинка, називалася “Швея-мотористка”, ну електрическа машинка. А підодіяльник ми по п’ятнадцять штук ми шили. Я їздила даже в Житомир, на ето, на сорівнованіє, ми зайняли, наш район заняв, вся Житомирська область, наш район заняв третє місце по всій області.
DW_A0044.docx	32	кажем	Надя. Батько Федір, Надежда Фьодоровна, ну Надя. Я всегда хочу бить молодую. Надежа, Вера і Любов в нас є празник. І мати Софа. О, так я була сначала швейною-мотористкою, ну як чотири роки провчилась. Ну в мене був ну як один парень мене подвів, ну як сказати, як всегда знаєте в молодості, трошки ж дур...Ну я родила сина, а він як я отак в селі, а я в районі, нас було семеро дівчат з нашого села.. Ну ...Бог з тобою, не нада так і всьо. Дак я ето самое...а тоді прийшов цей муж, ну мій чоловік щяс що живемо. А якраз було старшому сину Сірьюжка звать год і сім місяців. Ну він приїхав, він хотів до другої піти, а так як ми гостеприймні, ну як в нас було. Я з сім’ї двенадцять чоловік дітей, я восьма на щоту. Після мене ще брат і сестра, і менший самий брат, виродок ми на нього <b>кажем</b> .
DW_A0129.docx	8	височайшем	Шоб не гаварілі тут, як було в Союзі плохо, було і хароше. То шо наука біла на височайшем уровне – нікуда не денішся.

The results of our third input search do not completely correspond to what we aimed to find. That means that the results were composed by what we actually were supposed to find: four verbs of the Conjugation I in the first person plural, Present tense with a non-standard ending -м (*будем, ідем, будем, кажем*); and by the words that simply have the same ending as the verbs but they are Ukrainian nouns in Instrumental case in singular (*ножем*), Surzhyk adjectives in Locative case in singular (*височайшем*) or Surzhyk adverbs / introductory words (*вопшем*). At the end of this chapter we will

present possible solutions that could be useful in order to eliminate the results that do not correspond to our research question during testing phase.

Consequently, the search was repeated with the second possible ending of the non-standard Conjugation I in the first person plural, Present tense; what was obtained is presented in the Table 10.

Table 10.

File	Line	Word	Text
DW_A0027.docx	33	працюєм	Сутки ми тут <b>працюєм</b> . Завтра рано я сміюся. І мені подобається. Люди, приличні люди, я абажаю їх. Я сама їх люблю. Я любляща, я страшно люблю людей. Може тут є людина яка мене обідить, ви знаєте є такі, є люди, байдуже! Бог їм судья.
DW_A0027.docx	32	взнаєм	А ми <b>взнаєм</b> , я вже всіх знаю хто тут живе, ми ж вивчаємо жильців усіх по фамілії.
DW_A0031.docx	6	балакаєм	Та по п'ять, нечисто, ми чисто не <b>балакаєм</b> , хіба ми чисто балакаєм?
DW_A0031.docx	6	балакаєм	Та по п'ять, нечисто, ми чисто не балакаєм, хіба ми чисто <b>балакаєм</b> ?
DW_A0031.docx	20	отдаєм	Просимо шість і як хто за скільки, за п'ять <b>отдаєм</b> .
DW_A0031.docx	22	сієм	Самі <b>сієм</b> самі в'яжемо. Труд дуже важкий, коло його дуже багато роботи, пилява і все, але ще треба якось жити.
DW_A0037.docx	11	восєм	В «Смотрикиєвспатрою», на стройке, слюсар-сантехнік. Поробив, забрали в армію, забрали мене на Чорноморський флот. Прослужив чотири, три года і <b>восєм</b> місяців, демобілізувався і обратно пішов в Київ в то же саме старе место, де я робив до армії. І... пожив, пожив, пожив... женився, получил квартиру і так продовжаю жити.

DW_A0038.docx	23	сем	Сем...Не було можливості тогочасно навчатися, одразу після війни важко було. Сем'я велика, ну і вирішив...
DW_A0039.docx	1	сем	Сім'я була велика батько важко працював, мати важко працювала теж. Нас <b>сем</b> чоловік було дітей. Чотири родних, і троє матеріальної сестри, їх гром убив, знаєш це блискавка? Грім це гроза,гроза, в полі вбив...Так вони сироти залишилися, мати забрала до себе. Сім чоловік нас виховували. А я був найстарший. Сім класів я закінчив і пішов на свій хліб заробити, щоб прожити.
DW_A0039.docx	6	сем	<b>Сем</b> , вісім днів. Дні одягнувшись, так, через троє відпочиваю, знову дні троє. Вісім виходів роблю. Ну це сьогодні дні відпочиваю тут, вісімнадцятого, двадцять другого знову буду тут сидіти. Двісті гривень – це не гроші... Почти даром. Ну там же в районах роботи ні в одному районі нема. Хоч візьми в Попельні, Житомирська область, Сквирі, Київська область, у Вінницьку сюди візьми етап Гребіщанський – теж люди без роботи сидять. Тому що нема, в районах нема. Ранше були будівельні організації, п'ять, шість, там водоканали всілякі всілякі...
DW_A0039.docx	6	вісім	Сем, <b>вісім</b> днів. Дні одягнувшись, так, через троє відпочиваю, знову дні троє. Вісім виходів роблю. Ну це сьогодні дні відпочиваю тут, вісімнадцятого, двадцять другого знову буду тут сидіти. Двісті гривень – це не гроші... Почти даром. Ну там же в районах роботи ні в одному районі нема. Хоч візьми в Попельні, Житомирська область, Сквирі, Київська область, у Вінницьку сюди візьми етап Гребіщанський – теж люди без роботи сидять. Тому що нема, в районах нема. Ранше були будівельні організації, п'ять, шість, там водоканали всілякі всілякі...
DW_A0039.docx	6	вісім	Сем, вісім днів. Дні одягнувшись, так, через троє відпочиваю, знову дні троє. <b>Вісім</b> виходів роблю. Ну це сьогодні дні відпочиваю тут, вісімнадцятого, двадцять другого знову буду тут сидіти. Двісті гривень – це не гроші... Почти даром. Ну там же в районах роботи ні в одному районі нема. Хоч візьми в Попельні, Житомирська область, Сквирі, Київська область, у Вінницьку сюди візьми етап Гребіщанський – теж люди без роботи сидять. Тому що нема, в районах нема. Ранше були будівельні організації, п'ять, шість, там водоканали всілякі всілякі...
DW_A0044.docx	30	вообщем	Ну як... Я з п'ятнадцяти першого року, а в семидесяти четвертому сюди заміж вийшла. Мені було двадцять три роки, як я вийшла заміж, ну <b>вообщем</b>

			уже, да, двадцять три роки. Так я вже живу, мені пієят чотири, уже тридцять один рік. Вже сину моему тридцять три старшому...
DW_A0044.docx	38	чєм	Онъ стоїть протів нас, ото називається пирожок. І вона оце бідненька, в неї п'ятсот с чєм-то собак, і котів сто восьмидсят, представляєте? І вона викупила ферму там де корови були, знаєте шо таке ферма? Корови, корови, короварня.
DW_A0044.docx	45	тєм	Приходьте, пожалуста. Ну шо, непогано я розказала? Тєм болєс Вам нравиться на українській мові. Хорошо, приходьте, спосіба!
DW_A0129.docx	2	нічєм	І там швірьонка забита, замки поставили, ці двері ми забрали, ну я замки купив, два замка все я отдав... Ну прикріпити, зробити – значить ціна. «Коля, скільки це буде?» «Та потом, та потом.» «Коля, скільки буде коштувати? Потом то вже потом.» – «Та потом.» Я говорю «не, скільки буде?» – «Триста гривень.» В той час триста гривень мало по курсу долара- то були великі деньги. Ну харашо, всьо... Потом начить, вони мені зробили, я говорю – «Тобі сечас деньги оддати?» «Не...» Ну я дачі, а там дачі рядом. Я значить беру тих триста гривень, ну при мені не триста було – где-то мав чотиріста, но... «Там хлопці порахували, треба ж хлопцям ше шось додати.» Я говорю «Коля, ну ти ж називав ціну.» Ми сидимо, його жінка сидить, а жінка говорить – «Чо ти йому яйця морочиш? Ти ж сказав, то ти й всьо...» «Ну давай...» Шоб ти знав раз і навсїгда – де б ти не бил, чи в сібя в Італії, чи поїдеш ти в Мілан, чи ти поїдеш в Рим, чи поїдеш ти ше куда-то, понимаєш? І навіть в Неаполі. Єслі ти шота дагаварюєшся, то сразу оприділяйся – все розложить нужно по полочкам, де б ти не бил. В вас то же самос, то же нічєм не отлічяється. Ти просто приїхав до себе додому, ти к себе прїсхал. А єслі тобі б пришлось іскать жильїо і хто тобі скаже шо потом. Почему потом? Потому шо в нас в Росії так привикли – патом... А потом з тебе здеруть три шкури, понимаєш? Ну просто прийшло с Росії до нас, вот ета, ше та жилка ніяк не может оторватса. І вас то же то же самос...

The results obtained with the fourth input search did not completely correspond to what we were aimed to find. The outputs were composed by what we actually were supposed to find: six verbs of the Conjugation I in the first person plural, Present tense with a non-standard ending -м (*працюєм, визнаєм, балакаєм, балакаєм, отдаєм, сієм*); and by the words that simply have the same ending as the verbs but they are Surzhyk nouns

in Instrumental case in singular (*нічим*), Surzhyk numerals (*восем, сєм, сєм, сєм, восем, восем*), part of Surzhyk pronouns in Instrumental case (*чем-то*), Surzhyk conjunctions (*тєм*) or Surzhyk adverbs / introductory words (*вообщєм*). At the end of this chapter we will present possible solutions that could be useful in order to eliminate the results that do not correspond to our research question during testing phase.

For the first possible ending with the thematic vowel **-імо** characterising the Conjugation II in Ukrainian we had only one verb (*стоімо*) that perfectly corresponded to the defined qualities of the verb in Present tense of first person plural in standard Ukrainian language:

Table 11.

File	Line	Word	Text
DW_A0031.docx	23	стоімо	Електричкою. До вокзала їдем автобусом, а вже електричкою три часа нам їзди, ну так <b>стоімо</b> тут.

When we put the second possible form of the thematic vowel and ending of Ukrainian Conjugation II in the research bar **-имо** we got the following results:

Table 12.

File	Line	Word	Text
DW_A0027.docx	82	робимо	Це так і в книзі є сковзалися, сковзанки, ковзанка. А вона мені каже «Шо? Я не поняла». Вона сама по-руски, її чоловік руский [...] Це в нас є став, купасмя там, большой. В мене чоловік, всю жизнь рибу ловив, короп, отакій. А я ні одну не з'їла, не люблю. І зімой, є то в нас називають фургалка, зараз скажу. Забивають у лід бича такого, дерево, кол, і це за його такий є празник – сковзалка. І коло його всі хлопці і дівчата, а ввечері і чоловіки йдуть, і жінки. Всі беруть вірьовку таку большую, зав'язують за той же бич, і всі карусель <b>робимо</b> . Це часто таке в нас робилось. Карусель. Одна одну поштовхує, карусель помімає? І за мотузку держімося і кругом того ми катаємся читається. То таке було, хто дивиться, не подужає, то її тре виносить.



DW_A0031.docx	18	робимо	Самі <b>робимо</b> , це й і возим і торги.
DW_A0031.docx	20	просимо	<b>Просимо</b> шість і як хто за скільки, за п'ять отдаєм.
DW_A0129.docx	2	сидимо	І там швірьонка забита, замки поставили, ці двері ми забрали, ну я замки купив, два замка все я отдав... Ну прикріпити, зробити – значить ціна. «Коля, скільки це буде?» «Та потом, та потом.» «Коля, скільки буде коштувати? Потом то вже потом.» – «Та потом.» Я говорю «не, скільки буде?» – «Триста гривень.» В той час триста гривень мало по курсу долара- то були великі деньги. Ну харашо, всьо... Потом начить, вони мені зробили, я говорю – «Тобі сечас деньги оддати?» «Не...» Ну я дачі, а там дачі рядом. Я значить беру тих триста гривень, ну при мені не триста було – где-то мав чотириста, но... «Там хлопці порахували, треба ж хлопцям ше шось додати.» Я говорю «Коля, ну ти ж називав ціну.» Ми <b>сидимо</b> , його жінка сидить, а жінка говорить – «Чо ти йому яйця морочиш? Ти ж сказав, то ти й всьо...» «Ну давай...» Шоб ти знав раз і навсігда – де б ти не бил, чи в сібя в Італії, чи поїдеш ти в Мілан, чи ти поїдеш в Рим, чи поїдеш ти ше куда-то, понімасш? І навіть в Неаполі. Еслі ти шота дагаварюєшся, то сразу оприділяйся – все розложить нужно по полочкам, де б ти не бил. В вас то же самое, то же нічем не отлічяється. Ти просто приїхав до себе додому, ти к себе прієхал. А еслі тобі б пришлось іскать жильйо і хто тобі скаже шо потом. Почему потом? Потому шо в нас в Росії так привикли – патом... А потом з тебе здеруть три шури, понімасш? Ну просто прийшло с Росії до нас, вот ета, ше та жилка ніяк не может оторватса. І вас то же же самое..

All of the output words listed in the Table 12 belong to the research purpose, they are Ukrainian verbs in the first person plural in Present tense.

Finally, the same procedure with the possible endings of Conjugation II in **-їм** and **-им** was repeated. What was found for **-їм** was mainly composed by the Ukrainian plural personal pronoun in Dative **їм**. Actually there were three personal pronouns (*їм*) and only one Surzhyk verb (*устроїм*) that corresponded to the aimed search output:

Table 13.

File	Line	Word	Text
------	------	------	------

DW_A0027.docx	30	їм	Сутки ми тут працюєм. Завтра рано я смінюся. І мені нравиться. Люди, прилічні люди, я абажаю їх. Я сама їх люблю. Я любляща, я страшно люблю людей. Може тут є людина яка мене обідить, ви знаєте є такі, є люди, байдуже! Бог їм судья.
DW_A0027.docx	66	їм	А він в неї воєнний, в селі вешас був, де-то далеко, і в Чукотці був, де-то далеко був чоловік в неї. А вже як приїхали сюди в Київ, вони приїжають до мами, коло мене її мама жила. То як я їм показала, то вона каже: «Ми пол Кисва розказали, що сливи з коржами нада їсти.» А в неї чоловік такий, знаєте, крепкий, а вже 50, та й каже «Боже де це ми кумусю взяли? Йому нравляться ті коржі, зі сметаною, коржі з сиром, коржі з маком.»
DW_A0039.docx	3	устроїм	Я знаю, в Америке. Так він Союз розвалив. А при Союзі було... да... Він Германію соїдїнив, а Союз розвалив. А при Союзі нам тога було набагато лучше. Робота.. не роботаш, комісія візьме тебе і устроїть на роботу, еслі не роботаш. Йдеш по городу в такий день, в отпуску, спрашуют чо ти не на роботі, я говорю ето «в отпуску». Еслі не робиш, говоріт, то ми тебе <b>устроїм</b> . А щас не роби те нікто. Пішли тут роботу найшов, хотів роботать, посмотрели, говоріт, на пенсію, он ворота – пусть молодьож, молодьож должна роботать. А ти вже заробив пенсію, іди отдыхай. А шо зараз тої пенсії? Я за неї не проживу.
DW_A0039.docx	18	їм	То як длають комерцію... Присжают в колгоспи, в колгоспі там в селі народ остався самі пенсіонери. То районна йому їм всім везти поросята ілі теля, надо машину десь взять, випасать надо ходить, надо шоб хтось йому поміг продать. То вони не продають. До їх самі прижають на дом, і вдома, там наприклад за поросьонка він должен шиссот гривень заплатити, він дає їй чотириста, двісті має навару. Отаким путем він машини собі покупас. Він же бере у хазяїна без весов. І то він так, сьодні він приїхав, двоє поросят купив, привіз у нього вже есть там люди закололи, всьо...там...люди есть продають, а він опять поїхав других везе. І отак каждый день, оборот делає. Вот ті багаті люди, канешно, за щот кого-то вони живуть. Вони не заробили своїми руками, як я оцими мозолями.. А він за щот кого-то живе.

Final search input was the ending **-им**, the following Table 14 is the output with the list of words that mainly were not congruent to the search question:

Table 14.

File	Line	Word	Text
DW_A0027.docx	11	таким	Маладьож...Акруженіє кругом. Я ж в акно смотрю ідьот сюди до мене дамой і от курить, пляшку пива, а воно ж не нравиться вже <b>таким</b> как ми.
DW_A0027.docx	34	ним	Але я не обіжаюся, Бог з <b>ним</b> . Я собі записала, а вона мені не сказала.
DW_A0031.docx	18	возим	Самі робимо, це й і <b>возим</b> і торги.
DW_A0037.docx	13	молодим	Ну аякже, як <b>молодим</b> був, каншно про любов, тільки любов, шоб гарно ми жили. А шас шоб гарно жили, не ругалися, друг друга любили.
DW_A0039.docx	9	мусим	Я даже не знаю, спрашивал одного, він каже: «та шоб дома не сидіть, то» каже, « <b>мусим</b> ходить на роботу.» Не хотять говорить. А один сказав, то каже можна работать. Десь тищу, полтари мають, дві тищі....
DW_A0039.docx	10	плянським	Україна она большая страна, но толку. І все єсть на Україні: недри єсть там всякі, всьо єсть, землі хароші, чорнозьоми. Ну поїхать по колгоспах шас подивиться – пустують, не засівають, бур'яни ростуть. Знаєш шо таке бур'ян? Сорняки, тоже не знаєш? Ну як, еслі ти не посадив нічого, не посіяв, воно заростає. Так половина полей пустує, я проїжав. І по <b>Плянським</b> районам їхав, і сюди на Сквирщину їхав, то смотрів.
DW_A0039.docx	18	отаким	То як делають комерцію... Приєжают в колгоспи, в колгоспі там в селі народ остався самі пенсіонери. То районна йому їм всім взети поросята ілі теля, надо машину десь взять, випасать надо ходити, надо шоб хтось йому поміг продать. То вони не продають. До їх самі прижають на дом, і вдома, там наприклад за поросьонка він должен шиссот гривень заплатити, він дає їй чотириста, двісті має навару.

			<p><b>Отаким</b> путьом він машини собі покупас. Він же бере у хазяїна без весов. І то він так, сьодні він приїхав, двоє поросят купив, привіз у нього вже єсть там люди закололи, всьо...там...люди єсть продають, а він опять поїхав других везе. І отак каждый день, оборот делає. Вот ті багаті люди, канешно, за щот кого-то вони живуть. Вони не заробили своїми руками, як я оцими мозолями.. А він за щот кого-то живе.</p>
DW_A0039.docx	30	бачим	<p>Да, щас народ тежило на Україні. Не знаю, може оце будуть вибори може шось воно буде лучше, хто його знає, уже і так п'ятнацять лет пройшло, а ми лучшего на <b>бачим</b>, все хуже і хуже йде.</p>
DW_A0044.docx	13	цим	<p>Нє, я думала шо ви з <b>цим</b> їхнім зятьом родствєнніки.</p>
DW_A0044.docx	33	другим	<p>Уже померли старші. Дак ми так ну як, росли одне за <b>другим</b>, ті уїжжали на Київ, ну куди хто...а все ж таки.. Но в нас мама була гостеприймна, хто б не прийшов в нас хлїб на столі не виходив. Хтось прийшов в хату, ну як, в селі хата називається, так мати всегда вгощала всім, всегда.</p>
DW_A0044.docx	35	цим	<p>Такі ото рушники, полотно ткане, полотно. Не єто не такі ото шо щас, простинь оце такі білі, а то колись були полотно, полотняне, знаєте? І оце мати <b>цим</b> рушничком накрїє всегда на столі був хлїб і мати казала: «Дітки, шоб стїл голий шоб не був, на столі нада всегда шоб лежав хлїб». Так воно було колись. А щас голий. Я прийду додому «Мама, ну шо тут лежить?» Я кажу як шо лежить – хлїб лежить, хлїб нада шоб був. Я тако закрию салфєточкою. Нєт, токо убери шоб стол був чист.</p>
DW_A0129.docx	2	рим	<p>І там швірїонка забита, замки поставили, ці двері ми забрали, ну я замки купив, два замка все я отдав... Ну прикрїпити, зробити – значить ціна. «Коля, скїльки це буде?» «Та потом, та потом.» «Коля, скїльки буде коштувати? Потом то вже потом.» – «Та потом.» Я говорю «нє, скїльки буде?» – «Триста гривень.» В той час триста гривень мало по курсу долара- то були великі денги. Ну харашо, всьо... Потом начить, вони мені зробили, я говорю – «Тобі сєчас денги отдати?» «Нє...» Ну я дачі, а там дачі рядом. Я значить беру тих триста гривень, ну при мені не триста було – где-то мав чотириста, но... «Там хлопці порахували, треба ж хлопцям ше шось додати.» Я говорю «Коля, ну ти ж називав ціну.» Ми сидимо, його жінка сидить, а жінка говорить – «Чо ти йому яйця морочиш? Ти ж сказав, то ти й всьо...» «Ну давай...» Шоб ти знав раз і навсїгда – де б</p>

			ти не бил, чи в сібя в Італії, чи поїдеш ти в Мілан, чи ти поїдеш в <b>Рим</b> , чи поїдеш ти ше куда-то, понімаєш? І навіть в Неаполі. Єслі ти шота дагаварюєшся, то сразу оприділяйся –все розложить нужно по полочкам, де б ти не бил. В вас то же самос, тоже нічем не отлічяється. Ти просто приїхав до себе додому, ти к себе приїхал. А єслі тобі б пришлось іскать жильйо і хто тобі скаже шо потом. Почему потом? Потому шо в нас в Росії так привикли – патом... А потом з тебе здеруть три шкури, понімаєш? Ну просто прийшло с Росії до нас, вот ета, ше та жилка ніяк не может оторватса. І вас тоже то же самос...
DW_A0129.docx	9	чим	Нічо не кращий, ше хлеще <b>чим</b> в нас!

Only three results are actually verbs with a non-standard ending “-м” (*возим, мусим, бачим*). Others are: Ukrainian personal pronoun of the third person singular in Instrumental case (*ним*), Ukrainian adjectives in Dative case plural and Instrumental case singular masculine (*молодим, отаким, таким*), Ukrainian geographical name (*Рим*), Ukrainian hard group adjective in Instrumental case (*другим*), Ukrainian masculine demonstrative pronouns in Instrumental case (*цим, тим*), and an adjective “*Плянським*” that basically could be used in Ukrainian as adjective in Dative plural, but in the analysed text it appears as a syntactic Russism “по Плянським районам їхав...” that turns out to be incorrect according to the Ukrainian grammar.

To summarise, in this first phase of checking the results, all fifty three outputs contained the input pattern we searched for. Twenty seven of these outputs were considered false positive results: although they presented the pattern of input, they did not appertain to the category of verbs we aimed to find. The repetition was not taken into account, hence, even if repeated, every output contributes to the total amount. If we consider the repetition and count every repeated output once, the number becomes forty two, with eighteen false positive results. As the purpose of the search was to identify non-standard verbs of the first person plural in the Present tense, it was necessary then to analyse and try to exclude the outputs that were not in line with the purpose of the search. The total amount of elements with the non-standard first person plural verbs endings is forty,

fourteen of them are verbs, twenty six false positive results. If excluded the repetitions, the outputs count thirty elements of which eleven verbs and nineteen false positive results. On the basis of the irregularities analysis possible rules on elements recognition were outlined.

### **2. 2. 1. 3. Rules on identification**

Since the purpose of this research is the identification of mixed elements, it was decided to define some general rules for their automatic recognition where possible. Mixed elements are considered entities that present characteristics of two different languages, in case of Surzhyk the two participating languages are Russian and Ukrainian. The natural creation of the mixed elements is related to the structure of the mixed languages that in general are composed by some already formed parts of grammar systems of the two participating languages unified to the lexicon of the other language. (Golovko, 1994: 119). The rules are aimed on the identification of the mixed verbs of the first person plural in Present tense and are divided into two groups: general and specific.

#### **2. 2. 1. 3. 1. General rules**

##### **1) Copula verb in first person plural + infinitive**

The conjugated copula verb has to present the final letter of the ending **-м** while the infinitive form of the verb has to possess the letters **-ти** or **-ть** as endings and to be composed of more than two letters. It is important to underline that the infinitive form of the spoken language in question may be **-ть** or **-ти**. In the cases when a copula verb is “**будем**”, these forms express Future tense. The infinitive form of the verb has to follow the copula verb in the first person plural at a maximum distance of 3.

Examples:<sup>77</sup> *будем злазить, будем возвращать, мусим ходить.*

##### **2) Personal pronoun “ми” + verb on distance from 1 to 3 in first person plural with the final letter of the ending -м**

---

<sup>77</sup> All examples provided to every rule are exclusively from the output we had which could be found in the column Text in the presented before tables that are based on the testing of Del Gaudio registrations.

This rule implies that in cases where personal pronoun of the first person plural “**ми**” is expressed, it has to precede at maximum distance 3 the verb conjugated in the first person plural with the final letter of the ending **-м**. This rule has limits as the personal pronoun is not always expressed and can be implied, especially in the spoken language. However this rule on distance works on our material in examples where the personal pronoun “**ми**” is expressed, further studies should be conducted in order to understand if this rule can be used for other texts. Following examples can be found in texts: *ми визнаєм* (at the distance 1), *ми тут працюєм, ми часто балакаєм, ми тебе устроїм* (all at the distance 2), *ми на нього кажем, ми чисто не балакаєм, ми краще не бачим* (all at the distance 3).

### 3) Defining pronoun “**самі**” + verb on distance from 1 to 3 in the first person plural with the final letter of the ending **-м**

Defining pronoun “**сам**” changes in base of genre, number and case and its plural variant “**самі**” in Nominative case defines first, second and third person plural. When used with a verb it means that agent has conducted an action independently by himself / herself or alone. Defining pronoun “**самі**” is an indicator that there are more agents to act and when combined to the verb with the first person plural ending **-м** has surely to deal with the first person plural. “Defining pronouns have a function of a definition in the sentences and by transition into nouns can carry a function of subject and adjunct.” (Pentyliuk, 1993, 305).<sup>78</sup> This has to be considered especially in the cases when the subject expressed by personal pronoun is omitted. Certainly, “**самі**” can also express other forms but it would be combined to a verb with different endings respectively to that defined here above. Following example can be found in text: *самі сієм* (at distance 1).

However there are also examples of verbs that correspond to our input question but are not preceded neither by a personal pronoun nor by defining pronoun (*отдаєм, ідем*), in

---

<sup>78</sup> “В реченні означальні займенники виступають означеннями, а при переході в іменники виконують роль підметів і додатків.”

such cases other rules should be sufficient to exclude the outputs that do not correspond to our aim question.

#### 4) Numerals with the ending -ЄМ, in particular “ВОЄМ” and “ЄЄМ”

The numerals “ВОЄМ” and “ЄЄМ” (correspectively “eight” and “seven”) are part of the output to the search particle -ЄМ that is the ending of the verbs in the first person plural. They do not appertain to the standard Ukrainian nor to Russian and could be classified as Surzhyk numerals. However they are examples of Surzhyk, at this stage we have to exclude them from the list of true positive results since our research is currently aimed on the study of verbs. The numerals to exclude are: **ВОЄМ** and **ЄЄМ** that can be found in the analysed texts with the frequency three times each.

### 2. 2. 1. 3. 2. Specific rules

#### 1) Include “їМ” as a part of the ending and exclude it when it is a separate word

In order to separate the ending -їМ from personal pronoun of the third person plural in Dative case “їМ” we decided to define that this element cannot be present as single but should be part of a word and be connected to any vocal, consonant or apostrophe. The output we had to the input task consisted in identifying all the elements that have particle -їМ at the end, was actually a third person plural personal pronoun in Dative and it presented 3 results of 4 total.

Examples: *їМ* (personal pronoun of the third person plural in Dative repeated three times), *ми тебе устроїМ* (this last one to be processed by general rule number 2 which takes into account the distance from the personal pronoun of the first person plural “ми”).

#### 2) Exclude: “ЧЄМ”, “ТЄМ”, “НІЧЄМ”

There is also a non-standard group of words that do not correspond to what we were aimed to find but we cannot generalise them to the common rules, apart from the



common feature they have that is to be in the Instrumental case. For this reason we will exclude them.

Examples: *чєм, тєм, нічєм*.

**3) Exclude adverb / introductory word “вопшем” and its variant “вообщєм”**

A more detailed analysis will be needed, for example on the recognition of the parts of speech similar to the study of Adriaans and Katrenko (2007, 88-93) on Ukrainian but since our research is on a non-standard language more time will be needed to mark every word and to classify them in base of the affiliation to different part of speech.

**4) Exclude geographical names with the final letter of the ending -м, in particular “Рим”**

Some geographical names can present the letter -м as part of their ending, at this point of the study we will just exclude them. This rule can be effective and automatic when for example we would have all the elements tagged and named as entities in base of part of speech and under classifications. This process is called named entity recognition and it is already in progress for Ukrainian language.<sup>79</sup>

Examples: *Рим*.

**5) Exclude adjectives in Dative case plural and Instrumental case singular, in particular “молодим”, “отаким”, “таким”**

Ukrainian adjectives in Dative case plural and Instrumental case singular masculine can interfere to our research since they present symbols -им in their final endings. They have to be excluded one by one since we have no tagged parts of speech at the moment of research. Examples: *молодим, отаким, таким*.

**6) Exclude personal pronoun of the third person singular in Instrumental case “ним”**

---

<sup>79</sup> “Named entity recognition (NER) is a subtask of information extraction (IE) which can be used further on for different purposes.” (Adriaans, Katrenko, 2007: 88).

Personal pronoun of the third person singular “*він*” becomes “*ним*” in Instrumental case and has symbols **-им** in its final ending. Due to this reason at this stage of work we have to exclude it.

**7) Exclude masculine adjectives in Locative case with Russian soft group ending -ем**

Masculine adjectives in Locative case have the ending **-ому (-ьому)** in Ukrainian and **-ом (-ем)** in Russian. Due to the presence of an adjective with Russian ending **-ем** that characterise the flection of soft group adjectives in Locative, masculine, we will exclude these adjectives as the ending **-ем** can interfere to our analysis.

Examples: *височяйшем*.

**8) Exclude hard group adjectives in Dative plural and in Instrumental singular masculine and neutral that have ending -им**

In Ukrainian hard group adjectives in Dative plural and in Instrumental singular masculine and neutral have the same ending **-им**. We will exclude them due to the possible interference they caused during the testing process.

Examples: *Плянським, другим*.

**9) Exclude masculine demonstrative pronouns in Instrumental case “цим” and “тим” that end in -им**

Ukrainian demonstrative pronouns “*цей*” and “*той*” correspond to English “this” and “that”. When they take the flections of cases, in Instrumental case of masculine their endings are **-им**. In order to eliminate every possible interference to our input question we have to exclude them.

Examples: *цим* (present twice), *тим*.

### **2. 2. 1. 3. 3. General comments on rules**

During the description of the possible rules that could automate the process of Surzhyk recognition it was realised that such rules could work better when applied on the already

classified entities. Named entity recognition would allow to identify every entity as part of some group, for example adjective in singular feminine in Nominative case, and then when we would apply our general rules we could just mention that particular rule regards such group, without listing the words that are part of that group one by one. Studies on named entity recognition are common for standard languages, but when it comes to non-standard languages they are not developed sufficiently. At this point we cannot provide annotated corpus to the non-standard language in question due to time limits but we can describe the rules as presented before and provide a list of words to be excluded when analysing verbs of the first person plural in Present tense. Such rules and list of words are effective when applied to the texts we have analysed, but if proven could be applied even to other texts.

### 2. 2. 2. Second Phase

After the first phase of project development was concluded, based on the rules of automatic recognition, we run further tests. In this chapter, we first summarise the rules that were outlined at the end of the first phase. Then we present the results obtained after the testing of the rules number 3 and 4. The chapter ends with the final comments regarding the second phase of the project development.

#### 2. 2. 2. 1. Summary of the rules

As already discussed in the previous chapter, the non-standard endings of the first person plural in Present tense are **-ем** and **-ём** for the Conjugation I and **-им** and **-їм** for the Conjugation II. The specific group of rules on recognition of these verbs in first person plural could be summarised as follows:

1) **-ем / -ём / -им / -їм + -ти / -ть** on max distance 3.

The first element ends in “-ем” or “-ём” or “-им” or “-їм ” while the second one ends in “-ти” or “-ть” and is situated on maximum distance 3 from the first element.

2) **ми / самі + -ем / -ём / -им / -їм** on max distance 3.

The first element is a word “ми” or “самі” while the second element ends in “-ем” or “-ём” or “-им” or “-їм ” and is situated on maximum distance 3 from the first element.

In order to exclude false positive results that emerged during the first phase of the project development a list of words to be ignored was prepared. Notice that they emerged when we checked the possible endings of the non-standard verbs in the first person plural and they were searched singularly, with no combination to personal pronoun or to other rules listed before. These elements correspond to the input question; this means they contain the pattern of the search but they do not respect the search question in terms of part of speech affiliation, they are not verbs of the first person plural in Present tense. The elements that may compare as outputs are:

- 1) numerals with the ending **-ем**, in particular “**восем**” and “**сем**”;
- 2) “**їм**” when it is a separate word;

- 3) “чѐм”, “тѐм”, “нїчѐм”;
- 4) “вопшем” and its variant “вообщѐм”;
- 5) geographical names with the final letter of the ending -м, in particular “Рим”;
- 6) adjectives in Dative case plural and Instrumental case singular, in particular “молодим”, “отаким”, “таким”;
- 7) personal pronoun of the third person singular in Instrumental case “ним”;
- 8) masculine adjectives in Locative case with Russian soft group ending -ем, in particular “височяйшем”;
- 9) hard group adjectives in Dative plural and in Instrumental singular masculine and neutral that have ending -им, in particular “Плянським”, “другим”;
- 10) masculine demonstrative pronouns in Instrumental case “цим” and “тим” that end in -им.

A summary list of false positive results that should be ignored may be presented as follow:

1. воѐм
2. сѐм
3. їм
4. чѐм
5. тѐм
6. нїчѐм
7. вопшем
8. вообщѐм
9. Рим
10. молодим
11. отаким
12. таким
13. ним
14. височяйшем
15. Плянським

16. другим

17. цим

18. тим

#### 2. 2. 2. 2. Testing of the second group of specific rules in user mode

At this stage we analyse specific group of rules that present the following combinations of patterns that distance one from another max 3 positions:

- 1) ми + -ем
- 2) ми + -єм
- 3) ми + -им
- 4) ми + -їм
- 5) самі + -ем
- 6) самі + -єм
- 7) самі + -им
- 8) самі + -їм

The outputs that we individuated during the testing process could be presented as follows. For a “true positive” result is considered a term that corresponds to the aimed output criteria, that means it is a verb in the first person plural in Present tense.

#### 1) ми + -ем

Table 15. 1 result, 1 true positive.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A0044.doc x	3 2	ми	235	кажем	237	Надя. Батько Федір, Надежда Фьодоровна, ну Надя. Я всегда хочу быть молодой. Надежа, Вера і Любов в нас є празник. І мати Софа. О, так я була сначала швейю-мотористкою, ну як чотири роки провчилась. Ну в

							мене був ну як один парень мене подвів, ну як сказати, як завжди знаєте в молодості, трошки ж дур...Ну я родила сина, а він як я отак в селі, а я в районі, нас було семеро дівчат з нашого села.. Ну ...Бог з тобою, не нада так і всьо. Дак я ето самое...а тоді прийшов цей муж, ну мій чоловік щяс шо живемо. А якраз було старшому сину Сірьожка звать год і сім місяців. Ну він приїхав, він хотів до другої піти, а так як ми гостеприїмні, ну як в нас було. Я з сім'ї двенацять чоловік дітей, я восьма на щоту. Після мене ще брат і сестра, і менший самий брат, виродок ми на нього кажем.
--	--	--	--	--	--	--	--

## 2) ми + -єм

Table 16. 4 results, 4 true positive.

Ordinal number	File	Line	Word	Position	Word	Position	Text
1	DW_A0027.doc x	30	ми	2	працюєм	4	Сутки ми тут працюєм. Завтра рано я смінюся. І мені нравиться. Люди, прилічні люди, я абажаю їх. Я сама їх люблю. Я любляща, я страшно люблю людей. Може тут є людина яка мене обідить, ви знаєте є такі, є люди, байдуже! Бог їм судья.
2	DW_A0027.doc x	32	ми	2	взнаєм	3	А ми взнаєм, я вже всіх знаю хто тут живе, ми ж вивчаємо жильців усіх по фамілії.

3	DW_A0 031.doc x	6	ми	7	балакаєм	10	Та по п'ять, нечисто, ми чисто не балакаєм, хіба ми чисто балакаєм?
4	DW_A0 031.doc x	6	ми	12	балакаєм	14	Та по п'ять, нечисто, ми чисто не балакаєм, хіба ми чисто балакаєм?

### 3) ми + -им

Table 17. 1 result, 1 true positive.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A0 039.doc x	30	ми	110	бачим	113	Да, шас народ тежило на Україні. Не знаю, може оце будуть вибори може шось воно буде лучше, хто його знає, уже і так п'ятнацять лет пройшло, а ми лучшего на бачим, все хуже і хуже йде.

### 4) ми + -ім

Table 18. 1 result, 1 true positive.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A0 039.doc x	3	ми	148	устроїм	150	Я знаю, в Америке. Так він Союз розвалив. А при Союзі було...да... Він Германію соїдинив, а Союз розвалив. А при Союзі нам тога було набагато лучше. Работа.. не работаєш, комісія візьме тебе і устроїть на роботу, еслі



							не роботаєш. Йдеш по городу в такий день, в отпуску, спрашують чо ти не на роботі, я говорю ето «в отпуску». Єслі не робиш, говоріт, то ми тебе устроїм. А щас не роби те ніхто. Пішли тут роботу найшов, хотів роботать, посмотрели, говоріт, на пенсію, он ворота – пусть молодьож, молодьож должна роботать. А ти вже заробив пенсію, іди отдыхай. А шо зараз тої пенсії? Я за неї не проживу.
--	--	--	--	--	--	--	---

**5) самі + -ем**

No results.

**6) самі + -єм**

Table 19. 1 result, 1 true positive.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A0031.docx	2 2	самі	61	сієм	62	<b>Самі сієм</b> самі в'яжемо. Труд дуже важкий, коло його дуже багато роботи, пилява і все, але ше треба якось жити.

**7) самі + -им**

No results.

**8) самі + -їм**

No results.

This second group of specific rules, described as “ми / самі + -ем / -єм/ -им / -їм on max distance 3”, allowed to identify eight elements, all of them were true positive results.

### 2. 2. 2. 3. General rules

The rules tested previously (ми / самі + -ем / -єм/ -им / -їм on max distance 3) could also be generalised and it may be affirmed that in order to be non-standard verbs of the first person plural in Present tense these verbs have to present the final characterising consonant **-м**. Based on this, the specific rules on recognition could be summarised and generalised as follows:

1) -м + -ти / -ть on max distance 3.

The first element ends in “-м” while the second one ends in “-ти” or “-ть” and is situated on max distance 3 from the first element.

2) ми / самі + -м on max distance 3.

The first element is a word “ми” or “самі” while the second element ends in “-м” and is situated on max distance 3 from the first element.

### 2. 2. 2. 4. Testing of the second group of general rules (with a generalised final pattern “-м”) in user mode

It was decided to test the second group of general rules on automatic recognition that expects to have a word “ми” or “самі” in combination to another element with end pattern “-м”. The last element “-м” may be located at a maximum distance 3 compared to the first element “ми” or “самі”. Consequently the results of the input are presented in base of the search input.<sup>80</sup>

- **Input “-м” + “ми”**

When the input pattern “-м” is put in the search suffix bar and the element “ми” in the search word bar, then the output consists of ten results: seven of them are true positive and three are false positive.

---

<sup>80</sup> It was decided do add a column with ordinal number of example that does not regard the application output but is useful for organisational reasons in this chapter.

Table 20.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A 0027.d osx	30	ми	2	працюєм	4	Сутки <b>ми</b> тут <b>працюєм</b> . Завтра рано я смінюся. І мені нравиться. Люди, прилічні люди, я абажаю їх. Я сама їх люблю. Я любляща, я страшно люблю людей. Може тут є людина яка мене обідить, ви знаєте є такі, є люди, байдуже! Бог їм судья.
2	DW_A 0027.d osx	32	ми	2	взнаєм	3	А <b>ми визнаєм</b> , я вже всіх знаю хто тут живе, ми ж вивчаємо жильців усіх по фамілії.
3	DW_A 0031.d osx	6	ми	7	балакаєм	10	Та по п'ять, нечисто, <b>ми</b> чисто не <b>балакаєм</b> , хіба ми чисто балакаєм?
4	DW_A 0031.d osx	6	ми	7	балакаєм	10	Та по п'ять, нечисто, ми чисто не балакаєм, хіба <b>ми</b> чисто <b>балакаєм</b> ?
5	DW_A 0037.d osx	12	ми	13	столом	15	Ми за столом... А, як гості? Кажу своє. Я водки не п'ю, то в мене мови ніякої немає, а хто сто грам вип'є, то ну, кожен своє розкаже.
6	DW_A 0039.d osx	3	ми	148	устроїм	150	Я знаю, в Америці. Так він Союз розвалив. А при Союзі було...да... Він Германію соїдїнив, а Союз розвалив. А

							при Союзі нам тога було набагато лучше. Робота.. не роботаєш, комісія візьме тебе і устроїть на роботу, еслі не роботаєш. Йдеш по городу в такий день, в отпуску, спрашують чо ти не на роботі, я говорю ето «в отпуску». Еслі не робиш, говоріт, то <b>ми</b> тебе <b>устроїм</b> . А щас не роби те нікто. Пішли тут роботу найшов, хотів роботать, посмотрели, говоріт, на пенсію, он ворота – пусть молодьож, молодьож должна роботать. А ти вже заробив пенсію, іди отдыхай. А шо зараз тої пенсії? Я за неї не проживу.
7	DW_A 0039.d осх	30	ми	110	бачим	113	Да, щас народ тежило на Україні. Не знаю, може оце будуть вибори може шось воно буде лучше, хто його знає, уже і так п'ятнацить лет пройшло, а <b>ми</b> лучшего на <b>бачим</b> , все хуже і хуже йде.
8	DW_A 0044.d осх	24	ми	148	поїздом	150	До Білорусії ...оце наш Ємільчинський район,так? А тоді Олевський район і Білорусія, ми почті біля Білорусії, ну районний ж наш який Новоград-Волинський район, Ємільчинський , Олевський район і тоді вже Білорусія. Як я їхала із Києва в село, так я їхала Київ - Брест ілі Київ - Луцк,з Києва, і <b>ми</b> оце <b>поїздом</b> їхали п'ять годин до Олевского до станції, а тоді трицить кілометрів від станції до нашого села автобусом. Ну село канешно в нас красіве, но щяс же ця..ето..радіація так уже не то...

9	DW_A 0044.d осх	32	ми	234	кажем	237	Надя. Батько Федір, Надєжда Фьодоровна, ну Надя. Я всегда хочу бить молодою. Надєжа, Вера і Любов в нас є празник. І мати Софа. О, так я була сначала швейою-мотористкою, ну як чотири роки провчилась. Ну в мене був ну як один парень мене подвів, ну як сказать, як всегда знаєте в молодості, трошки ж дур...Ну я родила сина, а він як я отак в селі, а я в районі, нас було семеро дівчат з нашого села.. Ну ...Бог з тобою, не нада так і всьо. Дак я ето самое...а тоді прийшов цей муж, ну мій чоловік щяс шо живемо. А якраз було старшому сину Сірьожка звать год і сім місяців. Ну він приїхав, він хотів до другої піти, а так як ми гостеприймні, ну як в нас було. Я з сім'ї двенацять чоловік дітей, я восьма на щоту. Після мене ще брат і сестра, і менший самий брат, виродок <b>ми</b> на нього <b>кажем</b> .
10	DW_A 0129.d осх	1	ми	290	одном	293	А потом значіть, я кажу йому - Саша, ну понімаєш, я опитний человек. Я работал с людьмі, в мене було під руководством громадна кількість людєй. І мені патом ето дело не паждодіт. Я веду разговор – вот вот вот. Всьо, мені розложити надо по полочкам, іначе , говорю «Саша, сколько ето будет? Мені потом до лампочки. За ту работу буде коштувать васемсот долларов, всьо, приступай до работи». Ну начіт потом потом, вже значіть договорилися, заставив, люді сделали работу, все... Я розчитуюся,

						<p>ну за паркет, що спаскудили, я потом все зірвав і по новому переложили... Ну короче вона: «Ой, то малувато, надо девятсот доларов, не васемсот, а вже на сто...» «Саша, ми ж з тобою дагаварювались? Дагаварювались чи не дагаварювались? Дагаварювались. Тоді ти ж мені називав ціну, не я називав. Єслі б я тобі назвав ціну, ну ти сопротивлявся... Короче ми чітко дійшли до конкретного питання суми, денег скоко. Потом ти захочеш сказав – «Не, ми з тобою не дагаварілісь, ми ж то не оприділілісь суми, вот оно девятсот стоїть...» Но! Ви ж точно назвали. Я говорю «Саша, ми ж з тобою канкретно дагаварювались. Хоча ми з тобою в хороших отношениях», но я сказав отношения до этого дела не імеют. Та той же самий, которий і випивали, і зустрічались, друг до друга ходіли в квартири, <b>ми</b> жили в <b>одном</b> доме. І вот тебе результат. І я б де б не хаділ... Вот я собі на дачі там робив мені один, Коля. Короче говора, кажу: «скільки це буде стоїти?» «Та, потом.» Я говарю: «не, потом не будет!» А, і по дачі. Вот замки, двері ото бачив, ці входніє?</p>
--	--	--	--	--	--	--

The results that could be considered true positive are numbers 1, 2, 3, 4, 6, 7, 9 from the Table 20; they are all verbs of the first person plural in Present tense with a non-standard ending “-м”. The false positive results are numbers 5, 8, 10 that present a final pattern “-м” but are not verbs.

- **Input “-м” + “самі”**

When the input pattern “-м” is put in the “search suffix” bar and the element “самі” in the “search word” bar, then the output consists of two results: one of which is true positive, another is false positive. In the following table the first term corresponds to the aimed output as it is a verb, the second one does not as it is a noun:

Table 21.

Ordinal number	File	Line	Word x	Position x	Word y	Position y	Text
1	DW_A0031.docx	22	самі	61	сієм	62	<b>Самі сієм</b> самі в'яжемо. Труд дуже важкий, коло його дуже багато роботи, пилява і все, але ше треба якось жити.
2	DW_A0039.docx	18	самі	133	дом	136	То як делают комерцію... Прижають в колгоспи, в колгоспі там в селі народ остався самі пенсіонери. То районна йому їм всім везти поросята ілі теля, надо машину десь взять, випасать надо ходить, надо щоб хтось йому поміг продать. То вони не продають. До їх <b>самі</b> прижають на <b>дом</b> , і вдома, там наприклад за поросьонка він должен шиссот гривень заплатити, він дає їй чотириста, двісті має навару. Отаким путем він машини собі покупає. Він же бере у хазяїна без весов. І то він так, сьогодні він приїхав, двоє поросят купив, привіз у нього вже єсть там люди закололи, всьо...там...люди єсть продають, а він опять поїхав других везе. І отак каждый день, оборот делає. Вот ті багаті люди, канешно, за щот кого-то вони живуть. Вони не

							заробили своїми руками, як я оцими мозолями.. А він за щот кого-то живе.
--	--	--	--	--	--	--	--

At this point we can analyse outputs that do not correspond to the aimed results but were identified during the testing of rules. We will confront them with the list of words prepared before, that contained false positive results.

The false positive results in reference to the second group of general rule testing (with generalised ending “-м”) are following:

- “ми” + “столом”

A personal pronoun of the first person plural “ми” is followed by Ukrainian noun in Instrumental case “столом” at distance 2.

Example: “Ми за столом...”

- “ми” + “поїздом”

A personal pronoun of the first person plural “ми” is followed by Ukrainian noun in Instrumental case “поїздом” at distance 2.

Example: “...і ми оце поїздом їхали...”

- “ми” + “одном”

A personal pronoun of the first person plural “ми” is followed by a pronoun conjugated in a non-standard way “одном” at distance 3.

Example: “...ми жили в одном доме...”<sup>81</sup>

- “самі” + “дом”

A defining pronoun in plural “самі” is followed by a noun “дом” at distance 3.

The noun “дом” could be appointed to both Ukrainian and Russian languages if considered singularly, but combined to the preposition “на” becomes a Russian fixed expression.

---

<sup>81</sup> In the analysed text we have “ми жили в одном доме...” (pronounced: “my zhyly v odnom domie”) that would correspond to Russian “мы жили в одном доме” (pronounced: “my zhyli v adnom domie”) and Ukrainian “ми жили в одному домі” (pronounced: “my zhyly v odnomu domi”). Even though at this stage this example does not regard the research question, it could be considered as a Surzhyk example.



Example: “До їх самі прижають на дом...”<sup>82</sup>

It could be summarised that none of these outputs contains elements from the false positive list prepared before. The explanation consists in the fact that the results we obtained by searching a general ending “-м” include words that finish in “-м”, in this specific case all of false positive results present “-ом” as final letters. The first two outputs are to be considered standard Ukrainian: “*Ми за столом...*”, “*...і ми оце поїздом їхали...*” Relying on the structure of the last two examples it may be affirmed that they are in Surzhyk (“*...ми жили в одному домі...*”, “*До їх самі прижають на дом...*”), even though they are false positive results for this specific search question.

#### **2. 2. 2. 5. General considerations on the second phase**

In this second phase of project development, it was tested the second group of specific rules and a corresponding general rule in user mode. The user mode allowed to display immediately results and was useful in getting in acquaintance with the application. In the following phase the first general rule will be tested directly in R. At this point it could be premature to admit whether it is the group of the specific or of general rules that functions better. The lists of the false positive results should be ignored during the outputs analysis. In the future development of the application they may be useful, especially due to the fact that, even though they do not correspond to the search question, some of them are great examples of Surzhyk.

---

<sup>82</sup> It should be noticed that in this case a defining pronoun “самі” carries a function of a noun and expresses a third person plural agent, as it is combined to the verb with a characteristic third person plural ending -ють. Often the presence of defining pronoun “самі” is an indicator that there are more agents to act. Previously were analysed examples where this defining pronoun carried the function of subject and was combined to the verb with the first person plural ending “-м” that means that it expressed a first person plural agent “ми”. In the current example it deals with the third person plural and expresses the personal pronoun “вони”.

### 2. 2. 3. Third Phase

The third phase of the project consists in creating additional rules in RStudio and their direct testing by controlling the outputs. It could be said that this phase was conducted in the creator modality, as all the tests were carried out in the RStudio environment. In order to visualise better the output tables of RStudio that presented results for every rule and full text examples we adopted the file editor Sublime Text 3.

#### 2. 2. 3. 1. Additional rules

The specific rules on recognition of the first person plural verbs were summarised in the previous chapter. At this point we firstly check the first group of specific rules defined as follows: “-ем / -ѐм / -им / -їм + -ти / -ть” that means that the first element has to present final letters “-ем” or “-ѐм” or “-им” or “-їм ” and be followed at maximum distance 3 by the second element that ends in “-ти” or “-ть”. Secondly, we test the general rule “-м + -ти / -ть” that predicts the first element to end in “-м” and to be followed with maximum distance 3 by the second element ending in “-ти” or “-ть”. Both specific and general rules include different pattern combinations that distance one from the other maximum 3 positions. For the first group of specific rules the possible combinations are eight:

- 1) -ем + -ти
- 2) -ѐм + -ти
- 3) -им +-ти
- 4) -їм + -ти
- 5) -ем + -ть
- 6) -ѐм + -ть
- 7) -им +-ть
- 8) -їм + -ть

while for the first group of the general rules there are two possible combinations:

- 1) -м + -ти
- 2) -м + -ть.

In order to find the first pattern it is mandatory to specify that it should be in the final part of the word; this rule also applies to the second pattern. Moreover, we establish that the distance between these two patterns should be 3, this means that the second element must be located at distance 1, 2 or 3 from the first element. For some rules it is necessary to specify that the identified word should have more characters than a search pattern. Such addition is mandatory for the rules that contain particles “-ти” or “-ім” due to the fact that they may be also found as single words. The code written in RStudio is a representation of the rules and the general structure could be demonstrated as follows:

```
#choose first word suffix
first_suffix <-

#choose second word suffix
second_suffix <-

# choose distance between first suffix and second suffix
distance <-

# find first suffix
first_suffix_found <- tidy_interviews %>%
  filter(grepl(x = word,
              pattern = paste0(first_suffix, "$")))

# check the result
print(first_suffix_found)

# find second suffix
second_suffix_found <- tidy_interviews %>%
  filter(grepl(x = word,
              pattern = paste0(second_suffix, "$")))

# check the result
print(second_suffix_found)

# join results (with distance)
first_suffix_found %>%
  inner_join(y = second_suffix_found, by = c("file", "line")) %>%
  rename(pos_first_suf = position.x, first_word = word.x, # rename
         columns
         pos_second_suf = position.y, second_word = word.y) %>%
  filter(pos_second_suf - pos_first_suf <= distance & pos_second_suf -
         pos_first_suf > 0) %>%
  inner_join(y = interviews)
```

Notice that we adopted term “suffix” at the early stage of the code development and when searched in RStudio it may be affix, pattern, final particle and not necessarily suffix. In some cases at the place of “suffix” we may have “word”.

### **2. 2. 3. 2. Testing of the rules.**

The code running in RStudio allows to display if there are matches. Consequently the results for every combination are presented. It is important to underline that the results that respect the aimed search output are considered true positive, which means that one of the elements in combination have to be a copula verb of the first person plural, while another element has to be a verb in infinitive form. Similarly to the previous phases we firstly test the specific rules on recognition and secondly the general rules.

#### **2. 2. 3. 2. 1. Specific rules testing in creator mode**

The representation of these specific rules testing assumes a different interface in order to respect the RStudio outputs. Consequently follows a demonstration of the outputs for each specific rule:

1) -ем + -ти

No results.

2) -ем + -ти

No results.

3) -им +-ти

No results. Previously when we did not specify that the second element has to follow the first one at maximum distance 3, and that the second output word containing the pattern “-ти” has to be longer than the pattern “-ти” in terms of characters, we obtained different results and they were false positive. The first combination of results was a Ukrainian demonstrative pronoun in Instrumental case “*цим*” situated at position 93 and preceded by Ukrainian noun “*мату*” at position 92 in the file DW\_A0044.docx. The second combination of results was Ukrainian noun, actually a geographical name,

“*Рум*” located at position 276 preceded by Ukrainian personal pronoun “*ми*” at position 273 in the file DW\_A0129.docx. Both first elements of the two combinations were listed as false positive results after the first phase of project development, but with the perfection of the rules we managed to avoid this false positive outputs.

4) -їм + -ти

No results in this combination. Notice that if checked singularly the final pattern “-їм” is present once as part of the first person plural Surzhyk verb “*устроїм*” (DW\_A0039.docx). Moreover it may be identified three times as Ukrainian personal pronoun in Dative case “*їм*” if we do not define that the output word has to contain more characters than the search pattern “-їм”. The Surzhyk verb “*устроїм*” was already identified in the first phase of project development when searching for words ending in “-їм” and during the second phase while checking the rule that contained personal pronoun “*ми*” situated at distance 3 from verb ending in “-їм”.

5) -ем + -ть

Table 22. Two results, two true positive.

	file	line	first_word	pos_fi rst_suf	second_word	pos_se cond_s uf	text
1	DW_A0027.docx	20	будем	71	злазить	72	Ну от дивіться. В мене дочка приїхала сюди, я кажу «Подожди», або кажу «Сахар». А вона каже: «Мамо, не сахар – цукор»... Я ще говорю «нашо», а вона каже «навіщо». А я кажу ти вже мене не переведеш на свій лад. В мене ще таке слово є – «щас», а вона мені каже «зараз». Вона мені повторя, а я кажу «я же забуду». Ідемо в автобусі, я вибачаюся, а я кажу «Де <b>будем злазить?</b> » Вона така в мене дуже мужня. Ето

							виходить. А вона каже до мене «Мамо, не кричи».
2	DW_A0039.docx	28	будем	166	возвращать	167	Там сім. Ну тога у нас був як я ето працював у нас було шистідневка, в суботу ми робили до двух часов, а так сім часов работали. Зароботок там був непоганий, там я трошки і денег як то говориться скопив. А сюди приїхав збиріженія у нас пропали. Не знаю..може слухаеш радіо, кажуть <b>будем возвращать</b> , а коли вернуть ніхто не знає. Короче обдерли народ. А знаєш шо таке обдерли?

Both results correspond to the aimed output, they are examples of Surzhyk.

б) -ем + -ть

Table 23. One result, one false positive.

	file	line	first_word	pos_f irst_s uf	second_word	pos_se cond_s uf	text
1	DW_A0044.docx	30	вообщем	143	двадцять	146	Ну як... Я з пієят першого року, а в семисят четвертому сюда заміж вийшла. Мені було двадцять три роки, як я вийшла заміж, ну <b>вообщем</b> уже, да, <b>двадцять</b> три роки. Так я вже живу, мені пієят чотири, уже тридцять один рік. Вже сину мойому тридцять три старшому...

The output is composed by two strings as required, but the first element is a Surzhyk adverb while the second one is a Ukrainian numeral; thus they do not respect the aimed output and are to be considered false positive.

7) -им +-ТЬ

Table 24. One result, one true positive.

	file	line	first_word	pos_fir st_suf	second_wor d	pos_ seco nd_s uf	text
1	DW_A 0039.d осх	9	мусим	38	ходить	39	Я даже не знаю, спрашивал одного, він каже: «та щоб дома не сидить, то» каже, « <b>мусим ходить</b> на роботу.» Не хочять говорить. А один сказав, то каже можна работать. Десь тищу, полтари мають, дві тищі....

This result perfectly corresponds to the search question and may be considered an example of Surzhyk.

8) -їм + -ТЬ

No result was obtained.

### 2. 2. 3. 2. 2. General rules testing in creator mode

After the specific rules were tested, at this stage we check the general rules. Firstly, we analysed the outputs for the combination of the final pattern “м” situated at a maximum distance 3 from the final pattern “ть”. Secondly, the combination of the final pattern “м” located at a maximum distance 3 from the final pattern “ти” is checked. The

outputs that we individuated during this testing process are presented in the tables with a creator mode structure. As previously for a “true positive” result is considered a term that corresponds to the aimed output criteria, that means it is a verb in the first person plural in Present tense.

1) -м + -ть

39 results, 3 true positive. For space limits in the following table we present only the true positive results. The false positive results are 36.

Table 25.

	file	line	first_word	pos_ first_ suf	second_w ord	pos_ seco nd_s uf	text
1	DW_ A002 7.doc x	20	будем	71	злазить	72	Ну от дивіться. В мене дочка приїхала сюди, я кажу «Подожди», або кажу «Сахар». А вона каже: «Мамо, не сахар – цукор»... Я ще говорю «нащо», а вона каже «навіщо». А я кажу ти вже мене не переведеш на свій лад. В мене ще таке слово є – «щас», а вона мені каже «зараз». Вона мені повторяє, а я кажу «я же забуду». Їдемо в автобусі, я вибачаюся, а я кажу «Де <b>будем злазить?</b> » Вона така в мене дуже мужня. Ето виходить. А вона каже до мене «Мамо, не кричи».
2	DW_ A003 9.doc x	9	мусим	38	ходить	39	Я даже не знаю, спрашивал одного, він каже: «та щоб дома не сидить, то» каже, « <b>мусим ходить</b> на роботу.» Не хотять говорити. А



							один сказав, то каже можна работать. Десь тищу, полтари мають, дві тищі....
3	DW_A0039.docx	28	будем	166	возвраща ть	167	Там сім. Ну тога у нас був як я его працював у нас було шистіднєвка, в суботу ми робили до двух часов, а так сім часов работали. Зароботок там був непоганий, там я трошки і денег як то говориться скопив. А сюди приїхав збиріженія у нас пропали. Не знаю..може слухаеш радію, кажуть <b>будем возвращать</b> , а коли вернуть ніхто не знає. Короче обдерли народ. А знаєш шо таке обдерли?

A brief analysis of the false positive results led to the conclusion that they contain the ending pattern “-ть” but actually do not correspond to the aimed search question of being verbs of the first person plural. In the false positive outputs emerged some words that were defined previously as false positive results, appeared also verbs with the standard ending of the third person plural. Some of the outputs are standard Ukrainian elements, while others are new examples of Surzhyk. The emergence of the non-standard verb “*єсть*”, for example, is repeated 5 times in different files. This verb is present also in Russian language, written “*есть*” [jest] and corresponds to Ukrainian “*є*” [je]. In this case we define “*єсть*” [jest] as part of the spoken Surzhyk corpora, due to the reasons connected to the records transcribing methods that were explained in the introduction and in the preparatory phase of the project. Other false positive results were considered as not important for the current research. We briefly list all false positive outputs without giving specific details on their position or context of usage:

1. DW\_A0027.docx: “там” + “робить” repeated twice in the same file;
2. DW\_A0027.docx: “міністром” + “скажіть”;

3. DW\_A0027.docx: “пешком” + “ідуць”;
4. DW\_A0031.docx: “нам” + “їхать”;
5. DW\_A0031.docx: “там” + “шість”;
6. DW\_A0031.docx: “людям” + “робить”;
7. DW\_A0038.docx: “даром” + “сидить”;
8. DW\_A0038.docx: “там” + “есть” repeated twice in the same file;
9. DW\_A0038.docx: “там” + “выходить”;
10. DW\_A0039.docx: “там” + “есть” repeated 3 times in the same file;
11. DW\_A0039.docx: “там” + “сидять”;
12. DW\_A0039.docx: “там” + “продають”;
13. DW\_A0039.docx: “там” + “шиїсять”;
14. DW\_A0039.docx: “шестом” + “опять”;
15. DW\_A0039.docx: “будем” + “возвращать”;
16. DW\_A0044.docx: “там” + “писать”;
17. DW\_A0044.docx: “там” + “нажимать”;
18. DW\_A0044.docx: “поїздом” + “п’ять”;
19. DW\_A0044.docx: “вообщем” + “двадцать”;
20. DW\_A0044.docx: “вам” + “сказать”;
21. DW\_A0044.docx: “там” + “стоить”;
22. DW\_A0044.docx: “там” + “звонить”;
23. DW\_A0129.docx: “потом” + “значить”;
24. DW\_A0129.docx: “руководством” + “кількість”;
25. DW\_A0129.docx: “потом” + “значить” combination repeated twice in the same text;
26. DW\_A0129.docx: “потом” + “начить”;
27. DW\_A0129.docx: “рядом” + “значить”;
28. DW\_A0129.docx: “потом” + “здеруть”;
29. DW\_A0129.docx: “там” + “питають” combination repeated twice in the same text;
30. DW\_A0129.docx: “шум” + “наделают”;
31. DW\_A0129.docx: “там” + “сидить”.

2) -м + -ти

5 results, 5 false positive.

All the outputs are not related to the question of Surzhyk, they appertain to Ukrainian standard language and are not verbs of the first person plural; for these reasons we decided to not present them.

### **2. 2. 3. 3. Analysis of the results and comments**

The rules tested in this third phase of the project development are to be considered additional, as the most important group of rules was already tested in the second phase. These additional rules were aimed to identify more elements of Surzhyk verbs that were not recognised during the previous phase. From the one side, during this phase the first group of general rules allowed us to identify 44 elements and only 3 of them respected the search question. From the other side, it cannot be omitted that by testing the general rules we managed to identify other elements that may be attributed to Surzhyk corpora, even though they are not verbs and for this reason considered as false positive in relation to the search input. It is for this reason that we decided to selectively present some of them in the final part of this work as part of the Standard terminological table of Surzhyk.

The first group of specific rules let us identify four elements and three of them are true. When we compare the outputs of the specific and the general rules, we may notice that the true positive results are the same, the difference stays in the false positive results. If by specific rules there are three of four true positive outputs, by the general rules we get three of forty four true positive output. Specific rules allow to identify the elements properly, by respecting their affiliation to the category of the first person plural verbs in Present tense. General rules allow to find more Surzhyk words, though they are irrelevant in relation to the search input. In this phase of testing it was discovered that in order to identify the first person plural verbs in Present tense of Surzhyk it is recommended to adopt specific rules.



#### **2. 2. 4. Fourth phase. Testing on the Russian texts**

In this final phase, the application is tested on the new texts. We decided to select Russian texts of spoken language, in particular some transcribing of the real conversation. During this process we use the new texts for proving whether our rules will give some results when applied on Russian texts. On the one side, it may be supposed that no element will be identified as the rules we created are aimed to identify Surzhyk elements. On the other side, it should be considered that Surzhyk is a hybrid language between Russian and Ukrainian which means that its element may come also from Russian. One of the patterns that is characteristic as common final ending for Ukrainian, Russian and Surzhyk words is the final pattern “-ем”.

After selecting material for the final testing, we elaborated texts and prepared them for testing in RStudio, where we already created all rules on Surzhyk identification. We firstly test the general rules and consequently specific rules. Notice that for the first person plural verbs in the Present tense of Surzhyk the specific rules were more effective, if compared to the general rules, as they allowed to reduce the number of false positive results.

We firstly tested the general rules and we obtained 25 results; 20 of them were actually the outputs for the input search “-м” + “-ть” and 5 the results for the input “-м” + “-ти”. We did not analyse all the outputs in detail as it was concluded previously that the general rules were not effective for the identification of Surzhyk verbs of the first person plural. Secondly, it was decided to check if the specific rules may provide similar outputs. What was discovered by testing of the specific rules, was that the total number of the results was 10. As supposed, the highest number of the outputs (8 of 10) corresponded to the input combination containing the pattern “-ем”. Notice that “-ем” is one of the possible common endings that are present in Russian, Ukrainian languages and Surzhyk. Consequently it was decided to present a summary of Russian text outputs for the specific rules. Notice that the table contains only the input combinations of the

specific rules that gave results, that means Russian texts that contained the patterns created for Surzhyk verbs identification; all the specific rules that did not provide results were not listed in the table.

Table 26.

Input pattern combination	File	Output combination	Text
“ем” + “ти”	1 (1).docx	“временем” + “эти”	“Все-таки со временем все эти вопросы, наверное, будут решены мирным путем. Я считаю, что надо максимально избегать военных конфликтов. Все конфликты надо решать может быть экономическим путем.”
“ем” + “ть”	1 (1).docx	“беженцем” + “быть”	“Беженцем может быть лицо одного государства, оказавшееся на территории другого государства. Но лицо одного государства, которое оказалось на территории своего же государства, беженцем не может быть.”
	1 (1).docx	“беженцем” + “быть”	“Беженцем может быть лицо одного государства, оказавшееся на территории другого государства. Но лицо одного государства, которое оказалось на территории своего же государства, беженцем не может быть.”

	1 (1).docx	“берем” + “ответственность”	“С моей точки зрения, это, конечно, агрессия, потому что когда у нас стали раздавать паспорта кому ни попадя в Абхазии и в Осетии, то это означает, что мы как государство берем на себя ответственность за этих людей, а дальше мы их можем использовать как “наших бьют”, что и произошло.”
	1 (1).docx	“можем” + “использовать”	“С моей точки зрения, это, конечно, агрессия, потому что когда у нас стали раздавать паспорта кому ни попадя в Абхазии и в Осетии, то это означает, что мы как государство берем на себя ответственность за этих людей, а дальше мы их можем использовать как “наших бьют”, что и произошло.”
	1 (1).docx	“уйдем” + “есть”	“Это наша земля, мы здесь захватили, мы здесь поставили гарнизоны, мы отсюда не уйдем никогда. То есть это имперское самосознание, высокомерное, наглое.”
	1 (1).docx	“будем” + “давать”	“Если проанализировать наши обложки, там очень много есть

			пророческого. Мы сели на планерке и стали говорить, что же дальше, что мы будем давать на обложку.”
	1 (1).docx	“упрашиванием” + “остановить”	“Европа не старушка, она очень инфантильна, она думает, что словами и упрашиванием можно остановить агрессора.”
	Rus. 2.docx	“следующем” + “есть”	“Я знаю, что вы готовите проект для Венецианской биеннале в следующем году. Есть ли связь с фильмом?”
“им” + “ть”	Rus. 3.docx	“ним” + “спорить”	“Для меня он остался милый, красивый, обаятельный Никита, а сейчас это какой-то царь-батюшка, ужасный, совсем другой человек. Ходит, молится, говорит такие глупости. Уже с ним нельзя спорить, не стоит...”

It could be summarised that three of the specific rules on identification of Surzhyk verbs led to the identification of the elements in the standard Russian texts. Some of these elements were verbs with the infinitive ending “-ть”, others were irrelevant results because of their affiliation to the categories of Russian nouns, adjectives or personal pronouns that in different cases, such as Dative, Instrumental and Locative, present the final pattern “-ем” or “-им”. This final testing may, from one point of view, confirm the idea that if we were working with texts provided with POS corpus, the rules would be



more appropriate and maybe no standard Russian would be identified by applying Surzhyk specific rules. From another point of view, we have to consider that Surzhyk is mixing of Russian and Ukrainian standards and for this reason presenting elements of both languages is an important characteristic of its nature.

## **Conclusion**

Surzhyk is a Ukrainian-Russian hybrid language that presents mixed language characteristics. Its hybridity may be explained by the involvement not only of Ukrainian and Russian languages but also of the different dialects present in the Ukrainian territory. Moreover, it can vary depending on the social affiliation of the speakers, their age and language experience. According to the criterion aimed to define mixed language nature provided by Bakker in 1992, a language may be considered mixed if it can be assigned to two different language families with equal justification. The criterion proposed by Bakker in 1992 may be applied to the cases where the lexicon is taken from one language and the grammatical system from another (Bakker 1992 in Bakker, 1994: 26) and Surzhyk seems to be the case, especially if we consider that the process of “intertwining” mixes lexicon and grammar of different codes and leads to the creation of the new language that typically presents stems from one language and affixes from another (Bakker, 1996: 18). In the current research the idea of the Matrix Language Frame model (MLF) elaborated by Myers-Scotton (2002, 2003) and proposed by Kent for Surzhyk analysis (2010) is taken into consideration. The nature of Surzhyk is complex: in Ukrainian-Russian Surzhyk structure, Ukrainian is a matrix language whereas Russian is an embedded language. In addition it has a huge variety of possible mutation based on the geographical area and speaker’s characteristics. It may be supposed that in the nature of Ukrainian-Russian Surzhyk there are two or even three types of mixing: Ukrainian and Russian languages mixing as basis with ulterior mixing of the Ukrainian dialects and addition of the mutations connected to the personalisation of the speech. Although this is an important part in the question of Surzhyk, the topic is not discussed in great detail in this research because it is a large area and demands a separate treatment. Our purpose was to identify the first person plural verbs of Surzhyk with the help of computational language and environment R.

The current status of Surzhyk is characterised as a non prestigious spoken language that is commonly used in the whole Ukraine with the largest number of its speakers in the

central, southern and eastern parts of Ukraine. Since Surzhyk is more present in the spoken language and not in the written, analysis of spoken samples of Surzhyk was mandatory. The example based method adopted in this Master's thesis research allowed to identify various elements of the hybrid language Surzhyk. During the first stage of the research we analysed the interviews registered by Del Gaudio in different areas of Ukraine and published as a CD together with his PhD dissertation *On the nature of Surżyk* (2010). However, the interviews of Del Gaudio involved people from Kyiv, Chernihiv and Kharkiv areas, in the current Master's research were elaborated in detail only the first group of conversations regarding the area of Kyiv. The preparatory phase of manual transcribing of the records allowed from the beginning to notice the presence of Surzhyk in the samples. This led us, firstly, to create the corpus of Surzhyk in Microsoft Word, and consequently, to fill the representative tables with the elements of the text that did not appertain to the standard Ukrainian language, or that were recognised as incorrect Ukrainian by Google Drive Documents automatic corrector. For every record, that means also that for every transcribed text we created a simplified terminological table. The structure of this table was adopted to classify the terms that were identified in the registrations regarding the area of Kyiv and to understanding better their connection to Ukrainian and Russian languages. For seven registrations of Kyiv area collected by Del Gaudio were created seven tables. The total amount of the terms considered hybrid, non-standard Ukrainian, incorrect Ukrainian is of 1408 entities. All the terms were presented in the Simplified terminological tables of Surzhyk. Following there are figures for every record; we maintained the original titles of the audio files when giving names to our transcribed.

Table 27.

File	Terms found
DW_A0027_Kyjiv_BasicSurzyk_F	231
DW_A0031_Kyjiv_BasicSurzyk	80
DW_A0037_Kyjiv_BasicSurzyk_F	111

DW_A0038_Kyjiv_BasicSurzyk_M	140
DW_A0039_Kyjiv_BasicSurzyk_M	344
DW_A0044_Kyjiv_BasicSurzyk_F	252
DW_A0129_Kyjiv_DialectalSurzyk_M	250

In the appendices, we included the general tables with all the hybrid elements present in the analysed texts (Simplified terminological table of Surzhyk) and a model of the deeper description of the hybrid language terms (Standard terminological table of Surzhyk). The Simplified terminological tables of Surzhyk were designed in order to represent all non-standard Ukrainian terms found in the Kyiv audio records collected by Del Gaudio. The structure of the Standard terminological table of Surzhyk model was drafted with F. Vezzani in order to define important aspects of Surzhyk terms. Although creating terminological tables for Surzhyk was only the preparatory phase and not the main task of this Master's thesis, we consider that it may be useful for further studies on Surzhyk.

The preparatory phase required a significant amount of time and it was decided to identify a particular characteristic of Surzhyk verbs. Due to the recurring repetition of the final pattern of the ending “-м” in the first person plural of the Present tense in the processed texts it was decided to analyse the presence of this particle more in detail. The repetition of this pattern as final ending in the first person plural was noticed during the process of manual transcription of the registrations collected by Del Gaudio. To carry out this analysis it was necessary to compare the Russian and Ukrainian verbal systems, and, specifically, the endings of the first person plural in the Present tense. Together with the facts on historic development of the first person plural endings in Present tense in Slavonic languages, the present differences of the endings in both Russian and Ukrainian languages were taken into consideration.

In order to develop the project on hybrid elements identification we adopted the instruments of programming language R in the integrated development environment RStudio that facilitates data manipulation. Fundamental for our project were the word and pattern extraction and matching functions. RStudio allowed to write the code on the automatic identification of patterns characterising verbs of the first person plural and to check the results by code running. In order to visualise better the output tables of RStudio that presented results for every rule and full text examples we adopted the file editor Sublime Text 3.

Based on examples we provided a list of rules aimed to identify hybrid verbs of the first person plural. In the first phase of testing the text samples were checked on the presence of the possible endings of hybrid verbs. In the second and third phases we defined more precisely the rules on recognition and divided them into two groups: the general group and the specific group. The general group was a short summary of the specific rules, that means that 16 specific rules were summarised and generalised in 4 rules. The second group of the general rules identified 12 results in the second phase and the first group of the general rules gave 44 outputs in the third phase. The total amount of the second and third phase outputs for the general rules was of 56 elements but only 11 of them were true positive for our input. For true positive we intended results that respected the aimed output of being verbs with the endings of the first person plural in Present tense or copula verbs in the first person plural combined to the infinitive form. From one side, during these two phases the general rules allowed us to identify 56 elements, but only 11 of them respected the search question. From another side, it cannot be omitted that by testing the general rules we managed to identify other elements that may be attributed to Surzhyk corpora, even though they are not verbs and for this irrelevant in relation to our input. It was decided to selectively present some of them in the final part of this work as part of the Standard terminological table of Surzhyk in appendices.

During the second and third phases we also tested the specific rules that allowed us to identify 11 combinations of verbs with the first person plural ending. In a total amount of 12 outputs only one result was false positive. The eleven results obtained with the general rules corresponded to the results we had with the specific rules. The difference consisted in the number of false positive outputs that were limited to one by adoption of the specific rules. It can be concluded that the rules we provided were effective if applied to our texts, consequently we list them for practical reasons and present the number of outputs for every combination.

Table 28. Summarising table of results.

Rule num.	First pattern	Second pattern	Distance	Output word requests	Quantity of results	True pos.	False neg.
1	ми	-ем	max. 3		1	1	0
2	ми	-єм	max. 3		4	4	0
3	ми	-им	max. 3		1	1	0
4	ми	-їм	max. 3	2nd patt. > “їм”	1	1	0
5	самі	-ем	max. 3		0	0	0
6	самі	-єм	max. 3		1	1	0
7	самі	-им	max. 3		0	0	0
8	самі	-їм	max. 3	2nd patt. > “їм”	0	0	0
9	-ем	-ти	max. 3	2nd patt. > “ти”	0	0	0
10	-єм	-ти	max. 3	2nd patt. > “ти”	0	0	0
11	-им	-ти	max. 3	2nd patt. > “ти”	0	0	0

12	-їм	-ти	max. 3	1st. patt. > “їм” 2nd patt. > “ти”	0	0	0
13	-ем	-ть	max. 3		2	2	0
14	-єм	-ть	max. 3		1	0	1
15	-им	-ть	max. 3		1	1	0
16	-їм	-ть	max. 3	1st. patt. > “їм”	0	0	0

N.B.: all these combinations have to follow the basic rule: the second element of the combination has to follow the first element at a distance from 1 to 3. In addition, the output words regarding the patterns “-їм” or “-ти” have to present more letters than the input pattern. This additional condition is to be applied on all the rules that present “-їм” or “-ти” in the first or second pattern of the combination.

As final phase of project development it was decided to try whether the rules on identification will function on new texts. We selected Russian texts of spoken language, in particular some transcribing of the real conversation and interviews present on the website of the *Radio svoboda (Radio Liberty)*. On the one side, it could be supposed that no element would be identified as the rules we created are aimed to identify Surzhyk elements. On the other side, it should be considered possible to have Surzhyk pattern within Russian text, especially the final pattern “-ем”. After testing of the specific rules we noticed that three of the specific rules on Surzhyk verbs identification led to the recognition of the elements in the standard Russian texts. Some of these elements were verbs with the infinitive ending “-ть”, others were Russian nouns, adjectives or personal pronouns that in different cases, such as Dative, Instrumental and Locative, present the final pattern “-ем” or “-им”. This final testing may, from one point of view, confirm the idea that if we were dealing with texts provided with Part-of-Speech tagged corpus, the results would be more appropriate and maybe no standard Russian would be identified by applying Surzhyk specific rules. From another

point of view, we have to consider that Surzhyk is mixes Russian and Ukrainian standards and for this reason containing elements of both languages is part of its nature.

To conclude, it is important to underline that the purpose of this research was satisfied. We analysed the language situation in Ukraine (characterised by the prevalent presence of Ukrainian and Russian languages), the diffusion of the dialects in Ukrainian territory, the theories on language contact and its consequences. The acquaintance with the basis of Natural Language Processing and the study of Surzhyk samples allowed us to work properly. By adopting an example-based method we managed to create rules on the automatic identification of the first person plural verbs in RStudio. These rules were generalised in order to check if the outputs of true positive results could be improved. We also prepared a list of false positive results; it was expected that during the testing phase, independently of their morphological affiliation, all the words presenting a pattern corresponding to our search could emerge. By creating specific rules we managed to reduce the number of false positive results. During the first phase we expected that a minimum amount of 18 words, that corresponded to the input pattern but not to the aimed output, would emerge as results. After the second and third phases of the project development by testing of the specific rules only one result emerged as false positive. To summarise, the adoption of the specific rules led to the successful identification of the aimed elements. Moreover, the study of the endings of the first person plural of Surzhyk has demonstrated that there is an internal coherence in the Surzhyk verb phrase. Since this internal linguistic consistency is limited to our corpus, we propose to verify the hypothesis.

The identified elements concerning verbs with the ending of the first person plural were 11. The total amount of non-standard terms in the analysed texts were 1408. We should admit that not every entity that is part of 1408 may be considered Surzhyk, but surely more that a half of these terms appertain to Surzhyk, this could be checked by the comparison of the columns corresponding to the original term found in the text, its Ukrainian, and its Russian versions. Further studies in this direction should also develop



towards Named Entity Recognition. During the process of analysis it was realised that different types of rules could work better when applied on the already classified entities. Studies on Named Entity Recognition are common for standard languages, but when it comes to non-standard languages they are not developed enough. At this point we cannot provide an annotated corpus to the non-standard language in question due to time limits but we provide seven Simplified terminological tables of Surzhyk with a total amount of 1408 terms, a model of the deeper description of the hybrid language terms Standard terminological table of Surzhyk with five entities and sixteen specific rules on the automatic identification of the final pattern of the first person plural ending implemented in R.

## References

### In English:

- Allen, J. F. Natural Language Processing, *Encyclopedia of Computer Science 4th*, John Wiley and Sons Ltd. Chichester, UK, 2003: 1218-1222, [dl.acm.org/citation.cfm?id=1074630](http://dl.acm.org/citation.cfm?id=1074630). Accessed 2 Jul. 2018.
- Appel R., Muysken P. *Language Contact and bilingualism*. 1987. London, Arnold, 1993.
- “Artificial Stupidity”, *The Economist*, 324 (7770), 1 Aug. 1992: 14.
- Auer P., From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *The International Journal of Bilingualism* 3, 1999: 309-332, [ijb.sagepub.com/content/3/4/309.refs.html](http://ijb.sagepub.com/content/3/4/309.refs.html). Accessed 26 Jun. 2018.
- Bakker, P., “Michif, the Cree-French mixed language of the Métis buffalo hunters in Canada.” *Mixed languages: 15 Case Studies in Language Intertwining*, Bakker P., Mous M. (eds.), Amsterdam, Institute for Functional Research into Language and Language Use (IFOTT), 1994: 13-33.
- Bakker, P., Language intertwining and convergence: typological aspects of the genesis of mixed languages. *Sprachtypologie und Universalienforschung* 49, 1996: 9-20.
- Bakker P., Mous M., “Introduction.” *Mixed languages: 15 Case Studies in Language Intertwining*, Amsterdam, Institute for Functional Research into Language and Language Use (IFOTT), 1994: 1-11.
- Barman, U., Das, A., Wagner, J., Foster, J. Code Mixing: A Challenge for Language Identification in the Language of Social Media. Proceedings of The First Workshop on Computational Approaches to Code Switching, *Association for Computational Linguistics*, 2014: 13-23, [aclweb.org/anthology/W14-3902](http://aclweb.org/anthology/W14-3902). Accessed 3 Jul. 2018.
- Bilaniuk, L. “Purism and Pluralism: Language Use Trends in Popular Culture in Ukraine since Independence.” *The Battle for Ukrainian. A comparative perspective*, edited by Flier, M., Graziosi, A., Cambridge, Massachusetts, Harvard University Press for the Ukrainian Research Institute, Harvard University, 2017: 343-363.
- Chowdhury, G. G. Natural Language Processing. *Annual Review of Information Science and Technology* 37, 2003: 51-89, [onlinelibrary.wiley.com/doi/full/10.1002/aris.1440370103](http://onlinelibrary.wiley.com/doi/full/10.1002/aris.1440370103). Accessed 2 Jul. 2018.

- “Constitution of Ukraine. Official translation.” Official website of Ukrainian Government Portal, [www.kmu.gov.ua/storage/app/imported\\_content/document/110977042/Constitution\\_eng.doc](http://www.kmu.gov.ua/storage/app/imported_content/document/110977042/Constitution_eng.doc). Accessed 14 Jun. 2018.
- Cotterell, R., Duh, K. Low-Resource Named Entity Recognition with Cross-Lingual, Character-Level Neural Conditional Random Fields. Proceedings of the 8th International Joint Conference on Natural Language Processing, AFNLP, 2017: 91-96, [aclweb.org/anthology/117-2016](http://aclweb.org/anthology/117-2016). Accessed 3 Jul. 2018.
- Del Gaudio, S. *On the Nature of Surżyk: A Double Perspective*. Wiener Slawistischer Almanach Sonderband 75, München-Berlin-Wien, 2010.
- Ex Machina. Directed by Garland, A., A24 / Universal Pictures, 2014, [en.wikipedia.org/wiki/Ex\\_Machina\\_\(film\)](http://en.wikipedia.org/wiki/Ex_Machina_(film)). Accessed 19 Jul. 2018.
- Feinerer, I., Hornik, K., Meyer, D., Text Mining Infrastructure in R. *Journal of Statistical Software* 25, 5, 2008, [www.jstatsoft.org/article/view/v025i05](http://www.jstatsoft.org/article/view/v025i05). Accessed 12 Jul. 2018.
- Ferguson, C. A. Diglossia. 1959, *WORD* 2, 2015: 325-340, [www.tandfonline.com/doi/abs/10.1080/00437956.1959.11659702](http://www.tandfonline.com/doi/abs/10.1080/00437956.1959.11659702). Accessed 28 Jun. 2018.
- Flier, M. Surzhyk or Surzhyks? In Hentschel G., Zaprudski S. (eds.), *Belarusian trasjanka and ukrainian surzhyk: Structural and social aspects of their description and categorization*, Oldenburg, BIS-Verlag der Carl von Ossietzky Universität Oldenburg, 2008: 39-56.
- Gardiner, S. C. “Introduction.” *Old Church Slavonic. An elementary grammar*, Cambridge, Cambridge University Press, 1984: 1-17.
- Golovko, E. V., “Mednyj Aleut or Copper Island Aleut: an Aleut-Russian mixed language.” *Mixed languages: 15 Case Studies in Language Intertwining*, Bakker P., Mous M. (eds.), Amsterdam, Institute for Functional Research into Language and Language Use (IFOTT), 1994:113-121.
- Hammarström, Harald & Bank, Sebastian & Forkel, Robert & Haspelmath, Martin. Glottolog 3.2. Max Planck Institute for the Science of Human History. Jena, 2018, [www.glottolog.org](http://www.glottolog.org). Accessed 17 Jul. 2018.
- “Introduction to R Course”, [www.datacamp.com/courses/free-introduction-to-r](http://www.datacamp.com/courses/free-introduction-to-r). Accessed 17 Apr. 2018.
- Ivanov, V. V., Browne, W. “Slavic languages”. *Encyclopædia Britannica Online*, [www.britannica.com/topic/Slavic-languages](http://www.britannica.com/topic/Slavic-languages). Accessed 12 Jun. 2018.
- Jurafsky, D., Martin, J. H. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech*

Recognition. University of Colorado, Boulder. Prentice Hall, Upper Saddle River, New Jersey, 2000.

- Katrenko, S., Adriaans, P., Named Entity Recognition for Ukrainian: A Resource-Light Approach. *Balto-Slavonic Natural Language Processing*, Association for Computational Linguistics, 2007, 88-93, [aclweb.org/anthology/W/W07/W07-1712.pdf](http://aclweb.org/anthology/W/W07/W07-1712.pdf). Accessed 3 Jul. 2018.
- Kent, K. Language Contact: Morphosyntactic Analysis of Surzhyk Spoken in Central Ukraine, *LSO Working Papers in Linguistics* 8, *Proceedings of WIGL*, 2010: 33-53, [pdfs.semanticscholar.org/0705/54bea599cacdaac26bb3b753a0443a93026a.pdf](https://pdfs.semanticscholar.org/0705/54bea599cacdaac26bb3b753a0443a93026a.pdf). Accessed 28 Jun. 2018.
- Kent, K. “Morphosyntactic analysis of Surzhyk, a Russian-Ukrainian mixed lect.” Dissertation, University of Minnesota, 2012.
- Malmkjaer, K. *The Linguistics Encyclopedia*, Second edition, London-New York, Routledge, 2002.
- Myers-Scotton, C. *Contact linguistics: Bilingual encounters and grammatical outcomes*. New York, Oxford University Press, 2002.
- Myers-Scotton, C. What lies beneath: Split (mixed languages) as contact phenomena. In Matras Y., Bakker P. (eds.), *The mixed language debate*, 73-106. Berlin, Mouton de Gruyter, 2003.
- Olszański, T. A. *The language issue in Ukraine. An attempt at a new perspective*, Warsaw, Ośrodek Studiów Wschodnich im. Marka Karpia Centre for Eastern Studies, 2012.
- Roecker, S., Yoast, K., Wills, S., D’Avello, T. “Introduction to R and RStudio.” *Statistics For Soil Survey*. 2018, [ncss-tech.github.io/stats\\_for\\_soil\\_survey/chapters/1\\_introduction/1\\_introduction.html#3\\_rstudio:an\\_integrated\\_development\\_environment\\_ide\\_for\\_r](https://ncss-tech.github.io/stats_for_soil_survey/chapters/1_introduction/1_introduction.html#3_rstudio:an_integrated_development_environment_ide_for_r). Accessed 23 Jul. 2018.
- Sanchez, G. *Handling and Processing Strings in R*. Trowchez Editions. Berkley, 2013.
- Silge, J., Robinson, D., *Text Mining with R. A Tidy Approach*. Online version. O’REILLY, 2008, [www.tidytextmining.com](http://www.tidytextmining.com). Accessed 12 Jan. 2018.
- Surles, W. *Joining Data in R with dplyr*. 2017, [rpubs.com/williamsurles/293454](https://rpubs.com/williamsurles/293454). Accessed 25 Jul. 2018.
- “The R Base Package” created by R Core Team and contributors worldwide, [stat.ethz.ch/R-manual/R-devel/library/base/html](http://stat.ethz.ch/R-manual/R-devel/library/base/html). Accessed 25 Jul. 2018.

- Turing, A. M. Computing Machinery and Intelligence. *Mind*, 1950: 433-460, [demic.oup.com/mind/article/LIX/236/433/986238](http://demic.oup.com/mind/article/LIX/236/433/986238). Accessed 2 Jul. 2018.
- “What is R?”, *The Comprehensive R Archive Network (CRAN)*, [www.r-project.org/about.html](http://www.r-project.org/about.html). Accessed 11 Jul. 2018.

In Ukrainian:

- Грінченко, Б. Д. *Словарь української мови* / Упор. з дод. влас. матеріалу Б. Грінченко: в 4-х т. Київ, Вид-во Академії наук Української РСР, 1958. [www.hrinchenko.com](http://www.hrinchenko.com). Accessed 17 Jun. 2018.
- Дудик, П. С., Тараненко, І. Й. “Українська літературна мова у ХХ столітті.” *Українська мова: Підручник. Частина 1*. За редакцією П.С. Дудика, Київ, Вища школа, 1993: 26-30.
- “Закономірності розвитку укр. усного літ. мовлення.” *Енциклопедія Українська мова*. Київ, 2000, [litopys.org.ua/ukrmova/um107.htm](http://litopys.org.ua/ukrmova/um107.htm). Accessed 15 Jun. 2018.
- Запорожець, Л. М. “Про статус і прагматизм суржику.” *Studia Linguistica* 5, 2011: 147-153, [philology.knu.ua/php/4/7/Studia\\_Linguistica\\_5\\_1/147\\_153.pdf](http://philology.knu.ua/php/4/7/Studia_Linguistica_5_1/147_153.pdf). Accessed 28 Jun. 2018.
- “Засоби масової інформації та книговидання в Україні у 2017 році” за даними Державної наукової установи “Книжкова палата України імені Івана Федорова”, [www.ukrstat.gov.ua/operativ/operativ2012/cltr\\_rik/zas\\_mas\\_inf\\_knyg/zas\\_mas\\_inf\\_knyg\\_17.xls](http://www.ukrstat.gov.ua/operativ/operativ2012/cltr_rik/zas_mas_inf_knyg/zas_mas_inf_knyg_17.xls). Accessed 19 Jun. 2018.
- “Конституція України (чинна з 1996 р. зі змінами)”. Офіційний портал Верховної Ради України, [zakon0.rada.gov.ua/laws/show/254%D0%BA/96-%D0%B2%D1%80](http://zakon0.rada.gov.ua/laws/show/254%D0%BA/96-%D0%B2%D1%80). Accessed 14 Jun. 2018.
- Масенко, Л. Т. “Суржик: історія формування, сучасний стан, перспективи функціонування.” *Belarusian trasjanka and ukrainian surzhyk: Structural and social aspects of their description and categorization*, edited by Hentschel G. and Zaprudski S., Oldenburg, BIS-Verlag der Carl von Ossietzky Universität Oldenburg, 2008: 1-37.
- “Моя Україна”, [myukraine.info/uk/country/place](http://myukraine.info/uk/country/place). Accessed 19 Jun. 2018.
- Пентиліук, М. І., “Відмінювання займенників, співвідносних з іменниками.” *Українська мова: Підручник. Частина 1*. За редакцією П.С. Дудика, Київ, Вища школа, 1993: 303-306.

- Сердега, Р. В., Сагаровський, А. А. *Українська діалектологія навчальний посібник*. Харків, Міністерство освіти і науки, молоді та спорту України, Харківський національний університет імені В. Н. Каразіна, 2011, [www-philology.univer.kharkov.ua/katedras/prof\\_sites/serdega/serdega\\_sagarovskii\\_dialect.pdf](http://www-philology.univer.kharkov.ua/katedras/prof_sites/serdega/serdega_sagarovskii_dialect.pdf). Accessed 18 Jun. 2018.
- *Словник української мови: в 11 томах. — Том 9, 1978*, [www.sum.in.ua](http://www.sum.in.ua). Accessed 17 Jun. 2018.
- Таблиця транслітерації українського алфавіту латиницею. Затверджено постановою Кабінету Міністрів України від 27 січня 2010 р. № 55, [www.kmu.gov.ua/ua/npas/243262567](http://www.kmu.gov.ua/ua/npas/243262567). Accessed 23 Jul. 2018.
- Тараненко, О. О. “Українсько-російський суржик: статус, тенденції, оцінки, прогнози.” *Мовознавство* 1, 2008: 14-30.
- Токар, В. П. “Виникнення і розвиток письма у східних слов’ян.” *Українська мова: Підручник. Частина 1*. За редакцією П. С. Дудика, Київ, Вища школа, 1993: 18-20.
- Токар, В. П. “Діалекти української мови, їх взаємодія з літературною мовою.” *Українська мова: Підручник. Частина 1*. За редакцією П. С. Дудика, Київ, Вища школа, 1993: 30-35.
- Токар, В. П. “Становлення і розвиток української літературної мови.” *Українська мова: Підручник. Частина 1*. За редакцією П. С. Дудика, Київ, Вища школа, 1993: 24-26.
- Токар, В. П. “Українська мова серед інших слов’янських мов.” *Українська мова: Підручник. Частина 1*. За редакцією П. С. Дудика, Київ, Вища школа, 1993: 22-23.
- “Україна та її географічні дослідження.” *Фізична географія України*, 8 клас, [ukrmap.su/uk-g8/855.html](http://ukrmap.su/uk-g8/855.html). Accessed 5 Jul. 2018.

In Russian:

- *Лингвистический энциклопедический словарь*, Москва, Советская энциклопедия, 1990.
- Национальный корпус русского языка, [www.ruscorpora.ru](http://www.ruscorpora.ru). Accessed 17 Aug. 2018.
- Рыковцева, Е. “За время правления Путина самым эффективным была война.” *Радио свобода*, 08 Авг. 2018, [www.svoboda.org/a/29420499.html](http://www.svoboda.org/a/29420499.html). Accessed 20 Aug. 2018.

- Волчек, Д. “Смертельный помет бакланов. Птицы и люди в “Кислотном лесу.” *Радио свобода*, 20 Авг. 2018, [www.svoboda.org/a/29437182.html](http://www.svoboda.org/a/29437182.html). Accessed 20 Aug. 2018.
- Волчек, Д. “Сирота под статуей Сталина.” *Радио свобода*, 06 Мая 2017, [www.svoboda.org/a/28470012.html](http://www.svoboda.org/a/28470012.html). Accessed 20 Aug. 2018.

In German:

- Besters-Dilger, J. “Die Rezeption der russischen Sprachpolitik in der Ukraine und Weißrussland.” *Wechselbeziehungen zwischen slawischen Sprachen, Literaturen und Kulturen in Vergangenheit und Gegenwart*, edited by Ohnheiser, I., Innsbruck, 1996: 129-143.

Illustrations:

- The Slavic languages’ family tree. Illustration. *Encyclopædia Britannica Online*, [www.britannica.com/topic/Slavic-languages?oasmId=61368](http://www.britannica.com/topic/Slavic-languages?oasmId=61368). Accessed 12 Jun. 2018.
- Map of Ukrainian dialects. Primary source in Ukrainian [litopys.org.ua/ukrmova/um151.htm](http://litopys.org.ua/ukrmova/um151.htm) translated into English and published in the Wikipedia, [commons.wikimedia.org/wiki/File:Map\\_of\\_Ukrainian\\_dialects\\_en.png](https://commons.wikimedia.org/wiki/File:Map_of_Ukrainian_dialects_en.png), Accessed 14 Jun. 2018.

## **Identificazione automatica delle lingue ibride: il caso del Suržyk.**

### **Riassunto**

Il Suržyk è una lingua ibrida tra l'ucraino e il russo che presenta le caratteristiche delle lingue miste. La sua ibridità si spiega attraverso il coinvolgimento delle due lingue slave ucraino e russo nonché dei diversi dialetti presenti sul territorio ucraino; ulteriori variazioni possono aver luogo in base all'appartenenza sociale dei parlanti, loro età e esperienza linguistica. Secondo il criterio mirato a definire la natura delle lingue miste fornito da Bakker nel 1992 una lingua può essere considerata mista se con uguali giustificazioni può essere assegnata a due famiglie linguistiche diverse. Tale criterio può essere applicato ai casi in cui il lessico è tratto da una lingua e il sistema grammaticale da un'altra (Bakker 1992 in Bakker, 1994: 26). Il Suržyk sembra esserne un esempio, specialmente se si considera che il processo di contatto delle due lingue coinvolte mescola il lessico e la grammatica appartenenti ai diversi codici e porta alla creazione di una nuova lingua che tipicamente presenta radici da una lingua e affissi da un'altra. (Bakker, 1996: 18). Nell'attuale ricerca viene presa in considerazione l'idea del modello chiamato Matrix Language Frame (MLF) elaborato da Myers-Scotton (2002, 2003) e proposto da Kent per l'analisi del Suržyk (2010). La struttura del Suržyk è complessa: l'ucraino esprime il ruolo di una lingua matrice mentre il russo prende la parte della lingua incorporata. Inoltre si possono manifestare le variazioni aggiuntive in base all'area geografica in cui Suržyk viene registrato e in base alle caratteristiche dei parlanti. Si può presupporre che nella struttura del Suržyk ucraino-russo ci siano due o anche tre tipi di mescolanza: le lingue ucraina e russa si mescolano creando una base, poi segue un'ulteriore miscelazione dei dialetti ucraini e alla fine l'aggiunta delle mutazioni connesse alla personalizzazione del discorso da parte del parlante. Nonostante questa discussione sia importante nella questione di Suržyk, l'argomento non è discusso in dettaglio in questo lavoro di ricerca perché richiede un trattamento separato. Il nostro obiettivo era quello di identificare i verbi di prima persona plurale di Suržyk ricorrendo agli strumenti del linguaggio di programmazione e ambiente R.



Il Suržyk è caratterizzato come una lingua parlata non prestigiosa; è comunemente diffuso in tutta l'Ucraina con la maggior concentrazione dei parlanti nelle zone centrale, meridionale e orientale dell'Ucraina. Poiché il Suržyk è più diffuso nella lingua parlata rispetto a quella scritta, uno dei principi fondamentali di questo lavoro era l'analisi dei campioni del parlato.

Il metodo di ricerca basato sugli esempi ha permesso di identificare vari elementi del linguaggio ibrido Suržyk. Nella prima fase del lavoro sono state analizzate le interviste registrate da Del Gaudio in Ucraina e pubblicate come CD insieme alla sua tesi di dottorato *On the nature of Suržyk* (2010). Le interviste di Del Gaudio coinvolgevano le persone provenienti dalle zone di Kyjiv, Černihiv e Charkiv ma in questo lavoro di tesi magistrale è stato elaborato e analizzato in dettaglio solo il primo gruppo di audio riguardante i parlanti di Kyjiv. La fase preparatoria consisteva nella trascrizione manuale delle registrazioni e ha permesso di notare fin dall'inizio la presenza di Suržyk nei campioni. Questo ci ha portato, in primo luogo, a creare il corpus di Suržyk in Microsoft Word e, di conseguenza, a riempire le tabelle rappresentative con gli elementi del testo che non corrispondevano alla lingua ucraina standard, o che erano stati riconosciuti come errati ucraini dal correttore automatico del Google Drive Documents. Per ogni registrazione, che significa anche per ogni testo trascritto, è stata creata una tabella terminologica semplificata. La struttura delle tabelle semplificate è stata adottata per classificare i termini identificati nelle registrazioni e per comprendere al meglio la loro relazione con le lingue ucraine e russe. Per sette registrazioni dell'area di Kyjiv raccolte da Del Gaudio sono state create sette tabelle. Il numero totale di termini considerati ibridi, non appartenenti agli standard delle regole dell'ucraino è di 1408 entità. Tutti i termini sono stati presentati nelle tabelle terminologiche semplificate di Suržyk. Di conseguenza esponiamo i numeri per ogni registrazione, abbiamo mantenuto i titoli originali dei file audio.

Nome del file	Termini trovati
DW_A0027_Kyjiv_BasicSurzyk_F	231

DW_A0031_Kyjiv_BasicSurzyk	80
DW_A0037_Kyjiv_BasicSurzyk_F	111
DW_A0038_Kyjiv_BasicSurzyk_M	140
DW_A0039_Kyjiv_BasicSurzyk_M	344
DW_A0044_Kyjiv_BasicSurzyk_F	252
DW_A0129_Kyjiv_DialectalSurzyk_M	250

Le tabelle generali con tutti gli elementi ibridi presenti nei testi analizzati (Tabella terminologica semplificata del Suržyk) e un modello con descrizione più approfondita dei termini (Tabella terminologica standard del Suržyk) sono presenti nelle appendici di questa tesi. Le tabelle terminologiche semplificate del Suržyk sono state create per rappresentare tutti i termini non appartenenti all’ucraino standard trovati nelle registrazioni audio raccolti da Del Gaudio a Kyjiv. Il modello della tabella terminologica standard di Suržyk è stato elaborato con Federica Vezzani, una studentessa di dottorato del XXXIII ciclo dell’Università degli Studi di Padova del DISLL, per definire gli aspetti importanti della terminologia del Suržyk. Sebbene la creazione delle tabelle terminologiche del Suržyk fosse solo la fase preparatoria e non l’obiettivo principale della tesi, riteniamo che potrebbe essere utile per ulteriori studi sul Suržyk.

La fase preparatoria della tesi ha richiesto una quantità di tempo significativo e si è deciso di identificare una particolare caratteristica dei verbi del Suržyk. A causa della ricorrente ripetizione della lettera finale “-m” caratterizzante la desinenza nella prima persona plurale del presente indicativo è stato deciso di analizzare la presenza di questa particella in modo più dettagliato. La ripetizione della lettera finale “-m” alla fine dei verbi in prima persona plurale è stata notata durante il processo di trascrizione manuale delle registrazioni raccolte da Del Gaudio. Per effettuare l’analisi sulla desinenza

caratterizzante era necessario confrontare i sistemi verbali del russo e dell'ucraino e, in particolare, le possibili desinenze della prima persona plurale nel presente indicativo. Oltre all'evoluzione storica delle desinenze della prima persona plurale nelle lingue slave sono state prese in considerazione anche le differenze attuali di queste desinenze in entrambe le lingue russo e ucraino.

Per svolgere il progetto sull'identificazione di elementi ibridi sono stati adottati gli strumenti del linguaggio di programmazione R nell'ambiente di sviluppo integrato computazionale che facilita la manipolazione dei dati RStudio. Fondamentali per il nostro progetto sono state le funzioni di word/pattern extraction e word/pattern matching. L'applicazione di RStudio ha permesso di scrivere il codice sull'identificazione automatica dei pattern che caratterizzano i verbi della prima persona plurale e di controllare i risultati tramite l'esecuzione del codice. Per visualizzare meglio i risultati dei test per ogni regola e tutti gli esempi di testo completo, abbiamo adottato l'editor di file Sublime Text 3 con cui abbiamo aperto tutte tabelle di output create in RStudio.

Sulla base di esempi presenti nei testi abbiamo fornito un elenco di regole volte a identificare i verbi ibridi della prima persona plurale. Nella prima fase del test abbiamo controllato i campioni del testo sulla presenza delle possibili desinenze dei verbi ibridi. Nella seconda e nella terza fase abbiamo definito con più precisione le regole sulla identificazione e le abbiamo divise in due gruppi: il gruppo generale e il gruppo specifico. Il gruppo generale consisteva in un breve riassunto delle regole specifiche, il che significa che 16 regole specifiche sono state riassunte e generalizzate in 4 regole generali. Con l'aiuto delle regole generali sono stati identificati 12 risultati nella seconda fase e 44 nella terza fase, per un totale di 56 elementi solo 11 erano risultati veri positivi per il nostro input. Per risultati veri positivi si intendono risultati che rispettano l'output mirato a definire i verbi con le desinenze della prima persona plurale in presente oppure i verbi copulari della prima persona plurale combinati alla forma infinita. Da un punto di vista durante queste due fasi le regole generali hanno reso

possibile identificare 56 elementi, ma soltanto 11 di questi rispettavano l'input della ricerca. Dall'altro punto di vista non si può omettere che testando le regole generali siamo riusciti a identificare altri elementi che, anche non essendo verbi, potrebbero contribuire alla creazione del corpus del Suržyk. Per questo motivo abbiamo deciso di presentare selettivamente alcuni di essi nella parte finale del lavoro di tesi come parte della tabella terminologica standard del Suržyk.

Durante la seconda e la terza fase sono state testate anche le regole specifiche che ci hanno permesso di identificare 11 combinazioni di verbi con la desinenza di prima persona plurale. Per un totale di dodici output, solo un risultato era falso positivo. Gli 11 risultati del test delle regole generali corrispondono ai risultati ottenuti con le regole specifiche. La differenza consiste nella quantità di output falsi positivi che mediante l'adozione delle regole specifiche erano limitati a uno. Si può concludere che le regole specifiche fornite sono efficaci se applicate sui nostri campioni di testo; per ragioni pratiche le presentiamo di seguito con il numero di output per ogni combinazione di pattern.

Tabella riassuntiva dei risultati.

Num. regola	Primo pattern	Secondo pattern	Distanza	Requisiti della parola di output	Quantità di risultati	Di cui veri pos.	Di cui falsi pos.
1	ми	-ем	max. 3		1	1	0
2	ми	-ем	max. 3		4	4	0
3	ми	-им	max. 3		1	1	0
4	ми	-ім	max. 3	Il patt. > "ім"	1	1	0
5	самі	-ем	max. 3		0	0	0

6	самі	-єм	max. 3		1	1	0
7	самі	-им	max. 3		0	0	0
8	самі	-їм	max. 3	II patt. > “їм”	0	0	0
9	-єм	-ти	max. 3	II patt. > “ти”	0	0	0
10	-єм	-ти	max. 3	II patt. > “ти”	0	0	0
11	-им	-ти	max. 3	II patt. > “ти”	0	0	0
12	-їм	-ти	max. 3	I patt. > “їм” II patt. > “ти”	0	0	0
13	-єм	-ть	max. 3		2	2	0
14	-єм	-ть	max. 3		1	0	1
15	-им	-ть	max. 3		1	1	0
16	-їм	-ть	max. 3	I patt. > “їм”	0	0	0

N.B.: tutte queste combinazioni devono rispettare la regola di base: il secondo elemento della combinazione deve seguire il primo elemento a una distanza da 1 a 3. Inoltre, le parole di output relative ai pattern “-їм” o “-ти” devono presentare più lettere rispetto al pattern di input. Questa condizione aggiuntiva deve essere applicata su tutte le regole che presentano “-їм” o “-ти” nel primo o secondo pattern della combinazione.

Nella fase finale dello sviluppo del progetto si è deciso di provare se le regole di identificazione avrebbero funzionato su testi nuovi. Il test è stato svolto con una selezione dei testi russi di lingua parlata, in particolare le trascrizioni delle conversazioni vere. Da un lato, si poteva presupporre che nessun elemento sarebbe stato identificato in quanto le regole che abbiamo creato mirano a identificare gli elementi di Surżyk. Dall'altro lato, avrebbe potuto succedere che all'interno dei testi russi ci

sarebbero stati elementi che presentano gli elementi di Suržyk, in particolare il pattern finale “-ем”. Dopo aver verificato le regole specifiche, è stato notato che tre delle regole specifiche sull'identificazione dei verbi di Suržyk hanno portato al riconoscimento degli elementi nei testi russi. Alcuni di questi elementi erano verbi con la desinenza dell'infinito “-ть”, altri invece erano sostantivi russi, aggettivi o pronomi personali che in diversi casi, come Dativo, Strumentale e Locativo, presentano la desinenza finale “-ем” o “-им” . Questo test finale può, da un punto di vista, confermare l'idea che se stessimo lavorando con testi forniti del corpus POS, i risultati sarebbero stati più appropriati e forse nessuna parola della lingua russa standard sarebbe stata identificata applicando le regole specifiche mirate all'identificazione dei verbi di Suržyk. Da un altro punto di vista, è necessario prendere in considerazione che Suržyk mescola gli standard russo e ucraino e contenere gli elementi di entrambe le lingue fa parte della sua struttura e delle sue caratteristiche principali.

Per concludere, è importante sottolineare che l'obiettivo di questa ricerca è stato adempiuto. Abbiamo analizzato la situazione linguistica in Ucraina, caratterizzata dal prevalente dominio di due lingue slave ucraino e russo e dalla presenza dei dialetti. Inoltre è stata visionata la teoria sullo stretto contatto delle due lingue e le sue conseguenze. La conoscenza delle basi del Natural Language Processing e lo studio dei campioni del Suržyk ci hanno permesso di lavorare in modo appropriato. Adottando un metodo di studio basato sugli esempi siamo riusciti a creare regole sull'identificazione automatica dei verbi di prima persona plurale in RStudio. Queste regole sono state generalizzate per verificare se i risultati potessero essere migliorati in quantità di output veri. Dopo la prima fase dello sviluppo del progetto è stata preparata una lista di possibili risultati falsi positivi; ci si aspettava che durante le future fasi di test potessero emergere tutte le parole che presentano un pattern inserito come input ma che in base alla loro affiliazione morfologica non soddisfano i nostri output stimati. Con la creazione delle regole generali siamo riusciti a ridurre il numero dei risultati falsi positivi. Nella prima fase si presupponeva che almeno 18 parole presenti nei testi, contenenti nel pattern dell'input ma non corrispondenti all'output mirato, sarebbero

emerse come risultati. Successivamente, durante la seconda e la terza fase dello sviluppo del progetto testando le regole specifiche è apparso un solo risultato falso positivo. Riassumendo, l'adozione delle regole specifiche ha permesso di identificare con successo gli elementi mirati.

Gli elementi identificati riguardanti i verbi con il pattern finale caratterizzante la prima persona plurale erano 11. La quantità totale dei termini evidenziati nei testi analizzati e classificati come non appartenenti alla lingua standard era di 1408 unità. Bisogna ammettere che non tutte le entità che fanno parte di 1408 possono essere considerate Suržyk, ma sicuramente più della metà di questi termini appartengono a Suržyk, ciò potrebbe essere verificato confrontando le colonne corrispondenti al termine originale trovato nel testo, la sua versione ucraina e quella russa. Ulteriori studi dovrebbero svilupparsi verso il riconoscimento di entità nominali (Named Entities Recognition). Durante il processo dell'analisi è stato realizzato che diversi tipi di regole potrebbero funzionare meglio se applicate alle entità già classificate. Gli studi su Named Entity Recognition sono comuni per le lingue maggioritarie, ma quando si tratta delle lingue minoritarie, o addirittura delle lingue ibride o miste, il materiale del corpus annotato e pronto all'uso non è sufficiente, o addirittura assente. A questo punto non possiamo procedere alla creazione del corpus annotato di Suržyk ma forniamo sette tabelle terminologiche semplificate di Suržyk con un totale di 1408 termini, un modello della tabella terminologica standard di Suržyk con una descrizione approfondita di cinque termini della lingua ibrida e sedici regole specifiche sull'identificazione automatica del pattern finale caratterizzante le desinenze della prima persona plurale implementate in R.

## **Автоматическая идентификация гибридных языков: казус суржика.**

### **Краткое содержание дипломной работы**

В связи с глобализацией распространённость тесного контакта различных языков стала актуальным аргументом в современном обществе. Наличие двух или более языков в дискурсе может привести к таким явлениям как смешение кодов. В ситуациях контакта двух языков может возникнуть совершенно новый язык, в котором будут присутствовать элементы обоих языков вовлечённых в его создании. Гибридность смешанных языков заключается в том, что они перенимают свой лексикон из одного источника, а грамматику из другого. Классификация такого типа языков является сложной, поскольку на основе лексики они могут принадлежать одному языку, а на основе морфологии, синтаксиса и общих грамматических характеристик их можно отнести к другому языку. Явления гибридных языков представляются в первой части этой работы вместе с описанием гибридного украинско-русского языка суржика, который был создан в результате длительного сосуществования украинского и русского языков.

В курсе этой магистерской дипломной работы изучается во-первых языковая ситуация в Украине, известная присутствием двух славянских языков: украинского и русского. Как результат длительного сосуществования украинского и русского языков в Украине возник суржик — гибридный украинско-русский язык, представляющий характеристики смешанного народно-разговорного языка. Период возникновения суржика обозначен концом XVII века, когда в связи с политикой русификации украинскоязычные люди начали переходить или были вынуждены перейти на более престижный русский язык. Нынешний статус суржика весьма не престижный и его узус обычно ограничен разговорной речью. Распространение суржика было зарегистрировано на территории всей Украины, с самой большой концентрацией в центральной, южной и восточной частях страны.



Структура суржика весьма комплексная, в 2010 К. Кент предложила использовать модель под названием Matrix Language Frame, для его анализа. Следуя модели, украинский язык выступает в роли основного языка, в то время как русский язык входит в систему в качестве встроенного языка. Помимо этой идеи, мы предполагаем, что кроме смешивания русского и украинского языков, структура суржика допускает дополнительные изменения в зависимости от географического бытования и от характеристик его носителя. Можно предположить, что структура украинско-русского суржика включает два или даже три вида смешивания: украинский и русский языки смешиваются на ранней стадии, создавая базу; затем следует дополнительное смешивание украинских диалектов, и в конечном итоге добавляются некоторые мутации связаны с персонализацией речи со стороны носителя. Несмотря на то, что эта дискуссия очень важна в вопросе суржика, в данной работе она не изучается подробно, так как требует отдельного исследования. Наша цель состояла в том, чтобы идентифицировать глаголы первого лица множественного числа, используя инструменты языка программирования R и интегрированной среды разработки RStudio.

Основной темой этой дипломной работы является именно автоматическая идентификация элементов суржика. Для её достижения многое должно быть сделано в этом направлении, сначала можно идентифицировать присутствие гибридных элементов в разговорной речи вручную и уже впоследствии разработать систему их автоматической идентификации. Эта часть дипломной работы основана на области изучения, которая фокусируется на взаимодействии между человеческим языком и компьютерами и известна как обработка естественного языка (Natural Language Processing). Она является общим направлением между искусственным интеллектом и компьютерной лингвистикой и используется в областях машинного перевода, коррекции грамматики и автоматических ответов на вопросы. Глобальная цель нашего исследования может быть обозначена как возможность будущего автоматического распознавания украинско-русских гибридных элементов без вмешательства человека и

результаты данной дипломной работы могут послужить основой для дальнейшей эволюции этой темы.

Исследование связано с различными научными направлениями и должно рассматриваться как междисциплинарное. Первая часть дипломной работы теоретическая и состоит из трех глав. В первой главе изучается языковая ситуация Украины, основные диалекты распространенные на территории Украины, и положения о языковой политике из Конституции Украины; определяются отношения между украинским и русским языками и объясняется происхождение гибридного украинско-русского языка суржик. Во второй главе представлены различные явления, характерны ситуациям двуязычия и многоязычия. Вместе с обозначением гибридных языков мы представили и существующую классификацию суржика как гибридного языка с качествами смешанного языка. В третьей главе представлена теория обработки естественных языков, область информатики занимающаяся исследованием того, как компьютерные системы могут анализировать, понимать или производить естественные языки.

Вторая часть дипломной работы практическая и разделяется на четыре главы. В первой главе представлен этап подготовки проекта, состоящий в сборе материала для анализа и его обработке вручную; ознакомление с языком программирования R и с интегрированной средой разработки RStudio путем изучения их основных функций; описание элементов, которые необходимо проанализировать на этапах тестирования. Вторая глава второй части посвящена разработке проекта и делится на четыре этапа тестирования. Первые три этапа касаются образцов киевского суржика, которые были собраны Дель Гаудио в Украине и опубликованы на диске вместе с его докторской диссертацией «О природе Суржика» (2010). Последний этап представляет собой дополнительное тестирование прежде созданных правил на текстах русского языка. Во второй главе представлены все правила по автоматической идентификации глаголов первого лица множественного числа изъявительного наклонения. В заключении предоставляется список истинных

положительных результатов, полученных в ходе этого исследования, общие замечания о важности этой темы в многоязычном обществе и обсуждение возможных изменений.

В качестве заключительной части этой дипломной работы было решено предоставить приложения с терминологическими таблицами разговорного языка суржика. Они делятся на два типа: общие таблицы со всеми гибридными элементами, которые присутствовали в проанализированных текстах (Упрощенная терминологическая таблица суржика) и одна модель более глубокого описания терминов гибридного языка (Стандартная терминологическая таблица суржика). Упрощенные терминологические таблицы суржика были разработаны для предоставления всех терминов не соответствующих стандартному украинскому языку, которые мы нашли в киевских аудиозаписях собранных Дель Гаудио, или которые были обозначены автоматическим корректором в Google Drive Documents как неправильные элементы украинского языка. Структура модели Стандартной терминологической таблицы суржика была разработана со студенткой аспирантуры при кафедре Лингвистики, Филологии и Литературоведения Падуанского Университета Ф. Вещани для определения важных аспектов терминологии суржика. Несмотря на то, что создание терминологических таблиц суржика было только подготовительным этапом, а не основной задачей этой магистерской работы, мы считаем, что они все-таки могут быть полезными для дальнейших исследований в области суржика.

Суржик характеризуется как не престижный разговорный язык; он широко распространен по всей Украине с самой высокой концентрацией носителей в центральных, южных и восточных районах Украины. Поскольку Суржик более распространен в разговорной речи, чем в письменном языке, одним из основополагающих принципов этой работы был анализ образцов речи.

Метод исследования, основанный на примерах, позволил идентифицировать различные элементы суржика. На первом этапе были созданы документы и прописаны интервью, которые Дель Гауди записал в Украине как аудиофайлы и опубликовал вместе с докторской диссертацией «О природе Суржика» (2010). Дель Гауди собрал интервью людей в Киевской, Черниговской и Харьковской областях, представляя таким образом экземпляры разговорной речи и суржика. В этой магистерской дипломной работе была обработана и проанализирована в деталях только первая группа аудиофайлов касающаяся Киевских областей. Подготовительная фаза дипломной работы состояла в ручной транскрипции записей Дель Гауди и позволила с самого начала отметить присутствие суржика в образцах. Это привело нас, во-первых, к созданию текстов суржика в Microsoft Word, и в результате, к выделению элементов, которые не соответствуют стандартам украинского языка, или которые были признаны неправильными автоматическим корректором документов в Google Drive Documents. Впоследствии были заполнены представительные таблицы с этими элементами. Для каждой записи, то есть для каждого прописанного текста, была создана упрощенная терминологическая таблица. Структура упрощенных таблиц была принята для классификации терминов, найденных в записях, и для лучшего понимания их отношения к украинскому или русскому языку. В итоге было создано 7 таблиц для 7 записей Киевской области. Общее количество терминов, которые считаются гибридными, не относящимися к правилам стандартного украинского языка, составляет 1408 единиц. Все термины были представлены в Упрощенных терминологических таблицах суржика. Для каждой записи (мы сохранили исходные заголовки аудиофайлов) было найдено:

Аудиофайл	Количество найденных слов
DW_A0027_Kyjiv_BasicSurzyk_F	231
DW_A0031_Kyjiv_BasicSurzyk	80
DW_A0037_Kyjiv_BasicSurzyk_F	111

DW_A0038_Kyjiv_BasicSurzyk_M	140
DW_A0039_Kyjiv_BasicSurzyk_M	344
DW_A0044_Kyjiv_BasicSurzyk_F	252
DW_A0129_Kyjiv_DialectalSurzyk_M	250

В приложениях дипломной работы мы включили общие таблицы со всеми гибридными элементами.

На подготовительный этап работы потребовалось значительное количество времени, впоследствии было решено определить конкретную характеристику глаголов суржика. Ввиду повторяющегося окончания «-м» в первом лице множественного числа настоящего времени в обработанных текстах было решено проанализировать присутствие этой частицы подробнее. Повторение «-м» как последнего элемента окончания в первом лице множественного числа было замечено в процессе ручного прописывания регистраций собранных Дель Гаудио. Для проведения этого анализа необходимо было сравнить русские и украинские вербальные системы и, в частности, окончание первого лица множественного числа в настоящем времени. Вместе с фактами об историческом развитии в славянских языках окончаний первого лица множественного числа были учтены также и их настоящие различия в русском и украинском языках.

Для разработки проекта по идентификации гибридных элементов мы приняли инструменты языка программирования R и интегрированной среды разработки RStudio, которая облегчает манипулирование данными. RStudio позволил написать код для автоматической идентификации частиц характеризующих глаголы первого лица множественного числа, и проверить результаты. Для того, чтобы лучше визуализировать выходные таблицы RStudio, в которых были

представлены результаты каждого правила и примеры из текста, мы дополнительно использовали редактор файлов Sublime Text 3.

На основе примеров был представлен список правил, направленных на идентификацию гибридных глаголов первого лица множественного числа. На первом этапе тестирования образцы текста проверялись на наличие возможных окончаний гибридных глаголов. В курсе второй и третьей фаз мы более точно определили правила распознавания и разделили их на две группы: общую и специальную. Общая группа была кратким изложением специальных правил, это значит, что 16 специальных правил были обобщены в 4 общих правилах. Общие правила определили 12 результатов на втором этапе и 44 на третьем этапе, идентифицируя всего 56 элементов, но только 11 из них были истинными положительными для нашего ввода. Под “истинными положительными” имеются в виду результаты, которые уважают целевой выход, то есть они должны быть глаголами с окончаниями первого лица множественного числа в настоящем времени или копульными глаголами в первом лице множественного числа связанными с инфинитивной формой. С одной стороны, в течение этих двух фаз общие правила позволили нам идентифицировать 56 элементов, но только 11 из них соблюдали поисковый ввод. С другой стороны, нельзя не отметить, что при проверки общих правил, нам удалось идентифицировать другие элементы. Эти дополнительные элементы могут быть отнесены к корпусу суржика, несмотря на то, что они не являются глаголами, и таким образом не удовлетворяют наш поиск. Мы решили выборочно представить некоторые из них в заключительной части этой работы, а именно в Стандартной терминологической таблице суржика.

В ходе второй и третьей фаз мы также протестировали специальные правила, которые позволили нам идентифицировать 11 комбинаций глаголов с окончаниями первого лица множественного числа. В общей сумме 12 поисковых выходов только один результат был ложным положительным. 11 результатов, полученных с общими правилами, соответствуют результатам, которые мы

получили при помощи специальных правил. Разница заключалась в количестве ложных положительных результатов, которые были ограничены к 1 путем принятия специальных правил. Можно сделать вывод, что тестирования специальных правил на наших экземплярах текста было эффективным, поэтому по практическим соображениям мы представляем правило и количество выходов для каждой комбинации в следующей таблице.

Таблица результатов.

Правило	I элемент	II элемент	Дистанция	Запросы к выходному слову	Количество результатов	Ист. полож.	Ложные полож.
1	ми	-ем	макс. 3		1	1	0
2	ми	-єм	макс. 3		4	4	0
3	ми	-им	макс. 3		1	1	0
4	ми	-їм	макс. 3	II эл. > "їм"	1	1	0
5	самі	-ем	макс. 3		0	0	0
6	самі	-єм	макс. 3		1	1	0
7	самі	-им	макс. 3		0	0	0
8	самі	-їм	макс. 3	II эл. > "їм"	0	0	0
9	-ем	-ти	макс. 3	II эл. > "ти"	0	0	0
10	-єм	-ти	макс. 3	II эл. > "ти"	0	0	0
11	-им	-ти	макс. 3	II эл. > "ти"	0	0	0

12	-їм	-ти	макс. 3	I эл. > “їм” II эл. > “ти”	0	0	0
13	-ем	-ть	макс. 3		2	2	0
14	-єм	-ть	макс. 3		1	0	1
15	-им	-ть	макс. 3		1	1	0
16	-їм	-ть	макс. 3	I эл. > “їм”	0	0	0

Все вышеперечисленные комбинации должны уважать основное правило: второй элемент должен располагаться после первого элемента на дистанции от 1 до 3. Кроме того, выходные слова, касающиеся окончаний «-їм» или «-ти», должны представлять в своем составе больше букв, чем частица ввода. Это дополнительное условие касается всех правил, которые имеют «-їм» или «-ти» в первом или во втором элементе комбинации.

В заключительной фазе разработки проекта было решено проверить, будут ли правила идентификации работать на новых текстах. Было решено выбрать русские тексты разговорного языка, в частности, некоторые прописки устной речи. [С одной стороны, можно было бы предположить, что ни один элемент представленных правил не будет идентифицирован, поскольку созданные нами правила направлены на то, чтобы идентифицировать суржик. С другой стороны, также возможно, что в русских текстах будут присутствовать элементы, которые представляют некоторые окончания подобны образцу глаголов суржика, например окончание «-ем». После проверки общих правил было отмечено, что три общих правила идентификации глаголов суржика привели к распознаванию элементов в стандартных русских текстах. Некоторые из этих результатов были глаголами с инфинитивным окончанием “-ть”, другие же были русскими существительными, прилагательными или личными местоимениями, которые представляют окончание «-ем» или «-им» в разных падежах, например в



Дательном, Инструментальном или Местном падежах. Данное тестирование может, с одной стороны, подтвердить мысль о том, что если бы мы работали с текстами, представляющими автоматическую морфологическую разметку POS, результаты могли бы быть более уместным. Возможно стандартный русский язык не был бы идентифицирован с помощью общих правил идентификации глаголов суржика. С другой стороны, нужно учитывать, что суржик смешивает российские и украинские стандарты, и поэтому в своей структуре он представляет элементы обоих языков.

В заключение важно подчеркнуть, что цель этого исследования была удовлетворена. Мы проанализировали языковую ситуацию в Украине (характеризующуюся преобладанием украинского и русского языков), распространение диалектов на территории Украины, теории языкового контакта и их возможные последствия. Ознакомление с основами обработки естественного языка и изучение образцов суржика позволили нам работать в правильном направлении. Приняв метод основанный на примерах, мы сумели создать правила автоматической идентификации глаголов первого лица множественного числа в RStudio. Эти правила были также обобщены с целью проверки возможности улучшения верных результатов. В начале мы подготовили список ложных положительных результатов, так как ожидалось, что на разных этапах тестирования могли появиться слова, представляющие элементы поисковых частиц но не соответствующие нашему поисковому запросу, по причинам их морфологической категории например. Создав же специальные правила, нам удалось уменьшить количество ложных положительных результатов. На первом этапе мы ожидали, что общая минимальная сумма ложных положительных результатов будет 18 слов. После второго и третьего этапа разработки проекта, в процессе тестирования специальных правил, только один результат был ложно положительным, то есть он имел в своем составе частицу, которую мы искали, но не относился к категории глаголов. Подводя итог, принятие общих правил привело к успешной идентификации целевых элементов.

Количество идентифицированных элементов, касающихся глаголов с окончанием первого лица множественного числа, составило 11 единиц. Общее количество нестандартных терминов в проанализированных текстах было 1408. Нужно отметить, что не каждый элемент, входящий в сумму 1408, может считаться суржиком, но, безусловно, больше чем половина этих терминов таки относится к суржику. Принадлежит ли слово к словарю суржика можно проверить проанализировав его перевод на украинский и русский язык сопоставляя столбцы таблицы (Упрощенная терминологическая таблица суржика) соответствующие исходному термину найденному в тексте, его украинском и русском вариантах. Будущие исследования в этом направлении должны также развиваться в области Распознавания Именованных Элементов. В ходе нашего анализа было осознано, что функциональность различных правил идентификации может быть улучшена при их применении на уже классифицированные объекты. Научные исследования по распознаванию именованных элементов являются распространенными в области стандартных языков, но когда дело доходит до нестандартных языков, они почти не существуют. На данный момент мы не можем предоставить аннотированный корпус нестандартного языка суржик в связи с временными ограничениями, но мы разработали 7 упрощенных терминологических таблиц суржика с общей суммой 1408 терминов, дополнительную модель стандартной терминологической таблицы с более глубоким описанием пяти выбранных терминов суржика и 16 общих правил автоматической идентификации окончаний первого лица множественного числа суржика разработанные на языке программирования R.

## Appendices

### Standard terminological tables of Surzhyk<sup>83</sup>

Term	злазить
Pronunciation	[злáзиет']
Prefix	з
Root	лаз
Suffix	и
Ending	ть
Ukrainian translation	злазити
Russian translation	слезать
English translation	get off
Ukrainian pronunciation	[злáзиетие]
Russian pronunciation	[сл'иезáт']
English pronunciation	[get ɒf]
Gramm. category	v.inf.
Definition	Describes an active action of getting down from a higher surface, getting off the public transport, getting off the ladder.
Domain of usage	Transport
Geographical area	Bila Tserkva (Kyiv region)
Dominant language	Ukrainian
Context	«Де будем злазить?», «злазили з драбинки».
Status of certainty and reliability	+

<sup>83</sup> The structure of these terminological tables was elaborated with help of F. Vezzani.

Notice that the status of certainty and reliability could be of three types: “+” means high whereas “-” means low. Additionally “+/-” signs that more analysis may be required.

Term	почти
Pronunciation	[пocht'і]
Prefix	
Root	почти
Suffix	
Ending	
Ukrainian translation	майже
Russian translation	почти
English translation	almost
Ukrainian pronunciation	[мáйже]
Russian pronunciation	[пocht'і]
English pronunciation	['ɔ:lmeʊst]
Gramm. category	adverb
Definition	Expresses a similarity, closeness, proximity.
Domain of usage	General, money.
Geographical area	Skvyra (Kyiv region)
Dominant language	Ukrainian <sup>84</sup>
Context	«Двесті гривень – это не деньги... Почти даром. Ну там же в районах роботи ні в одному районі нема.»
Status of certainty and reliability	+/-

<sup>84</sup> It was concluded that Ukrainian is a dominant language of the speaker, but his lasting permanence in Russian-speaking environment could have signed his vocabulary, especially the lexicon connected to his job.

Term	вопшем
Pronunciation	[вóпшием]
Prefix	в
Root	опш
Suffix	е
Ending	м
Ukrainian translation	в загальному
Russian translation	в общем
English translation	generally
Ukrainian pronunciation	[в загáлноуму]
Russian pronunciation	[в óпш'ьм]
English pronunciation	['dʒenərəli]
Gramm. category	adverb
Definition	Considering the whole of something generally; summarising.
Domain of usage	General, profession.
Geographical area	Kyiv
Dominant language	Ukrainian
Context	«Вопшем підодіяльники шили, труси шили мужські, я в смену шила по трицять пар трусів, представляєте?»
Status of certainty and reliability	+

Term	пїєсять
Pronunciation	[пїйіс'át]
Prefix	
Root	пїс
Suffix	ять
Ending	
Ukrainian translation	п'ятдесят
Russian translation	пятьдесят
English translation	fifty
Ukrainian pronunciation	[пїятдеис'át]
Russian pronunciation	[п'ьд':ис'át]
English pronunciation	['fifti]
Gramm. category	noun
Definition	Expresses a quantity of fifty entities.
Domain of usage	Numerals, dates.
Geographical area	Skvyra (Kyiv region)
Dominant language	Ukrainian <sup>85</sup>
Context	«Да, я всю жизнь проработал. С пїєсять четвертого года по дві тисячі год отработал, камінщиком, кірпічіки складував.»
Status of certainty and reliability	+/-

<sup>85</sup> It was concluded that Ukrainian is a dominant language of the speaker, but his lasting permanence in Russian-speaking environment could have signed his vocabulary, especially the lexicon connected to his job.

Term	ПОТОМ
Pronunciation	[пoтoм]* In other speakers [пaтoм] may be possible
Prefix	
Root	ПОТОМ
Suffix	
Ending	
Ukrainian translation	пoтiм
Russian translation	пoтoм
English translation	later, then
Ukrainian pronunciation	[пoтiм]
Russian pronunciation	[пaтoм]
English pronunciation	[ˈleɪtə], [ðen]
Gramm. category	adverb
Definition	By a sequence designation expresses an action that happens later, comes after.
Domain of usage	General, conversation.
Geographical area	Kyiv region
Dominant language	Ukrainian
Context	«А потом значить, я кажу йому...»
Status of certainty and reliability	+

**Simplified terminological tables of Surzhyk**



здрaсте	здpaстуйте	здpaвстуйте	здpaвстуйте			Greetings		
двaцять	двaцять	двaдцaть	двaцaть			twenty		
кілометрoв	кілометрів	кілометрoв	кілометрaв			kilometers		
oтсюдa	звідси	oтсюдa	aтсюдa			from here		
пo-укрaїнськи	укрaїнськoю	пo-укрaїнськи	пo-укрaїнськi			in Ukrainian		
через тo	через тe	из-зa тoгo	із-зa тaвo			because of that		
бoльшeнькi	бiльшeнькi	бoльшeньки	бoльшeнькi			big		
кoнчилaсь	кiнчилaсь	кoнчилaсь	кoнчiлaсь			to be finished		
пoїхaть	пoїхaти	пoeхaть	пaeхaть			to go		
шo	щo	чтo	штo			what		
пaдpужки	пoдpужки	пoдpужки	пaдpужкi			friends (fem.)		
Київськaя	Київськa	Киeвськaя	Київськaя			of Kyjiv		
Бeлaя Цepкoвь	Бiлa Цepквa	Бeлaя Цepкoвь	Бeлaя Цepкaвь			Bila Tserkva		
рядoм	пopяд	рядoм	рядaм			near		
дa	тaк	дa	дa			yes		
рaбoтaлa	пpaцoвaлa	рaбoтaлa	рaбoтaлa			to work (Past, 3 pers. sing.)		
сюдa	сюди	сюдa	сюдa			here (direction)		
нpaвитьcя	пoдoбaється	нpaвитьcя	нpaвитьcя			to like (3 pers. sing.)		
oт	від	oт	oт		зaбpaли oт зaбoт	from		
зaбoт	тyрбoт	зaбoт	зaбoт			worries (Genitive)		
aгaрoд	гoрoд	oгoрoд	aгaрoд			garden		
нeмнoжкo	нeбaгaтo	нeмнoжкo	нeмнoжкa			a little bit		
дoм	дiм	дoм	дoм			house		
всьo	вce	всe	всьo			everything		
нaпpoтiв	нaвпaки	нaoбoрoт	нaaбaрoт			on the contrary		
нo	aлe	нo	нo			but		
рaбoтy	рoбoтy	рaбoтy	рaбoтy			work (Genitive)		
мaлaдьoж	мoлoдь	мoлoдeжь	мaлaдьoж			youth		
дaмoй	дoдoмy	дoмoй	дaмoй		я пpийду дaмoй	home (Genitive)		
нe мaгy	нe мoжy	нe мoгy	нe мaгy			to be not able (1 pers. sing.)		
нepви	нepви	нepвы	нepви			nerves		
кyрaть	кyрaть	кyрaт	кyрaт			to smoke (3 pers. plur.)		
aкpужeнiє	oтoчeння	oкpужeнiє	aкpужeнiє			surrounding		
кpyгoм	кpyгoм	кpyгoм	кpyгoм			around		
aкнo	вiкнo	oкнo	aкнo			window		
cмoтpю	дивлoсь	cмoтpю	cмaтpю			to look (1 pers. sing.)		
iдьoт	iдe	iдeт	iдьoт			to walk (3 pers. sing.)		

как	як	как	как		таким как ми	as		
прімєрно	преблизно	пребмерно	пребмерно			approximately		
отак во	отак от	так вот	так вот			so thus		
она	вона	она	ана			she		
большая	велика	большая	большая			big (f.)		
маленькій	маленький	маленький	маленький			small (m., Nominative)		
радосно	радїсно	радосно	радасна			joyfully		
понїмаєте	розумїєте	понимаєте	панїмаєте			to understand (Interrogative, plur.)		
свєокр	свєкор	свєкор	свєокр			father-in-law		
анї	вони	они	анї			they		
девочки	дївчатка	девочки	девачкї			girls		
замуж	замїж	замуж	замуж		вийшли замуж	marriage (Accusative)		
две	двї	две	две			two		
продалї	продали	продали	прадалї			to sell (Past, plur.)		
прїєхалї	прїїхали	прїєхали	прїєхалї			to come (Past, plur.)		
нє	нї	нет	нет			no		
кумов'я	куми	кумовья	кумавья			godfathers		
друзья	друзї	друзья	друзья			friends		
трошечки	трошечки	немножечко	немножечка			a little bit		
подожди	почекай	подожди	падаждї			to wait (Imperativ, 2 pers. sing.)		
сахар	цукор	сахар	сахар			sugar		
говору	говору	говору	гаварю			speek (1 pers. sing.)		
нащо	навїщо	зачем	зачем			what for		
щас	зараз	сейчас	сєйчас			now		
повтора	повторює	повторяет	павтаряет			repeat (3 pers. sing.)		
будем	будемо	будем	будем			to be (Future, 1 pers. sing.)		
злазить	злазити	слезать	слезать			to get off		
ето	це	это	ета			this		
виходїть	виходити	выходить	вихадїть			to get out		
коркуль	куркуль	куркуль	куркуль			kulak		
ще	ще	ещє	сщю			again		
зажиточний	заможний	зажиточный	зажитачный			wealthy (m.)		
челавек	людина	человек	чславек			person		
бульо	було	было	била			to be (Past, sing. f.)		
строїть	строїти	строить	строїть			to build		
дедушка	дїдусь	дедушка	дедушка			grandfather		
шоб	шоб	чтобы	штоби			in order to		

Рождество	Різдво	Рождество	Раждество			Christmas		
нада	потрібно	надо	нада			need (3 pers. sing.)		
полотенця	полотенця	полотенца	палатенца			towels		
тоже	також	также	также			also		
панімаєте	розумієте	понимаєте	панімаєте			understand (Interrogative, plur.)		
веретениця	веретено	веретено	верстено			spindle		
толкачка	товкачка	толкушка	талкушка			masher		
раньше	раніше	раньше	раньше			earlier		
зажарку	засмажку	зажарку	зажарку		на зажарку	roast		
отак	ось так от	вот так вот	вот так вот			so like this		
безпе́кості	безпе́ки	безопасности	безопасна́сті			security (Genitive)		
тоже	теж	тоже	тоже			also		
хароша	хороша	хорошая	харошая			good		
в їх	в них	в них	в ніх			by them		
коло його	біля нього	возле него	возле него			near to him		
другий офіс	інший офіс	другой офис	другой офіс			another office		
двохкомнатній	двокімнатній	двохкомнатной	двохкомнатнай		в двухкомнатній квартирі	two-room (Lokative)		
квартирі	квартирі	квартире	квартире			apartment (Lokative)		
харашо	добре	хорошо	харашо			good / fine		
двохкомнатну	двокімнатну	двохкомнатну	двохкомнатну		в двухкомнатну квартиру	two-room (Accusative)		
квартиру	квартиру	квартиру	квартиру			apartment (Accusative)		
домік	будиночок	домик	домік			small house		
місяць	місяць	месяц	місяць			month		
етаж	поверх	этаж	етаж			floor		
тоска	нудьга	тоска	таска			wistfulness		
кумов'я	куми	кумовья	кумов'я			godfathers		
сутки	доба	сутки	суткі			twenty-four hours		
працюєм	працюємо	работаем	работаем			to work (1 pers. pl.)		
змінюся	змінюся	изменюсь	ізмєнюсь			to change (Fut., 1 pers. sing.)		
прилічні	пристойна	приличная	прілічна			decent		
абажаю	обожаю	обожаю	абажаю			to adore (1 pers. sing.)		
любляща	любяча	любящая	любяща			loving		
обідить	образить	обидит	абідить			to offend		
судья	судья	судья	судья			judge		
більш любіші	любіші	более ласковые	болєє ласкавіє			more tender		
жінщина	жінка	женщина	жінщина			woman		
очень	дуже	очень	очень			a lot		

куда	куди	куда	куда			where (direction)		
врач	лікар	врач	врач			doctor		
обідно	прикро	обидно	абідна			offensively		
знаєм	знаємо	узнаем	узнаем			get to know (Fut. 1 pers. plur.)		
фамілії	прізвищу	фаміліи	фамілії		по фамілії	by surname		
по чотири	по чотири	по четьре	по четири		по четири человека	by four		
человека	чоловіка	человека	человека			person (Genitive)		
есть	є	есть	есть			to be (3 pers. sing.)		
больше	більше	больше	больше			more		
человек	людина	человек	человек			persons		
квартиранти	квартиранти	квартиранты	квартиранти			tenants		
електричкою	електричкою	електричкой	електричкою		електричкою їду	to be offended (1 pers. sing.)		
рубля	рублі / гривні	рубля	рубля	The interview take place a	два рубля	rubli		
нічо	нічого	ничего	нічево			nothing		
прелесна	чарівна	прелестная	прелесная			adorable		
гармоню	гармонь	гармонь	гармонь			accordion (Accusative)		
співа	співає	поет	пайот			to sing (3 pers. sing.)		
повидавала	повидавала	повыдавать	павидавать		повидавала заміж	to give away, when in collocation with "заміж" than to marry off (Past)		
пійсят	п'ятьдесят	пятьдесят	п'ятьдесят		пійсят шість	fifty		
пійсять	п'ятьдесят	пятьдесят	п'ятьдесят		пійсять сім	fifty		
ана	вона	она	ана			she		
создає	створює	создаёт	сздають			to create (3 pers. sing.)		
наверно	напевно	наверно	наверна			probably		
отдала	віддала	отдала	отдала			to give away (Past)		
поотдавала	повіддавала	поотдавала	паатдавала			to give away (Past)		
решотка	решітка	решётка	решотка			lattice		
входная	вхідна	входная	входная			of the entrance (adj.)		
сібе	собі	себе	себе		само по сібе	for oneself		
отключается	відключається	отключается	атключаетса			to switch off (impersonal)		
автоматически	автоматично	автоматически	автаматическі			automatically		
автаматика	автоматика	автоматика	автаматика			automatics		
первий	перший	первый	первий			first		
пеша	пішки	пешком	пешком			on foot / by walking		
пешка	пішки	пешком	пешком			by foot / by walking		
пешком	пішки	пешком	пешком			by foot / by walking		
спасіба	спасибі	спасибо	спасіба			thank You		

такіє	такі	такіє	такіє			such		
діла	справи	дела	дела			affairs		
скрізь?	скрізь	езде	езде			everywhere		
фестиваль	фестиваль	фестиваль	фестиваль			festival		
годах	роках	годах	годах		в яких годах	years		
даже	навіть	даже	даже			even		
красіві	красиві	красивые	красивієє			beautiful (plur.)		
налісники	налісники	блины	бліни	No perfect match in E.		filled pancakes		
бліни	млинці	блины	бліни	No perfect match in E.		pancakes		
макітра	макітра	макітра	макітра	Used in Eastern European kitchens		makitra / clay bowl		
розтреш	розтреш	растрешь	растрьош	не розтреш за раз		to knead (2 pers. sing.)		
мукой	мукою	мукой	мукой	з мукой		(with) flour		
мужу	мужу	мужу	мужу			husband (Dative)		
соль	сіть	соль	соль			salt		
вкосу	смаком	вкосу	вкосу		по вкосу VS за смаком	on taste		
в їх	в них	у них	у ніх			by them		
нічю	нічого	ничего	нічево			nothing		
начинить	начинити	начинить	начініть			to fill		
конвертіком	конвертиком	конвертиком	канвертікам	No exact match in E.		like envelope		
макітру	макітру	макітру	макітру		в макітру	makitra / clay bowl (Acusative)		
обезательно	обов'язково	обязательно	абязательна			necessarily		
тре	потрібно	нужно	нужна			to need (3 pers. sing.)		
собираю	збираю	собираю	сабіраю			to collect (1 pers. sing.)		
отдельно	окремо	отдельно	атдельна			separately		
отварюю	відварюю	отвариваю	атваріваю			to boil smth. (1 pers. sing.)		
тіки	тільки	только	толька			just		
вещас	весь час	всё время	всьо время			all time		
де-то	десь там	где-то	где-та			somewhere		
в Чукутці	в Чукутці	в Чукутке	в Чукутке			in Chukotka		
приїжають	приїжджають	приезжают	прієзжают			to come (3 pers. plur.)		
пол	пів	пол	пол			half		
крепкий	міцний	крепкий	крепкий			strong		
кумусю	кумусю	кумушка	кумушка			godmother (Vocative)		
нравлятся	подобаються	нравятся	нравятса			to like		
тугеньке	тугеньке	тугое	тугоє	"туге" is more used		tight		
ложете	ложите	ложите	ложите			to lie down (2 pers. plur.)		
сторон	сторін	сторон	старон			sides (Genitive)		

двух	двох	двух	двух		з двух	two (Genitive)		
помить	помити	помыть	памить			to wash (inf.)		
косточкою	кісточкою	косточкой	костачкай		з косточкою	(with) bone		
косточка	кісточка	косточка	костачка			bone (of fruits)		
отходить	відходить	отходит	атходіт			to go away (3 pers. sing.)		
сладке	солодке	сладкое	сладкає			sweet		
считается	вважається	считается	счітаєтса			to be considered (3 pers. sing.)		
згльогне	загусне	загустеет	загустєєт	Probably an old dialectal form		to thicken	?	
звертиться	звертиться	свертится	свертітса	Rarely used in U, R, no E match		become sour (in this Speaker speech)	?	
всеодно	все одно	все равно	всьо равно			anyway		
біологіческіє	біологічні	биологические	біалагіческіє			biological		
сладость	солодкість	сладость	сладасьт			sweetness		
прокисше	прокисле	прокисшее	пракисшеє			sour		
лучше	краще	лучше	лучше			better		
получается	виходить	получается	палучаєтса			come out (3 pers. plur.)		
хоть	хоть	хоть	хоть			though		
творог	творог	творог	твораг			cottage cheese		
зара	зараз	сейчас	сейчас			now		
смачнющу	смачнющу	вкуснейшую	вкуснейшую			very tasty (Accusative)		
шиїсят	шістдесят	шестьдесят	шестьдесят			sixty		
сковзалися	сковзалися	скользили	скальзілі			to slid (Past, plur.)		
сковзанки	ковзанка / сковзанка / каток	каток	каток	In U "сковзанка" is rarely used.		rink (plur.)	?	
поняла	зрозуміла	поняла	паняла			to understand (1 pers. sing.)		
по-русски	по-російськи	по-русски	па-русскі			in Russian		
русский	російський	русский	русскій			Russian (m.)		
купаемся	купаємося	купаемся	купаємся			to swim (1 pers. plur.)		
большой	великий	большой	бальшой			big		
жизнь	життя	жизнь	жизнь			life		
зимой	взимку	зимой	зімой			in winter		
фургалка	фургалка	фургалка	фургалка	no exact match in E., an ancient analogue of a modern		carousel		
кол	кіл	кол	кол			pale, s.		
за його	через нього	из-за него	із-за него			because of him		
празник	свято	праздник	празднік			holiday		
сковзалка	сковзалка	скользанка	скальзанка	no exact matching in E.		slide / slope		
його	нього	его	него	Only if preceded by "колс коло його		near to him		
вірвовку	вірвовка / мотузка	веревку	верьовку	In U rarely used but exists <a href="http://sum.in.ua/s/virjovka">http://sum.in.ua/s/virjovka</a> ,		rope (Dative, fem.)		
большую	велику	большую	бальшую			big (Dative, fem.)		

карусель	карусель	карусель	карусель			carousel		
поштовхує	поштовхує	подталкивать	падталківать изредка	no exact matching in E. and R.		to push with certain frequency (3 pers. sing.)		
держімося	держимося	держимся	держимся			to held (1 pers. plur.)		
катаємся	катаємося	катаемся	катаємся			to ride with amusement (3 pers. sing.)		
щитається	вважається	считается	счітається			to be considered (3 pers. sing.)		
наквасяться	нап'ються	напьются	нап'ються	jargon term		to get drunk (Fut., 3 pers. plur.)		
пажалуста	будь-ласка	пожалуйста	пажалуста			please / You are welcome		
пока	бувай	пока	пака			bye		
времні	часу	времени	времні			time (Genitive)		

нада	потрібно / необхідно	надо	нада			need		
ета	ця	эта	ета			this		
да	так	да	да			yes		
шо	що	что	што			what		
скіки	скільки	сколько	сколька			how/ what		
время	часу	время	время			time		
їхать	їхати	ехать	ехать			to ride (inf.)		
часа	годин	часа	часа			hours		
Київська	Київська	Киевская	Кієвская			of Kyjiv		
на українській мові	українською мовою	на украинском языке	на українськам языке			in Ukrainian language		
балакаєм	балакаємо	разговариваем	разгаваріваєм	"балакати" is a U. verb generally more used in the spoken language		to speak (1 pers. plur.)		
розказать	розказати	розказать	розказать			to tell (inf)		
раньше	раніше	раньше	раньше			earlier		
своїлася	освоювалася	осваивалась	асваївалась			to get common with (1 pers. sing)		
на проіздстві	на виробництві	на производстве	на праіздстве			at the production		
з університета	з університету	с университета	с універсітета			from university		
по обмену	по обміну	по обмену	па абмену		по обмену опитом	on exchange		
опитом	досвідом	опытом	опитам			experience		
шоб	щоб	чтоб	штоб			in order to		
работают	працюють / роблять	работают	роботають			work (3 pers. plur.)		
года	роки	года	года		на три года	years (for three years)		
мог	міг	мог	мог		шоб він мог	(that he) could		
пообщаться	поспілкуватися	пообщаться	паабщаться			to conversate (inf.)		
написать	написати	написать	напісать			to write (inf.)		
не	ні	нет	нет			no		
даже	навіть	даже	даже			even		
высказаться	висказатися	высказаться	висказаться			to express (inf.)		
з дедушкой	з дідусем	с дедушкой	с дедушкой			with grandfather		
з бабушкой	з бабусею	с бабушкой	с бабушкой			with grandmother		
есть	є	есть	єсть			to be (3 pers. sing.)		
вмію	вмію	умею	умею			be able (1 pers. sing)		
писать	писати	писать	пісать			to write (inf.)		
читать	читати	читать	чітать			to read (inf.)		
звать	звати	звать	звать			to call (inf.)		
сколько	скільки	сколько	сколька		сколько лет	how many		
лет	років	лет	лет			years (for three years)		



сьодні	сьогодні	сегодня	сегодня			today		
привикла	привикла	привыкла	прівикла			to get used to (1 pers. sing.)		
трудностей	труднощів	трудностей	труднастей		я не боюся трудностей	difficulties		
вставать	вставати	вставать	вставать			to get up (inf.)		
робить	робити	делать	делать			to do (inf.)		
маладец	молодець	молодец	маладец			brave/ well done		
шось	щось	что-то	што-то			something		
курі	кури	куры	кури			chicken		
неділі	тижня	неделе	неделе		дві неділі	weeks		
обої	шпалери	обои	абої			wallpaper		
клеїли	клеїли	клеили	клеїли			to glue (1 pers. plur. Past)		
возим	возимо	возим	возім			to transport (1 pers. plur.)		
занимаесть	займаесть	занимаесть	занимаесть			to occupy (2 pers. plur.)		
отдаем	віддаємо	отдаем	атдайом			give away (1 pers. plur.)		
сіем	сіємо	сеем	сеєм			to sow (1 pers. plur.)		
коло його	біля нього	возле него	возле него			around it		
пилява	пил	пыл	пил	This meaning was adopted	considered the context	dust	?	
електричкою	електричкою	электричкой	електричкай		електричкою їдем	electric railway (Instrumental case)		
до вокзала	до вокзалу	до вокзала	до вокзала		До вокзала їдем	to the station		
їдем	їдемо	едем	едем			to drive (1 pers. plur.)		
часа	години	часа	часа		три часа	hours		
спродать	спродати	распродать	распрадать			to sell out		
в чотирі часа	в четвертій годині	в четыре часа	в четире часа			at four o'clock		
електричка	електричка	электричка	електричка			electric railway (Nominative case)		
в час	в першій	в час	в час		в час їде	at one [o'clock]		
мінут	хвилин	минут	мінут			minutes		
последня	остання	последняя	паследняя			the last		
варе	варить	варит	варіт			to cook, to stew (3 pers. sing.)		
в університеті	в університеті	в университете	в університете			at university		
готове	готує	готовит	готовіт			to prepare, to cook (3 pers. sing.)		
дом	дім	дом	дом			house		
собственний	приватний	собственный	собственный			private		
кирпичный	цегляний	кирпичный	кірпічний			[made of] brick		
промышленный	промисловий	промышленный	промишленний			industrial		
сахарный	цукровий	сахарный	сахарний			sugary		
строительства	будівництва	строительства	страїтельства		для строїтельства	[for] construction		

детский	дитячий	детский	детській	kindergarten	детский сад	infant		
сад	садок	сад	сад	kindergarten, in Ukrainian r	детский сад	kindergarten		
по чом	по скільки	по чём	по чьом			how much		
канєшно	звичайно	конечно	канєшна			of course		
робе	робить	делает	делает			to do (3 pers. sing.)		
жить	жити	жить	жить			to live (inf.)		
спасїба	спасибї	спасибо	спасїба			to thank (1 pers. sing.)		
їзвініть	вибачте	извините	їзвініте			to excuse (Imperative, 2 pers. plur., attract attention)		

не	ні	нет	нет			no		
пообщатся	поспілкуватися	пообщаться	пааобщатся			to conversate / to chat		
пожалуста	будь ласка	пожалуйста	пажалуста			please		
минут	хвилин	минут	мінут			minutes		
побалакать	побалакати	поболтать	побалтать			to speak		
канешна	звичайно	конечно	канешна			of course		
присажуйтесь	сідайте / присідайте	присаживайтесь	прісаживайтесь			"Take a seat"		
страни	країни	страны	страни			country (Genitive)		
інастранець	іноземець	иностранец	інастранец			foreigner		
учіца	учиться	учится	учіца			to study (3 pers. sing.)		
універсітеті	універсітеті	университете	універсітеті			in University		
обмєну	обміну	обмену	абмєну		по обмєну опитом	exchange		
опитом	досвідом	опытом	опитам			experience (Instrumental)		
студєнтов	студєнтів	студентов	студєнтав			students (Genitive)		
собирає	збирає	собирает	сабіраєт			to collect (3 pers. sing.)		
розговорну	розмовну	разговорную	разговорную			spoken		
рєчь	мову	речь	рєчь			speech		
шоб	щоб	чтобы	штоби			in order to		
українська	українська	украинская	українська			Ukrainian		
руська	російська	русская	русская			Russian		
разговарюють	розмовляють	разговаривают	разгаварівають	Same speaker different pronunciation		to speak (3 pers. plur.)		
поняв	зрозумів	понял	понял			to understand (1 pers. sing. m.)		
записать	записати	записать	записать			to write down		
заданіє	завдання	задание	заданіє			homework (in this context)		
собрать	зібрати	собрать	сабрать			to collect (inf.)		
человек	людей	человек	чєлавєк			people		
разговарюют	розмовляють	разговаривают	разгаварівають	Same speaker different pronunciation		to speak (3 pers. plur.)		
фамілію	фамілію / прізвище	фамилия	фамілія	The term "фамілія" is colloquial but possible in U		surname (Accusative)		
нада	потрібно / необхідно	нужно	нужна			to need (impers.)		
шо	що	что	што			what		
балакать	годин	болтать	балтать			to speak (inf.)		
дєцтві	дитинстві	детстве	дєтствє		в дєцтві	in childhood		
інтерєсне	інтерєсне	интересное	інтерєсне			interesting		
разговорним	розмовним	разговорным	разговорним		разговорним рєчью	spoken		
рєчью	мовою	речью	рєчью			speech		

умно	розумно	умно	умна			smart		
харашо	добре	хорошо	харашо			good		
Житомирської	Житомирської	Житомирской	Житомирській			of Zhytomyr		
п'яст	п'ятдесят	пятьдесят	п'ятдесят		с п'яст третього года / с	fifty		
года	року	года	года			year (Genitive)		
чудесна	чудесно	чудесно	чудесна			wonderful		
Чорноморський	Чорноморський	Черноморский	Черноморській		Чорноморський флот	Of Black Sea		
кончив	кінчив	окончил	акончил			to finish (1 pers. sing. m.)		
ремесляне	ремесляне	ремесленное	ремесленнає			industrial		
Смотрикиєвспатрою	?	?	?	Name of agency	В «Смотрикиєвспатрою»	?		
стройке	будівництві	стройке	стройке		на стройке	at a construction site		
слюсар-сантехнік	слюсар-сантехнік	слесарь-сантехник	слесарь-сантехнік			plumber technician		
восем	вісім	восем	восем		восем місяців	eight		
місяців	місяців	месяцев	месяцев		восем місяців	months		
обратно	назад	обратно	абратна			back		
место	місце	место	мєста		пішов в Київ в то же сам	place		
получив	отримав	получил	палучил			to get / to receive (1 pers. sing. m.)		
квартиру	квартиру	квартиру	квартиру			apartment (Accusative)		
продолжаю	продовжую	продолжаю	прадалжаю			to continue (1 pers. sing.)		
жить	жити	жить	жить			to live (inf.)		
інтересне	цікаве	интересное	інтереснає			interesting		
жизні	житті	жизни	жизні		у жизни	(in) life		
вспомнить	згадати	вспомнить	вспомнить			to remember / to bethink		
приятно	приємно	приятно	приятна			pleasantly		
водки	горівки	водки	водкі		водки не п'ю	vodka (Accusative)		
каждый	кожен	каждый	каждий			everyone		
приятно	звичайно	конечно	канєшна			of course		
щас	зараз	сейчас	сейчас			now		
ругались	лаялися	ругались	ругались		шоб не ругались	to swear (Past, plur.)		
друг друга	один одного	друг друга	друг друга		друг друга любили	each other		
встречались	зустрічалися	встречались	встречались			to meet (Past, plur.)		
кругом	всюди	езде	езде		Ходив кругом	everywhere		
Фокстрот	Фокстрот	Фокстрот	Фокстрот			Foxtrot		
маладец	молодець	молодец	маладец			"well done"		
лучше	краще	лучше	лучше		Саме лучше	better		

саме	най-	самое	самае	*only if followed by a comparative adj.		Term describes the highest level of adj. comparison		
потому	тому	потому	патаму		потому що	because		
свободне	вільне	свободное	свабоднас			free		
время	час	время	время			time		
собираете	збираєте	собираете	сабіраеце			to collect (2 pers. plur.)		
жизнь	життя	жизнь	жизнь		всю жизнь	life		
стройке	будівництві	стройке	стройке		на стройке	construction		
часов	годин	часов	часов		в п'ять часов	hours (Locative)		
поднимаюся	піднімаюсь	поднимаюсь	паднімаюсь			to wake up (1 pers. sing.)		
сна	сну	сна	сна		сна немає	sleep (Genitive)		
небудь	небудь	нибудь	нібудь			any		
всё	все	всё	всьо			all		
ума	ума / розуму	ума	ума		с ума зйду	mind (Genitive)		
скука	скука / нудьга	скука	скука			boredom		
вопше	взагалі	вообще	ваабще			at all		
помню	пам'ятаю	помню	помню			to remember (1 pers. sing.)		
слοжилася	зложилася	слοжилась	слажилась	Generally used "it happend like that"		to happen somehow	?	
это	це	это	ета		до это	this		
перестройки	перестройки	перестройки	перестройкі		до перестройки	perestroika (Genitive)		
после	після	после	после			after		
заработаш	заробиш	заработаешь	заработаш			to earn (Fut., 2 pers. sing.)		
хватает	хватает	хватает	хватает			to be / have enough		
все	всі	все	всс			everyone		
історическіє	історичні	исторические	істаріческіє			historical		
хороші	хороші	хорошие	хорошие			good (plur.)		
да	так	да	да			yes		
родіні	батьківщині	родине	родіне		на родіні	in homeland		
класов	класів	классов	классав	1 class = 1 year of school	сім класов	classes (Genitive)		
раньше	раніше	раньше	раньше			earlier		
года	року	года	года		года раждєнія	year (Genitive)		
раждєнія	народження	рождения	раждєнія			birth		
тридцять	тридцять	тридцать	трідцять			thirty		
корабле	кораблі	корабле	карабле		на корабле	(on) ship		
Одесса	Одеса	Одесса	Адесса			Odesa		
нічю	нічого	ничего	нічого			nothing		

один	один	один	адін			one		
виходілі	виходили	выходили	вихаділі			to go out (Past, plur.)		
біше	більше	больше	більше			more		
Дерибасовську	Дерибасівську	Дерибасовскую	Дерібасавську	Name of the street	Дерибасовську знаю	Deribasovskaya (Accusative)		
понравилось	сподобалось	понравилось	панравілась			to like (impers.)		
спасіба	спасибі	спасибо	спасіба			thank you		

звать	звати	звать	звать			to be called
понимаеш	розумієш	понимаешь	панімаєш			to understand (inter. 2 pers. sing.)
по-українськи	українською	по-украински	па-українські			in Ukrainian
да	так	да	да			yes
понимаю	розумію	понимаю	панімаю			to understand ( 1 pers. sing.)
Житомирській	Житомирській	Житомирской	Житомирській			of Zhytomyr
город	місто	город	горад			town
Сквира	Сквира	Сквира	Сквіра			Skvyra
шоб	щоб	чтобы	штоби			in order to
шо	що	что	што			what / that
сидіть	сидіти	сидеть	сідєть			to sit (inf.)
решил	рішив	решил	решил			decided (Past. sing. m.)
сюда	сюди	сюда	сюда			here (direction)
работу	роботу	работу	работу		на роботу	work (Accusative)
немножко	небагато	немножко	немножка			a little bit
подработать	підробити	подработать	падробатать		No perfect match in E	to work part time in order to earn extra money
щяс	зараз	сейчас	сейчас			now
время	час	время	время			time
тяжеле	тяжке	тяжелое	тяжелое			heavy / hard
жить	жити	жить	жить			to live
нада	потрібно / необхідно	нада	нада			to need (impers.)
это	це	это	ето			this
второй	другий	второй	второй		второй год	second
год	рік	год	год			year
сюдою	сюдою	сюдой	сюдой			this way
если	якщо	если	єслі			if
Сквиру	Сквиру	Сквиру	Сквіру		через Сквиру	Skvyra (Accusative)
Сквири	Сквири	Сквири	Сквіри		до Сквири	Skvyra (Genitive)
двести	двісті	двести	двєсті		двести сорок	two hundred
навкружки	навкруги	вокруг	вакруг			around
есть	є	есть	єсть			to be (3 pers. sing.)
западне	західне	западное	западнає		западне направлєніє	western (adj.n.)
направленіє	напрямок	направление	направлєніє			direction
заход	захід	запад	запад		на заход	west (s.)
Ружинський	Ружинський	Ружинский	Ружинській			of Ruzhyn

Вербівка	Вербівка	Вербовка	Вербовка			Verbivka	
сказать	сказати	сказать	сказать			to tell	
глуш	глуш	глушь	глушь	"an uncultivated, uninhabited, and inhospitable locality"		wilderness	
поняв	зрозумів	понял	понял			to understand (1 pers. sing. m.)	
дев'ятсот	дев'ятсот	девятьсот	дев'ятсот			nine hundred	
трицят	тридцять	тридцать	тридцать			thirty	
дев'ятий	дев'ятий	девятый	дев'ятий			ninth	
сижу	сиджу	сижу	сїжу			to sit (1 pers. sing.)	
консьєржкой	консьєржкою	консьєржкой	кансьєржкой			conciierge (Instrumental)	
падрабатываю	підробляю	подрабатываю	падрабатываю		No perfect match in E	to work part time in order to earn extra money	
сем'я	сім'я	семья	сем'я			family	
шось	щось	что-то	што-та			something	
пенсії	пенсії	пенсии	пенсії			pensions	
токо	тільки	только	только			only	
заплатить	заплатити	заплатить	заплатіть			to pay	
трицить	тридцять	тридцать	тридцать			thirty	
но	але	но	но			but	
шото	щось	что-то	што-та			something	
шіїсять	шітдесять	шестьдесят	шестьдесят			sixty	
получается	получается	получается	палучаетса			to come out / to succeed	?
раньше	рініше	раньше	раньше			earlier	
строїтельом	будівельником	строителем	страїтелем			builder (Instrumental)	
тогда	тоді	тогда	тагда			then	
лучше	краще	лучше	лучше			better	
работаеш	працюєш	работаешь	работаеш			to work (2 pers. sing.)	
пїтаніє	харчування	питание	пїтаніє		на пїтаніє	alimentation	
хватало	хватало	хватало	хватала			to be enough (Past, sing., n.)	
одівалися	одівалися / одягалися	одевались	адєвались			to dress up (impers.)	
випить	випити	выпить	випіть			to drink	
збірнїжкку	ощаднїжкку	сбернїжкку	сбернїжкку		на / в збернїжкку	savings book /account	
потому	тому	потому	патаму		потому шо	because	
хватас	хватас	хватает	хватает			to be enough (Past, sing., n.)	
со	зі	со	са		со всей України	from	
всей	всієї	всей	всєї		со всей України	all over (Genitive)	
строїтелі	будівельники	строители	страїтелі			builders	



работать	робити / працювати	работать	работать			to work	
нема	немає	нету	нету			to be not (3 pers. sing.)	
все	всі	все	все			all / everyone	
вот	ось	вот	вот			here / "here we are"	?
здесь	тут	здесь	здесь			here	
осьо	ось	вот	вот			here / "here we are"	
всьо	все	всё	всьо			everything	
прієжджі	пріїжджі	приезжие	прієзжие			newcomers	
месного	місцевого	местного	меснава			local	
стройка	будівництво	стройка	стройка			Construction	
с	з	с	с		с Західної України	from	
западної	західної	западной	западной		с Західної України	western	
Житомирської	Житомирської	Житомирской	Житомирской		з Житомирської	of Zhytomyr	
Чернігівської	Чернігівської	Черниговской	Чернігавской		з Чернігівської	of Chernihiv	
областя	області	области	області			regions	
любого	будь-якого	любого	любова		можеш найти любого з	whatever (+m.)	
любої	будь-якої	любой	любой		можеш найти любого з	whatever (+f.)	
найти	найти / знайти	найти	найти			to find	
каждий	кожен	каждый	каждий			everyone	
заработать	заробити	заработать	заработать			to earn	
крупных	крупних	крупных	крупних		в таких крупных города	massive	
городах	містах	городах	гарадах		в городах	towns	
узять	узяти	взять	взять			to take	
напрімер	наприклад	например	напрімер			for example	
ілі	або	или	ілі			or	
взять	узяти / взяти	взять	взять			to take	
нету	немає	нету	нету			to be not (3 pes. sing.)	
Сквири	Сквири	Сквире	Сквире		в / у Сквири	in Skvyra (Lokative)	
напрімер	наприклад	например	напрімер			for example	
остальні	останні	остальные	астальние			last (plur.)	
предприятия	підприємства	предприятия	предприятия			enterprises	
ше	ще	ещё	іщо			more	
работают	роблять / працюють	работают	работают			to work (3 pers. plur.)	
отопляют	опалюють	отопливают	атапливають			to heat up (3 pers. plur.)	
такое	таке	такое	такое			such as	

перепродать	перепродати	перепродать	перепрадать			to resell	
отак	ось так	вот так	вот так			like this	
сім'юй	сім'єю	семьей	семьей		с сім'юй	family (Instrumental)	
квартира	квартира	квартира	квартира			apartment	
жена	жона	жена	жена			wife	
Чокалки	Чокалки	Чокалки	Чокалки	No correspondance found,	на Чокалки	Chocalky	?
рядом	поряд	рядом	рядом			near	
Каравай	Каравай	Каравай	Каравай	Karavai village	Каравай дач	Karavai (village)	
железна	залізна	железная	железная		Железна дорога	iron	
дом	дім	дом	дом			house	
одірваний	відірваний	оторванный	аторванный			stripped / separed from others	
вокзала	вокзалу	вокзала	вакзала		з вокзала	station (Genitive)	
леву	ліву	леву	леву		на леву сторону	left	
єдинственний	єдиний	единственный	єдинственний			the only one	
чтирнацятий	чотирнадцятий	четырнацатый	чтирнацятий			fourteenth	
етаж	поверх	этаж	етаж			floor	
етом	цьому	этом	етам		в етом	this (Lokative, m.)	
часа	години	часа	часа		два часа	hours	
етой	цій	этой	етай		по етой	this (Lokative, f.)	
железної	залізній	железной	железной		по железной дороге	iron (Lokative, f.)	
дороге	дорозі	дороге	дароге		по железной дороге	road (Lokative, f.)	
електричка	електричка	электричка	електричка			electric railway	
маршруткою	маршруткою	маршруткой	маршруткой			by bus	
електричку	електричку	электричку	електричка		пересідаю на електричку	electric railway	
приїжджаю	приїжджаю	приезжаю	прієзжаю			to come (1 pers. sing.)	
кончав	кінчав	окончал	аканчав			to finish ( Past, m.)	
сем	сім	семь	семь			seven	
важності	можливості	возможности	важності			possibility (Genitive)	
тога	тоді	тогда	тагда			then	
учитса	учитися	учиться	учітса			to study	
сразу	зразу	сразу	сразу			immediately	
после	після	после	после		после войны	after	
войни	війни	войны	вайни			war (Genitive)	
тяжело	тяжко	тяжело	тяжело			heavy /hard	
большая	велика	большая	большая			big	

била	була	была	била			to be (Past, sing., f.)		
большая	велика	большая	большая			big (f.)		
отца	батько	отец	отец			father		
тяжело	тяжко	тяжело	тяжело			hard /heavy		
работал	робив працював	работал	работал			to work (Past, sing., m.)		
мать	мати	мать	мать			mother		
работала	робила / працювала	работала	работала			to work (Past, sing., f.)		
тоже	також	также	также			also		
сем	сім	семь	семь			seven		
человек	чоловік людей	человек	человек			people		
дѣтей	дїтей	детей	детей			children (Genitive)		
Четири	чотири	четыре	чотири			four		
родных	рідних	родных	родных			native		
матерінай	матеріної	материной	матерінай			maternal (Genitive)		
гром	грім	гром	гром			thunder		
это	це	это	это			this		
молнія	блискавка	молния	молнія			lightning		
вот	ось	вот	вот			here		
они	вони	они	ані			they		
воспитывали	виховували	воспитывали	васпїтивалї			brought up (Past, plur.)		
самий	най-	самый	самий		самий старший	highest level of comparation		
кончив	кінчив	кончил	кончїл			to finish (Past, sing.)		
зарабатывать	заробляти	зарабатывать	зарабативать			to earn		
чтоб	щоб	чтобы	штоби			in order to		
прожить	прожити	прожить	пражить			to survive		
не	ні	нет	нет			no		
помощ	допомога	помощь	помащ			help		
шо	що	что	што			what / that		
взять	узяти / взяти	взять	взять			to take		
последние	останні	последние	паследніє			the last (plur.)		
времена	часи	времена	времена			times		
пока	поки	пока	пака			still		
щас	зараз	сейчас	сейчас			now		
Америке	Америці	Америке	Америке		в Америке	America		
находиться	знаходиться	находиться	находїтса			to be		

да	так	да	да			yes		
Германію	Німеччину	Германию	Германію			Germany (Accusative)		
со'єднав	з'єднав	соединил	саєдніл			to connect (Past, sing.)		
тога	тоді	тогда	тагда			then		
лучше	краще	лучше	лучше			better		
роботаєш	робиш / працюєш	работаешь	роботаєш		не роботаєш	to work (2 pers. sing.)		
устроїть	улаштувати / влаштув	устроитъ	устроїть		устроїть на роботу	to arrange		
єслі	якщо	если	єслі			if		
городу	місту	городу	гораду		Йдеш по городу	town (Accusative)		
отпуску	відпустці	отпуске	отпуске		в отпуску	in holidays		
спрашуют	питаються	спрашивают	спрашивають			to ask (3 pers. plur.)		
чо	чому	почему	пачему			why		
говоріт	говорить	говорит	гаваріт			to say (3 pers. sing.)		
устроїм	улаштуємо / влаштуєм	устроим	устроїм			to arrange (1 pers. plur.)		
роботать	робити / працювати	работать	роботать			to work		
посмотрели	подивились	посмотрели	пасматрелі			to look (Past, plur.)		
пенсію	пенсію	пенсию	пенсію		на пенсію	in pension		
пусть	хай / нехай	пусть	пусть			let		
молодьож	молодь	молодежь	маладьож			youth		
должна	повинна	должна	далжна			should (3 pers. sing. f.)		
отдыхай	відпочивай	отдыхай	атдыхай			to rest (2 pers. sing.)		
пенсії	пенсії	пенсии	пенсії			pension (Genitive)		
Правильно	правильно	правильно	правільна			correctly		
тіх	тих	тех	тех		для тіх	that		
нічо	нічого	ничего	нічево			nothing		
сделали	зробили	сделали	сделалі			to have done (Past, plur.)		
достаточно	достатньо	достаточно	дастатачна			enough		
удочкою	вудочкою	удочкой	удочкой		с удочкою	fishing rod (Instrumental)		
рибалке	рибалці	рыбалке	рибалке		на рибалке	on fishing		
сизу	сиджу	сизу	сіжу			to sit (1 pers. sing.)		
нада	потібно / необхідно	надо	нада			to need (impersonal)		
мать	мати	иметь	іметь			to have		
получаю	отримую	получаю	палучаю			to get (1 pers. sing.)		
восем	вісім	восемь	восемь			eight		
суток	дїб	суток	сутак			24-hours (Genitive)		

сутки	доба	сутки	суткі			24-hours	
оддежури́в	відчегрува́в	отдежури́л	атдежури́л			to finish the turn at work (Past, : ?	
а́пять	зно́ву	о́пять	а́пять			again	
вихо́дов	вихо́дів	вихо́дов	виха́дов			exits (Genitive)	
делаю́	роблю́	делаю́	делаю́			to do (1 pers. sing.)	
сего́дне	сьо́годні	сего́дня	сего́дня			today	
здесь	тут	здесь	здесь			here	
отсе́дел	відсиді́в	отсиде́л	атсе́дел	Used commonly for "finish working" especially if		to finish sitting (Past, sing. m.)	
восемна́дцатого	ві́сімна́дцятого	восемна́дцатого	васемна́дцатава			eighteenth (Genitive)	
в́тарого	дру́ого	второго	в́тарова		двадцять в́тарого	second	
о́пять	зно́ву	о́пять	а́пять			again	
сіде́ть	сиді́ти	сиде́ть	сіде́ть			to sit	
двесе́ті	двісе́ті	двесе́ти	двесе́ті			two hundred	
де́ньги	гроші́	де́ньги	де́ньгі			money	
не́	не	нет	нет		не де́ньги	not	
почті́	майже́	почти́	пачті́			almost	
даром	даром	даром	даром			for free	
не́ма	нема́є	нету́	нету́			not to be (impers.)	
возьми́	візьми́	возьми́	вазьмі́			to take (Imperative, 2 pers. sing.)	
Попельні́	Попі́льні	Попельне́	Папельне́		в Попельні́	in Popilnia	
ета́т	цей	э́тот	ета́т			this	
ра́ньше	рі́ніше	ра́ньше	ра́ньше			earlier	
стро́ительні́	буді́вельні́	строительные́	страі́тельние́			constructional	
водокана́ли	водокана́ли	водокана́лы	вадакана́ли			water channels	
стро́ят	буду́ють	стро́ят	стро́ят			to build (3 pers. plur.)	
везде́	повсю́ди	везде́	везде́			everywhere	
стро́йку	буді́вництво	стро́йку / строите́льстве́	стро́йку / страі́тельства́		на стро́йку	for construction	
возьму́ть	візьму́ть	возьму́т	вазьму́т		не возьму́ть	to take (Fut, 3 pers. plur.)	
пѝ́сяти	п'я́тдеся́ти	пятидеся́ти	п'я́тдеся́ті		до пѝ́сяти п'я́ті ле́т	fifty (Genitive)	
п'я́ті	п'я́ти	пяти́	п'я́ті			five (Genitive)	
ле́т	рокі́в	ле́т	ле́т			years (Genitive)	
сви́ше	пона́д	сви́ше	сви́ше			over	
прора́ботал	проробив́ / пропрацюе́	прора́ботал	пра́работал			to work for a long time (Past, sing. m.)	
пѝ́сять	п'я́тдеся́т	пятьдеся́т	п'я́тдеся́т		С пѝ́сять че́твертого го́д	fifty	
с	з	с	с			from	

года	року	года	года			year (Genitive)		
год	рік	год	год			year		
отрабатал	відпрацював	отрабатал	атрабатал			to work for a long time and finish (Past, sing. m.)		
камінщиком	каменярем	каменщиком	камєнщїкам			bricklayer (Instrumental)		
кірпичики	цеглинки	кирпичики	кірпїчікі			bricks		
складував	складав	складывал	складивал			to add together (Past, sing. m.)		
спрашивал	питався	спрашивал	спрашивал			to ask ( Past, sing. m.)		
даже	навіть	даже	даже			even		
сидіть	сидіти	сидеть	сїдеть		шоб не сидіть	to sit		
мусим	мусимо	должны	далжны			must (1 pers. plur.)		
ходить	ходити	ходить	хадіть			to walk		
хотять	хочуть	хотят	хатят		Не хотять	to want (3 pers. plur.)		
говорить	говорити	говорить	гаваріть			to talk		
работать	робити / працювати	работать	работать			to work		
тыщу	тысячу	тысячу	тысячу			thousand		
полтари	півтори	полторы	палтари			one and a half		
тыщі	тысячі	тысячи	тысячі		дві тыщі	thousand (Genitive)		
она	вона	она	ана			she		
большая	велика	большая	бальшая			big		
страна	країна	страна	страна			country		
есть	є	есть	есть			to be (impers.)		
но	але	но	но			but		
Україні	Україні	Украине	Україне		на Україні	Ukraine (Lokative)		
недри	недри	недри	недрї			bowels of the earth		
всьо	все	всё	всьо			everything		
хороші	хороші	хорошие	харошие			good		
чорнозьоми	чорноземи	черноземы	черназьоми			black earth		
поїхать	поїхати	поехать	паєхать			to go		
сорняки	бур'яни	сорняки	сарнякі			weeds		
полей	полів	полей	палей		половина полей	fields (Genitive)		
проїжав	проїжджав	проезжал	праєзжал			to pass (Past, sing., m.)		
смотрів	дивився	смотрел	сматрел			to look (Past, sing., m.)		
домік	домик	домик	домік			small house		
туда	туди	туда	туда			in that direction		
транспорта	транспорту	транспорта	транспарта			transport (Genitive)		

государственный	державний	государственный	государственный		государственный дом	of the state (adj.)		
дом	дім	дом	дом			house		
государственному	державному	государственному	государственному		в государственном доме	of the state (adj. Dative)		
дому	домі	дому	дому			house (Lokative)		
Второй	другий	второй	второй		Второй год	second		
год	рік	год	год			year		
будет	бути	будет	будет			to be		
рыбалка	рибалку	рыбалку	рыбалку			on fishing		
удочкой	вудочкою	удочкой	удачкай		с удочкой	with fishing rod (Instrumental)		
или	або	или	ілі			or		
сеточкой	сіточкою	сеточкой	сетачкай		з сеточкой	mesh (Instrumental)		
что-то	щось	что-то	што-та			something		
поймал	піймав	поймал	паймал		поймал бы	to catch / to caught (Past, sing., m.)		
отдыхаешь	відпочиваєш	отдыхаешь	атдыхаєш			to rest (2 pers. sing.)		
дежурю	чергую	дежурю	дежурю			to work on turn (1 pers. sing.)		
отдохну	відпочину	отдохну	атдахну			to rest (Fut., 1 pers. sing.)		
здесь	сюди	здесь	сюда			here		
отдыхаю	відпочиваю	отдыхаю	атдыхаю			to rest (1 pers. sing.)		
делаю	роблю	делаю	делаю			to do (1 pers. sing.)		
сделать	зробити	сделать	сделать			to do		
поставить	ставити	поставить	ставити			to put		
помогал	допомагав	помогал	памагал			to help (Past, sing., m.)		
знакомые	знайомі	знакомые	знакомие		давно знакомі	acquaintances		
помоги	допоможи	помоги	памагі			to help (imperative, 2 pers. sing.)		
потому	тому	потому	патаму		потому что	because		
жизнь	життя	жизнь	жизнь			life		
Брежнев	Брежнев	Брежнев	Брежнев			Brezhnev		
наслыхан	чув	слышал	слишал			to hear about (Past, sing., m.)		
Брежнев	Брежнев	Брежнев	Брежнев			Brezhnev (Genitive)		
здесь	ось	здесь	здесь			here	?	
тогда	тоді	тогда	тагда			then		
почувствовал	відчув / почув	почувствовал	пачуствавал			to feel (Past, sing., m.)		
работай	роби / працюй	работай	работай			to work (Imperative, 2 pers., sing.)		
будут	будуть	будут	будут			to be (Fut., 3 pers. plur.)		
сбережники	ощадкнижки	сбережники	сберкніжки		на ощадкнижки	on savings books		

складували	складали	складывали	складивали			to accumulate (Past, 1 pers. plur.)	
отак	ось так	вот так	вот так			in this way	
полож	положити	положить	палажить			to put	
конци	кінці	концы	канци		конци з кінцями звести	ends	
концями	кінцями	концами	канцями			ends (Instrumental)	
возьми	візьми	возьми	вазьмі			to take (Imperative, 2 pers. sing.)	
шіїсят	шістдесят	шестьдесят	шестьдесят		триста шіїсят гривень /	sixty	
восемнацать	вісімнадцять	восемнацать	васемнацать			eighteen	
сколько	скільки	сколько	сколька		Сколь разніца какая?	how many	
разница	різниця	разница	разніца			difference	
какая	яка	какая	какая			which	
получае	отримуе	получает	палучаєт			to get (3 pers. sing.)	
получаю	отримую	получаю	палучаю			to get (1 pers. sing.)	
Кто	хто	кто	кто			who	
руководство	керівництво	руководство	руководства			direction	
воровав	крав	воровал	варавал			to grab (Past, sing. m.)	
нет	ні	нет	нет			no	
ворует	краде	ворует	варует			to grab (3 pers. sing.)	
воровані	крадені	ворованные	варованнис			stolen (plur.)	
покупают	купають	покупают	пакупают			to buy (3 pers. plur.)	
некоторые	деякі	некоторые	некаториес			some / someone	
хватает	хватає	хватает	хватаєт			to be enough (impersonal)	
меньше	менше	меньше	меньше			less	
движения	руху	движения	двіженія			movement (Genitive)	
наворовав	накрыв	наворовал	наваравал			to grab (Past, sing., m.)	
считались	вважались	считались	счіталісь			to be considered (Past, plur.)	
каждого	каждого	каждого	каждава		в каждого	each (Genitive)	
делают	роблять	делают	делають			to do (3 pers. plur.)	
коммерцию	комерцію	коммерцию	каммерцию			commerce (Accusative)	
приезжают	приїжджають	приезжают	прієзжають			to come (3 pers. plur.)	
пенсионеры	пенсіонери	пенсионеры	пенсіанери			pensioners	
надо	потібно / необхідно	надо	нада			to need (impersonal)	
выпасать	выпасати	выпасать	выпасать			to pasture	
продать	продати	продать	прадать			to sell	
їх	них	ним	нім	In R. another case will b	До їх	them	



прижають	приїжджають	приезжают	прієжають			to come (3 pers. plur.)		
на дом	додому	на дом	на дам		на дом	at home		
наприклад	наприклад	например	наприклад			for example		
поросьонка	поросся	поросенка	парасьонка		за поросьонка	pig (Genitive)		
должен	повинен	должен	должен			should (3 pers. sing.)		
шіссот	шістсот	шестьсот	шестьсот			six hundred		
заплатить	заплатити	заплатить	заплатить			to pay		
навару	навару	навару	навару		No exact match in E	difference between the basic pri ?		
путьом	шляхом	путём	путьом		отаким путьом	way (Instrumental)		
покупає	купляє	покупает	пакупаєт			to buy (3 pers. sing.)		
весов	ваги / ваг	весов	весов		без весов	scales (Genitive)		
сьодні	сььогодні	сегодня	сегодня			today		
других	інших	других	другіх		других везе / по других	another		
кожний	кожен	каждый	каждий		каждий день	every		
оборот	оборот	оборот	абарот			turnover		
делає	робить	делает	делаєт			to do (3 pers. sing.)		
за щот	за рахунок	за счет	за щьот			at the expense of		
кого-то	когось	кого-то	каво-та		за щот кого-то	somebody (Genitive)		
канешно	звичайно	конечно	канешна			of course		
двацить	двадцять	двадцать	двадцать		Двацить шостого	twenty		
етого	цього	этого	етава		етого месяца	this (Genitive)		
месяца	місяця	месяца	месяца			month (Genitive)		
васкрісєня	неділі	воскресенья	васкресєня			Sunday (Genitive)		
следущє	наступне	следующее	следующее		в следущє воскресєніє	next / the following		
воскрісєніє	неділя	воскресенье	васкресєніє			Sunday		
по-українськи	українською	по-украински	па-українські			in Ukrainian		
расказать	розказати	рассказать	расказать			to tell		
Житомирській	Житомирській	Житомирской	Житомирській		в Житомирській області	Zhytomyr (adj., Lokative)		
п'яст	п'ятдесят	пятьдесят	п'ятдесят		В п'яст четвертому році	fifty		
юг	південь	юг	юг			south		
кончив	кінчив	кончил	кончил			to finish (Past, sing., m.)		
ГПШ	ГПШ	ГПШ	ГПШ		горно-промишльонне уч	mining-industrial school		
горно-промишльонне	гірничо-промислове	горно-промышленное			горно-промишльонне уч	mining-industrial		
учіліще	училище	училище	учіліще			high school		
годічне	річне	годовое	гадічнає			annual		

горний	гірський	горный	горний		Горний мастер	mining		
мастер	майстер	мастер	мастер			master		
кончили	кінчили	кончили	кончили			to finish (Past, plur.)		
проїхатса	проїхатись	проехаться	праехатса			to make a ride		
отправляют	відправляють	отправляют	атправляють			to send (3 pers. plur.)		
Казахская	Казахська	Казахская	Казахская		Казахская ССР	Kazakh (adj., f.)		
большая	велика	большая	большая			big		
вremenі	часі	времени	вremenі		у вremenі разлиця	time (Dative)		
разница	різниця	разница	разніца			difference		
работал	робив працював	работал	работал			to work (Past, sing., m.)		
харашо	добре	хорошо	харашо			good		
пїсят	п'ятдесят	пятьдесят	п'ятьдесят		с пїсят четвертого	fifty		
шестой	шестий	шестой	шестой			sixth		
год	рік	год	год			year		
отработав	відпрацював / відпр	отрабатывал	атрабативал			to finish working / to work (Past, sing., m.)		
остався	остався	остался	астался			to stay (Past, sing., m.)		
после	після	после	после			after		
шестого	шостого	шестого	шестова		до шїсїать шестого годк	sixth (Genitive)		
шестом	шостому	шестом	шестом		в шїсїать шестом году	sixth (Dative)		
году	році	году	гаду			year (Dative)		
вернувся	вернувся	вернулся	вернулся			to return (Past, sing., m.)		
Україну	Україну	Украину	Україну		на Україну	in Ukraine		
заработували	заробляли	зарабатывали	зарабативали			to earn (Past, plur.)		
государство	держава	государство	гасударства			state		
отработать	відпрацювати / відпр	отрабатывать	атрабативать			to finish working / to work		
некоторые	деякі	некоторые	некаторие			some / someone		
побегали	повтікали	побегали	паубегали			to run away (Past, plur.)		
сразу	зразу	сразу	сразу			at once / immediately		
женился	женився	женился	женілся			to get married (Past, sing., m.)		
жену	дружину / жінку	жену	жену			wife (Accusative)		
квартиру	квартиру	квартиру	квартиру			apartment (Accusative)		
считай	вважай	считай	щітай			to consider (imperative, 2 pers. sing.)		
десять	десять	десять	десять			ten		
двенадцать	дванадцять	двенадцать	двенадцать			twelve		
город	місто	город	горад			town		

Тімертаул	Теміртау	Темиртау	Теміртау			Temirtau		
где	де	где	где			where		
тежолa	важка	тяжелая	тяжолая		тежолa промишленност	heavy		
промишленность	промисловість	промышленность	прамишленнасть			industry		
КарГРЭС	КарГРЕС	КарГРЭС	КарГРЕС	(Карагандинская ГРЭС)		KarGRES Power Station		
треста	трести	треста	треста			trust		
МЖС	МТС	МЖС	МЖС	Few interpretations of the abbreviation are possible		Machine and animal breeding st ?		
строїли	строїли	строили	строїлі			to build (Past, plur.)		
шістїднєвка	шестиденка	шестидневка	шестїднєвка			six-day		
двух	двох	двух	двух	до двух часов		two (Genitive)		
часов	годин	часов	часов			hours (Genitive)		
работали	робили / працювали	работали	работалі			to earn (Past, plur.)		
заработок	заробіток	заработок	заработак			income		
денег	грошей	денег	денєг			money (Genitive)		
скопив	скопив	скопил	скопїл			to accumulate (Past, sing. m.)		
збиріженія	збереження	сохранения	сохраненія			savings		
будем	будемо	будем	будем			to be (Fut., 1 pers. plur.)		
возвращать	повертати	возвращать	вазвращать			to return		
короче	коротше	короче	кароче			shorter / "in short words"		
деньги	грошей	деньги	деньгі			money		
сделали	зробили	сделали	сделалі			to have done (Past, plur.)		
токи	тільки	только	толька			only		
нігде	ніде	нигде	нігде			nowhere		
прослухуватся	прослухуватися	прослушиватся	праслушиватса			to listen to (impersonal)		
тежело	тяжко	тяжело	тяжело			heavy		
пятнадцать	п'ятнадцять	пятнадцать	п'ятнадцать		пятнадцать лет	fifteen		
лучшего	кращого	лучшего	лучшева		лучшего на бачим	better (Genitive)		
бачим	бачимо	видим	відім			to see (1 pers. plur.)		
хуже	гірше	хуже	хуже		хуже і хуже йде	worse		
розстраується	розстроується	расстраивается	расстраиваетса			to be upset (3 pers. sing.)		
посмотри	подивись	посмотри	пасматрі			to look (Imperative, 2 pers. sing.)		
делается	робиться	делается	дслается			to do / to happen (impersonal)		
европейську	європейську	европейскую	єврапейскую		європейську столицу	European		
столицу	столицю	столицу	сталіцу			capital (Accusative)		
везде	всюди	везде	везде			everywhere		

говору	говорю	говорю	гаварю			to say (1 pers. sing.)		
етом	цьому	этом	етам		в етом дому	this (Locative)		
дому	домі	доме	доме		в етом дому	house (Locative)		
рабочих	робочих	рабочих	рабочіх			working (adj., Genitive)		
Казахстана	Казахстану	Казахстана	Казахстана		После Казахстана	Kazakhstan (Genitive)		
Сквира	Сквира	Сквира	Сквира			Skvyra		
встроївся	влаштувався	встроился	встроїлся		встроївся на роботу	to settle down (Past, sing. m.)	?	
рядом	поруч	рядом	рядам			near		
оно	воно	оно	ано			it		
четири	чотири	четыре	честире		четири годи	four		
годи	роки	года	года			years		
получив	получив	получил	палучіл			to get (Past, sing., m.)		
однокомнатну	однокімнатну	однокомнатную	аднакомнатную		однокомнатну квартиру	one-room		
квартиру	квартиру	квартиру	квартиру			apartment		
двацять	двадцять	двадцать	двадцять			twenty		
ещё	ще	ещё	ещё			still /more		
двохкомнатну	двокімнатну	двохкомнатную	двохкомнатную		двохкомнатну квартиру	two-room		
планіровка	планування	планировка	планіровка			layout / plan		
хароша	хороша	хорошая	харошая			good		
комнати	кімнати	комнаты	комнати			rooms		
должно	повинно	должно	должно			should (3 pers. sing. n.)		
буть	бути	быть	бить			to be		
квартири	квартири	квартиры	квартири			apartments		
годами	роками	годами	гадами			years (Instrumental)		
получить	получити	получить	получить		не получить	to become		
другіх	інших	других	другіх		з другіх областей	anothers (Genitive)		
вплочені	вплачені	оплачены	аплачени			payed (plur.)		
долларов	долларів	долларов	долларав		до шести тысяч долларов	dollars (Genitive)		
вроді	начебто	вроде	вроде			such as		

пожалуста	будь-ласка	пожалуйста	пажалуста			please
присажуйтесь	присідайте сідайте	присаживайтесь	прісаживайтесь			"take a sit"
на українській	українською	на украинском	на українском			in Ukrainian
по-руськи	російською	по-руськи	па-руські			in Russian
шо	що	что	што			what
шось	щось	что-то	што-та			something
писать	писати	писать	пісать			to write
ізвиняюсь	вибачаюсь	извиняюсь	ізвиняюсь			sorry
вечером	вечором	вечером	вечерам			evening (Instrumental)
туалет	туалет	туалет	туалет		у туалет	toilet
п'ісят	п'ятдесят	пятьдесят	п'ятьдесят			fifty
щітай	вважай рахуй	считай	щітай			to consider (Imperative, 2 pers. sing.)
чую	відчуваю	чувствую	чую			to feel (1 pers. sing.)
всьо	все	всё	всьо		так і всьо	everything
ето	це	это	ето			this
сміщення	зміщення	смещение	сміщення		перелом без зміщення	displacement
неділі	тижні	недели	неділі		чотири неділі	weeks
гіпсу	гіпсі	гипсе	гіпсе		в гіпсу	gypsum (Lokative)
таксіст	таксист	таксист	таксіст			taxist
їх	в них	в них	в ніх		в їх	they
мобілки	мобілки	мобилки	мабілки			cell phones
мобілках	мобілках	мобилках	мабілках		по мобілках	cell phones (Lokative)
понімаю	розумію	понимаю	панімаю		не понімаю	to understand (1 pers. sing.)
мобілка	мобілка	мобилка	мабілка			cell phone
телефону	телефоні	телефону	телефону		По телефону	telephone (Lokative)
да	так	да	да			yes
мобілці	мобілці	мобилке	мабілке		по мобілці	cell phone (Lokative)
нажимать	нажимати	нажимать	нажимать			to press
кнопочки	кнопочки	кнопочки	кнопачкі			bottons
ладжку	щиколотку	лодыжку	ладжку		поломала ладжку	ankle (Accusative)
бінт	бинт	бинт	бінт			bandage
чух	чух	чух	чух	Particular exclamation of the speaker adopted t		chuh ?

восемнацать	вісімнадцять	восемнадцать	васемнадцать		в восемнацять років	eighteen	
тоже	також	также	также			also	
опухша	опухла	опухшая	апухшая			swollen	
поломала	поламала	поломала	паламала		поломала ногу	to break down (Past, sing. f.)	
полностью	повністю	полностью	полнастью			completely	
ше	ще	ещё	іщю			more	
чо	чого	чего	чево			why what	
сидить	сидіти	сидеть	сідеть			to sit	
рядом	поруч	рядом	рядам	May have also another meaning "side by side"		near	
Искровской	Іскрівській	Искровской	Искровскай		на Искровской	Iskrivskaya (Lokative)	
оттуда	звідти	оттуда	аттуда			from there	
откуда	звідки	откуда	аткуда			where from	
понятно	зрозуміло	понятно	панятна			clear	
родственники	родичі	родственники	родственніки			relatives	
зятьом	зятем	зятем	зятем		з зятьом	son-in-law (Instrumental)	
не	ні	нет	нет			no	
хороший	хороший	хороший	хороший			good	
парень	парубок	парень	парень			boy	
приятна	приємно	приятно	пріятна			nicely	
только	тільки	только	толька			only	
взять	взяти	взять	взять			to take	
трёхкомнатна	трикімнатна	трехкомнатная	трёхкомнатная		трёхкомнатна кварти	three-room	
брать	брати	брать	брать			to take	
кормитесь	кормитесь	кормитесь	кормітєсь	Verb normally used in U, R, to describe the pro		to feed (2 pers. plur.)	?
готовите	готуєте	готовите	гатівіте			to cook	
молодец	молодець	молодец	маладец			brave / "well done"	
бутербродіки	бутербродики	бутербродики	бутербродікі			sandwiches	
Колюня	Колюня	Колюня	Калюня	Personal name		Kolyunya	
супчика	супчика	супчика	супчіка			soup (Accusative)	
потом	потім	потом	патом			after	
ночью	ніччю	ночью	ночью			night (Instrumental)	

вот	ось	вот	вот			here	
детском	дитячому	детском	детскам	kindergarten	детском садіку	infant (Lokative)	
садіку	садочку	садике	садіке	kindergarten + in Ukrainian means also "garder		kindergarten (Lokative)	
больнице	лікарні	больнице	бальнице		в больнице	hospital (Lokative)	
кладовщиком	кладовщиком	кладовщиком	кладовщікам			storekeeper (Instrumental)	
Ємільчинський	Ємільчинський	Емилченский	Ємільченській	region	Ємільчинський район	Emilchinsky	
Рудня Івановська	Рудня Іванівська	Рудня Ивановская	Рудня Іванавская		село Рудня Івановська	Rudnya Ivanivska	
Житомирська	Житомирська	Житомирская	Житомирська		Житомирська область	of Zhytomyr (adj., f.)	
сільно	сильно	сильно	сільна		не сільно велике	a lot / very	
красіве	красиве	красивое	красіває		красіве село	beautiful	
черніка	чорниця	черника	черніка			blueberries	
Олевский	Олевський	Олевский	Алевській	Olevsk region	Олевский район	of Olevsk (adj. m. Nominative)	
почті	майже	почти	пачті			nearly / almost	
Новоград-Волинський	Новоград-Волинський	Новоград-Волинский	Новаград-Валинській		Новоград-Волинський	Novograd-Volynsky	
ілі	або	или	ілі			or	
Луцк	Луцьк	Луцк	Луцк			Lutsk	
Олевского	Олевського	Олевскому	Алевскаму	Different cases and pre	до Олевского	Olevsk (adj. m. Genitive)	
трицять	тридцять	тридцать	трідцять			thirty	
канешно	звичайно	конечно	канешна			certainly	
щяс	зараз	сейчас	сейчас			now	
Олевска	Олевська	Олевску	Олевску	Different cases and pre	до Олевска	Olevsk	?
семдісят	сімдесят	семьдесят	семьдесят			seventy	
от	від	от	от		сто кілометрів от Чор	from	
щітаєца	вважається	считается	щітаєца			to be considered (impersonal)	
самое	саме	самое	самає		ето самое	most / "exactly this"	
получае	получае	получает	палучаєт			to get (3 pers. sing.)	
український	українська	украинский	українській		Український язык	Ukrainian	
язык	мова	язык	язык		Український язык	language	
но	але	но	но			but	
русский	російська	русский	русскій		русский язык	Russian (adj. m.)	
русска	російська	русская	русская		русска література	Russian (adj. f.)	

література	література	литература	література		руська література	literature
лучше	краще	лучше	лутше			better
десятом	десятому	десятом	десятам		в десятом класе	tenth
класе	класі	класе	класе		в десятом класе	class
пятьорку	п'ятірку	пятерку	пятьорку		на пятьорку	five (Accusative)
український	український	украинский	українській		український здала	Ukrainian
літературу	літературу	литературу	літературу		Я любила літературу	literature
семидесят	сімдесят	семьдесят	семьдесят		С семисят четвертого	seventy
іюне	червні	июне	іюне		в іюне	June (Dative)
года	року	года	года			year (Genitive)
сюда	сюди	сюда	сюда			here
вообщем	в загальному	в общем	в общем			generally
мойому	моєму	моему	маєму			mine (Dative)
раньше	раніше	раньше	раньше			earlier
вопшем	в загальному	в общем	в общем			generally
деятельность	діяльність	деятельность	деятељнасть		трудова діятељнасть	activity
швея-мотористка	швея-мотористка	швея-мотористка	швея-матарістка	a worker at sewing or shoe enterprise, performi		seamstress
вопше	взагалі	вообще	вааобще			in general
мужскі	чоловічі	мужские	мужськіє			men's
смену	зміну	смену	смену		в смену шила	turn
трицять	тридцять	тридцать	трідцять		шила по трицять пар	thirty
била	була	была	била			to be ( Past, sing., f.)
електріческа	електрична	электрическая	електріческа		електріческа машинка	electric
п'ятнацять	п'ятнадцять	пятнадцать	п'ятнадцять		по пятнацять штук	fifteen
даже	навіть	даже	даже			even
Житомир	Житомир	Житомир	Житомир		в Житомир	Zhytomyr
сорівнованіє	змагання	соревнование	саревнованіє		на сорівнованіє	competition
заняв	зайняв	занял	занял			to take (Past, sing. m.)
место	місце	место	места			place
Надежда	Надія	Надежда	Надежда		Надежда Фьодоровна	Nadezda
Фьодоровна	Федорівна	Федоровна	Фьодаравна			Fodorovna



Надя	Надія	Надя	Надя			Nadya	
всегда	завжди	всегда	всегда			always	
бить	бути	быть	бить			to be	
Вера	Віра	Вера	Вера			Vera	
празник	празник	праздник	празднік			holiday	
мать	мати	мать	мать			mother	
сначала	спочатку	сначала	сначала			firstly	
швейною-мотористкою	швеєю-мотористкою	швейей-мотористкой	швейой-матарісткай	a worker at sewing or shoe enterprise, performi		seamstress (Instrumental)	
провчилась	провчилась	проучилась	праучілась			to study (Past, f.)	
парень	парубок	парень	парень			boy	
подвѣл	підвів	подвёл	падвѣол	Figurative sense		to betray / to trick (Past, m.)	
сказать	сказати	сказать	сказать		ну як сказати	to say	
отак	ось так	вот так	вот так			in this way	
нада	потібно / необхідно	надо	нада			to need (impersonal)	
дак	так	так	так	Could be used in the main sentence with meani		then / in that case	
муж	чоловік	муж	муж			husband	
Сірьожка	Сергійко	Серёжака	Серёжка			Sergiy (diminutive)	
звать	звати	звать	звать			to call (impersonal)	
год	рік	год	год			year	
гостеприімні	гостеприемні / гостинні	гостеприимные	гостепріимніе	"гостеприемный" is part of U. lang. but it is a		hospitable (adj. plur.)	
двенацать	дванадцять	двенадцать	дванадцять			twelve	
щоту	рахунку	счету	щету		на щоту	on the account	
кажем	кажемо	говорим	гаварім			to say (1 pers. plur.)	
уїзжали	уїжджали	уезжали	уезжалі			to go away (Past, plur.)	
гостеприімна	гостеприемна	гостеприимная	гостепріимная			hospitable (adj. f.)	
угощала	угощала	угощала	угащала	no exaxt match in E, describes an action of "giv		to give to eat / to drink (Past. f.) ?	
городі	місті	городе	гораде		в городі	town (Lokative)	
квартира	квартира	квартира	квартира			apartment	
квартирою	квартирою	квартирой	квартирай		з квартирою	apartment (Instrumental)	
почтальйонша	поштарка	почтальонша	пачтальйонша			postwomen	
пенсію	пенсію	пенсию	пенсію			pension (Accusative)	

в курсе	курсі	кurse	курсє	When used "в курсе де не в курсе дела		informed	
дела	діла	дела	дела	When used "в курсе де не в курсе дела		business (Genitive)	
щас	зараз	сейчас	сейчас			now	
простинь	простиня	простынь	простинь			sheet	
шоб	щоб	чтобы	штоби			in order to	
так	так	так	так	Dialectal form ?	так закрию	in this way	
салфеточкой	салфеточкою	салфеточкой	салфетачкай			napkin (Instrumental)	
нет	ні	нет	нет			no	
токо	тільки	только	толька			only	
убери	убири	убери	убері	U version is rarely used		to take away (imperative, 2 pers. sing.)	
стол	стіл	стол	стол			table	
чист	чистий	чистый	чістий			clean	
трицить	тридцять	тридцать	трідцять		трицить три	thirty	
трицать	тридцять	тридцать	трідцять		трицать один	thirty	
сказать	сказати	сказать	сказать			to say	
четвертом	четвертому	четвертом	чєтвѣортам		на четвертом етажі	fourth (Lokative)	
етажі	поверсі	этаже	етаже		на четвертом етажі	floor (Lokative)	
держить	тримає	держит	держит			to keep (3 pers. sing.)	
етіх	цих	этих	етіх			these (Accusative)	
приют	приют	приют	пріют	In U also "притулок", "приют" found in U. of		shelter	
между	між	между	между		между нами	between	
собирає	збирає	собирает	сабіраєт			to collect (3 pers. sing.)	
красенькая	червоненька	красенькая	красенькая			red	
пирожок	пиріжок	пирожок	піражок	Speaker uses this similiarity to a pie in order to		small pie	
онь	ось	вон	вон			there / "There it is"	
протів	навпроти	напротив	напротів		стоїть протів нас	opposite	
чем-то	чимось	чем-то	чєм-та		п'ятсот с чєм-то	something (Instrumental)	
восемдсят	вісімдесят	восемьдесят	восемьдєсят			eighty	
короварня	короварня	коровник	каровнік			cowhouse	
помещеніє	приміщення	помещение	памєщеніє			building	
лічно	особисто	лично	лічна		за свої лічно деньги	personally	

деньги	гроші	деньги	деньгі		за свої лічно деньги	money	
государство	державна	государство	гасударства			state	
благотворительну	благодичинну	благотворительную	благатварительную		благотворительну пом	charitable	
помощ	допомога	помощь	помашь		благотворительну пом	help	
дают	дають	дают	дают			to give (3 pers. plur.)	
кладовщиком	комірником	кладовщиком	кладовщiком			warehouseman (Instrumental)	?
последне	останне	последнее	паследнеє		последне время	last	
время	час	время	время			time	
медосмотри	медогляди	медосмотры	медасмотри			medical examinations	
тридцать	тридцять	тридцать	трiдцать			thirty	
пищевые	харчоблоці	пищевые	пищевые		на пищевые	cooking area	?
кресло	крісло	кресло	кресла		на кресло	chair (Accusative)	
женщины	жінки	женщин	женщiни			women	
половыми	статевим	половой	палавой	Speaker uses it in plur.	половыми жизнью	sexual (Instrumental)	
жизню	життям	жизнью	жизнью		половыми жизнью	life	
кровать	ліжку	кровате	кравате		на кровати	bed (Lokative)	
ничего	нічого	ничего	нічево			nothing	
Евгеновна	Євгенівна	Евгеновна	Євгенавна	Patronymic name	Ася Евгеновна	Evgenievna	
выгодно	вигідно	выгодно	вигадна			profitably	
хоть	хоть	хоть	хоть			though	
сотрудников	співробітників	сотрудников	сатрудніків		для сотрудников	employees (Genitiv, plur.)	
понравилось	сподобалось	понравилось	панравілась			to like (Past, impersonal)	
наготовлю	наготовлю	наготовлю	наготовлю			to prepare (Fut., 1 pers. sing.)	
сотрудникам	співробітникам	сотрудникам	сатруднікам			employees (Dative, plur.)	
сутки	добы	сутки	суткі			24-hours	
переночевала	переночувала	переночевала	переночевала			to spend the night (Past, f.)	
помыться	помитися	помыться	памитса			to wash / to take a shower	
посижу	посиджу	посижу	пасіжу			to sit (Future, 1 pers. sing.)	
отдохну	відпочину	отдохну	атдахну			to rest (Future, 1 pers. sing.)	
вовремя	вчасно	вовремя	вовремя		во время платила	in time	
во время	під час	во время	ва время			during	

советского	радянського	советского	саветскава		Советского союза	soviet (adj., Genitive)	
союза	союзу	союза	саюза		Советского союза	union (Genitive)	
лучше	краще	лучше	лучше			better	
получала	получала	получала	палучала	U. usage of "получала" registered in 1956, cur		to get (Past, sing. f.)	
семдїсят	сімдесят	семьдесят	сємьдєсят		семдїсят два рубля	seventy	
рубля	рублі	рубля	рубля		семдїсят два рубля	rybli	
советскі	радянські	советские	саветскіє		на советскі	soviet (adj.)	
рострочку	розстрочку	рассрочку	рассрочку		рострочку платила	mortgage (Accusative)	
квартиру	квартиру	квартиру	квартиру			apartment (Accusative)	
двенадцать	дванадцять	двенадцать	дванадцять			twelve	
яслі	ясла	ясли	яслі		за яслі	nursery school	
садік	садочок	садик	садік	kindergarten + in Ukraini	за садік	kindergarten	
гривнів	гривень	гривен	грівен		десять гривнів	hryvni	
скіки	скільки	сколько	сколька			how much	
плотять	платять	платят	платят			to pay (3 pers. plur.)	
шеснадцять	шістнадцять	шестнадцать	шестнадцять			sixteen	
рубль	рубль	рубль	рубль			rybl'	
шіїсят	шістдесят	шестьдесят	шестьдесят			sixty	
подорожало	подорожчало	подортжало	падаражала			to become more expensive	
магазінах	магазінах	магазінах	магазінах		в магази́нах	shops (Lokative)	
ізобіліє	надлишок	изобилие	ізабіліє			excess	
купить	купити	купить	купіть			to buy	
капйкі	копійки	копейки	капйкі			pennies	
еті	ці	эти	еті			these	
напрімер	наприклад	например	напрімер			for example	
гостей	гостей	гостей	гастей			guests (Genitive)	
хороше	хороше	хорошее	харошає			good	
дальше	далі	дальше	дальше			further	
жить	жити	жить	жить			to live	
тем	тим	тем	тєм	When followed by "бол	тем более	that (Dative, plur.)	
более	більше	более	болеє	When preceded by "те	тем более	more	

нравиться	подобается	нравится	нравітса			to like (impersonal)	
хорошо	добре	хорошо	харашо	Used as a reply to a question, expresses an agree		good / ok	
спасіба	спасибі	спасибо	спасіба			thank you	

потом	потім	потом	патом			then	
значить	значить	значит	значіт			to mean (impersonal)	
понимаешь	розумієш	понимаешь	панімаєш			to understand (Inter. 2 pers. sing.)	
опытный	досвідчений	опытный	опитний		я опытный человек	with experience (adj. m.)	
человек	людина	человек	челавек		я опытный человек	person	
работал	робив / працював	работал	работал			to work (Past, sing. m.)	
людьми	людьми	людьми	людьми		Я работал с людьми	people (Instrumental)	
руководством	керівництвом	руководством	руководствам	under the guidance	під руководством	guidance (Instrumental)	
громадна	величезна	огромная	агромная		громадна кількість лю	huge (adj., f.)	
людей	людей	людей	людей			people (Genitive)	
потом	потім	потом	патом			then	
это	це	это	ета			this	
дело	діло	дело	дела			bussiness	
не	не	не	не		не подходит	not	
подходит	підходить	подходит	падходить			to be suitable (impers.)	
веду	веду	веду	веду		Я веду разговор	to conduct (1 pers. sing.)	
разговор	розмову	разговор	разговор			conversation (Accusative)	
вот	ось	вот	вот			here	
всё	все	всё	всё			everything	
разложить	розложити	разложить	разлажить		разложить по полочкам	to decompose	
полочкам	поличках	полочкам	полачкам		разложить по полочкам	shelves (Dative)	
иначе	інакше	иначе	іначе			otherwise	
надо	потрібно / необхідно	надо	нада			to need (impers.)	
сколько	скільки	сколько	сколька			how much	
будет	буде	будет	будет			to be (Fut., 3 pers. sing.)	
работу	роботу	работу	работу			work (Accusative)	
коштувать	коштувати	стоит	стоить		буде коштувать	to cost	
васемсот	вісімсот	восемьсот	васемьсот		васемсот долларов	eight hundred	
долларов	доларів	долларов	долларав		васемсот долларов	dollars (Genitive)	
работы	роботи	работы	работи		приступай до работы	work (Genitive)	
начит	значить	значит	значіт			to mean (impersonal)	
значить	значить	значит	значіт			to mean (impersonal)	
договорилися	договорилися / домови	договорились	дагаварілись			to have agreed (Past, plur.)	
люди	люди	люди	люди			people	

сделали	зробили	сделали	сделали			to do (Past, plur.)	
розчитуюся	розраховуюся	рассчитываюсь	рассчитиваюсь			to count / to pay (1 pers. sing.)	
шо	що	что	што			what	
спаскудили	спаскудили	испортили	іспортіли			to ruin (Past, plur.)	
переложили	переклали	переложили	перелажилі			to shift (Past, plur.)	
короче	коротше	короче	кароче			shorter	
дев'ятсот	дев'ятсот	девятьсот	дев'ятсот			nine hundred	
васемсот	вісімсот	восемьсот	восемсот			eight hundred	
дагаварювались	договорювались / домов	договаривались	дагаварівались			to have agreed (Past, plur.)	
еслі	якщо	если	еслі			if	
сопротівлявся	супротивлявся	сопротивлялся	сапратівлявся			to withstand (Past, sing. m.)	
дойшли	дійшли	дошли	дашли			to reach / to arrive (Past, plur.)	
денег	грошей	денег	денег			money (pl. Genitive)	
скоко	скільки	сколько	сколька			how much	
не	ні	нет	нет			no	
дагаварілись	договорились / домови	договорились	дагаварілись			to have agreed (Past, plur.)	
оприділілись	визначилися	определились	апрелілісь	не оприділілись		to decide / to determinate (Past, plur.)	
оно	воно	оно	ано			it	
но	але	но	но			but	
канкретно	конкретно	конкретно	канкретна			specifically	
дагаварювались	договорювались / домов	договаривались	дагаварівались			to have agreed (Past, plur.)	
хороших	хороших	хороших	хороших	в хороших отношениях		good (Lokative)	
отношениях	відносинах	отношениях	атнашениях	в хороших отношениях		relationships (Lokative)	
отношения	відносини	отношения	атнашения			relationships	
этого	цього	этого	етава			this (Genitive)	
дела	діла	дела	дела			bussiness (Genitive)	
имеют	мають	имеют	імеют			to have (3 pers. plur.)	
который	який	который	каторий			that / which	
друг до друга	один до одного	друг к другу	друг к другу	друг до друга		one to another	
ходили	ходили	ходили	хаділі			to walk (Past, plur.)	
квартиры	квартири	квартиры	квартири	в квартири		apartment	
доме	домі	доме	доме	в одном доме		house (Lokative)	
тебе	тобі	тебе	тебе			you (Dative)	
хаділ	ходив	ходил	хаділ	де б не хаділ		to walk (Past, sing. m.)	

говора	говорячи	говоря	гаваря		Короче говора	to say (gerund)		
стоїти	коштувати	стоит	стоїть			to cost		
говарю	говорю	говорю	гаварю			to say (1 pers. sing.)		
входниє	вхідні	входные	вхадниє		входниє замки	entrance (adj. plur.)		
швірьонка	фіранка	занавеска	занавеска			curtain	?	
отдав	відав	отдал	атдал			to give away (Past, sing. m.)		
деньги	гроші	деньги	деньгі			money		
харашо	добре	хорошо	харашо			good / ok		
начить	значить	значит	значіт			to mean		
сечас	зараз	сейчас	сейчас			now		
отдати	віддати	отдать	атдать			to give away		
рядом	поряд	рядом	рядам			near		
где-то	десь	где-то	где-та			somewhere		
ше	ще	ещё	іщю			more		
шось	щось	что-то	што-та			something		
Чо	чому	почему	пачему			why		
Шоб	щоб	чтобы	штоби			in order to		
навсігда	назавжди	навсегда	навсігда			forever		
бил	був	был	бил		де б ти не бил	to be (Past, sing. m.)		
сібя	себе	себя	себя		в сібя	oneself (Accusative)		
Мілан	Мілан	Милан	Мілан		в Мілан	Milan		
куда-то	кудись	куда-то	куда-та			somewhere		
понімаєш	розумієш	понимаешь	панімаєш			to understand (inter. 2 pers. sing.)		
шота	щось	что-то	што-та			something		
дагаварюєшся	договорюєшся / домов	договариваешься	дагаваріваєшся			to have agreed (2 pers. sing.)		
сразу	зразу	сразу	сразу			immediately		
оприділяйся	визначися	определись	апрелелісь			to decide / to determinate (Imperative, 2 pers. sing.)		
Неаполі	Неаполі	Неаполе	Неапале		в Неаполі	Neaples (Lokative)		
нужно	потрібно / необхідно	нужно	нужна			to need (impers.)		
самое	саме	самое	самое		то же самое	same		
тоже	також	тоже	тоже			also		
нічем	нічим	ничем	нічем			nothing (Instrumental)		
отлічяється	відрізняється	отличается	атлічяєтса			to differ (3 pers. sing.)		
себе	себе	себе	себе		к себе приехал	oneself (Dative)		



приїхав	приїхав	приехал	приїхав			to arrive (Past, sing. m)		
пришлось	довелось	пришлось	пришлось			to have to / to happen to (Past, impersonal)		
іскать	шукати	искать	іскать			to search		
жильйо	житло	жилье	жильйо			habitation		
почему	чому	почему	пачему			why		
потому	тому	потому	патаму		Потому що	because		
привикли	привикли	привыкли	прівиклі			to get used to (Past, plur.)		
может	може	может	може			to can (3 pers. sing.)		
оторватса	відірватись	оторваться	атарватса			to break away		
сталківался	стикався	сталкивался	сталківалса			to encounter (Past, sing. m.)		
відику	відику	видику	відику	Both U. and R. terms are	на відику	video player		
Норвегії	Норвегії	Норвегии	Нарвегії		в Норвегії	Norway (Lokative)		
Германії	Німеччині	Германии	Германії		в Германії	Germany (Lokative)		
нелегалом	нелегалом	нелегалом	нелегалам			illegal (Instrumental)		
девочька	дівчинка	девочка	дєвачька			girl (diminutive)		
планірував	планував	планировал	планіравал			to plan (Past, sing. m.)		
туда	туди	туда	туда			there		
уїхати	уїхати	уехать	уехать			to move away		
она	вона	она	ана			she		
кончила	кінчила	кончила	кончіла		кончила університет	to finish (Past, sing. f.)		
універсітет	університет	университет	універсітет			university (Accusative)		
есть	є	есть	есть			to be (3 pers. sing.)		
адін	один	один	адін			one		
начальнік	начальник	начальник	начальнік		начальнік проіздствє	boss		
проіздствєнного	виробничого	производственного	праіздствєннава			production (Genitive)		
атдела	відділу	отдела	атдела			department (Genitive)		
прораб	прораб	прораб	прараб			craftsman		
рабочі	робочі	рабочие	рабочіє			workers		
ка	до	ко	ка		ка мнє	to		
мнє	мене	мне	мнє	In Ukrainian "I" in Dative	ка мнє	I (Dative)		
прістал	пристав	пристал	прістал			to stuck (Past, sing. m.)		
діплом	диплом	диплом	діплом			diploma		
купіл	купив	купил	купіл			to buy (Past, sing. m.)		
он	він	он	он			he		

сьогодня	сьогодні	сегодня	сегодня			today	
много	багато	много	много			a lot	
виходит	виходить	выходит	виходит			to go out	
спеціалістів	спеціалістів	специалистов	спеціалістав			specialists (Genitive)	
каторие	які	которые	каторие			that /which	
дійствительно	насправді	действительно	дійствительна			really	
деньгі	гроші	деньги	деньгі		за деньгі	money	
купляють	купляють	покупают	пакупають			to buy (3 pers. plur.)	
кагда	коли	когда	кагда			when	
учілся	учився	учился	учілся			to study (Past, sing. m.)	
ізбав	збав	избавь	ізбавь			to relieve (Imperative, 2 pers. sing.)	
ісключат	виключать	исключат	ісключат		ісключат з інститута	to expel (Future, 3 pers. plur.)	
інститута	інституту	института	інститута		з інститута	institute (Genitive)	
сметуть	зметуть	сметут	сметут		Figurative sense of "sweep away" that means "expel"	sweep away (Future, 3 pers. plur.)	
заместітеля	заступника	заместителя	заместітеля		заместітеля по учебной	vice / deputy	
учебной	навчальній	учебной	учебнай		заместітеля по учебной части		
часті	частині	части	часті		заместітеля по учебной части		
вигонять	вигонять	выгонять	виганят			to expel (Future, 3 pers. plur.)	
дальше	далі	далее	далес		так дальше	further	
тіпа	типу	типа	тіпа			type	
наделают	нароблять	наделают	наделают			to make (Future, 3 pers. plur.)	
но	але	но	но			but	
нічево	нічого	ничего	нічево			nothing (Accusative)	
делают	роблять	делают	делают			to do (3 pers. plur.)	
єдинствене	єдине	единственное	єдинственнос			the only one	
смели	змели	смели	смелі		Figurative sense of "sweep away" that means "expel"	sweep away (Past, plur.)	
авіаціонного	авіаційного	авиационного	авіаціоннава		авіаціонного інститута	aviation (adj., Genitive)	
інститута	інституту	института	інститута		авіаціонного інститута	institute (Genitive)	
тебе	тобі	тебе	тебе			you (Dative)	
говору	говорю	говорю	гаварю			to say (1 pers. sing.)	
то	те	то	то		то, што	that	
што	що	что	што		то, што	that	
сьогодне	сьогодні	сегодня	сьогодня			today	
Україне	Україні	Украине	Україне		на Україне	Ukraine (Lokative)	

как	як	как	как			as	
наукі	науки	науки	наукі			science (Accusative)	
нету	немає	нету	нету			not to be (3 pers. sing., impers.)	
панімаєш	розумієш	понимаешь	панімаєш			to understand (inter. 2 pers. sing.)	
признавать	признавати	признавать	признавать			to admit	
старие	старі	старые	старие		старие кадри	old (plur.)	
шото	щось	что-то	што-та			something	
молодьож	молодь	молодежь	маладьож			youth	
ідьот	іде / йде	идет	ідьот			to go (3 pers. sing.)	
плотят	платять	платят	платят			to pay (3 pers. plur.)	
профессори	професори	профессоры	прафессары			professors	
столько	стільки	столько	столька			so many	
шо	що	что	што			that	
напрімер	наприклад	например	напрімер			for example	
взять	взяти	взять	взять			to take	
вместе	разом	вместе	вместе			together	
свести	звести	свести	свести			to bring together	
профессуру	професуру	профессуру	прафессуру			professorship (Accusative)	
профессури	професури	профессуры	прафессури			professorship (Genitive)	
професора	професора	профессора	прафессара		до професора	professor (Genitive)	
Шоби	щоб	чтобы	штоби			in order to	
было	було	было	била			to be (Past, sing. n.)	
получить	получили	получить	палучіть			to become	
технических	технічних	технических	технічеськіх		кандидатом технічеськіх	technical (adj. Genitive)	
фізико-математическіх	фізико-математичних	физико-математическіх	фізико-математическіх		кандидатом фізико-математическіх	physical and mathematical (adj. Genitive)	
педагогическіх	педагогічних	педагогических	педагогічеськіх		кандидатом педагогічеськіх	pedagogical (adj. Genitive)	
скільки	скільки	сколько	сколька			how much	
тудов	тудів / праць	тудов	тудов	In U. also a term "труд" can be used but for "science"		works (Accusative)	
написать	написати	написать	написати			to write	
оно	воно	оно	ано			it	
п'яти	п'яти	пяти	п'яти		в п'яти бібліотеках	five (Genitive)	
должно	повинне	должно	далжно			should (3 pers. sing. n.)	
бить	бути	быть	бить			to be	
Нью-Йорке	Нью-Йорку	Нью-Йорке	Нью-Йорке		В Нью-Йорке	New York (Lokative)	

Лондоне	Лондоні	Лондоне	Лондане		в Лондоні	London (Lokative)		
Пекіне	Пекіні	Пекине	Пекіне		в Пекіне	Pekin (Lokative)		
Союзе	Союзі	Союзе	Саюзе	Soviet Union	при Союзе	Union (Lokative)		
Москве	Москві	Москве	Маскве		в Москве	Moscow (Lokative)		
должни	повинні	должны	далжни			should (3 pers. plur.)		
должен	повинен	должен	должен			should (2 pers. sing.)		
стіки-то	стільки-то	столько-то	столька-та			so much		
статтів	статей	статей	статей	In this context: "research"	статтів написати	papers		
стіки	стільки	столько	столька			so much		
получити	получити	получить	палучіть			to become		
токо	тільки	только	толька			only		
знает	знає	знает	знает			to know (3 pers. sing.)		
ілі	або	или	ілі			or		
підсідатель	голова	председатель	председатель		підсідатель комісії	chairperson		
практически	практично	практически	практически			practically		
ізвіщається	звіщається	извещается	ізвещается			to be announced (Pres, sing.)		
од	від	от	от			from		
підсідателя	голови	председателя	председателя			chairperson (Genitive)		
приїхал	приїхав	приехал	приїхал			to arrive (Past, sing. m.)		
с	з	с	с		с Москви	from		
тебя	тебе	тебя	тебя		тебя не знає	you (Accusative)		
такой	такий	такой	такой	"what for a one"	хто ти такой	such		
били	були	были	били			to be (Past, plur.)		
дествітельно	дійсно	действительно	дійствительна			really		
учьоние	учені	ученые	учьоние			scientists		
височайшево	найвищого / височезно	высочайшего	висачайшева		височайшево класу	highest (Genitive)		
міре	світі	мире	міре		в міре	world (Lokative)		
учьоних	учених	ученых	учьоних			scientists (Genitive / Accusative)		
гаварілі	говорили	говорили	гаварілі			to say (Past, plur.)		
плохо	погано	плохо	плоха			bad		
хорошее	добре	хорошее	харошее			good		
височайшем	найвищому / височезно	высочайшем	висачайшем		на височайшем уровне	highest (Lokative)		
уровне	рівні	уровне	уравне		на височайшем уровне	level (Lokative)		
нікуда	нікуди	никуда	нікуда		нікуда не деніся	nowhere		

денішся	дінешся	денешся	денешся		не денішся	to get away / to escape (2 pers. sing.)	
нічо	нічо	ничё	нічю			nothing	
хлеще	гірше	хуже	хуже			worse	
Росіяне	росіяни	россияне	расіяне			Russians	
харашо	добре	хорошо	харашо			good / correctly	
делают	роблять	делают	делают			to do (3 pers. plur.)	
рекламу	рекламу	рекламу	рекламу			advertise (Accusative)	
себе	собі	себе	себе			oneself (Dative)	
больше	більше	больше	більше			more	
нічого	нічого	ничего	нічево			nothing	
їздишь	їздиш	їздишь	їздиш			to ride (2 pers. sing.)	
Україне	Україні	Украине	Україне		по Україне	Ukraine	