FAKULTÄT FÜR !NFORMATIK

Providing Electronic Assistance for Autodidacts of Agglutinative Languages

Using the Example of the Endangered Finno-Ugric Mari Language

MASTERARBEIT

zur Erlangung des akademischen Grades

Mag.rer.soc.oec.

im Rahmen des Studiums

Informatikmanagement

eingereicht von

Jeremy Bradley

Matrikelnummer 8971060

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung:
Betreuerin: Ao. Univ.Prof. Mag. Dr. Margit Pohl

Wien, 22.10.2009

(Unterschrift Verfasser) (Unterschrift Betreuerin)

Contents

Contents 2
1. Preface 5
1.1 Objectives 5
1.2 Acknowledgements 6
2. The Mari Web Project 8
2.1 Purpose 8
2.2 Agenda 10
2.2.1 Keyboard Layouts, Fonts, Unicode 11
2.2.2 Mari for Everyone 14
2.2.3 Reading Texts 15
2.2.4 The Mari Dictionary Project 15
2.2.5 Morphology Games 17
2.2.6 Interconnections Between Elements 18
2.2.7 Handling Language Variants 18
3. The Mari Language 19
3.1 Sociological Context 19
3.2 Orthography 20
3.2.1 Phonemes of the Mari Language 21
3.2.2 Cyrillic Orthography 22
3.2.2.1 Difficulties 23
3.2.2.2 Defects 28
3.3 Grammar 31
3.3.1 Word Classes 31
3.3.1.1 Nominals 31
3.3.1.2 Verbs – First and Second Conjugation 32
3.3.1.3 Postpositions and Other Word Classes 33
3.3.2 Agglutination 33
3.3.3 Stem Changes 35
3.3.3.1 Regular Processes 35
3.3.3.2 Irregular Word Forms 36
3.3.4 Vowel Harmony 37

3.3.5 Deletion, Reduction 38
3.3.6 Declension of Nominals 41
3.3.7 Verbal Conjugation 43
3.3.7.1 Finite Verb Forms 43
3.3.7.2 Non-Finite Verb Forms 45
3.3.8 Derivation 45
3.3.9 Arrangement of Suffixes 47
4. Constructing and Breaking Down Morphology 49
4.1 Resources for a Kindred Language – Estonian 49
4.1.1 Eesti Keele Süntesaator 50
4.1.2 Eesti Keele Lemmatiseerija 54
4.1.3 Õigekeelsussõnaraamat 55
4.1.4 General Observations 56
4.2 The Morphology Generator 57
4.3 The Morphology Analyzer 60
4.3.1 An Example 61
4.3.2 Ambiguous Forms 66
4.3.3 Cutting Down on SQL Queries 67
4.3.4 Testing 69
5. Application of the Tools Created 71
5.1 Presentation of Information 71
5.2 Textbook Exercises 76
5.3 Assisted Reading 77
5.4 Personalized Vocabulary Sheets 79
5.5 Spelling Checker 80
5.6 Scholarly Uses 81
6. Usability 84
6.1 User Interface Design 84
6.1.1 Paradigms of User Interface Design 84
6.1.1.1 Designers ≠ Users 84
6.1.1.2 Expectations and Affordances 85
6.1.2 Measures of Usability 89
6.1.2.1 Learnability 90
6.1.2.2 Efficiency 90

6.1.2.3 Memorablity 90
6.1.2.4 Errors 90
6.2 Conclusions 91
A Tables 93
A.1 Suffixes 93
A.1.1 Enclitics, etc. 93
A.1.2 Nominal Declension 94
A.1.2.1 Case Suffixes 94
A.1.2.2 Number Markers 94
A.1.2.3 Possessive Suffixes 95
A.1.3 Verbal Conjugation 95
A.1.3.1 Finite Verb Forms 95
A.1.3.1.1 Indicative Present 95
A.1.3.1.2 Indicative First Preterite 96
A.1.3.1.3 Indicative Second Preterite 97
A.1.3.1.4 Imperative 97
A.1.3.1.5 Desiderative 98
A.1.3.2 Non-Finite Verb Forms 99
A.1.3.2.1 Infinitives and Gerunds 99
A.2 Productive Derivations 100
A.2.1 Nominal → Nominal 100
A.2.2 Verb → Nominal 100
A.2.3 Nominal → Verb 100
A.2.4 Verb → Verb 101
A.3 Arrangement of Suffixes 101
Figures 102
Sources 104

1. Preface

1.1 Objectives

At the Department of Finno-Ugric Languages of the University of Vienna, I am currently involved in a major project dedicated to making the Mari language, a Finno-Ugric language spoken in the Russian Federation, more accessible to the world as a whole and making the world more accessible to the speakers of this language. Spoken by roughly half a million people, the Mari language is far from being moribund, but a drastic decline in usage is evident. This is facilitated by the perception many Maris have that their language is a “useless” one that is not worth teaching to their children – a perception that is intensified by the almost complete absence of modern linguistic materials. Moreover, the lack of dictionaries makes it necessary to use Russian as a proxy for international communication. Reporters, students, journalists and translators alike suffer from this situation.

At the same time, the Mari language is of interest to a small, but growing number of people outside the Russian Federation. From my own personal correspondence alone, I can vouch for the fact that there are scholars interested in the language in Finland, Estonia, Hungary, Austria, Italy, Japan, Sweden, the United States and Basque areas. All of these individuals suffer from the dearth of learning and lexicographical materials on Mari. Where these exist, they are predominately in Russian, dated and essentially impossible to obtain, even in the native areas of the Maris.

These problems motivated the launch of a project aimed at creating a web-based platform offering some of these much needed resources for free to anyone interested, anywhere in the world. Our agenda includes the creation of an English-language textbook on the Mari language, a Mari-English dictionary and tools aimed at making the study of the Mari language easier for autodidacts. These tools – a Morphology Analyzer and a Morphology Generator as well as applications based on them, such as a Reading Aid –

are the focus of this thesis. In order to understand the context and relevance of these tools, it is necessary to have a very basic grasp of the structure of the Mari language. Thus a brief overview of its grammar will be provided, focusing on morphology.

Like its kindred languages in the European Union – Hungarian, Finnish and Estonian – but in contrast to the overwhelming majority of languages spoken in Europe – Mari is an agglutinative language. This means that a large number of morphemes carrying meaning are combined in individual words, with the result that words have a lot of internal structure. In order to become proficient in Mari, one must learn to understand and use its rich morphology. I have created tools capable of constructing and breaking down complex word forms in the Mari language and will discuss applications of these from a didactic point of view, exploring how a student of such a language might profit from them.

One of the exciting possibilities offered by this set of tools is a principle I call “assisted reading”, which refers to the annotation of Mari texts with structural information that is invisible to users by default, but can be accessed by demand for specific words with which they are having difficulty. This application also allows users to create customized vocabulary checklists and gives them the option of accessing relevant dictionary and textbook entries when necessary. Further applications of the tools created here, such as in a spelling checker, will be discussed as well.

1.2 Acknowledgements

A large number of people have been helpful to us in the launch of our project and continue to assist us. A complete and regularly updated list of these can be found at our website, www.mari-language.com. With respect to the present thesis, I would like to thank Professor Timothy Riese of the University of Vienna, my colleague in the Mari Web Project, for supplying me with grammatical tables in the course of his lectures and aiding me

greatly in my attempts to make a computer grasp the morphology of the Mari language. I also owe a debt of gratitude to Ilona Soukup, who kindly offered to serve as a software tester.

2. The Mari Web Project

When this thesis was printed, our Mari Web Project was still at a very early stage of development. Readers at a later date can see what progress has been made by visiting www.mari-language.com.

2.1 Purpose

At the present time, opportunities to study the Mari language for those who have not grown up as native speakers are basically non-existent. While an excellent textbook for the Mari language exists, it has been nearly two decades since the two volumes of "Марийский язык для всех" ("Marijskij jazyk dlja vseh" – "The Mari Language for Everyone"– **Yakimova et al. 1990, Yakimova et al. 1991**) were published. This sadly makes the book's title a falsity, simply due to the low print numbers of Mari books. As a rule Soviet and Russian publications always include information on print numbers and for publications related to Mari these are rarely over 1000. Considering that Mari still has roughly half a million speakers, these numbers are quite low and such books become unavailable within a very short period of time. If one does not have access to a library that was fortunate enough to purchase copies of the volumes of this valuable textbook when they were published, it is virtually impossible to even borrow or photocopy them.

Even if one was able to get a hold of a copy of this textbook, there would still be another major problem: like all linguistic resources on Mari published in the Russian Federation, it is in Russian. This naturally means that it cannot be used by people who lack extensive Russian skills or do not have a teacher to assist them.

There are some foreign publications on Mari written in languages other than Russian, primarily in Finnish, Hungarian and very rarely in German. However, none worth mentioning are in English.

As the Russian-language Mari textbook was based on materials compiled before the fall of the Soviet Union, many of its entries reflects aspects of life under Communism and are thus rather dated. The linguistic resources on Mari in German, however, go back even further in time, as they are primarily based on fieldwork conducted in the late 19th century. They can tell you what every single part of a loom is called, but not what word Maris use for an automobile.

To summarize, existing linguistic resources on Mari are:

- hard to get
- not in English
- in most cases, quite dated

Our aim is to do what we can to counteract all three of these weaknesses. A web platform was the logical choice , as it will be readily and indefinitely available online at no cost and can be easily expanded and updated. With respect to English, we are in a unique position to carry out this project at the University of Vienna, as Professor Riese and I are both native speakers.

This thesis focuses on the tools we intend to offer to people who want to learn or improve their understanding of the Mari language on their own. Competent teachers of Mari are rare worldwide. Whereas countries like Finland and Estonia have cultural institutions that provide teachers to foreign universities, thus enabling them to offer comprehensive courses on their languages, the Maris have no such resources. People interested in the Mari language who do not live in a Mari hotspot like Vienna must either go to Russia for instruction in the language or learn it on their own. Ideally, students of Mari would do both – they would in fact go to Russia to attend the language courses offered to foreigners in the native regions of the Mari,

but would not want to go there unprepared or to forget everything they learned when the course was over.

Learning a language by autodidactic means is difficult in any case. Students of Mari face the additional challenge posed by the agglutinative structure of Mari, as was mentioned above and will be discussed in more detail in Section 3.3.2. For now it suffices to say that in addition to learning vocabulary, which is naturally necessary for any language, students of agglutinative languages must also learn to deal with the internal structure of words – initially so that they can understand the language and eventually so that they can actively use it themselves. Contrary to popular belief, this does not make agglutinative languages harder by default; it just makes them different. In many respects, Mari is quite an easy language. Its pronunciation and spelling rules are extremely simple. It has no irregular plurals and only two irregular verbs. This thesis will explore how these differences in language types affect didactic resources dedicated to them.

2.2 Agenda

Some parts of our web platform can already be used today, but years of work will be necessary before others reach maturity. This section will give a brief overview of all the different things our web platform already includes and will eventually encompass.

Fig. 1: The Mari web platform

2.2.1 Keyboard Layouts, Fonts, Unicode

The Mari alphabet is a variation of the Russian Cyrillic alphabet, using a few characters in addition to those found in Russian. In much the same way that someone who wants to write a letter in German with an English keyboard will run into problems when he or she gets to the first ä, ö, ü or ß, people who want to write texts in Mari soon encounter problems with the special Mari characters.

While many fonts lack realizations of these characters, all special characters found in Mari have their own entries in Unicode, the computing industry standard with its extensive repertoire of characters. Given the right font, it is easy to find the special characters used in Mari. This thesis uses the “DejaVu Serif” font, part of the free DejaVu font pack (**Roh 2006**) included in some Linux installations, which can be easily installed on any computer.

This font set was designed for linguists and works very well for any writing system included in Unicode. The Mari letters Ӈ, Ö and Ӱ are thus no problem. Some of the font alternatives provided in standard Windows installations (XP or later) include a font capable of handling the Mari characters, namely Microsoft Sans Serif. We have used this font for many of our HTML online resources, where we must use a font we can assume all users will have installed.

Keyboard layouts remain a problem. Even if one has the font to display these letters, the keys needed to type them are generally not available. The special Mari characters are not included on the Russian keyboard layout and there is no standardized Mari keyboard layout. As a result, we have designed two sets of keyboard layouts, based on existing designs, which are easy to install.

The first set is aimed at users of QWERTY/QWERTZ keyboard layouts, such as the English, German, Scandinavian or Hungarian layouts. These layouts, where possible, phonetically transcribe the base layout into Mari. The German ü key, for example, houses the Mari letter “Ӱ”, which is pronounced like a German ü.

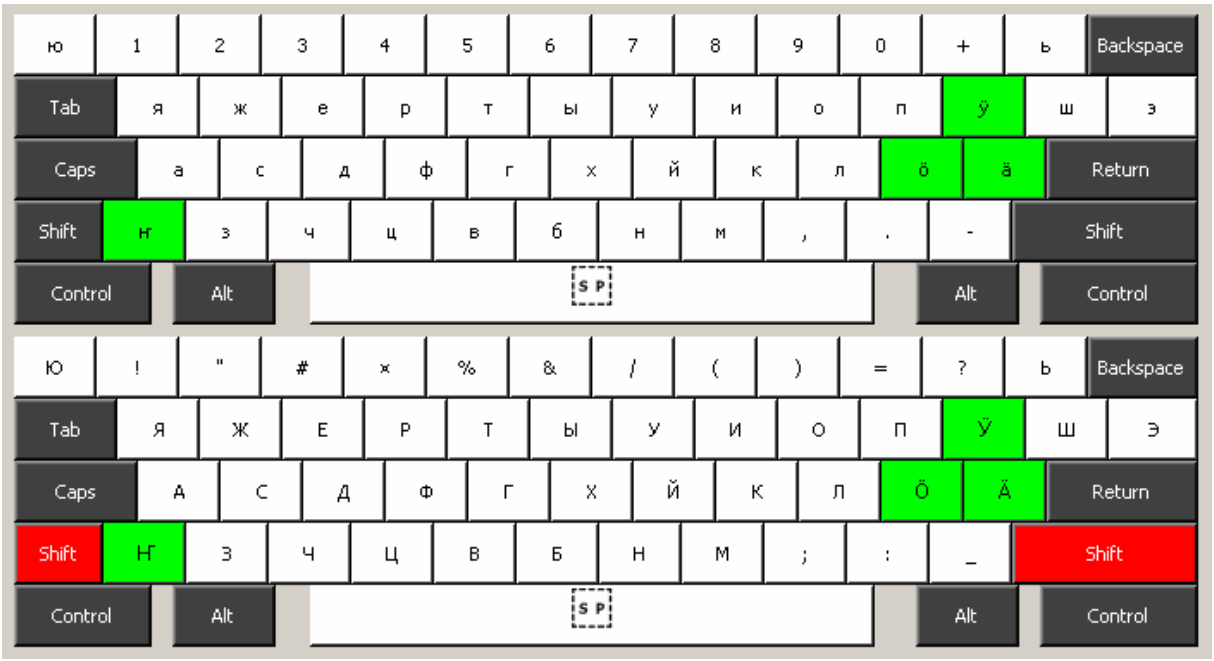

Fig. 2 Layout for users of Scandinavian, Finnish and Estonian keyboards

As perfect transcription is not possible between these two alphabets, some liberties had to be taken. The Latin letter q, which has no equivalent in Mari, is assigned a Mari letter that has no equivalent in German. As a result, the user will initially need some time to get used to this layout, but the similarities with conventional layouts should speed up this learning process.

This is the reason why we have designed individual German, English, Hungarian and Scandinavian/Finnish/Estonian layouts. In addition to including keys for ä, ö, ü and ß, the German layout also differs from the English one in that the letters y and z are switched. The equivalent Mari letters, Ы and З, will therefore also be switched on the respective Mari layouts.

In addition to these layouts aimed at foreign students of Mari, we have also created a layout for Maris in Russia, which has to meet entirely different criteria. The arrangement of keys on the Russian layout does not resemble the QWERTY/QWERTZ keyboard layout in any way. Our Mari-Russian layout follows the Russian arrangement, making it useless for foreign students of Mari, as they would have to learn to type all over again in order to use this keyboard.

Maris need to use Russian a great deal in everyday life. Thus, the layout prepared for them must be fully usable for Russian as well and, when installed, it should not cause them any problems when they want to type something in Russian.

Because the Russian layout is already cramped as it is, we were forced to use Alt-Gr for the special Mari characters. Russian characters not found in Mari but used in loan words, which we put on Alt-Gr keys in our layouts for foreign students, need keys of their own, thus taking up space needed for the special Mari characters. Fortunately, all Mari special characters closely resemble letters of the standard Russian alphabet. It is relatively easy to grasp that Alt-Gr + О is Ӧ, Alt-GR + У is Ӱ and Alt-Gr + Н is Ҥ.

With the right font and the right keyboard layout, handling Mari on a computer is no problem. Unfortunately, many Maris are not aware of the options Unicode offers and believe that Windows is by default incapable of handling Mari characters. A petition was drawn up in the native areas of the Maris in 2008, pleading with Microsoft to add the Mari characters to Windows, even though any installation of Microsoft Windows XP is capable of handling the Mari characters.

We plan to create a “handbook” on these problems. It will offer our keyboard layouts, explain the differences between character encodings and fonts, and provide instructions on how to install and use custom fonts and custom keyboard layouts.

2.2.2 Mari for Everyone

The Russian-language textbook on the Mari language, published in two volumes almost two decades ago (**Yakimova et al. 1990, Yakimova et al. 1991**) was mentioned in Section 2.1. This continues to be an asset for any student of the Mari language who can find copies and is competent in Russian. The University of Vienna has used these volumes in its Mari courses. Students there who lack Russian skills rely on the professor's ad hoc translations in their use of the book.

In cooperation with the authors of the original volumes, two native speakers of Mari and old acquaintances of his, Professor Riese is currently working on an English-language adaptation of this textbook. He is translating the parts that still seem appropriate today, updating sections that focus too strongly on life under Communism and adding explanations in places he has found them to be lacking, both in his own studies of Mari and while using the book in his lessons.

This new revised textbook will be published in print in the Russian Federation. We also plan to make it available online, in a downloadable pdf form. The tools described in this thesis will be integrated in the online version, as discussed below.

2.2.3 Reading Texts

For students of Mari who have already learned the basics, we plan to compile reading texts of various levels of difficulty and upload them on the Mari Web Platform. We plan to find newspaper articles that would be as interesting to foreigners as they are to native Maris, to select texts from classical literature that are comprehensible and meaningful for students of Mari, etc. Tools designed to aid non-native speakers of Mari with the interpretation of these texts will be discussed in Section 2.2.5.

2.2.4 The Mari Dictionary Project

The most ambitious item on our agenda is the creation of the world's first Mari-English dictionary. A grant proposal for this project is currently being evaluated by the Austrian Science Fund (FWF) and a decision will be made by November 2009. Should we receive funding for this project, we will start working on the dictionary in 2010, after Timothy Riese has completed the primary work on his textbook. Four people will spend three years compiling and translating some 50 000 entries and a number of subentries that has not yet been specified.

This dictionary will use software similar to a dictionary application I have created for small dictionaries for a different project at the University of Vienna. These dictionaries on the Nganasan and Nenets languages of Siberia can be found at http://www.univie.ac.at/negation.

A web dictionary has many advantages over a printed one. Some of these – such as speed, cost and accessibility – need little or no explanation. Others will become apparent in later sections of this thesis, when I illustrate the integration of the tools explained here and other linguistic applications into our dictionary. A web approach will also allow us to make sections of the dictionary available online in stages, before the entire project is completed. As the Maris' need for this dictionary is truly urgent and an incomplete dictionary would be better than no dictionary at all, we plan to upload our work letter by letter as soon as the respective final editing is completed.

An online presentation will also enable us to continue updating our database after the dictionary has been completed. Mari vocabulary, like the vocabulary of any living language, is constantly changing. Dictionaries printed in the 19th century have no entries on automobiles and dictionaries printed in the 1980s have no entries on computers. While we intend to make our dictionary as up to date as possible, it is to be expected that essential modern words will be missing a few years down the road. Thanks to the Internet, only our printed dictionary will be destined to become outdated. As long as our project can be continued in some form, we can continuously add new entries, as is done in other online resources such as the LEO German-English/English-German dictionary and Wiktionary. We could also give registered users the opportunity to suggest new entries, and accept or reject these depending on their validity and source.

The English translations provided in all of the entries of the dictionary will be based on Finnish, German, Russian and Hungarian sources. Between the members of our proposed project, we share competencies in all of these languages and several members of our team have experience as translators. For very specific terms that we will not know in our native languages (such as a species of duck native to the Volga basin and several hundred different

mushrooms), we will use a wide variety of existing dictionaries from the source languages into English.

In addition to the web-based and print versions of the dictionary, we will eventually also be able to make CD copies of the online dictionary available on request, both easily and cheaply. This service might be interesting to Mari native speakers who have computers at home, but no access to the Internet.

As discussed below, our Mari-English dictionary will support the electronic tools presented in this thesis, as well as be enhanced by them.

2.2.5 Morphology Games

When learning an agglutinative language, it is essential to learn to recognize both the form and meaning of the various morphemes attached to word stems. When encountering a Mari word such as /olalaštak/, native speakers of Mari have no trouble identifying the word stem /ola-/ (city), the plural marker /-la-/, the inessive case marker /-št-/ (corresponding to the English preposition “in”) and the so-called enclitic suffix /-ak/, used to stress an element. In this way they immediately understand the meaning of this word, which can be roughly translated into English as “especially in the cities”. Students of Mari might not see the forest for the tree when they first see such words. Forming words like this can likewise be difficult, as different suffixes connect with different stems in various ways and because not all arrangements of suffixes are legitimate.

Tools demonstrating Mari morphology in action can be quite useful for students. A morphology generator illustrates how to construct the literally thousands of different forms Mari words can assume. A morphological Analyzer helps students break down complex word forms into the stem and the suffixes.

Applications of these tools for more advanced students and linguists will also be presented in this thesis. In particular, an “assisted reading” application enables users to request additional information on particular words in texts.

2.2.6 Interconnections Between Elements

The various elements of our Mari web platform will be relevant to each other. Therefore, we plan to interconnect our resources and tools to the greatest degree possible. For example, the morphology generator might be interesting to someone using the dictionary who wants to know what a Mari word found in the dictionary might look like in some specific form. To create the reading aid, both the morphological Analyzer and the dictionary will be essential. People interested in reading texts should be given instant access to the reading aid, if they wish to use it.

2.2.7 Handling Language Variants

As will be discussed in later sections, attempts to create one unified written standard for the Mari language have not yet been successful. While one dominant variant, known as Meadow Mari, exists and is used by the majority of Maris, a small group of Maris continue to use a second written norm, known as Hill Mari. This second norm is quite similar to the dominant one and shares most of its vocabulary, but has some orthographical differences, etc.

While we do not intend to offer didactic resources on variants of Mari other than the dominant written norm, we do want to make our materials usable for people attempting to read texts written in the Hill Mari variant. Our dictionary will understand Hill Mari passively: For example, if a user searches for the Hill Mari word /näläš/ (to take), the dictionary should direct him/her to this word's equivalent in the dominant language variant – /nalaš/.

3. The Mari Language

3.1 Sociological Context

The Mari language, referred to as Cheremis in older materials, is one of hundreds of minority languages spoken in the Russian Federation. It is a Finno-Ugric, or Uralic, language, unrelated to Russian, but related to the Finnish and Estonian languages and more distantly to Hungarian.

Fig. 3: The Uralic world (Wikipedia 2007) – Mari is marked in dark red

Centuries of migration and foreign conquest have given the Uralic language family a rather scattered appearance when pictured on a map. The closely

related Baltic Finnic languages, Finnish and Estonian, are spoken on the eastern shores of the Baltic Sea. The Saami, or Lappic, languages are spoken in northern Scandinavia, and Hungarian is spoken in central Europe, especially along the shores of the Danube River. Mari is spoken in the Volga basin and Ural region of the Russian Federation, roughly a thousand kilometers east of Moscow.

It is difficult to make exact estimates of the number of speakers of the Mari language. In the most recent Russian census (**Federal State Statistics Service of Russia 2002**), some 604 298 residents of the Russian Federation declared themselves to be ethnic Maris. While minor émigré communities exist in Finland, Estonia and Hungary, these are statistically negligible. It is very hard to estimate, however, to what degree ethnic self-identification is reflected in language usage. Complete assimilation by the dominant Russian-speaking community in the relatively near future is a very real danger (**UNESCO 2009**).

As discussed above, there are two written norms of the Mari language – the dominant Meadow Mari variant, and the much smaller Hill Mari variant. All references to Mari in this thesis pertain to the Meadow Mari norm, unless explicitly stated otherwise.

3.2 Orthography

While a Latin orthography exists for the Mari language and has been used by Western linguists in the past, contemporary Mari exclusively uses an adaptation of the Russian Cyrillic orthography. As the focus of our project is contemporary Mari and we want to make our resources accessible to the Mari community, we use this Cyrillic orthography in all our projects.

This same principle of accessibility makes the exclusive use of the Cyrillic orthography problematic for this thesis. Whereas any linguist dealing with the Mari language must quickly become proficient in the Cyrillic alphabet,

such a skill cannot be expected of readers at a Technical University. Thus, all examples will be written in a phonological transcription of the Mari language. Any Cyrillic content found on screenshots will be transcribed into the Latin alphabet in the image's annotations.

3.2.1 Phonemes of the Mari Language

Phonemes are the smallest linguistically distinctive units of sound. (**Bünting 1996**). One phoneme can have several realizations that do not change the meaning or interpretation of this unit of sound. For example, the realization of the letter *r* differs from dialect to dialect in English and German, but the meaning carried by this letter is not altered by different variants. In both languages /r/ is one linguistic unit. Likewise, the realization of the "ch" sound in the German differs depending on the vowels surrounding it ("ich" vs. "ach"), but these alterations have no effect on meaning. As this thesis focuses on morphology and not phonology or phonetics, details of pronunciation are not of interest here and will not be discussed.

The following lists contain all phonemes found in the Mari language, with examples of equivalent or similar sounds found in English where this is possible and in other major European languages where it is not. Phonemes marked in grey are only found in loan words, not in native Mari words. The letters used here to mark phonemes will be used in all Mari examples sentences and words hereafter.

Vowels:

/a/	like a in 'sp**a**'	/ö/	no English equivalent. Like ö in German 'sch**ö**n'
/e/	like e in 'm**e**n'		
/i/	like i in 'p**i**n'	/ü/	no English equivalent. Like ü in German 'Bl**ü**te'
/o/	like o in English 'sh**o**w'		
/u/	like oo in 'b**oo**t'	/ə/	like u in 'pl**u**s'

Consonants:

/b/	*like b in ‘boot’*	/m/	like m in ‘**m**an’
/c/	*like ts in ‘ha**ts**’*	/n/	like n in ‘**n**ame’
/č/	like ch in ‘**ch**air’	/ń/	like ni in ‘o**ni**on’
/d/	like d in ‘**d**oor’	/ŋ/	like ng in ‘si**ng**’
/f/	*like f in ‘**f**arm’*	/p/	like p in ‘**p**ill’
/g/	like g in ‘**g**ood’	/r/	like r in German ‘**r**eden’
/x/	*no English equivalent.*	/s/	like s in ‘**s**oon’
	*like ch in Scots ‘Lo**ch**’*	/š/	like sh in ‘**sh**ame’
	*like ch German ‘Da**ch**’*	/t/	like t in ‘**t**able’
/j/	like y in ‘**y**ellow’	/v/	like v in ‘**v**isit’
/k/	like c in ‘**c**amp’	/z/	like z in ‘**z**oo’
/l/	like l in ‘**l**ife’	/ž/	like si in ‘vi**si**on’
/ľ/	no English equivalent. Like gli in Italian ‘fi**gli**o’		

3.2.2 Cyrillic Orthography

Like all other Finno-Ugric languages spoken in Russia, the Mari language uses a variation of the Russian Cyrillic orthography. Latin orthographies have been created, but the implementation of these has failed. An ideal orthography for a language would have a 1:1 relationship between phonemes and letters of the alphabet – a principle the Cyrillic alphabet is by default not conducive to. Thanks to the introduction of special characters for phonemes that Mari has and Russian does not (/ö/, /ü/, /ŋ/), Mari fares better in this regard than many other Finno-Ugric languages spoken in Russia. Nevertheless, some problems exist with the orthography. Those relevant to our software will be briefly discussed.

The alphabet of the Mari language is as follows, with characters found only in loan words marked in grey italics, and characters not found in Russian marked in bold blue.

А а – /a/	Л л – /l/, /ľ/	*Ф ф – /f/*
Б б – /b/	М м – /m/	*Х х – /x/*
В в – /v/	Н н – /n/	Ц ц – /c/
Г г – /g/	**Ӈ ӈ – /ŋ/**	Ч ч – /č/
Д д – /d/	О о – /o/	Ш ш – /š/
Е е – /e/, /je/	**Ӧ ӧ – /ö/**	*Щ щ – /štš/*
Ё ё – accented /jo/	П п – /p/	Ъ ъ – hard sign
Ж ж – /ž/	Р р – /r/	Ы ы – /ə/
З з – /z/	С с – /c/	Ь ь – soft sign
И и – /i/	Т т – /t/	Э э – /e/
Й й – /j/	У у – /u/	Ю ю – /u/, /ju/
К к – /k/	**Ӱ ӱ – /ü/**	Я я – /a/, /ja/

3.2.2.1 Difficulties

The Cyrillic alphabet was created as a writing system for Slavic languages. When applied to languages outside of this family, certain aspects perfectly reasonable and functional for Slavic languages become little more than annoyances. One such feature is the manner in which some vowels have two variants, one of which is used after palatalized consonants and the other after non-palatalized consonants. In the Latin transcription of Russian words used in this thesis, palatalized consonants are consistently marked with the acute accent sign (´). For example, t is not palatalized and ť is palatalized.

Phoneme	Letter	Example	Transcription	Translation
/a/:	а	Москв**а**	Mosvk**a**	Moscow
	я	мен**я**	me̲ńa	me
/e/:	э	с**э**р	s**e**r	sir
	е	н**е**т	ń**e**t	no
/o/:	о	шк**о**ла	šk**o**la	school
	ё	вс**ё**	vś̲**o**	everything
/u/:	у	ст**у**л	st**u**l	chair
	ю	л**ю**бл**ю**	l̲**u**bl̲**u**	I love

Fig. 4: Vowels in Russian

As Russian has a palatalized and non-palatalized variant of almost every consonant, such a system works quite well here. The vowels written after palatalized consonants – я, е, ё, ю – are referred to as “soft” vowels, а, э, о, у are called “hard” vowels. When soft vowels appear after other vowels or at the beginning of a word, they are pronounced as /jV/, where V is the vowel in question. Should difficulties arise, the Russian version of the Cyrillic alphabet uses a so-called “soft sign” (ь) or “hard sign” (ъ), to explicitly indicate whether or not the consonant preceding the sign is palatalized. This is necessary for example when a palatalized consonant appears at the end of a word where there is no vowel following it to mark the palatalization or when the combination /jV/ appears after a consonant.

Sign	Example	Transcription	Translation
ь:	брат	bra**t**	brother
	брать	bra**ť**	to take
ъ:	**те**л**е**фон	**ťeľe**fon	telephone
	о**бъе**кт	o**bje**kt	object

Fig. 5: Soft signs and hard signs in Russian

As Mari also has palatalized consonants, but does not have as wide a range of these as Russian does, this system was adopted inconsequentially. For /a/

and /u/ the rules in Mari are equivalent to the Russian rules just discussed. In Russian the letter ё – the soft o – is always stressed. As word stress in Mari follows a very different system, such a rule would make little sense in Mari and this letter is not used at all, except in Russian loan words. Should a palatalized consonant appear before /o/, it must be marked with the soft sign. The same applies to palatalized letters before /ö/, /ü/ and /ə/ – vowels that either do not exist in Russian or do not influence the pronounciation of the preceding consonant.

/e/ and /i/ are problematic. Problems related to /i/ do not occur in places that they might interfere with any of the software discussed in this thesis and will thus be ignored. The vowel /e/, however, causes real problems.

/e/ in Mari is generally realized with the soft variant, e, regardless of the character of the consonant preceding it. Only at the beginning of words and after other vowels does Mari stick to the Russian rules – е is /je/, э is /e/.

Position	**Example**	**Transcription**	**Translation**
After consonant	в**е**р	v**e**r	place
	имн**е**	imń**e**	horse
After vowel	ку**э**	ku**e**	birch
	ву**е**ш	vu**je**š	into the head
Initial letter	**е**ҥ	j**e**ŋ	person
	эҥер	**e**ŋer	river

Fig. 6: /e/ in Mari

As a result, palatalized consonants are not marked if they precede /e/. Problems caused by this will be discussed in the following section.

The opposition between soft and hard vowels must be kept in mind both in generating Mari morphology and in breaking it down. When, for example, the suffix “-em”, the possessive suffix of the first person singular, is added to

a stem, the spelling of the suffix or of the stem can change, even if the pronunciation or transcription does not.

Ending	Example	Transcription	Translation
Consonant (not –j)	пӧрт	pört	house
+em	пӧрт**е**м	pört**e**m	my house
Vowel	изи	izi	little
+em	изи**э**м	izi**e**m	my little (one)
-j	вуй	vuj	head
+em	ву**е**м	vu**je**m	my head
Soft sign	мыскынь	məskəńь	unhappy
+em	мыскынем	məskəń**e**m	my unhappy (one)

Fig. 7: The suffix “-em”

Endings with /j/ or the soft sign are “absorbed” by the suffix, so to speak.

An entirely different problem is posed primarily by the letter д (/d/). It is subject to a process called final obstruent devoicing, which does not exist in English, but is, for example, found in German or Russian. As in English and German, Mari distinguishes voiced consonants from voiceless consonants. A consonant is voiced if a speaker’s vocal chords vibrate in its pronunciation. Vibrating vocal chords are what set /b/ apart from /p/, /d/ apart from /t/, /g/ apart from /k/, /z/ apart from /s/, /ž/ apart from /š/, /v/ apart from /f/ and /ð/ (th in “then”) apart from /θ/ (th in “thin”).

Mari is subject to a rule that also applies in German: Certain voiced consonants at the coda of a syllable become voiceless. It is due to this rule that the German words “Rad” and “Rat” are indistinguishable in spoken German – the d at the end of “Rad” is pronounced as /t/, resulting in both words having the identical pronunciation /rat/. Likewise, a д at the end of a word is not pronounced as /d/, but as /t/ in Mari. The same applies to other consonants having voiceless counterparts; unlike д, however, these rarely occur in the final position of a word or only do so in Russian loan words.

Only voiced consonants lacking a voiceless counterpart – /j/, /l/, /m/, /n/, /ŋ/, /r/ – are voiced when in the final position in a word.

Even if these shifts are not marked orthographically, they are relevant for our software. Certain suffixes vary in their realization depending on whether they are added after a vowel, a voiced consonant or a voiceless consonant. One such example is the possessive suffix of the third person singular, which generally is -же (/že/), -жо (/žo/) or жö (/žö/), depending on vowel harmony (see 3.3.4). When added to a stem ending with a voiceless consonant, the ж (/ž/) is assimilated and becomes an ш (/š/), leading to three options -ше (/še/), -шо (/šo/) and шö (/šö/). Unlike consonant shifts affecting the stem, these shifts are marked orthographically. As a result, the software must be misinformed about the phonetic nature of voiced obstruents – it must consider them to be voiceless, as they will always be pronounced as such when in positions relevant to us.

Ending	**Example**	**Transcription**	**Translation**
Vowel	ун**а**	un**a**	visitor
+že/žo/žö/še/šo/šö	ун**аж**е	un**až**e	his/her visitor
Voiced consonant	лӱ**м**	lü**m**	name
+že/žo/žö/še/šo/šö	лӱ**мж**ö	lü**mž**ö	his/her name
Voiceless consonant	пör**т**	pör**t**	house
+že/žo/žö/še/šo/šö	пör**тш**ö	pör**tš**ö	his/her house
Final obstruent	кид	ki**t**	hand
+že/žo/žö/še/šo/šö	кид**ш**е	ki**tš**e	his/her hand

Fig. 8: Final obstruents

As this process is quite regular, it is not difficult to handle. It only leads to some linguistically confusing classifications of letters in the software.

3.2.2.2 Defects

Whereas the previous section was concerned with aspects of Mari orthography that make life more complicated for a programmer dealing with software intended to master this orthography, the present section deals with actual defects in the orthography – that is, cases in which the orthography does not carry all the information necessary.

One such defect has been briefly addressed: Palatalization is not marked before the vowel /e/. Another is that two consonants have palatalized and non-palatalized versions in Mari – л (/l/, /ľ/) and н (/n/, /ń/). Should one encounter the letter combination “не”, it is impossible to know whether it should be pronounced /ne/ or /ńe/. The same applies to the letter combination “ле”, which can be both /le/ or /ľe/.

	Example	Transcription	Translation
не	нер	**n**er	nose
	имне	im**ń**e	horse
ле	неле	ne**l**e	hard
	ыле	ə**ľ**e	was

Fig. 9: Palatalization in Mari

For our morphology software, this defect would not be of any relevance if it occurred consistently for all vowels. As this is not the case, problems arise when a suffix causes the vowel at the end of a word to be replaced with another vowel. Take, for example, the case suffix of the accusative case, -m. This suffix causes an unstressed vowel at the end of a word to become an /ə/. If the consonant preceding this vowel is palatalized, a soft sign must appear in the accusative case, as palatalized consonants are marked when they appear before /ə/.

	Example	Transcription	Translation
Not palatalized	неле	nele	difficult (one)
	нелым	neləm	difficult (one) – ACC
Palatalized	имне	imńe	horse
	имн**ь**ым	imńəm	horse – ACC

Fig. 10: Marking of palatalization depending on the vowel

If the computer is not told in some way that the consonant before an ultimate vowel is palatalized, in certain cases it will not know whether or not it must insert the soft sign.

The same problem in reverse occurs when a word ends with the soft sign and a suffix starting with an /e/ is added to it. In the previous section, it was shown that the word мыскынь (/məskəń/) becomes мыскынем (/məskəńem/) when endowed with the possessive suffix of the first person singular – the ambiguity of the letter e towards palatalization allows the soft sign to be absorbed into the suffix. If a morphological analyzer was to encounter the word мыскынем, it would not be able to tell if it was /məskə**ń**em/ or /məskə**n**em/. As a result, it would also not know if this form was derived from the word мыскы**нь** (/məskə**ń**/) or мыскы**н** (/məskə**n**/). It would have to attempt to create both forms. The ambiguity here can only be resolved by a computer if it can check to see whether either of the possible forms can be found in the dictionary.

Word stress is a problem as well. When a suffix is added to a word ending with a vowel, it is necessary to know whether this vowel is stressed or not. Unlike English, German and Russian, stress generally follows strict rules in Mari. Under normal conditions, /a/, /i/, /u/ and /ü/ are stressed in the final position, /e/, /o/ and /ö/ are not and /ə/ generally does not appear in the final position. However, there are exceptions to these rules that are not marked orthographically. Occasionally, a final /e/ will be stressed. Many Russian words end with an unstressed final /a/. Most Russian loan words in Mari

have been adapted to fit Mari rules here. For some, the stress has moved to the final /a/. For example, the word машина (car), though orthographically identical in Mari and Russian, is pronounced /maši•na/ in Russian and /mašina•/ in Mari. In other Russian words used in Mari, the stress remains unchanged, but the final /a/ is replaced by an /e/ or an /o/ – a letter that is normally unstressed – in the final position, depending on vowel harmony. For example, the Russian word форма (/fo•rma/ - form) is формо (/fo•rmo/) in the most recent Mari orthography. Some Russian loan words have not been adapted in this manner and retain an unstressed /a/ in the final position.

Irregularities regarding stress are not marked orthographically. To illustrate problems caused by this, let us revisit the accusative suffix –m, which was discussed above. As we have seen, unstressed vowels preceding the accusative suffix become /ə/. Stressed vowels remain unchanged. If a word followed the standard orthographical rules, /a/ would remain /a/, /e/ and /o/ would become /ə/. However, irregularly stressed vowels in the final position, such as a stressed /e/ or /o/, remain unaltered by the accusative suffix. Likewise, /a/ becomes /ə/ when irregularly unstressed.

	Example	Transcription	Translation
/e/ – standard	теле	te•le	winter
	телым	te•ləm	winter – ACC
/e/ – irregular	теӊге	teŋge•	ruble
	теӊгем	teŋge•m	ruble – ACC
/a/ – standard	уна	una•	visitor
	унам	una•m	visitor – ACC
/a/ – irregular	ту•ндра	tu•ndra	tundra
	ту•ндрым	tu•ndrəm	tundra – ACC

Fig. 11: Stressed and unstressed final vowels

In the software it was necessary to compensate for both of these defects in Mari orthography: Word stress and palatalized consonants before /e/ are marked. When a word taken from the dictionary is inflected with the Morphology Generator, these markings give the software all the information it needs. When users enter a word into the Generator, failure to mark irregularities will lead to incorrectly inflected forms.

3.3 Grammar

The software that is the core of this thesis is designed to handle morphology. The following brief grammatical overview will ignore syntax and focus exclusively on the creation of word forms in Mari.

3.3.1 Word Classes

Like English, Mari has nouns, adjectives, verbs, adverbs, conjunctions and interjections. There are, however, some important differences in the classification of Mari words and the range of options the language's morphology offers for the inflection of words.

3.3.1.1 Nominals

As is typical of Finno-Ugric languages, the line between adjectives, nouns, pronouns and numerals is rather thin in Mari. Nouns can often be used as adjectives without any further alterations. This is also sometimes possible in English as well. For example, the noun "iron" is used as an adjective in the phrase "Iron Curtain" without any further alteration. In German, however, the suffix "-ern" must be added to the noun in question to create an adjective ("Eisen" → "der Eiserne Vorhang"). In Mari it is also possible to use any adjective as a noun without altering it. This is a process that is allowed in German ("der alte Mann" → "der Alte"), but not in English. As a

result, it makes no sense to differentiate between such words morphologically – nouns, adjectives, pronouns and numerals are grouped into a category that we call “nominals”.

3.3.1.2 Verbs – First and Second Conjugation

Mari verbs are strictly split into two conjugations, referred to as the first and second conjugations. Neither in function nor in semantics can a clear distinction be found between these groups of verbs and there is no other explanation for the split. Thus the differences between these verbs are only morphological – the two conjugations use different suffixes in the same situation. To the utter annoyance of lexographers and students of Mari alike, the one verb form that is the same in both conjugations is the infinitive – the standard form of verbs used in dictionaries, in keeping with international conventions.

Often, two verbs with no semantic connection whatsoever will be identical in the infinitive from, but differ in all other verb forms.

Form	Conj. I	Translation	Conj. II	Translation
Infinitive	kolaš	to hear	kolaš	to die
1.P.Sg.	kol**a**m	I hear	kol**e**n	I die
2.P.Sg.	kol**a**t	You hear	kol**e**t	You die
3.P.Sg.	kol**e**š	He/she/it hears	kol**a**	He/she/it dies
1.P.Pl.	kol**ə**na	We hear	kol**e**na	We die
2.P.Pl.	kol**ə**da	You hear	kol**e**da	You die
3.P.Pl.	kol**ə**t	They hear	kol**a**t	They die

Fig. 12: Conjugations I & II

Any serious Mari dictionary must note every verb’s conjugation. If users want to conjugate a verb not taken from the dictionary, the conjugation to which it belongs must be specified.

3.3.1.3 Postpositions and Other Word Classes

Mari does not have prepositions at all. Instead it uses grammatical cases and postpositions. A postposition is a word that serves the same function as a preposition, but comes after a nominal. English and German only have very few postpositions (“ago” – “five years ago”, “away” – “five miles away”, “through” – “all night through”; “wegen” – “des Geldes wegen”, “gleich” – “einem Engel gleich”, “nach” – “meiner Meinung nach”). While postpositions cannot be declined as nominals are, they are not morphologically inert as English prepositions are. A possessive suffix can be used to mark the object of a postposition instead of, or in addition to, a pronoun.

/voktene/	– beside
/(**məjən**) voktene**m**/	– beside **me**
/(**təjən**) voktene**t**/	– beside **you**

Thus postpositions form a group in their own right in our analysis of the Mari morphology. Most adverbs have a comparative degree and are thus not completely inert. Postpositions, adverbs and conjunctions can receive the same enclitic suffixes that nouns and verbs can. These will be discussed later.

3.3.2 Agglutination

As has already been stressed, Mari is an agglutinative language. Stemming from the Latin word “agglutinare” meaning “to glue together”, agglutinative languages use a large number of affixes. Each of these affixes carries one “unit of meaning” and is referred to as a morpheme. In contrast to such languages are so-called isolating languages, in which one word carries one unit of meaning, and fusional languages, in which alterations of a word’s stem express meaning. This classification is not a clear-cut one and no language falls into only one category. English is generally considered to be

more of an isolating language, whereas German has very strong fusional tendencies. However, affixes are not unkown in either of these language.

Nevertheless, the differences between these three types of languages can be illustrated as follows.

pört	house	Haus
pörtem	my house	mein Haus
pörtemvlak	my houses	meine Häuser
pörtemvlaklan	to my houses	meinen Häusern
pörtemvlaklanat	to my houses, also	auch meinen Häusern

Fig. 13: Agglutinative, isolating, fusional

In this example English uses one affix, –s, to mark the plural. All the other information is encoded through separate words. German's fusional tendencies manifest themselves through the so-called umlaut – through the alteration of stem vowels, the meaning of a word is altered. Often, a suffix is added as well, but an umlaut can also suffice to change the meaning of a word ("die M**u**tter" – the mother, "die M**ü**tter" – the mothers).

This example illustrates the usefulness of the kinds of tools we have created. Whereas in English one nominal has only two different forms – singular and plural – in Mari, it can have literally thousands, disregarding derivations and the inflection of these. No dictionary could ever include all forms of all words, nor would such a dictionary make any sense even if possible, as a competent speaker of Mari is familiar with the suffixes of the language and can assemble and take apart these building blocks quite freely. When one is learning the language, this is much more difficult. Even someone familiar with all the suffixes can get confused, especially since suffixation is not always as unambiguous as it is in this example, as will be discussed later in this section.

3.3.3 Stem Changes

In an ideal agglutinating language, a word's stem does not change when affixes are added to the stem. While Mari is closer to being an ideal agglutinating language than Finnish or Estonian (see 4.1), stem changes do occur. As these are highly regular, they are not a major obstacle. They must, however, be taken into consideration.

3.3.3.1 Regular Processes

Infinitives in Mari always end with –aš. For first conjugation verbs (3.3.1.2), the imperative is created by removing this ending.

/tolaš/ (to come)	>	/tol/ (come!)
/lijaš/ (to be)	>	/lij/ (be!)

When the infinitive ending is removed, consonant clusters formerly between two vowels can end up in the final position in a word – where such consonant combinations are not allowed. For example, no Mari word can end with the consonant combination /kt/. If one was to create the imperative of the first conjugation verb /lektaš/ (to go) by conventional means, the resulting form */lekt/ would be invalid. As a result, one of the consonants in the consonant combination is dropped – in this case the /k/. The imperative of /lektaš/ is /lek/.

Only four such consonant combinations exist and each combination always reacts in the same way.

1:	/kt/	>	/k/	/lektaš/	>	/lek/ (to go > go!)
2:	/šk/	>	/š/	/muškaš/	>	/muš/ (to wash > wash!)
3:	/čk/	>	/č/	/kočkaš/	>	/koč/ (to eat > eat!)

4: /nč/ > /č/ /sinčaš/ > /sič/ (to sit > sit!)

Many other forms of verbs, in which a suffix starting with a consonant is added to the verb stem, use these altered versions of the verb stem.

/z/ in the final position of a first conjugation verb stem becomes /č/ in the same situation.

/vozaš/ (to fall) > /voč/ (fall!)

In second conjugation verbs, the imperative is formed by removing the –aš ending from the infinitive and adding an /e/, /o/ or /ö/, depending on vowel harmony (3.3.4). Consonant combinations in the stem are not altered.

/mondaš/ (to forget) > /mondo/ (forget!)
/malaš/ (to sleep) > /male/ (sleep!)

A few second conjugation verbs – verbs with single syllable stems ending with a vowel – do not get this extra /e/, /o/ or /ö/.

/puaš/ (to give) > /pu/ (give!)
/šuaš/ (to throw) > /šu/ (throw!)

3.3.3.2 Irregular Word Forms

Only very few words in Mari are truly irregular. Pronouns have inflected forms that cannot be derived by the standard morphological means offered by Mari and the language has (only) two irregular verbs – the verb “to be” and the so-called negation verb.

As there are so few irregular word forms in Mari, there are special entries for these in the dictionary. The morphological generator has been supplied

with tables listing these few irregular forms. No further thought must be given to these words.

3.3.4 Vowel Harmony

Like many other Finno-Ugric languages, Mari is bound by rules of vowel harmony that require assimilation of vowels in a word on the basis of certain criteria.

As mentioned above, unstressed final vowels in Mari are generally /e/, /o/ or /ö/. Which one of these three they are in a specific case depends on the nature of the word's stressed vowel. If the stressed vowel of a word is either /ö/ or /ü/, the unstressed letter in the word's final position will be /ö/. If the stressed vowel is /o/ or /u/, it will be /o/. If it is /a/, /e/, /i/ or /ə/, it will be /e/.

Stressed Vowel	Word	Translation
ö	p**ö**•rtəšt**ö**	in the house
ü	š**ü**•d**ö**	hundred
o	**o**•nč**o**	look!
u	l**u**•d**o**	duck
a	p**a**•l**e**	sign
e	l**e**•v**e**	warm
i	t**i**•d**e**	this
ə	j**ə**•lm**e**	tongue; language

Fig. 14: Vowel harmony

Vowel harmony affects words stems and suffixes alike. As a result, suffixes ending with unstressed vowels will always have three alternatives – one ending in /e/, one ending in /o/ and one ending in /ö/.

Stressed Vowel	Nominative	Inessive	Translation
ö	p**ö**•rt	p**ö**•rtəšt**ö**	house – in the house
ü	küt**ü**•	küt**ü**•št**ö**	herd – in the herd
o	**o**•lək	**o**•ləkəšt**o**	meadow – in the meadow
u	k**u**•do	k**u**•dəšt**o**	hut – in the hut
a	ol**a**•	ol**a**•št**e**	city – in the city
e	pöl**e**•m	pöl**e**•məšt**e**	room – in the room
i	k**i**•t	k**i**•dəšt**e**	hand – in the hand
ə	jə•lm**e**	jə•ləšte	language – in a language

Fig. 15: Inessive suffix –šte/-što/-štö

Exceptions to vowel harmony can be found in Russian loan words. Many of these have been adapted to comply with vowel harmony. For example, the Russian word /kńaže•stvo/ is /kńaže•stve/ according to modern Mari orthography. However, such adaptations have not been universal. The Russian word /ko•fe/ (coffee) remains /ko•fe/ in Mari; it does not become */ko•fo/.

It should be noted that Mari does not require consister palatal-velar vowel harmony in the sense that Finnish and Hungarian do. Vowel harmony only dictates the nature of unstressed final vowels. Other vowels in a word do not have to follow any such rules, regardless of whether they are stressed or not – e.g. /pöle•m/ (room).

3.3.5 Deletion, Reduction

Different suffixes are connected with word stems in different ways. They can roughly be grouped into three categories.

1. Suffixes that delete unstressed final vowels.

One example of these is the possessive suffix of the first person singular, “-em”. When attached to a stem ending with a reduced vowel, this reduced vowel disappears completely.

/kü•zö/ (knife) > /küz**e•m**/ (my knife)

When attached to a stem ending with a stressed /a/, the /e/ in the suffix is dropped.

/ola•/ (city) > /ola•**m**/ (my city)

When attached to a stem ending with any other stressed vowel or with a consonant, the stem and suffix remain unaltered.

/izi•/ (little) > /izi**e•m**/ (my little one)
/pö•rt/ (house) > /pört**e•m**/ (my house)

2. Suffixes that reduce unstressed final vowels (with epenthesis)

An example of these is the accusative suffix “-m”. When attached to a stem ending with an unstressed vowel, this unstressed vowel becomes /ə/.

/kü•zö/ (knife) > /kü•z**əm**/ (my knife)

When attached to a stem ending with any stressed vowel, the stem and suffix remain unaltered.

/ola•/ (city) > /ola•**m**/ (my city)
/izi•/ (little) > /izi**e•m**/ (my little one)

Note that when connected to a word ending with a stressed /a/, such as /ola•/, the accusative suffix is identical to the possessive suffix of the first person singular.

When attached to a word ending with a consonant, epenthesis occurs – a vowel /ə/ appears between the stem and the suffix.

/pö•rt/ (house) > /pö•rt**əm**/ (my house)

3. Suffixes that reduce unstressed final vowels (without epenthesis)

The ending “-dəme”/“-dəmo”/“-dəmö”, which is roughly equivalent to the English “-less” suffix, belongs to this group. When attached to stems ending with vowels, the formations are the same as in the previous group.

/kü•zö/ (knife) > /kü•z**ədəmö**/ (without a knife)
/ola•/ (city) > /ola•**dəme**/ (without a city)
/lu•/ (bone) > /lu•**dəmo**/ (boneless)

When attached to a stem ending with a consonant, no epenthesis occurs – the suffix is connected directly to the stem.

/pö•rt/ (house) > /pö•rt**dəmö**/ (my house)

By way of comparison, the following figure shows how these three types of suffixes are attached to different types of stems:

Ending		**-em**	**-m**	**-dəm(e/o/ö)**	**Translation**
unstressed /e/	va•te	vate•m	vatə•m	va•tədəme	wife
stressed /e/	teŋge•	teŋge•m	teŋge•m	teŋge•dəme	rouble
unstressed /o/	šu•do	šude•m	šu•dəm	šu•dədəmo	grass
stressed /o/	depo•	depoe•m	depo•m	depo•dəmo	depot
stressed /a/	ola•	ola•m	ola•m	ola•dəme	city
unstressed /a/	fu•ga	fuge•m	fu•gəm	fu•gədəmo	fugue
stressed /u/	lu•	lue•m	lu•m	lu•dəmo	bone
consonant	lü•m	lüme•m	lü•məm	lü•mdəmö	name

Fig. 16: Suffix types

As highlighted in the figure, in some situations different suffixes can lead to the same results. This causes problems for the morphological analyzer, as when it encounters the form /ola•m/ it cannot know if this means “my city” or if it is the accusative of “city”.

Similarly, when it encounters the word /vate•m/ – “my wife” – and tries to extract the stem, the Morphology Analyzer has several seemingly valid interpretations to choose from, if it is not connected to a dictionary. Theoretically, /vate•m/ could be:

- /va•t/ + possessive suffix first person singular
- /va•te/ + possessive suffix first person singular
- /vate•/ + possessive suffix first person singular
- /vate•/ + accusative
- /va•to/ + possessive suffix first person singular
- /va•ta/ + possessive suffix first person singular

The third option offered here would be irregular; the last two options offered would have to be Russian loan words. Access to the dictionary would allow the Morphology Analyzer to determine that only the word /va•te/ actually exists in Mari and it would then be able to determine the correct interpretation.

3.3.6 Declension of Nominals

Mari has nine grammatical cases and six grammatical persons. Each grammatical person has its own possessive suffix, used to express ownership. Pronouns are used in this capacity in English. Unlike English, Mari also has multiple options to mark plurals. Aside from three roughly equivalent options for a “standard” plural, there is also a so-called sociative plural, used for people only.

Furthermore, there is the comparative degree's suffix, used like the suffix "-er" in English (tall → tall**er**). It is also used in other contexts in Mari and can be connected to other nominals, such as nouns.

In addition, there are also two so-called enclitics – grammatically independent suffixes that can be attached to the end of a word.

Mari nominals can be given the following suffixes:

1. one of four plural markers
2. one of eight case endings (excluding the nominative)
3. one of six possessive suffixes
4. the comparative marker
5. one of two enclitics

With certain limitations, any combination of these suffixes can be used – one can pick one – or no – suffix from each category and combine them at will. A complete overview of the suffixes can be found in the appendix.

pört	**-em**	**-vlak**	**-lan**	**-at**
house	1PSg.	PL	DAT	ENCL-and
	(3)	(1)	(2)	(5)

"to my houses, also"

saj	**-rak**	**-əm**	**-ak**
good	COMP	ACC	too
	(4)	(2)	(5)

"the better one, indeed"

In some rare cases it is also possible to add more than one case suffix to a word. For example, genitive forms of words ("my father's") can themselves be treated as fully functional nominals which can be declined. This is not done in English, but a somewhat similar situation occurs in German when

one uses a posessive pronoun, used to express ownership, as a noun, e.g. “Mein Erfolg hat Deinem geholfen.” Here, the word “Deinem” is in the dative case, but also serves the function of a genitive.

3.3.7 Verbal Conjugation

When one inflects a verb according to person, tense and grammatical mood, one speaks of finite verb forms. Gerunds, participles and infinitives make up a rather substantial group of non-finite verbs. These two categories will be handled separately.

3.3.7.1 Finite Verb Forms

Mari has six grammatical persons, seven grammatical tenses and three grammatical moods. The semantic differences between the six different past tenses does not need to be discussed.

The three moods that Mari has are the indicative mood (“I go.”), the imperative mood (“Go!”), and the desiderative mood (“I want to go.”). In contrast to the declension of nominals, verbal conjugation does not allow for all combinations of all options. For example, the imperative has no past tense.

Of the 21 potential combinations theoretically possible in this system, only 11 actually exist and only five of these are formed morphologically. All other forms are periphrastic – that is, they are formed using auxiliary verbs. In English, for example, the simple past is formed morphologically (“to plant” → “I planted”), the present perfect is formed periphrastically, using the auxiliary verb “to have” and the past participle (“to plant” → “I have planted”).

All negated verb forms are formed periphrastically, using the so-called negation verb, which itself has its own forms in all tenses and moods that are formed morphologically.

	Indicative	Imperative	Desiderative
Present	M	M	M
First Preterite	M	X	X
Second Preterite	M	X	X
First Periphrastic Imperfect	P	X	P
Second Periphrastic Imperfect	P	X	P
First Periphrastic Perfect	P	X	X
Second Periphrastic Perfect	P	X	X

Fig. 17: Finite verb forms

M denotes morphological forms, P denotes periphrastic forms, X denotes forms that do not exist. The periphrastic forms are not of interest to us in the morphological analysis of words. Nothing speaks against including the periphrastic forms in the morphological generator, as the generator can aid a learner here just as it can provide assistance in the learning of morphological forms.

Every finite verb form can receive one of the two enclitics discussed under nominal declension. Verbs can also be given the comparative degree marker in Mari.

3.3.7.2 Non-Finite Verb Forms

Mari has:

- Three infinitives
- Four participles
- Five gerunds
- Two nominalizations

Possessive suffixes can be added to all non-finite forms. The main infinitive, in some situations, can be put into the dative case. Such oddities will not be treated in detail here, but it was necessary to make sure that the software is at least passively capable of understanding such things.

Like the finite forms, non-finite forms can receive the comparative degree marker and one of the two enclitics.

The participles and nominalizations are nominal forms, and thus can be declined in the same manner as any other nominal. For purely pragmatic reasons, they will be grouped with derivations as far as the software is concerned.

3.3.8 Derivation

Derivation is the process of creating a fully functional new word from an existing word. For example, the English word “happiness” is a derivative of the word “happy”.

The line between derivation and inflection can be thin at times. For example, one could consider participles to be derived forms of verbs. They are fully functional nominals which, in many cases, have entries of their own in dictionaries. The past participle of the English word “to tire”, “tired”, is treated as an adjective in its own right in English dictionaries. For this

reason, we have grouped participles with derivations for the software, even if this is not linguistically clean.

As in English, one must differentiate between productive and non-productive derivational suffixes in Mari. A productive suffix is a suffix that can be used to spontaneously create new meaningful words. One reasonably productive suffix in English is “-able”, which can be added to any verb to create an adjective that will generally be understood, even if it cannot be found in a dictionary.

“to read”	>	“readable”
“to download”	>	“downloadable”
“to investigate”	>	“investigatable”

An example of a non-productive suffix is the suffix “-dom”, which is used to create nouns from adjectives or nouns. The English language has words derived from other words using this suffix, but attempts to create new words using this suffix would have comical results.

“king”	>	“kingdom”
“free”	>	“freedom”
“serf”	>	“serfdom”
“modem”	>	*”modemdom”
“internet”	>	*”internetdom”

Our morphological tools will completely disregard Mari’s many non-productive derivations. Mari words derived using non-productive suffixes must have dictionary entries of their own.

Derivational suffixes can create nominals from verbs, verbs from nominals, nominals from other nominals, verbs from other verbs and adverbs from either group of words. The resulting words are fully functional words and can be declined or conjugated freely. A complete overview of productive derivations can be found in the appendix.

3.3.9 Arrangement of Suffixes

Up to six flectional suffixes and any given number of derivational ones can be added to a nominal stem. The arrangement of these suffixes is not random. For example, the enclitic particles are always in the final position in a word.

On the other hand, the arrangement of suffixes is not completely rigid in Mari. Some grammatical case suffixes must come after the possessive suffixes (group 1), some must come before the case suffixes (group 2). And some allow for both possibilities (group 3). The placement of plural suffixes is labile too, allowing for several possibilities.

In nominal declension, the following arrangements are valid:

stem + [der] + [comp] + [gen] + [poss] + [plur] + [case-g1] + [enc]
stem + [der] + [comp] + [gen] + [poss] + [plur] + [case-g3] + [enc]
stem + [der] + [comp] + [gen] + [plur] + [poss] + [case-g1] + [enc]
stem + [der] + [comp] + [gen] + [plur] + [poss] + [case-g3] + [enc]
stem + [der] + [comp] + [gen] + [plur] + [case-g2] + [poss] + [enc]
stem + [der] + [comp] + [gen] + [plur] + [case-g3] + [poss] + [enc]

Every element in these rows, with the exception of the stem, is optional. If there is at least one derivational suffix, the stem does not have to be a nominal – it could also be a verb.

The Morphology Generator only needs to be taught one alternative, as any person learning Mari as a foreign language will not bother to learn all theoretically possible arrangements of suffixes, but will just learn the one used most commonly used. The Morphology Analyzer can take no such liberties – it must be familiar with all possible suffix arrangements.

Verbal conjugation is somewhat simpler, as there are no multiple arrangements. Time, mood and person are all marked with one suffix:

stem + [der] + time/mood/person + [comp] + [enc]

If a finite form is being constructed, time, mood and/or person must be marked. Again, if at least one derivation occurs, the stem can be either a verb stem or a noun stem.

Gerunds and infinitives can have possessive suffixes and, again, the comparative marker and an enclitic:

stem + inf/ger marker + [comp] + [poss] + [enc]

Postpositions can have possessive suffixes and enclitics:

stem + [poss] + [enc]

Adverbs can have a comparative marker and one of the two enclitics:

stem + [derr] + [comp] + [enc]

Words not included in any of these groups – such as conjunctions – can only have enclitics:

stem + [enc]

4. Constructing and Breaking Down Morphology

On the basis of the rudimentary knowledge of Mari grammar outlined in the previous chapter, I will now explain the software I have developed to construct and break down Mari words.

I decided to use Java for these tools. PHP is a fitting choice when a means of displaying dynamic content in a static manner is needed, which is why we chose PHP for our dictionary's software. For intensely interactive software, like the tools, PHP is less fitting.

4.1 Resources for a Kindred Language – Estonian

While major European languages are predominantly non-agglutinative, Mari is not the only European language of this type. Other Finno-Ugric languages such as Hungarian, Finnish and Estonian are classified as agglutinative as well. Non-related languages such as Turkish, Basque and Georgian are also agglutinative. It therefore made sense to explore resources available for such comparatively large languages on the Internet. As a tutor of the Estonian language, I chose Estonian as my example language here.

Estonian is not much larger than Mari with respect to the number of speakers – it is spoken by roughly a million people. However, its presence on the Internet is considerably greater than that of Mari. There are many reasons for this. Estonian is the sole national language of Estonia, an industrialized nation and EU Member State. Estonia has also been a pioneer in the use of the Internet by government authorities, which has earned it the nickname "e-stonia" in international publications (**Basu 2008**). It was the first country in the world to implement e-voting in 2005 and free wireless LAN is available throughout central areas of all major Estonian cities. Estonia's Internet-friendliness is plainly visible with respect to Wikipedia as well. Whereas Estonian is only the 243rd-largest language in the world, in a ranking of languages by their number of native speakers, the Estonian

Wikipedia, with its 63 769 edits (on 6 June 2009) is the 34th largest Wikipedia. This makes it considerably larger than, for example, the Greek Wikipedia, even though Greek is spoken by 15 times as many people. (**Wikipedia 2009**). One can also find relatively many free linguistic resources for Estonian on the Internet.

4.1.1 Eesti Keele Süntesaator

The Estonian Language Synthesizer (**Filisoft 2007 I**) is similar to the Mari Morphology Generator, which will be discussed later in this section.

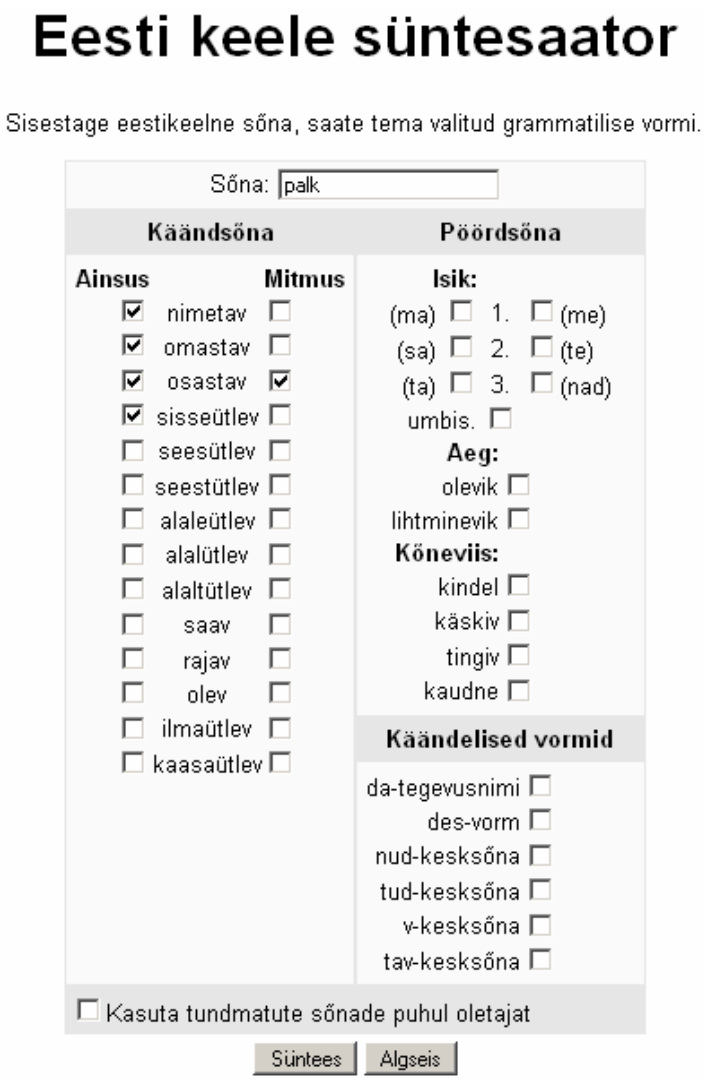

Fig. 18: Estonian Language Synthesizer

Unlike my Java-based application, this application uses a server-side application, making the application nothing more than a dynamic HTML page for users. This has some definite advantages, but also some clear disadvantages.

The synthesizer's control panel is split down the middle: On the left-hand side, one can pick and choose from the various forms a nominal ("Käändsõna") has and on the right-hand side, one can choose from the various forms a verb ("Pöördsõna") has. The verb side is split again, into finite verb forms, and non-finite forms ("Käändelised vormid").

For nominals, this application allows users to choose from a wide variety of forms, each of which has a checkbox. This application differs from my application in that it allows users to pick and choose from forms such as "inessive plural" ("seesütlev" in the "mitmus" column) and "genitive singular" ("omastav" in the "ainsus" column), whereas I offer users one set of radio buttons for each attribute a finished word should have. Whereas the Estonian application has 28 checkboxes (14 cases, in singular and plural each), my approach would have 14 radio buttons allowing users to choose the case, and 2 allowing them to pick the number, for the same application.

The approach chosen by the developers just barely works for Estonian, which is not quite as agglutinative as Mari is – it has fewer suffixes and suffix types than Mari has. The fact that it just barely works is demonstrated, for example, by the fact that the preparers had to disregard enclitic particles, which are found in Estonian, and derivations. In Mari, where one has to consider a number of additional suffix categories, it would not make any sense to offer users one checkbox for every combination of suffixes, as this would result in thousands of checkboxes.

The Estonian application also differs from mine in that it uses checkboxes whereas I use radio buttons. The Estonian application is capable of showing multiple grammatical forms of one word at the same time – if one clicks 5 checkboxes, the output will consist of five word forms.

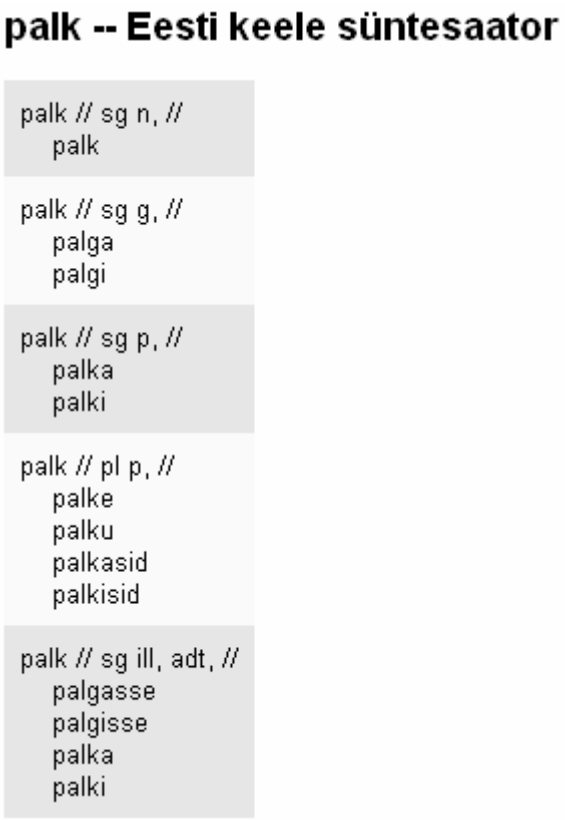

Fig. 19: Estonian Language Synthesizer – output

These results are displayed on a new page. The software gives all variants of a word form when there are several equivalent options – "palgasse" and "palga" are equivalents, for example.

It also shows inflected forms of words that are homonyms in the nominative singular case, but are not homonyms in all inflected forms. For example, the word chosen here – "palk" – has two meanings in the nominative singular. It can either be an original Estonian word meaning "salary", or it can be a loan word from German ("der Balken"), meaning girder. In the genitive singular, for example, these two words differ – the genitive singular of salary is "palga" and the genitive singular of the girder is "palgi". The inflected forms of both words are displayed.

This information is displayed in a disorganized manner, however. The software does not, for instance, clarify which word the form "palke" belongs to – palk meaning salary or palk meaning girder. Users must determine this

on their own. Fortunately, Mari does not have such annoyances – homonyms in the nominative singular remain homonyms in all word forms. Thus, it is not necessary to be concerned about such matters.

I can report from personal experience that users also have problems with the finite forms of verbs. One must click at least one checkbox each to choose the desired person ("Isik"), time ("Aeg") and grammatical mood ("Kõneviis"). If one does not do this, the software will generate an error message.

Fig. 20 “I am”

My approach using radio buttons should lead to fewer irritations, as radio buttons have a default initial state. For the generation of finite verb forms, by default my software will be set on first person singular, present and indicative. Users can alter their choices as desired from this point on, but they do not have worry about factors that do not concern them, such as grammatical mood, unless they are interested in forms other than the indicative forms.

A clear advantage the Estonian approach has over my approach is that it allows users to access multiple word forms at the same time. In return, my

software is faster for users interested in the same form of multiple words. Since I use Java for my applications, I am not constrained by the limits of an HTML page and can make the application fully interactive. Users are not forced to go back and forth between two HTML pages to enter queries and to view results.

4.1.2 Eesti Keele Lemmatiseerija

The Estonian Language Lemma Machine, created by the same people who designed the synthesizer discussed in the previous section, finds the canonical form or lemma (nominative or infinitive) of any word it is given, allowing users to check this word in a dictionary (**Filisoft 2007 II**).

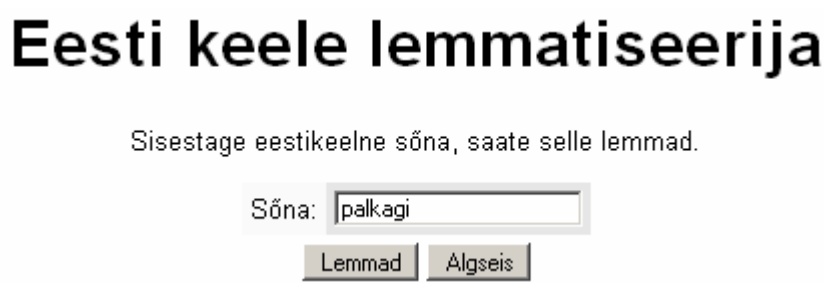

Fig. 21: The Estonian Language Lemma Machine

Unlike the synthesizer, the Lemma Machine can handle enclitics. Like my Morphology Analyzer, it can display the various stems from which an inflected word form could be derived.

Fig. 22: Output of the Lemma Machine

It does not, however, give any information on how the input was derived from one of the possible lemmas. It also does not in any way deal with the problem of homonyms. As has been discussed, the Estonian word “palk”, in

the nominative, can mean both “salary” and “girder”. The inflected form “palgagi”, however, can only be derived from the stem meaning “salary”. While the tool does tell us that the inflected word's stem is “palk”, it does not tell us which “palk”.

Neither of Filisoft's applications offer any type of translations to languages other than Estonian.

4.1.3 Õigekeelsussõnaraamat

The Estonian dictionary “Õigekeelsussõnaraamat”, which is the highest authority on Estonian vocabulary and in many ways serves the same function for Estonian that the “Duden” dictionary does for German, is available online, as are many other essential Estonian dictionaries (**Keelevara 2006**). Just as any quality dictionary on English includes information on irregular verbs, this dictionary provides information on the inflection of the words it includes.

If you search for the word used in the previous sections – “palk” – you will find both words, with definitions in Estonian and some key grammatical forms. Anyone who understands Estonian will be able to tell the two words apart here. Students of Estonian will have more problems.

If you want to get more information on the inflection of a word and related words, you can click on a little number beside the dictionary entry (22 and 20 in the example below).

Fig. 23: Online Estonian dictionary

These explanations, however, are complex and in Estonian. They are not useful to students of Estonian who are not yet at an advanced level.

4.1.4 General Observations

All of the resources just discussed are quite useful to students of Estonian. It is obvious, however, that they were not primarily designed with foreign students of Estonian in mind, but rather were designed by Estonians for Estonians – no native speaker of Estonian will have any problems with the shortcomings I have pointed out. While Estonian might be a "bigger" language than Mari, it is not big enough for its speakers to be accustomed to foreigners attempting to learn their language.

These resources have served as an inspiration, but I have not attempted to replicate them or to use them as strict templates for my tools. I have,

however, borne in mind those aspects of these applications that students of Estonian in Vienna have had trouble with and have attempted to find improvements in the tools I have created for Mari.

4.2 The Morphology Generator

Generating Mari morphology is by far easier than breaking it down. Ambiguity can be disregarded here – even where alternative word forms exist, it is not necessary to take these into consideration. It is enough to make the software offer the stylistically optimal variant and to disregard all others.

The first step in the design of the Morphology Generator was to create individual functions capable of connecting various types of suffixes (see Section 3.3.5) to stems. All irregularities discussed in the previous chapter had to be taken into consideration. The software also had to take into account factors not marked orthographically, but relevant morphologically (see Section 3.2.2.2). The same symbols used for the dictionary are also used here; palatalization is marked with an apostrophe, where necessary, and word stress is indicated by a big dot.

Once functions capable of adding all relevant suffixes to a stem were created and tested, the next task was to enable users to pick and choose from these suffixes.

I decided to depart from designs used in previous, similar projects (4.1) and to use radio buttons where possible (see screen shots in the figures below). When suffixes are only loosely connected to each other, checkboxes are used. Suffixes falling into the same grammatical category are grouped together – one block has grammatical cases, one column has possessive suffixes, et cetera.

Note that every set of choices has a “standard” choice activated by default – for a nominal, this will be the nominative case and for a verb, this will be the present tense. If one enters a nominal and clicks the inflection button (>), the software displays an unaltered form – transcribed into the Cyrillic alphabet, if it was entered in the Latin alphabet.

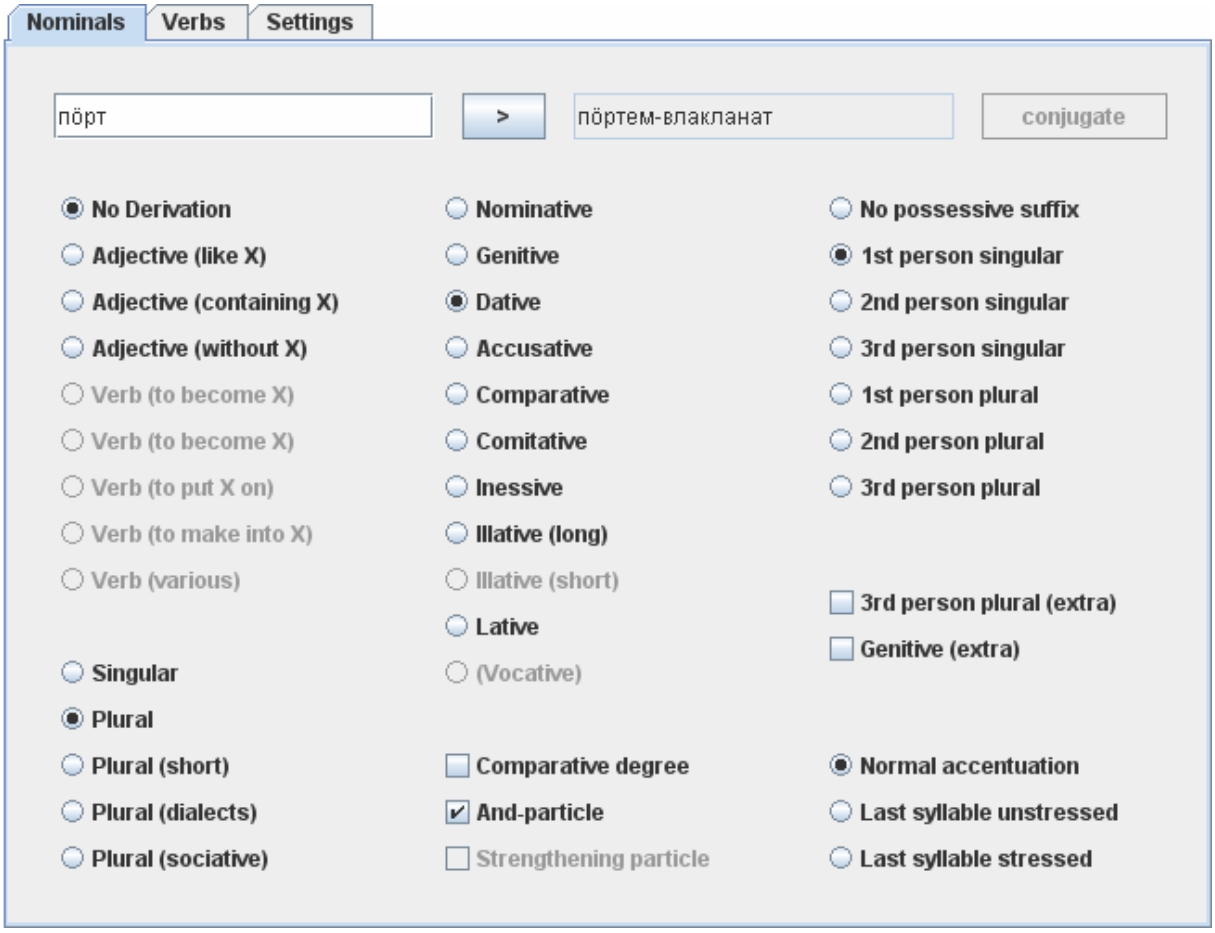

Fig. 24: The Morphology Generator – Nominals

As users should only be able to choose from grammatically valid choices, I used the Java options to grey out invalid possibilities. If a user picks an option that makes other choices invalid, these will become disabled until the original choice is changed. For example, if a user has picked a grammatical case – such as the dative case, as in Fig. 24 – derivations resulting in verbs will be disabled and cannot be selected. This happens because these options would conflict with each other, as verbs cannot be put into grammatical cases.

In the lower right corner, the user can specify where the word stress lies, should he/she be dealing with an irregular word. Palatalization must be entered manually, using a special character; this is clarified in the instructions found on the applet's web page. Should one be inflecting a word taken from the dictionary, this information will already be known.

Also note the button "Conjugate", which is disabled here. If the derivation results in a verb, one can push this button and thereby be redirected to the "Verbs" tab, where the verb can then be conjugated.

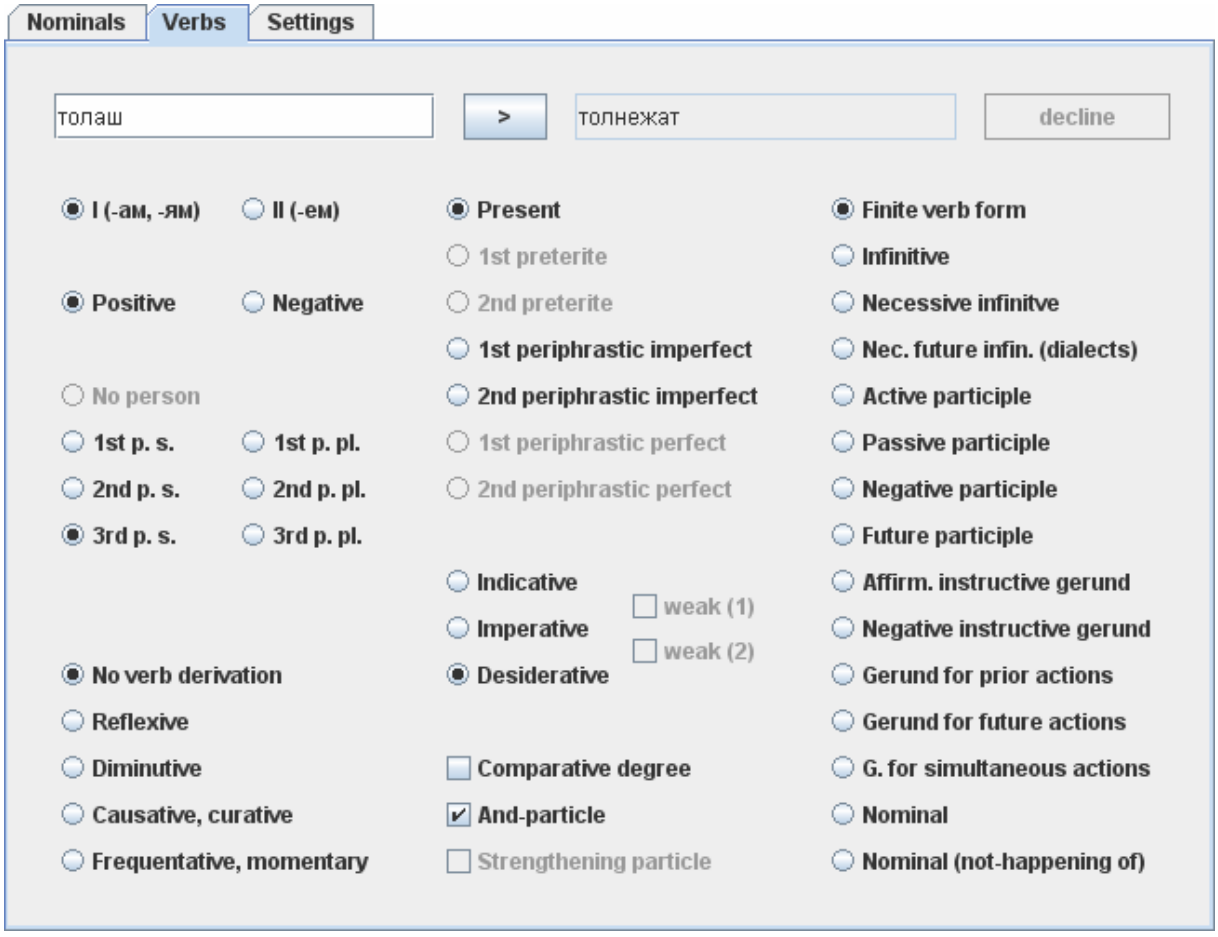

Fig. 25 The Morphology Generator – Verbs

The same process is possible in reverse as well. If a nominal is derived from a verb, one can click on "Decline" to be redirected to the nominal tab.

The settings tab can be ignored by users who do not have any special desires, in accordance with the instructions found directly under the applet

on the web page. The settings tab allows users to change the transcription schemes used by the applet for text entered in the Latin alphabet. The sound /š/, for example, is generally written as “sh” in English, “sch” in German, “s” in Hungarian and “š” in Finnish and Estonian. Transcription systems optimized for speakers of specific languages or users of specific keyboard layouts can be selected on this tab. Tables on these transcriptions can be found on the website as well, under “HELP!”.

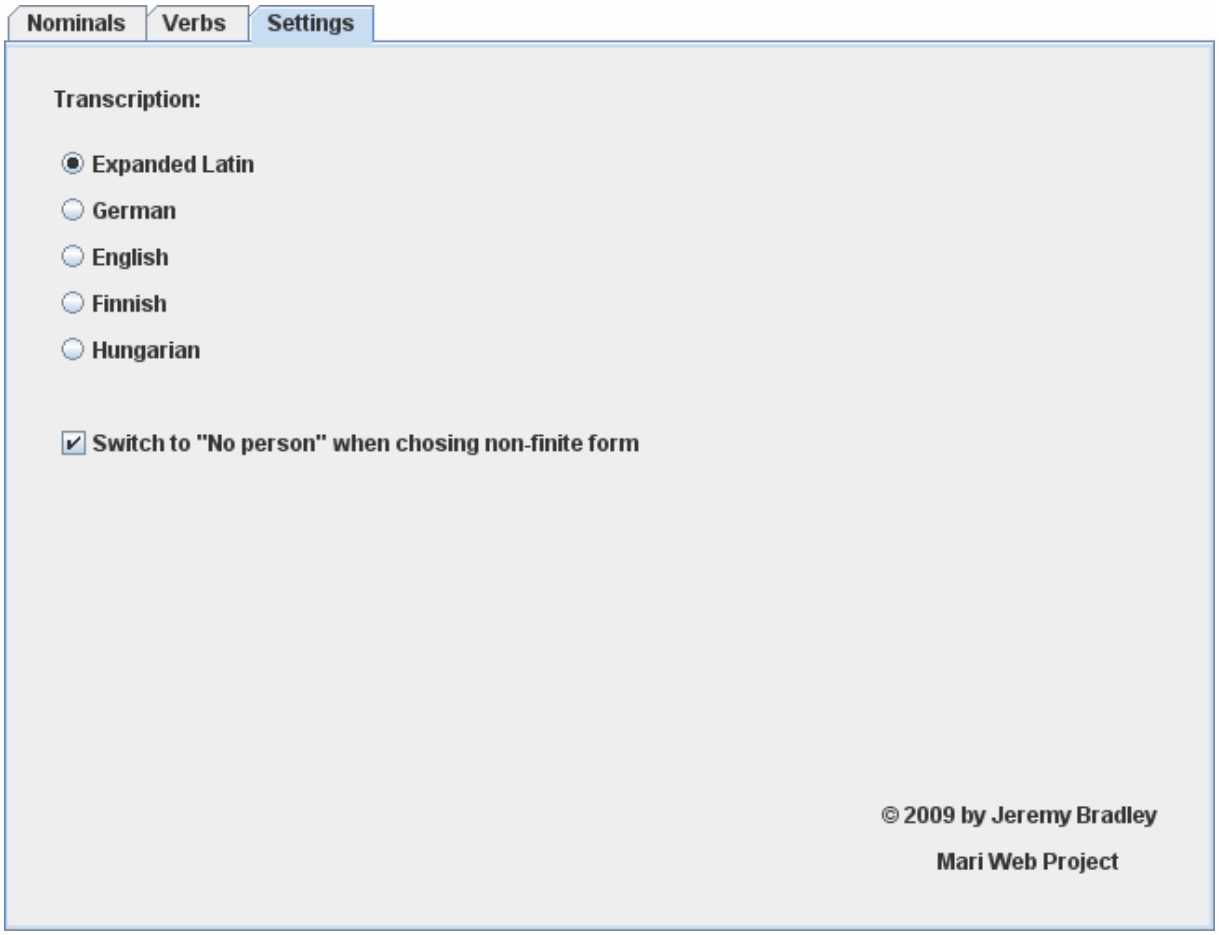

Fig. 26 The Morphology Generator – Settings

4.3 The Morphology Analyzer

This section is several times as difficult and long, as the previous one. When breaking a word down, users have no knowledge of what they are looking at or for. They cannot disregard any unusual suffix arrangements. It is possible that a stem change has occurred, but it cannot be assumed that this has

happened. It is also possible that the word one is attempting to find among all these suffixes is a Russian loan word and thus does not comply with the rules of Mari orthography.

To best illustrate the deductive process the software must go through, it is easiest to simply show its modus operandi in action.

4.3.1 An Example

Let us revisit the inflected Mari word /pörtemvlaklanat/, which we have already seen a number of times. Four suffixes have been attached to the word stem and the Analyzer must recognize every single one of them:

pört	**-em**	**-vlak**	**-lan**	**-at**
house	1PSg.	PL	DAT	ENCL-and

"to my houses, also"

Of course, the information displayed here will not already be known to the Analyzer – figuring this out is its task. As discussed in the previous section, in certain situations different suffixes can look identical. The suffix –at, here an enclitic particle, is also the ending of the second person indicative for first conjugation verbs – /tolaš/ (to come) → /tolat/ (you come). As no verb /pörtemvlaklanaš/ exists, it certainly cannot be such a suffix in the present situation. Our software has no way of knowing this, however, and must thus check this possibility out. It must also check to see if a verb /pörtemvlaklanaš/ could be a derived verb of some sort. Here it must be borne in mind that only after the removal of four suffixes can the correct dictionary form of the word be found. The software must examine every possibility, even if this means that it will go down many dead ends.

In order to provide an overview of all the options the Morphology Analyzer still has to check out, a queue lists all the options that must be considered. Every element in this queue must be looked up in the dictionary; the

software must attempt to detach more suffixes from it. As the analysis process proceeds, every new possible interpretation of a word is added to the end of the queue. Once a word has been looked up in the dictionary and has been checked for all possible suffixes, it will be taken out of the queue.

Section 3.3.9 discussed the arrangement of suffixes in Mari. All 12 possible arrangements of suffixes are listed in Appendix A.3. When the analysis of a word is begun, it is not known whether it is a verb, a noun or some other type of word. Thus, none of the potential arrangements can be disregarded from the beginning. A list of all suffixes can be found in Appendix A.1 and derivations are in Appendix A.2.

When the analysis of the word in question begins, the queue has only one entry.

```
01 /pörtemvlaklanat/  (?)
```

We have no information on the word's classification at this point. The Analyzer will check the dictionary for /pörtemvlaklanat/ and will find no entries. We must then assume that there is some internal structure that must be analyzed. It should be noted that even if the word had been found in the dictionary, further analysis would have been necessary – see 4.3.2 below.

When looking for the first suffix, the Analyzer has no context whatsoever. With the exception of the endings of finite verbs, all suffixes in Mari are optional. This means that the final suffix in a word could be a case ending, a possessive suffix, a plural marker, an enclitic particle, et cetera. All possibilities must be considered. The Analyzer thus break downs the word in several ways, with each possibility being added at the end of the queue:

~~01 /pörtemvlaklanat/~~	~~(?)~~	
02 /pörtemvlaklanaš/	(verb conj. 1)	+ personal ending 2. person singular indicative
03 /pörtemvlaklanataš/	(verb conj. 1)	+ personal ending 2. person singular imperative
04 /pörtemvlaklana•/	(nominal)	+ possessive suffix 2. person singular
05 /pörtemvlaklana/	(?)	+ enc. particle – “and”
06 /pörtemvlaklane/	(?)	+ enc. particle – “and”
07 /pörtemvlaklano/	(?)	+ enc. particle – “and”
08 /pörtemvlaklanö/	(?)	+ enc. particle – “and”
09 /pörtemvlaklan/	(?)	+ enc. particle – “and”

Note that the first entry of the queue, the starting point, has now been discarded: It has been checked in the dictionary, where no entry per se was found, and all possible interpretations have been listed. Nothing remains to be done with the original entry.

Also note that entries 02 through 08 are meaningless from a linguistic point of view. But, at this point, for the Analyzer nothing distinguishes any one of these entries from the correct interpretation, 09.

The Analyzer then continues to work through the list. Entries 02 and 03 can be discarded fairly quickly, as the hypothetical verbs cannot be found in the dictionary. It is still possible that we are dealing with a derived verb, but the software fails to find derivational suffixes matching the endings we have here. These lines are discarded completely.

Entry 04 can also be discarded relatively fast. /pörtemvlaklana•/ could hypothetically be the third person singular of the conjugation 2 verb /pörtemvlaklanaš/. However, the assumption entry 04 is based on – that the /t/ we took off the end was a possessive suffix – presupposes that we are dealing with a nominal – thus not a verb. na could also be the possessive suffix of the first person plural. However, since we have already had a possessive suffix in this interpretation, this cannot be true either.

Entries 05 and 06 can broken down even further – incorrectly, but again, the software cannot know this in advance. These interpretations are added to

the end of the queue as entries 10 and 11. Our correct entry, 09, can also be interpreted in several ways, all but one of which will soon prove to be erroneous.

...

~~09 /pörtemvlaklan/~~	~~(?)~~	~~+ enc. Particle – “and”~~
10 /pörtemvlaklanaš/	(verb conj. 2)	+ personal ending 3. person singular indicative + enc. Particle – “and”
11 /pörtemvlaklanaš/	(verb conj. 2)	+ personal ending 2. person singular imperative + enc. Particle – “and”
12 /pörtemvlak/	(nominal)	+ dative + enc. Particle – “and”
13 /pörtemvlakl/	(nominal)	+ derr. Suffix “with” + enc. Particle – “and”
14 /pörtemvlakla/	(nominal)	+ derr. Suffix “with” + enc. Particle – “and”
15 /pörtemvlakle/	(nominal)	+ derr. Suffix “with” + enc. Particle – “and”
16 /pörtemvlaklo/	(nominal)	+ derr. Suffix “with” + enc. Particle – “and”
17 /pörtemvlaklö/	(nominal)	+ derr. Suffix “with” + enc. Particle – “and”

Entries 13-17 quickly prove to be dead ends, when the hypothetical words cannot be found in the dictionary and we can find no other derivational suffixes allowing us to look deeper into these words. The only option remaining is 12.

When analyzing entry 12 further, there is some morphological context. When we started analyzing our word, we did not know what kind of word we were analyzing. As we now know that our word includes a dative suffix, we know that we are analyzing a nominal (or a nominal derivation of a verb). Only these two derivational patterns remain potentially valid:

stem + [der] + [comp] + [gen] + [plur] + [poss] + [p3] + [**case-g3**] + [**enc**]
stem + [der] + [comp] + [gen] + [poss] + [p3] + [plur] + [**case-g3**] + [**enc**]

The elements in bold are those that have already been identified for the interpration in question. Only elements to the left of these are still to be taken into consideration.

Thus, when attempting to analyze /pörtemvlak/, we can disregard the possibility of any further enclitics, since these can only appear at the end of a word. Even if the enclitic particle -ak exists in Mari, we know that this cannot be what we are seeing at the end of the form we are currently analyzing.

~~17 /pörtemvlaklö/~~	~~(nominal)~~	~~+ derr. Suffix "with"~~ ~~+ enc. Particle – "and"~~
18 /pörtem/	(nominal)	+ plural + dative + enc. Particle – "and"

The Analyzer is now on the right track. It will manage to unambiguously identify the suffix found at the end of /pörtem/ as the possessive suffix of the first person singular, but will know of several theoretical forms from which it could be derived. It will only be able to find one of these in the dictionary.

~~18 /pörtem/~~	~~(nominal)~~	~~+ plural~~ ~~+ dative~~ ~~+ enc. Particle – "and"~~
19 /pörta/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – "and"
20 /pörte/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – "and"
21 /pörto/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – "and"
22 /pörtö/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – "and"
23 /pört/	**(nominal)**	**+ possessive suffix first person singular** **+ plural** **+ dative** **+ enc. Particle – "and"**

At long last, one of our possibilities is found in the dictionary. After the software has determined that /pört/ itself cannot be derived from anything else, the analysis of this word is complete. The stem of our word means house and, with its suffixes, the word /pörtemvlaklanat/ means “to my houses, also”.

Section 4.3.3 will cover further means of identifying and discarding some implausible forms in advance. Taking these into consideration, a tree graph of our software’s deductive process would look like this:

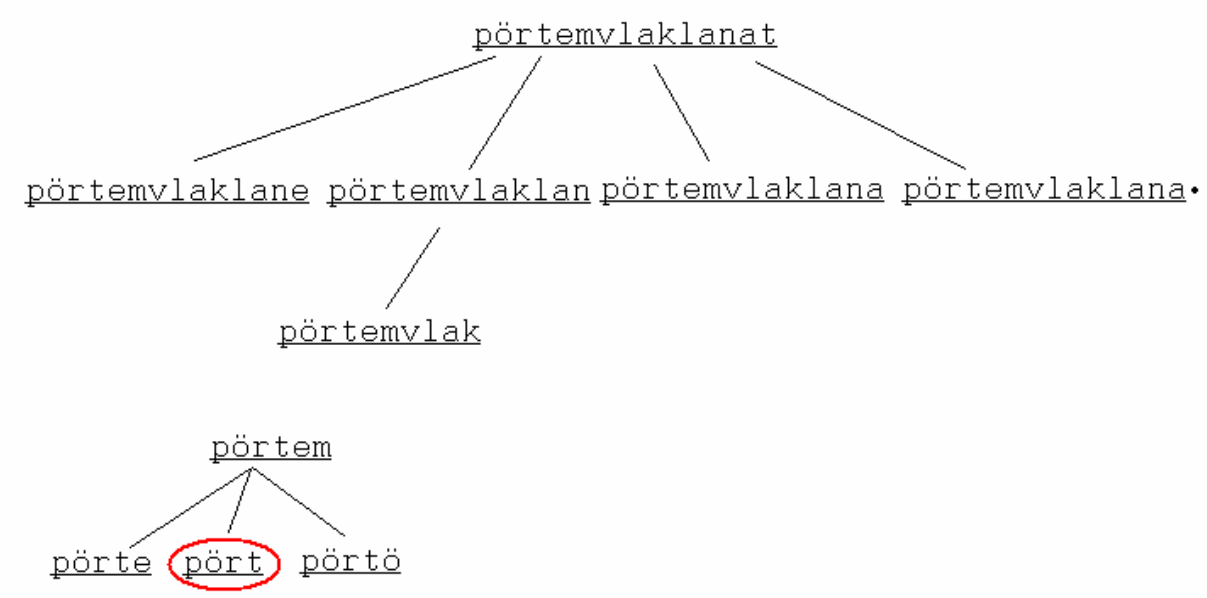

Fig. 27: Tree graph of word analysis

4.3.2 Ambiguous Forms

The Analyer cannot simply quit once one valid interpretation of a word is found – one word can have multiple interpretations. These can come from the same entry, from related entries or from completely independent entries.

As an example of two interpretations coming from the same entry, take the form /olam/ of the word /ola/ (city). This can either be the accusative of the

word or the word with the possessive suffix of the first person singular. Both interpretations are equally valid and must be presented.

As an example of ambiguity stemming from multiple entries, let's look at two Mari words – /urem/, meaning "street"; and /ur/, meaning "squirrel". If the word /ur/ is given the possessive suffix of the first person singular, it becomes /urem/ and is thus identical in appearance to the word meaning street. Again, both derivations must be presented.

Not all ambiguity encountered will be morphological. Like English, Mari has a certain degree of lexical ambiguity – different words can be completely identical in appearance. In English, one example of lexical ambiguity is the word "bark", which can either be the outer covering of a tree or a sound made by a dog. In Mari, one example would be the word /lu/, meaning both "ten" and "bone". When there are several entries in our dictionary on a word that is derived, all of these must be listed.

4.3.3 Cutting Down on SQL Queries

Avoiding unnecessary strain on the SQL server would certainly be a bonus. In the example above, before we reached our word's final form we discarded 22 bogus possibilities, all of which had to be checked with the database. If it was possible to avoid checking forms that are obviously not going to be dictionary entries in their own right, we might be able to reduce this number.

One easy approach here is to make no SQL queries as long as we know we are dealing with a plural form. A quick search in the world list we are using to compile our dictionary revealed that the combination of letters /vlak/ never appears in the Mari language in any capacity other than as the plural marker. The same can be said of the alternative plural marker /šaməč/. As all dictionary entries are in the singular, it is not necessary to search any entry still known to contain such a suffix.

As these plural markers remain unaltered under all circumstances, it is very easy to search for them. When one is found, this allows us to instantly classify a given word as a nominal and thus to rule out all verbal interpretations in the deductive process. We can also disregard derivational suffixes as long as we know that we have still got a plural suffix to extract – derivational suffixes are attached to the stem directly and cannot appear after any other markers.

This move alone would cut our SQL queries down to 6. In this case, one could reduce the count even further by disregarding the entries /pörta/ and /pörto/. The final /a/ would have to be unstressed – which violates Mari accentuation rules. As has been stated, words violating these accentuation rules exist in Mari, but they are exclusively Russian words. As /ö/ does not exist in Russian, we know that we are not dealing with a Russian word. /pörto/ would violate vowel harmony – which again could only happen in a Russian word.

This brings the number of SQL queries down to four. Our queue, greatly simplified at this point, would only have 10 entries. For the first 6 entries, the SQL database would not be used at all.

01 /pörtemvlaklanat/	*(?)*	
02 /pörtemvlaklana•/	*(nominal)*	*+ possessive suffix 2. person singular*
03 /pörtemvlaklana/	*(?)*	*+ enc. particle – “and”*
04 /pörtemvlaklane/	*(?)*	*+ enc. particle – “and”*
05 /pörtemvlaklan/	*(?)*	*+ enc. particle – “and”*
06 /pörtemvlak/	*(nominal)*	*+ dative* *+ enc. Particle – “and”*
07 /pörtem/	(nominal)	+ plural + dative + enc. Particle – “and”
08 /pörte/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – “and”
09 /pörtö/	(nominal)	+ possessive suffix first person singular + plural + dative + enc. Particle – “and”
10 /pört/	**(nominal)**	**+ possessive suffix first person singular** **+ plural** **+ dative** **+ enc. Particle – “and”**

4.3.4 Testing

The Analyzer worked quite well for our example word /pörtemvlaklanat/. However, as /pört/ is a regular Mari word, not subject to any stem changes, not containing any violations of vowel harmony or word stressing, it made sense to test the Analyzer on more irregular words – words subject to stem changes, Russian loan words not conforming to Mari pronunciation rules and words containing unmarked palatalized consonants.

Thus, the first testing step was to attempt to analyze inflected forms of words which might be considered difficult. The following words were chosen for the tests:

- /šüľö/ (шӱльӧ)
- /kočmaš/ (кочмаш)
- /teŋge•/ (теӈге)
- /imńe/ (имне)

- /tu•ndra/ (тундра)
- /ko•fe/ (кофе)
- /glasnosť/ (гласность)

After some fine-tuning, the Analyzer managed to correctly interpret all of these words.

A more extensive test run, testing both our dictionary and the Morphology Analyzer, will be discussed in Section 5.6.

5. Application of the Tools Created

While the previous chapters discussed the basic tools designed for the Mari Web Platform, this chapter will discuss how they will be implemented in reality.

5.1 Presentation of Information

Lexicographical work inevitably involves the managing of large amounts of data. For a computerized dictionary, it is especially important to adhere to a consistent data structure, as irregular organisation would make it very difficult to enter data into a database. Every entry and subentry must be in exactly the same form. The structure we have chosen is as follows (optional fields are in brackets):

```
HEADWORD–TYPE–ENTRY–WORD CLASS–[NOTES]–TRANSLATION–[LATIN]–[PARENT]–[REFERENCES]
```

- HEADWORD is the lexeme with which the entry is associated.
- TYPE is a number denoting whether the entry is a main entry (0), a subentry or an example sentence. Subentries and example sentences may be associated with one of the meanings of the main entry.
- ENTRY is the actual Mari word, phrase or example sentence.
- WORD CLASS denotes whether the word in question is a noun, verb, adjective, postposition, etc.
- [NOTES], when provided, give information on contexts in which the word is used, etc.
- TRANSLATION is straightforward. Should an entry have several related meanings, these will be numbered.
- [LATIN] translations will be provided for biological terms.

- [PARENT] refers to the word from which an entry is derived. For example, the parent of “teacher” is “to teach”, the parent of “kingdom” is “king”, etc.
- [REFERENCES], when provided, will direct users either to words with the same meaning that are considered to be better style for some reason or to a relevant appendix. The appendices will cover various topics, including grammatical tables for both Mari and English, lists of cardinal and ordinal numbers, geographical names, etc.

In order to stick to such a rigid scheme, we will use Microsoft Excel for our actual lexicographical work. The following example, which presents a simplified entry for the Mari word “ышташ (-ем)” (to do; to build; to work as), illustrates the structure of our Excel template.

HW	T	ENTRY	WC	[N]	TRANSLATION	[L]	[P]	[R]
ышташ (-ем)	0	ышта•ш (-е•м)	vb2	tr	[1] to do; [2] to build; [3] <DAT> to work as			
ышташ (-ем)	1	паша•м ышта•ш (-е•м)	vb2		to do work			
ышташ (-ем)	2	пӧ•ртым ышта•ш (-е•м)	vb2		to build a house			
ышташ (-ем)	3	врачла•н ышта•ш (-е•м)	vb2		to work as a doctor			
ышташ (-ем)	4	ыште•н лукта•ш (-а•м)	vb1		to manufacture			

Fig. 28: Sample entry as seen by the project team

In this example, the optional fields have mostly been left empty. For example, Latin translations are not necessary here, as the English meanings of the word are self evident. Also, no parent is cited, as “to do” is not derived from any other word.

One can also see how we use a set of shorthand abbreviations in our work. For example, the abbreviation “tr” denotes a transitive verb and <DAT> signifies that this specific usage of the word requires the use of the dative case.

Neither the tabular organization of data nor such abbreviations would be optimal for those who want to use our dictionary. This is not necessary, however, as the software takes the relevant data out of the SQL database and dynamically creates a profile in a format that is more pleasing to the eye.

For every subentry, the software will find the corresponding meaning under which it should appear. Related subentries that are not assigned to a specific meaning in the word’s translation will be listed separately after all other meanings and their subentries.

ышт**а**ш (-**е**м)

verb – conjugation 2, transitive

1. to do

    ♦ паш**а**м ышт**а**ш (-**е**м) *(verb conj. 2)* to do work

2. to build

    ♦ п**ӧ**ртым ышт**а**ш (-**е**м) *(verb conj. 2)* to build a house

3. [+ DATIVE CASE] to work as

    ♦ врачл**а**н ышт**а**ш (-**е**м) *(verb conj. 2)* to work as a doctor

* *

| ♦ ышт**е**н лукт**а**ш (-**а**м) *(verb conj. 1)* to manufacture

Fig. 29: The same entry, as seen by users

In addition to all the data found in an entry itself, the software will also display a list of all the terms derived from a word. When looking at a Mari entry meaning "good", a table on the profile will also offer links to words meaning "well", "goodness" and "not good".

A similar but more concise layout will be used for the printed dictionary. The online dictionary can be used in two primary ways – one can browse it or search for specific entries.

When users access the dictionary, they are presented with a clickable Mari alphabet and a search field. If they click on any of the letters, they will be given an alphabetical list of all words starting with this letter. Note in Fig. 30 that some letters are greyed out. This is because they are either not allowed at the beginning of a word (like the German "ch" sound is not allowed at the beginning of a word) or because they have not yet been uploaded onto the web page.

Fig. 30: The dictionary's main page

Using the search field will usually be quicker, especially as students of Mari, while capable of using the Cyrillic alphabet, might not yet have a good feel for the correct alphabetical order. Also, the search field offers a range of additional options:

- One can use it "in reverse", that is, one can search all occurrences of an English word and in this way use the dictionary as a makeshift English-Mari dictionary.
- One can search subentries.
- One can search for partial matches.
- One can use wildcards (see Fig. 31).
- One can let the software compensate for recent changes in Mari orthography.

Note that the user does not have to enter Mari entries in the Cyrillic alphabet. The dictionary software is capable of transcribing search entries into the Mari version of the Cyrillic alphabet.

The following example shows the results a user would get when searching for "_rach", where "_" is the wildcard character meaning "any one character".

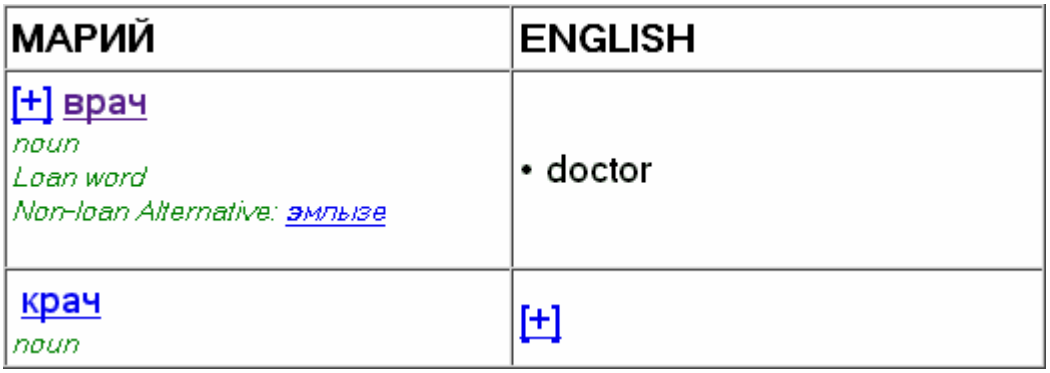

Fig. 31: Search results for "_rach"

The two words found by the dictionary and displayed in this manner are /vrač/ – a Russian loan word meaning "doctor" – and /krač/ – a lexeme for which we have yet to enter a translation into our database.

By clicking on the respective entries, users can access a full profile of the word /vrač/. By clicking on the [+] symbol to the left of the word, they can

display all the example sentences and subentries on the word /vrač/ in the database. Additional information on the word itself is given under the word: We are told that it is a noun and a loan word, and that there is an alternative to the word – /emləze/ – which is not a loan word. The latter would be of interest to people who want to avoid overusing Russian loans where they are not necessary. Should a user click on this link, the software will search for this word. Similar links are offered to parents of derived words, alternatives, etc.

Should one search for words in variants of Mari other than the dominant Meadow Mari variant, links are offered to the equivalent Meadow Mari words.

Hill Mari, Dialect	Meadow Mari
ырды	рӱдӧ

Fig. 32: Searching for Hill Mari words

5.2 Textbook Exercises

Mari morphology can be somewhat overwhelming for students at first. As a result, the Mari textbook now being prepared in our project does not confront students with all 9 grammatical cases, all possessive suffixes and all derivational suffixes at once. Every chapter introduces new grammatical concepts and students only become familiar with all the suffixes found in Mari when they have worked through the entire book.

As a result, the Mari Morphology Generator discussed in Section 4.2 will be overkill for students of Mari at first, exposing them to suffixes and grammatical concepts they are not yet aware of.

To avoid causing too much initial panic, I have made it possible to flexibly "slim down" the Morphology Generator so that it only offers those options a student will be aware of at any given time. A slimmed-down version of the

generator will be offered to students after each chapter of the textbook, allowing them to "test" new grammatical concepts they have learned, in combination with all the morphology they are already supposed to know.

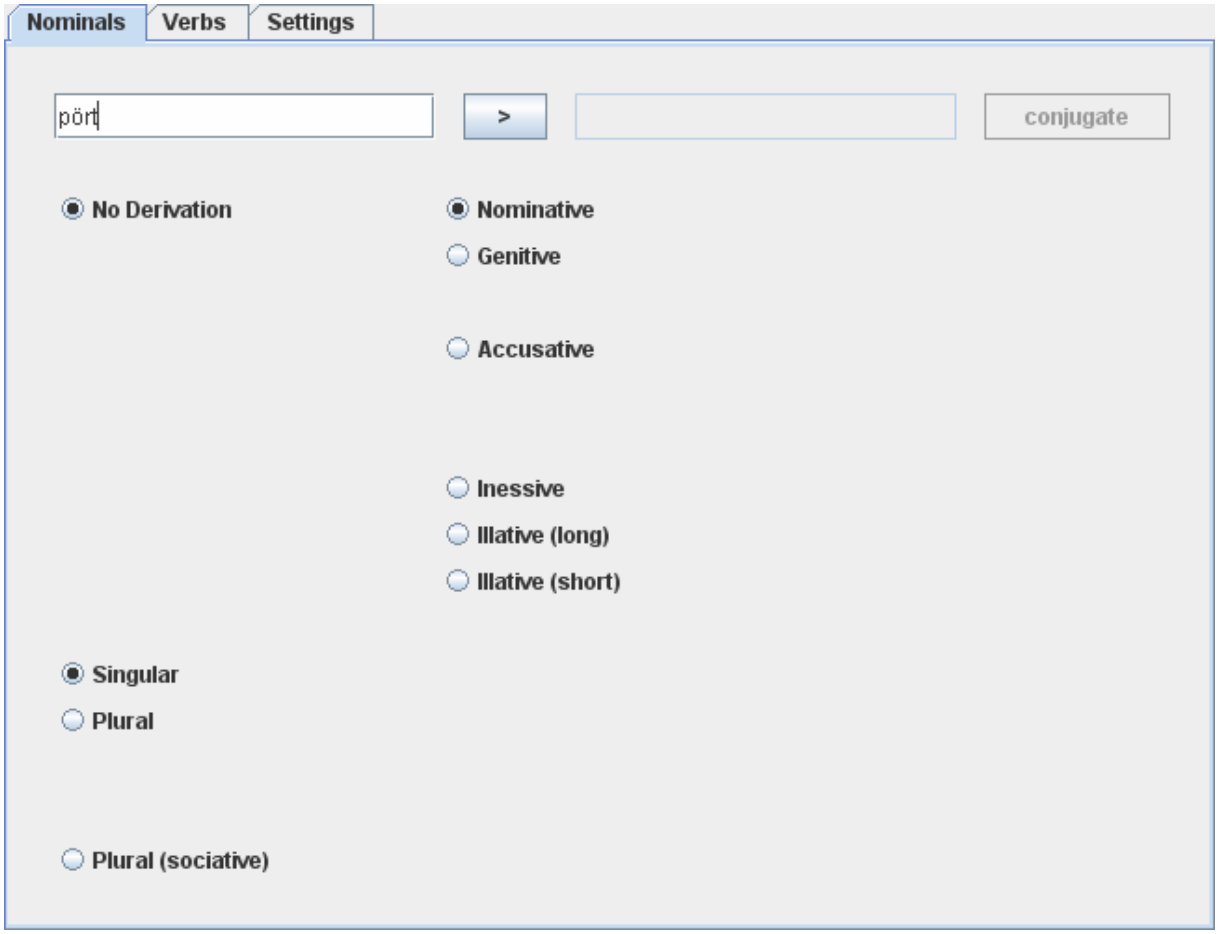

Fig. 33: Customized grammatical generator for absolute beginners

5.3 Assisted Reading

One rather exciting application of the Morphology Analyzer presented in section 4.3 is the creation of a tool allowing so-called "assisted reading". Similar software exists for non-agglutinative languages such as English, but is considerably less elaborate. Various software dictionaries can be run in the background and allow little pop-ups to appear when the user hovers over words for some time.

For Mari, there will be too much information to comfortably hover in pop-ups in such a manner.

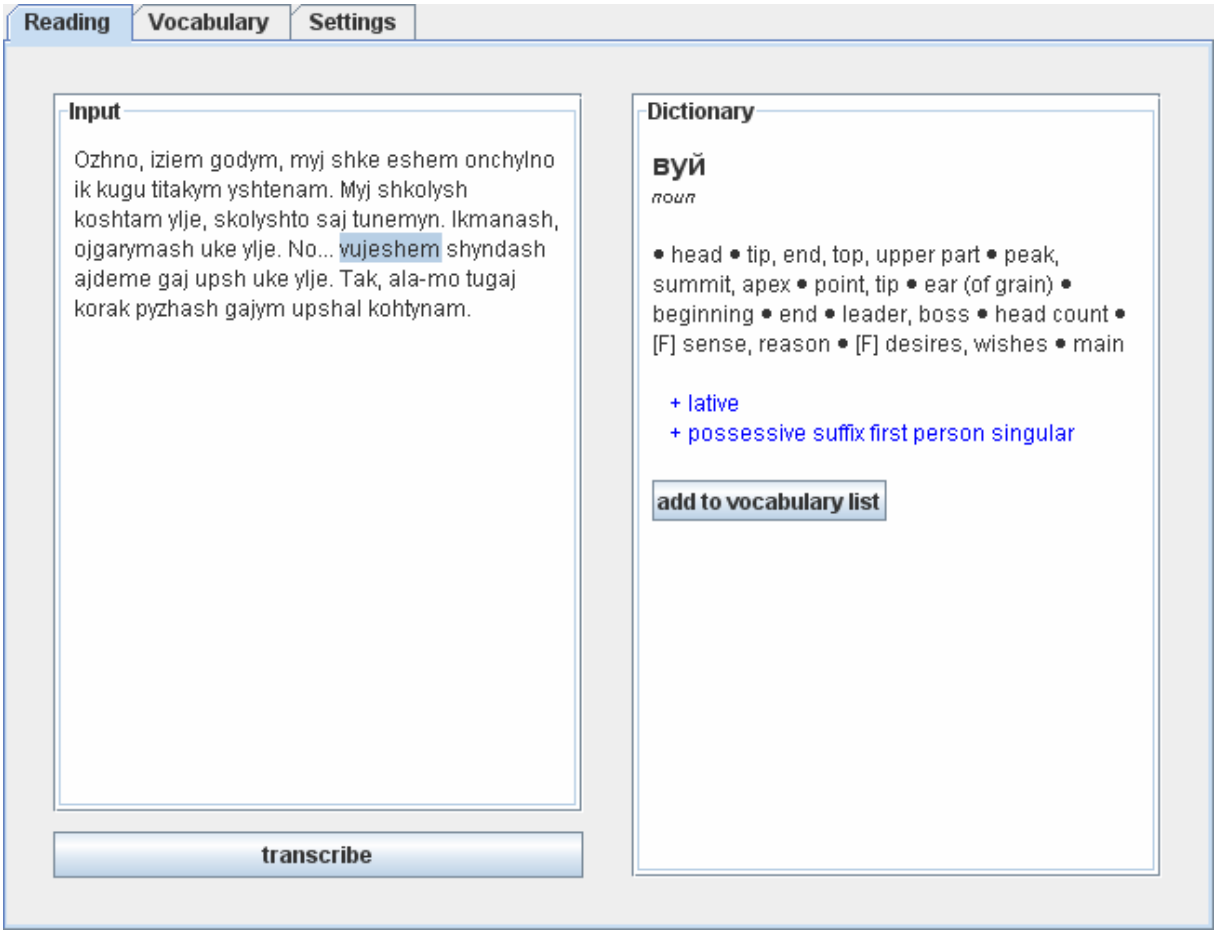

Fig. 34: Mari Reading Aid

The Reading Aid is a Java-based application that accepts input in both Latin and the Cyrillic alphabets. A "transcribe" button allows users to convert Latin input into Cyrillic. This is not necessary for the software to function; it only serves cosmetic purposes.

The left half of the screen is the so-called reading field, the right half the dictionary field. A user can click on any word on the left-hand side and the program will display on the right all legitimate interpretations of the word selected. These interpretations might be derived from the same stem or from different stems.

Each derivation includes a list of suffixes that have been attached to the stem to create the final form. Every suffix in the list will, eventually, link to the respective chapter of Timothy Riese's upcoming Mari textbook, allowing users to access more information on a suffix that has been attached to a stem, should they not be familiar with all the suffixes added.

It should be noted that the Reading Aid is blind to context. Some of the interpretations it finds for words may seem comical. It is up to users to decide which interpretation seems most valid – common sense will be necessary and the context will have to be taken into consideration.

While one can click on single words to get further information, it is possible to activate groups of words, to check whether they are included in the dictionary as subentries. For example, the Mari word for question mark is /jodəš pale/, where /jodəš/ means question, and /pale/ means symbol. Should this word be inflected, all suffixes will be attached to the second word, /pale/. Users can activate phrases like /jodəš palədəme/, meaning "without a question mark", by dragging the mouse over it. Should any entry in the dictionary have /jodəš pale/ as a subentry, this will be displayed, with the same lists of suffixes that are displayed when users activate single words by clicking on them.

5.4 Personalized Vocabulary Sheets

Every interpretation of a word found in a text comes with an "add to vocabulary list" button. Should the user click this, the word, in its canonical form, will be copied to the "vocabulary sheet", along with its English translation and the sentence in which it occurred. Using these buttons, users can easily create lists of all the new words they have encountered in a text.

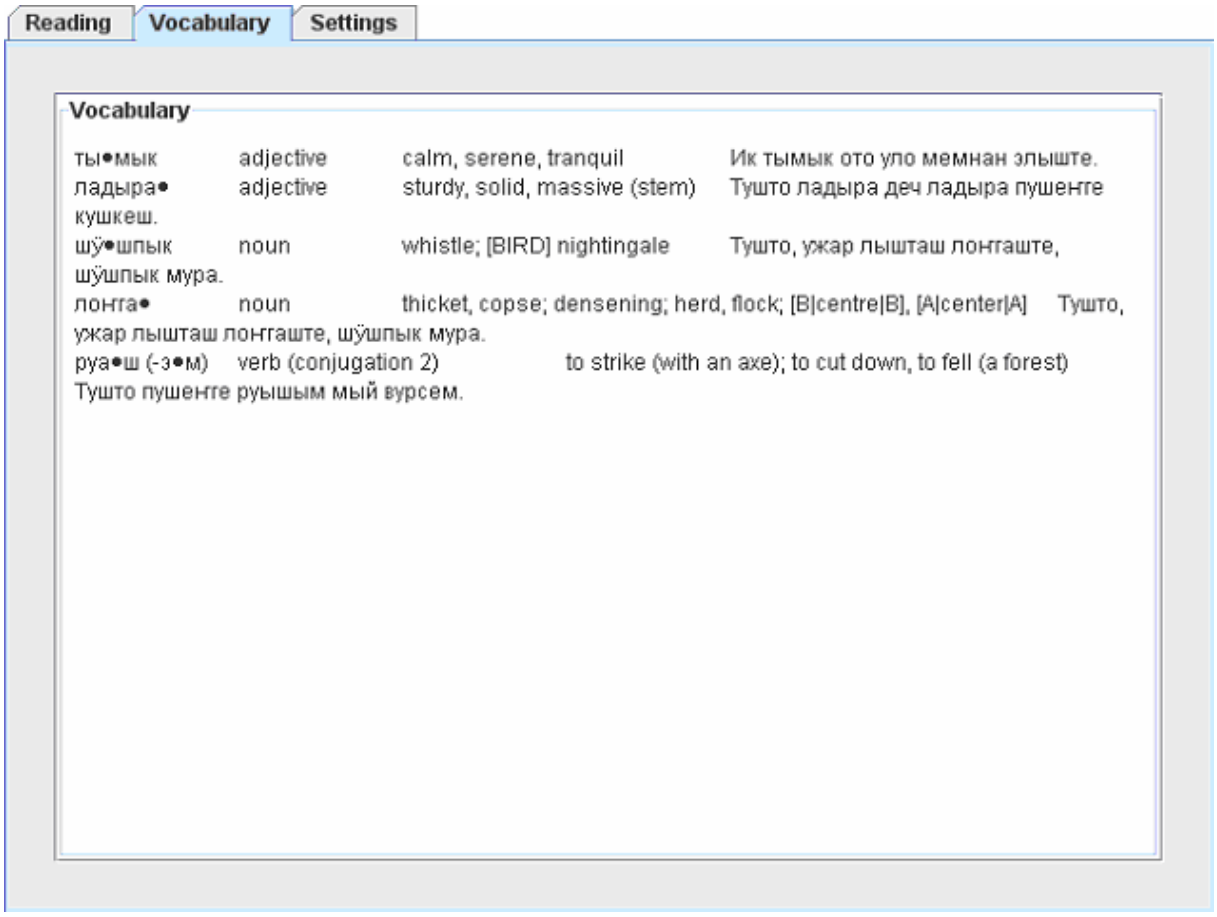

Fig. 35: A personalized vocabulary list

At this time, I have not yet created a special "testing" application allowing users to test themselves on words encountered. The personalized vocabulary list is only displayed in plain text. When users have finished reading the text, they can copy their vocabulary lists out of the application and paste them in spreadsheet applications or text processing applications. They can keep these lists as lists of words they still need to learn. Users can also add English translations of the sentences the respective words were found in, if they want to have more context.

5.5 Spelling Checker

Another possibility offered by the Morphological Analyzer is that of creating a spelling checker. Conventional spelling checkers are difficult to implement for intensely agglutinative languages. Whereas a spelling checker for

English, using only a list of legitimate English words, would need to know only two forms for each noun (singular and plural), the same approach for Mari would require lists of millions of word forms.

The Morphology Analyzer can, however, analyze entire texts, searching them for words that cannot be derived from any known stems in the dictionary and mark these non-derivable words in a certain colour.

We do not expect this application to be very popular. In order to gain any popularity, a Mari spelling checker would have to be integrated into Microsoft Word. If the possibility ever presents itself to create a spelling checker for Microsoft Word using our dictionary's word corpus, we would be very happy to do this, but applying to Microsoft to carry out such a task is not on our short-term agenda.

5.6 Scholarly Uses

This thesis focuses on applications of our set of tools for students of Mari. Our tools offer possibilities for scholars researching certain aspects as well. The Morphology Analyzer can be used to analyze syntactic structures of Mari sentences, to search for certain grammatical constructions, to create statistics on the usage of certain gerunds, etc.

One possibility we certainly intend to use in Vienna is a variation on the theme presented in the previous section, where we used the Morphology Analyzer as a spelling checker, finding words that cannot be derived from any stems in the dictionary.

If we have a reliable corpus that we do not expect to contain many spelling mistakes, we could use the same approach to detect gaps in our dictionary (or, deficiencies in the Morphology Analyzer). We plan to provide this "detector of unknown words" with many modern Mari texts, taken from contemporary publications. The software will read through these texts and

copy out all the words it does not know, along with their location in the text and the sentence in which they occur. In the following example, a quote from the Maris' national poem was interlaced with an English sentence, "These are no Mari words at all."

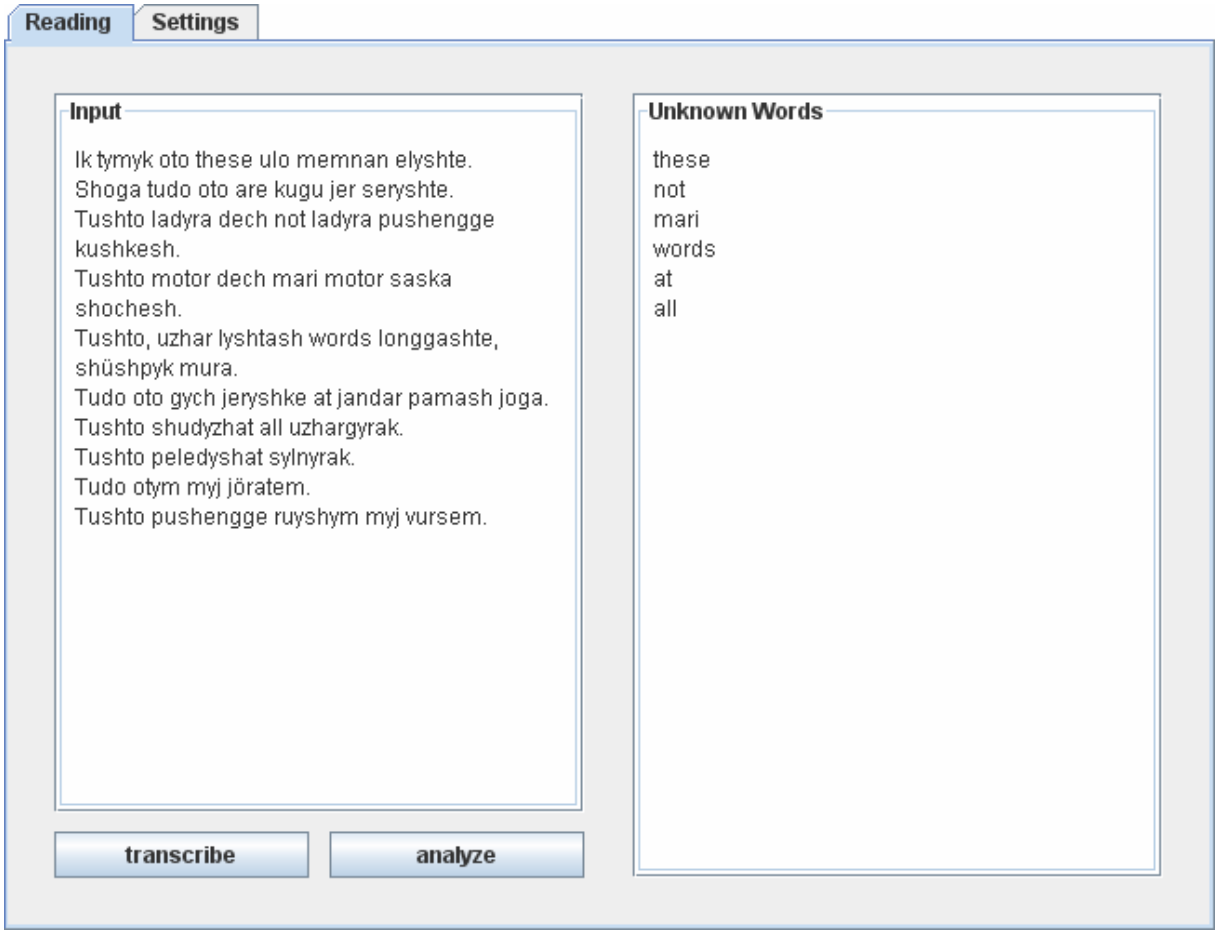

Fig. 36: Detector of unknown words

The only English word the software did not catch is "are", as this happens to be a legitimate Mari word. Other than that, the software correctly identified all non-Mari words and did not incorrectly reject any of the Mari words, which is encouraging both in regard to the Morphology Analyzer's capabilities and the scope of our database.

We will then be left with the task out of determining what these words actually mean – whether they are in fact "real" words, whether they need to be included in the dictionary, et cetera.

Not all words found by these means will be new words for our dictionary. The tool will find proper names it does not recognize, Russian phrases quoted in Mari texts, onomatopoeic expressions and, of course, any typos that might have crept into any of the materials. But it certainly will narrow things down for us in our search for new lexemes to add to our database.

6. Usability

Questions of usability are essential in the design of any piece of software, be it a website, a computer game or a text editor. For any software product, it is desirable for the actual functionalities of the product to match users' expectations to the greatest degree possible, given technological limitations.

This chapter will seek to analyse the expectations users might have of our software and will demonstrate in what ways guidelines for user interface design were taken into consideration in the creation of our tools.

6.1 User Interface Design

6.1.1 Paradigms of User Interface Design

Variations on the same paradigms of interface design are discussed in several definitive books concerned with usability engineering (**Dix et al. 2006, Nielsen 2000, Norman 2006, Shneiderman et al. 2005, Tognazzini 1996**). This section will elaborate on how these principles were taken into consideration in the development of our software.

6.1.1.1 Designers ≠ Users

A recurring theme in the literature of usability is the rift between those who design software and those who use it. Designers naturally have more experience with computers than potential users, and as a result things that may seem obvious to them might be difficult to understand or confusing to those who do not share their background. In general, people are prone to take their own knowledge for granted. In particular, when creating an interface software designers have a hard time remembering that they have computer proficiencies that others lack.

It can also happen that designers are inadequately familiar with the tasks users of the software intend to execute and that their ideas about the product may diverge from those of those who are actually planning to use the software. It is less difficult to avoid problems like this, as it is easier for people to acquire new information than it is for them to disregard knowledge they already have.

We were able to bypass the cliché of interface designers who are completely out of touch with what users want, as our interfaces were designed by someone who actually intended to use the software in question himself – in our case, the designer was a user. Nonetheless, our software was designed primarily for people with different educational and cultural backgrounds who could not have been expected to have detailed knowledge of computers. Fortunately, the software designer was the only member of our project team who had a background in computer science, which meant that other team members and various acquaintances interested in our software could offer opinions from different points of view.

6.1.1.2 Expectations and Affordances

The noted cognitive scientist Donald Norman regards an object's affordances to be the interactions it suggests to its users. Through their physical appearance, objects imply functionalities and put ideas into people's minds regarding how they should be used. This principle applies to physical objects as well as to computer interfaces. As an example from the physical world, if one gives a small child a toy hammer, there is no need to explain the purpose of this object – its physical appearance implies that it can be used to hit things. The shape and size of the handle, which match the child's hand, make it clear that this part of the hammer is supposed to be gripped. An example in software design is the manner in which buttons in user interfaces are modelled in a mock-3D manner, making it look as though they can be physically pushed.

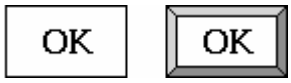

Fig. 37: Flat and sculptured buttons

If one fails to take affordances into consideration when designing user interfaces, one runs the risk of creating incorrect affordances that can lead to incorrect actions. The photograph below depicts lift controls in a Viennese metro station, where a status lamp indicating whether the lift is in operation or not is designed in the same way as the button used to call the lift. The wear and fingerprints on the status lamp indicate that many people have pushed this status lamp, expecting some sort of functionality.

Fig. 38: Bad affordances in the Viennese metro

A number of things influence our expectations of the functionality of objects and interface elements. In addition to psychological factors, cultural factors are also relevant. In Western culture, for example, the colour pink arbitrarily represents femininity and blue arbitrarily represents masculinity. Such cultural conventions can differ greatly, even among the users of a small web application.

Moreover, our experiences with other objects or interfaces that are in some way similar to the object or interface currently being handled will influence our expectations regarding the functionality of the latter. People who have spent a great deal of time working on a PC will expect the key combination Ctrl-Z to mean “undo” in any application they use. People who are accustomed to the QWERTY keyboard layout will have trouble typing on a French AZERTY layout.

It is not always easy to draw a sharp line between these factors. For example, the cultural convention that red stands for “hot” and blue represents “cold” is less arbitrary than the assignment of colours to genders, as the association of fire with red and water with blue are global constants. A psychological factor is certainly involved here. Likewise, specific interfaces can become so ubiquitous in a society that the conventions related to them can be considered a cultural factor.

Disregarding cultural factors in interface design can have dire consequences. The American custom pictured here of labelling hot water taps with the letter H and cold water taps with the letter C is not very intuitive for native speakers of the United States’ second-most commonly spoken language, Spanish, as the Spanish word for “hot” is “**c**aliente”. Anecdotes about Hispanic plumbers putting the C on the hot water tap are manifold in some parts of the United States.

Fig. 39: Cold or caliente?

Because our software should be equally usable by people in the West and people who live in rural areas of the Russian Federation, it was essential for diverging expectations based on the cultural backgrounds of potential users to be taken into consideration. As was discussed in Section 2.2.1, it was impossible to design one Mari keyboard layout that would equally satisfy everyone no matter what their native language is and/or what keyboard layout they are accustomed to. One based on the Russian layout, which has no similarities to the QWERTY layouts or related designs used for languages written in the Latin alphabet, would be very difficult for people to learn who are used to working with American, German or Finnish keyboard layouts. A QWERTY-based layout, on the other hand, would be rejected within Russia. A similar situation exists for Russian itself: QWERTY-based keyboard layouts do in fact exist for people who live outside of Russia and do not want to learn to type from scratch in order to be able to type in Russian.

For the design of the Morphology Generator (4.2), there were hardly any existing applications to take into consideration when designing the interface. While similar applications do exist, e.g. for Estonian, only very few users of our software will be familiar with these. It was, however, important to maintain consistency with existing grammar books (**Alhoniemi 1985, Bereczki 2004**). Russian linguistic traditions differ from Western Finno-Ugric ones. As a result, certain elements of the language are described in a slightly different manner in Russian textbooks than they are in the Finnish

and Hungarian books upon which we based our interface. However, these differences are not dramatic enough to have justified the creation of a second interface, which would not only have involved a lot of work but would also have been quite confusing to users.

For the design of the dictionary (2.2.4), we took designs from several notable Western and Russian dictionaries into account. We primarily aimed, however, to achieve consistency with printed Mari dictionaries, except in cases where we felt improvements were necessary.

In the design of our Reading Aid (5.3), there is a certain amount of consistency with Google Translate. Although our software's functionality differs greatly from that of this well-known online translation service, the existing consistency is nonetheless a positive factor, as Google Translate is a linguistic tool that people interested in this kind of software are quite likely to be familiar with.

It was fortunate for us that the Cyrillic alphabet, like the Latin alphabet, is written from left to right and from top to bottom. As a result, we were able to design our software under the assumption that users will process the interface starting at the top left. Input fields could be consistently placed at the left or top and output fields at the right or bottom, depending on what made more sense in a given situation. If we had been designing software for a language using the Arabic or Hebrew alphabet, it would not have been possible to make this assumption. For example, the Israeli version of Google Translate has all of its interface elements switched around – the input field is on the right and the output field is on the left.

6.1.2 Measures of Usability

Just as books on usability contain many guidelines for the design of user interfaces, they also present factors by means of which the quality of a user interface can be evaluated. This section will briefly summarize our efforts in this area. As we do not have a sufficient number of people to gather statistics on how well our interfaces score in regard to the individual factors

and no software designer can fairly assess his or her own software, no actual evaluations will be given.

6.1.2.1 Learnability

The faster that users become proficient in using an interface, the better. The consistency with other software interfaces and printed materials discussed in the previous section should be an asset with respect to our software's learnability.

6.1.2.2 Efficiency

This factor is rather straightforward. Our Reading Aid should offer some great improvements regarding efficiency compared to traditional pen-and-paper methods using printed dictionaries and grammar tables. It is a prerequisite here, however, that the text in question exists in a digital form and does not have to be typed in its entirety by the user.

It was possible to further increase our dictionary's efficiency by creating a plugin for Mozilla Firefox and possibly other browsers that allows users to directly search for words in our dictionary via their browser's search bar, without having to find a specific bookmark or having to navigate through the web platform.

6.1.2.3 Memorablity

Casual users of software should not forget how it functions if they are away from it for an extended period of time. As our interfaces use no key combinations or other complexities of this nature, we hope that we have successfully satisfied this criterion.

6.1.2.4 Errors

Users should naturally make as few errors as possible, but software should be forgiving when errors do occur and these should be easily reversible.

Contemplations regarding this factor led us to create a dynamic radio-button based interface for our Morphology Generator instead a static interface using check boxes such as that of the Estonian Language Synthesizer discussed in Section 4.1.1. The dynamic interface allows us to prevent users from choosing illegitimate grammatical combinations, such as a past-tense form of the imperative. Choices contradicting an option already selected are disabled by the software. Thanks to radio buttons, it is also not possible for users to make no choice at all or to give the computer insufficient information when making a request in our software. This can happen quite easily with the Estonian software, where queries submitted without any check boxes ticked will return empty result sets.

We made our transcription systems as flexible as possible. Mari is transcribed differently by speakers of various languages using the Latin alphabet. All of our applications attempt to understand as many transcriptions as possible. For example, the Mari word чын (/čən/) can be entered as “chyn”, “tshyn”, “tšyn”, “čyn”, “tšõn”, etc. There are, however, limitations on how many transcription systems can be accommodated by one program, due to conflicts existing between different transcription systems. The letter combination “ch”, for example, represents /č/ in English transcriptions of Cyrillic texts, but /x/ in German transcriptions. In such situations, the software has to make a choice, because both transcription systems cannot be accommodated at the same time.

We hope we have made our Reading Aid software error-safe by allowing users to either click or activate individual words in order to check them in the dictionary. The result is the same in both situations.

6.2 Conclusions

As of November 2009, when this thesis was completed, our project was still at a very early stage. It will be some time before we will be able to publish our software for a wider audience.

The practical use made of our software in the small community of Mari enthusiasts at the University of Vienna has been encouraging with respect to the prospects of the approach we have chosen. The raw version of the dictionary available to students at the Department of Finno-Ugric languages has been of great assistance to us all, in particular to those who are not adequately familiar with Russian or Finnish or who lack access to the printed dictionaries. Fellow students have managed to handle the Reading Aid quite well, as long as they were given a digital copy of the text they wanted to use it on.

A challenge for the future is to seek to get other people interested in the Mari language – whether native speakers or foreign scholars – to adopt the digital approach we are developing and using. In spite of the obvious advantages, we expect to encounter some scepticism. Our current monopoly on online linguistic resources for the Mari language should be a valuable asset to us in this effort.

A Tables

A.1 Suffixes

Suffixes are split into several types, which were discussed in section 3.3.5. Here, we will use the following abbreviations:

U Suffixes that leave stem unaltered
D Suffixes that delete unstressed final vowels.
E Suffixes that reduce unstressed final vowels (with epenthesis)
R Suffixes that reduce unstressed final vowels (without epenthesis)

Suffixes attached to verb stems will be split into suffixes attached to the infinitive stem (INF) and verbs attached to the imperative stem (IMP). See 3.3.3.1 for the difference between these two stems. Note that these stems only differ for some First Conjugation verbs.

For details on the meaning of all of these suffixes, see (**Riese et al. 2010**).

A.1.1 Enclitics, etc.

The suffixes listed here appear both in nominal declension and in verbal conjugation.

Name	Suffix	Suffix (LAT)	Suffix Type	Meaning
Comparative degree	-рак	-rak	R	“more”
“and”-Enclitic	-ат, -ят	-at	D	“and”
Strengthening Enclitic	-ак, -як	-ak	D	“especially”

A.1.2 Nominal Declension

A.1.2.1 Case Suffixes

Case name	Suffix	Suffix (LAT)	Suffix Type	Meaning
Nominative	-	-	-	-
Genitive	-н	-n	E	Possession
Dative	-лан	-lan	R	Indirect object
Accusative	-м	-m	E	Direct object
Comparative	-ла	-la	R	“like”
Comitative	-ге	-ge	R	“with”
Inessive	-шт(е/о/ӧ)	-št(e/o/ö)	E	“in”
Illative	-шк(е/о/ӧ), -ш	-šk(e/o/ö), -š	E	“into”
Lative	-еш, -ш	-eš, -š	D	“into, onto”

Fig. 40: Case suffixes

A.1.2.2 Number Markers

Name	Suffix	Suffix (LAT)	Suffix Type
Singular	-	-	-
Plural	-влак	-vlak	U
Short Plural	-ла	-la	R
Alternate Plural	-шамыч	-šaməč	U
Sociative Plural	-мыт	-mət	R

Fig. 41: Plural suffixes

A.1.2.3 Possessive Suffixes

Name	Suffix	Suffix (LAT)	Suffix Type
1.P.Sg.	-ем, -эм	-em	D
2.P.Sg.	-ем, -эт	-et	D
3.P.Sg.	-ж(е/о/ӧ), -ш(е/о/ӧ)	-ž(e/o/ö), -š(e/o/ö)	R, E
1.P.Pl.	-на	-na	R
2.P.Pl.	-да	-da	R
3.P.Pl.	-шт	-št	E

Fig. 42: Possessive suffixes

A.1.3 Verbal Conjugation

A.1.3.1 Finite Verb Forms

A.1.3.1.1 Indicative Present

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-ам, -ям	-am	INF
2.P.Sg.	-ат, -ят	-at	INF
3.P.Sg.	-еш, -эш	-eš	INF
1.P.Pl.	-ына	-əna	INF
2.P.Pl.	-ыда	-əda	INF
3.P.Pl.	-ыт	-ət	INF

Fig. 43: Indicative Present – First Conjugation

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-ем, -эм	-em	INF
2.P.Sg.	-ет, -эт	-et	INF
3.P.Sg.	-а, -я	-a	INF
1.P.Pl.	-ена, -эна	-ena	INF
2.P.Pl.	-ыда, -эда	-eda	INF
3.P.Pl.	-ат, -ят	-at	INF

Fig. 44: Indicative Present – Second Conjugation

A.1.3.1.2 Indicative First Preterite

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-(')ым	-(')əm	INF
2.P.Sg.	-(')ыч	-(')əč	INF
3.P.Sg.	-(')(e/o/ö)	-(')(e/o/ö)	INF
1.P.Pl.	-на	-na	IMP
2.P.Pl.	-да	-da	IMP
3.P.Pl.	-(')ыч	-(')əč	INF

Fig. 45: Indicative First Preterite – First Conjugation

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-ышым	-əšəm	INF
2.P.Sg.	-ышыч	-əšəč	INF
3.P.Sg.	-ыш	-əš	INF
1.P.Pl.	-ышна	-əšna	INF
2.P.Pl.	-ышда	-əšda	INF
3.P.Pl.	-ышт	-əšt	INF

Fig. 46: Indicative First Preterite – Second Conjugation

A.1.3.1.3 Indicative Second Preterite

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-ынам	-ənam	INF
2.P.Sg.	-ынат	-ənat	INF
3.P.Sg.	-ын	-ən	INF
1.P.Pl.	-ынна	-ənna	INF
2.P.Pl.	-ында	-ənda	INF
3.P.Pl.	-ыныт	-ənət	INF

Fig. 47: Indicative Second Preterite – First Conjugation

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-енам, -энам	-enam	INF
2.P.Sg.	-енат, -энат	-enat	INF
3.P.Sg.	-ен, -эн	-en	INF
1.P.Pl.	-енна, -энна	-enna	INF
2.P.Pl.	-енда, -энда	-enda	INF
3.P.Pl.	-еныт, -эныт	-enət	INF

Fig. 48: Indicative Second Preterite – Second Conjugation

A.1.3.1.4 Imperative

Name	Suffix	Suffix (LAT)	Stem Type
2.P.Sg.	-	-	IMP
3.P.Sg.	-ж(e/o/ö), -ш(e/o/ö)	-ž(e/o/ö), -š(e/o/ö)	IMP
2.P.Pl.	-за, -са	-za, -sa	IMP
3.P.Pl.	-ышт	-əšt	INF

Fig. 49: Imperative – First Conjugation

Name	Suffix	Suffix (LAT)	Stem Type
2.P.Sg.	-(e/o/ö)	-(e/o/ö)	INF
3.P.Sg.	-ыж(e/o/ö)	-əž(e/o/ö)	INF
2.P.Pl.	-ыза	-əza	INF
3.P.Pl.	-ышт	-əšt	INF

Fig. 50: Imperative – Second Conjugation

A.1.3.1.5 Desiderative

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-нем	-nem	IMP
2.P.Sg.	-нет	-net	IMP
3.P.Sg.	-неж(e)	-než(e)	IMP
1.P.Pl.	-нена	-nena	IMP
2.P.Pl.	-неда	-neda	IMP
3.P.Pl.	-нешт	-nešt	IMP

Fig. 51: Desiderative – First Conjugation

Name	Suffix	Suffix (LAT)	Stem Type
1.P.Sg.	-ынем	-ənem	INF
2.P.Sg.	-ынет	-ənet	INF
3.P.Sg.	-ынеж(e)	-ənež(e)	INF
1.P.Pl.	-ынена	-ənena	INF
2.P.Pl.	-ынеда	-əneda	INF
3.P.Pl.	-ынешт	-ənešt	INF

Fig. 52: Desiderative – Second Conjugation

A.1.3.2 Non-Finite Verb Forms

Participles have been grouped together with nominal derivations for the software. While this is linguistically unclean, it makes sense on a pragmatic level, as participles follow the same formation rules as derivations do.

A.1.3.2.1 Infinitives and Gerunds

Name	Suffix	Suffix (LAT)	Stem Type
Infinitive	-аш	-aš	INF
Inf + Dative	-ашлан	-ašlan	INF
Necessive Inf.	-ман	-man	IMP
Nec. Future Inf.	-мыла	-məla	IMP
NFI (1.P.Sg)	-мемла	-memla	IMP
NFI (2.P.Sg)	-метла	-metla	IMP
NFI (3.P.Sg)	-мыжла	-məžla	IMP
NFI (1.P.Pl)	-мынала	-mənala	IMP
NFI (2.P.Pl)	-мыдала	-mədala	IMP
NFI (3.P.Pl)	-мыштла	-məštla	IMP
Affirmative Ger.	-ын / -ен, -эн	-ən, -en	INF
Negative Ger.	-де	-de	IMP
G. for Prior Act.	-мек(е)	-mek(e)	IMP
G. for Fut. Act.	-меш(ке)	-meš(ke)	IMP
G. for Sim. Act.	-шыла	-šəla	IMP
GSA (1.P.Sg)	-шемла	-šemla	IMP
GSA (2.P.Sg)	-шетла	-šetla	IMP
GSA (3.P.Sg)	-шыжла	-šəžla	IMP
GSA (1.P.Pl)	-шынала	-šənala	IMP
GSA (2.P.Pl)	-шыдала	-šədala	IMP
GSA (3.P.Pl)	-шыштла	-šəštla	IMP

Fig. 53: Non-Finite Verb Forms

A.2 Productive Derivations

A.2.1 Nominal → Nominal

Case Name	Suffix	Suffix (LAT)	Suffix Type	Meaning
Adjective 1	-ан, ян	-an	D	with X
Adjective 2	-с(е/о/ö)	-s(e/o/ö)	E	like X
Adjective 3	-дым(е/о/ö)	-dəm(e/o/ö)	R	without X

Fig. 54: Derivations: Nominal → Nominal

A.2.2 Verb → Nominal

Name	Suffix	Suffix (LAT)	Stem Type
Active Participle	-ш(е/о/ö)	-š(e/o/ö)	IMP
Passive Participle	-м(е/о/ö)	-m(e/o/ö)	IMP
Negative Participle	-дым(е/о/ö)	-dəm(e/o/ö)	IMP
Future Participle	-шаш	-šaš	IMP
Nominalization	-маш	-maš	IMP
Negated Nominalization	-дымаш	-dəmaš	IMP

Fig. 55: Derivations: Verb → Nominal

A.2.3 Nominal → Verb

Meaning	Suffix	Suffix (LAT)	Suffix Type
To Become X (1)	-аҥ-^{I}, -яҥ^{I}	-aŋ-^{I}	D
To Become X (2)	-ем-^{I}, -эм-^{I}	-em-^{I}	D
To Put X on	-ал-^{I}, -ял-^{I}	-al-^{I}	D
To Make into X	-кт-^{II}	-kt-^{II}	E
(Various)	-л-^{II}	-l-^{II}	E

Fig. 56: Derivations: Nominal → Verb

A.2.4 Verb → Verb

Attribute	Suffix	Suffix (LAT)	Stem Type
Reflexive	-калт-[I]	-kalt-[I]	INF
Diminutive	-ал-[I], -ялаш[I]	-al-[I]	INF
Causative	-ыкт-[II]	-əkt-[II]	INF
Frequentative	-ыл-[I]	-əl-[I]	INF

Fig. 57: Derivations: Verb → Verb

A.3 Arrangement of Suffixes

Nominal:

stem + [derN] + [comp] + [gen] + [poss] + [p3] + [plur] + [case-g1] + [enc]
stem + [derN] + [comp] + [gen] + [poss] + [p3] + [plur] + [case-g3] + [enc]
stem + [derN] + [comp] + [gen] + [plur] + [poss] + [p3] + [case-g1] + [enc]
stem + [derN] + [comp] + [gen] + [plur] + [poss] + [p3] + [case-g3] + [enc]
stem + [derN] + [comp] + [gen] + [plur] + [case-g2] + [poss] + [p3] + [enc]
stem + [derN] + [comp] + [gen] + [plur] + [shILL] + [enc]
stem + [derN] + [comp] + [gen] + [plur] + [case-g3] + [poss] + [p3] + [enc]

Group one cases are the genitive, accusative and comitative cases. Group two cases are the inessive, illative and lative cases. Group three cases are the dative and the comparative.

Finite Verb:	stem + [derV] + time/mood/person + [comp] + [enc]
Non-Finite Verb:	stem + [derV] + inf/ger + [comp] + [poss] + [enc]
Postposition:	stem + [poss] + [enc]
Adverb:	stem + [derA] + [comp] + [enc]
Miscellaneous:	stem + [enc]

Figures

Fig. 1: The Mari Web Platform 11
Fig. 2 Layout for users of Scandinavian, Finnish and Estonian keyboards .. 12
Fig. 3: The Uralic world (Wikipedia 2007) – Mari is marked in dark red 19
Fig. 4: Vowels in Russian 24
Fig. 5: Soft signs and hard signs in Russian 24
Fig. 6: /e/ in Mari 25
Fig. 7: The suffix “-em” 26
Fig. 8: Final obstruents 27
Fig. 9: Palatalization in Mari 28
Fig. 10: Marking of palatalization depending on vowel 29
Fig. 11: Stressed and unstressed final vowels 30
Fig. 12: Conjugations I & II 32
Fig. 13: Agglutinative, isolating, fusional 34
Fig. 14: Vowel harmony 37
Fig. 15: Inessive suffix –šte/-što/-štö 38
Fig. 16: Suffix types 40
Fig. 17: Finite verb forms 44
Fig. 18: Estonian Language Synthesizer 50
Fig. 19: Estonian Language Synthesizer – output 52
Fig. 20 “I am” 53
Fig. 21: The Estonian Language Lemma Machine 54
Fig. 22: Output of the Lemma Machine 54
Fig. 23: Online Estonian dictionary 56
Fig. 24: The Morphology Generator – Nominals 58
Fig. 25 The Morphology Generator – Verbs 59
Fig. 26 The Morphology Generator – Settings 60
Fig. 27: Tree graph of word analysis 66
Fig. 28: Sample entry as seen by the project team 72
Fig. 29: The same entry, as seen by users 73
Fig. 30: The dictionary's main page 74
Fig. 31: Search results for "_rach" 75
Fig. 32: Searching for Hill Mari words 76

Fig. 33: Customized grammatical generator for absolute beginners 77
Fig. 34: Mari Reading Aid 78
Fig. 35: A personalized vocabulary list 80
Fig. 36: Detector of unknown words 82
Fig. 37: Flat and Sculptured Buttons 86
Fig. 38: Bad Affordances in the Viennese Metro 86
Fig. 39: Cold, Caliente? 88
Fig. 40: Case suffixes 94
Fig. 41: Plural suffixes 94
Fig. 42: Possessive suffixes 95
Fig. 43: Indicative Present – First Conjugation 95
Fig. 44: Indicative Present – Second Conjugation 96
Fig. 45: Indicative First Preterite – First Conjugation 96
Fig. 46: Indicative First Preterite – Second Conjugation 96
Fig. 47: Indicative Second Preterite – First Conjugation 97
Fig. 48: Indicative Second Preterite – Second Conjugation 97
Fig. 49: Imperative – First Conjugation 97
Fig. 50: Imperative – Second Conjugation 98
Fig. 51: Desiderative – First Conjugation 98
Fig. 52: Desiderative – Second Conjugation 98
Fig. 53: Non-Finite Verb Forms 99
Fig. 54: Derivations: Nominal → Nominal 100
Fig. 55: Derivations: Verb → Nominal 100
Fig. 56: Derivations: Nominal → Verb 100
Fig. 57: Derivations: Verb → Verb 101

Sources

Alhoniemi, A. 1985: *Marin kielioppi*, Suomalais-ugrilainen seura, Helsinki, Finland.

Alhoniemi, A. 1986: *Marin kielen lukemisto sanastoineen*, Suomalais-ugrilainen seura, Helsinki, Finland.

Basu I. 2008: *Estonia becomes E-stonia*, in: Government Technology http://www.govtech.com/dc/articles/284564, Folsom, California, USA. [accessed 6. June 2009]

Bereczki, A. 2004: *The Mari – A historical overview*, in: *The Finno-Ugric World*, (ed.) György Nanovfszky, Teleki Lásyló Foundation, Budapest, Hungary.

Bereczki, G. 1990: *Chrestomathia Ceremissica*, Tankönyvkiadó, Budapest, Hungary.

Bünting, K.: *Einführung in die Linguistik*, 15. Auflage, Beltz Athenäum, Weinheim, Germany.

Dix, A. et al. 2006: *Human-Computer Interaction*, 3rd Edition, Pearson Prentice Hall, Harlow, United Kingdom.

Federal State Statistics Service of Russia 2002: *Всероссийская перепись населения*, http://www.perepis2002.ru/, Moscow, Russian Federation. [accessed 30. April 2009]

Filosoft 2007: *Eesti keele lemmatiseerija*, http://www.filosoft.ee/lemma_et/, Tartu, Estonia. [accessed 30. April 2009]

Filosoft 2007: *Eesti keele morfanalüsaator*, http://www.filosoft.ee/html_morf_et/, Tartu, Estonia. [accessed 30. April 2009]

Filosoft 2007: *Eesti keele speller*, http://www.filosoft.ee/ html_speller_et/, Tartu, Estonia. [accessed 30. April 2009]

Filosoft 2007: *Eesti keele süntesaator*, http://www.filosoft.ee/gene_et/, Tartu, Estonia. [accessed 30. April 2009]

Kahrs, U.; Schötschel, M. 2005: *Mari und Mordwinen im heutigen Russland*, Harrassowitz Verlag, Wiesbaden, Germany.

Keelevara 2006: Õigekeelsussõnaraamat 2006, http://www.keelevara.ee/, Tallinn, Estonia. [accessed 6. June 2009]

Luutonen, J.; Moisio, A.; Saarinen, S. et al. 2007: *Electronic Word Lists: Mari, Mordvin, Udmurt*, Suomalais-ugrilainen seura, Helsinki, Finland.

Minakova-Boblest, E. 2007: *Praktisches Lehrbuch Russisch*, Langenscheidt, Berlin, Germany.

Moisio, A. 1992: *Marilais-suomalainen sanakirja*, Publications of the Department of Finnish and General Linguistics of the University of Turku, Turku, Finland.

Nielsen, J. 1999: *Designing Web Usability: The Practice of Simplicity,* Peachpit Press, Berkeley, California, USA.

Nielsen, J. 2000: *Usability Engineering*, 9th Edition, Morgan Kaufmann, San Diego, California, USA.

Norman, D. 2006: *The Design of Everyday Things*, 14th Edition, Basic Books, New York City, New York, USA.

Pomozi, P. 2004: *The Mari language*, in: *The Finno-Ugric World*, (ed.) György Nanovfszky, Teleki Lásyló Foundation, Budapest, Hungary.

Project NOS Negation 2009: Typology of Negation in Ob-Ugric and Samoyedic Languages, http://www.univie.ac.at/negation/, University of Vienna, Vienna, Austria.

Riese, T.; Bradley, J. 2009: *Mari Web Resources*, http://www.mari-language.com, Vienna, Austria.

Roh, Štěpán 2006: *DejaVu Fonts*, http://dejavu-fonts.org/. [accessed 2. May 2009]

Shneiderman, B.; Plaisant, C. 2005: *Designing the User Interface*, 4th Edition, Addison Wesley, Boston, Massachusetts, USA.

Tognazzini, B. 1996: *Tog on Interface*, 5th Edition, Addison Wesley, Reading, Massachusetts, USA.

UNESCO 2009: *Safeguarding endangered languages*, http://www.unesco.org/culture/en/endangeredlanguages, Paris, France. [accessed 30. April 2009]

Wikipedia 2007: *Distribution of Uralic languages and peoples*, http://en.wikipedia.org/wiki/File:Fenno-Ugrian_people.png, released into the public domain by its creator. [accessed 30. April 2009]

Yakimova, E. S.; Zorina Z. G.; Зорина, Krylova, G. S. *1990: Марийский язык для всех I*, Марийское книжное издательство, Yoshkar-Ola, Russian Federation.

Yakimova, E. S.; Krylova, G. S. 1991: *Марийский язык для всех II*, Марийское книжное издательство, Yoshkar-Ola, Russian Federation.