apPILcation: an Android-based Tool for Learning Mansi

Gábor Bobály[1]**, Csilla Horváth**[2,3]**, Veronika Vincze**[4]
[1]IT Services Hungary
[2]Research Institute for Linguistics, Hungarian Academy of Sciences
[3]University of Helsinki
[4]MTA-SZTE Research Group on Artificial Intelligence
bobalygabor@gmail.com, naj.agi@gmail.com, vinczev@inf.u-szeged.hu

Abstract

Here we introduce our Android application for Mansi language learning, called apPILcation / PILozhenie / apPILkáció. Learners can select Hungarian, English or Russian as the source language while learning Mansi. Currently, the application offers a general vocabulary practising session as well as a thematic word guessing game. apPILcation is primarily supposed to be used by learners of the Mansi language but language teachers and linguists may also be interested in it. The application can be freely used for anyone interested, and will be soon made available for download.

A cikkben bemutatjuk az apPILkáció / apPILcation / PILozhenie nevű Android-alkalmazásunkat, mely a manysi nyelv elsajátítását, konkrétabban manysi szavak tanítását célozza. Az alkalmazás nyelve választhatóan magyar, orosz vagy angol. Moduljai között találunk általános szókincsre épülő, véletlenszerűen kiválasztott manysi szavakat tanító modult, illetve meghatározott szemantikai mezőhöz (pl. állatok) tartozó szavakra épülő szókitaláló modult. Az alkalmazás elsődlegesen a manysi nyelvtanulók érdeklődésére tarthat számot, de hasznos lehet nyelvészeknek vagy nyelvtanároknak is. Az alkalmazást mindenki számára ingyenesen elérhetővé tesszük.

1 Introduction

Nowadays, the widespread use of the internet and digital technologies offers a variety of possibilities for real-time communication among people around the world. Such interaction is further supported by several language technology tools such as speech to text systems, spellcheckers and machine translation systems, just to name a few. However, minority languages often lack these tools, which might lead to the loss of such languages in the digital space. On the other hand, there are some efforts to revitalize endangered languages, which aim at constructing tools and resources for such languages to be used in digital communication.

In this paper, we focus on Mansi, an endangered language spoken in Western Siberia. Although the number of Mansi native speakers decreases, the prestige of language proficiency and language use is rising, also there is a growing interest towards language acquisition and heritage language acquisition, with special focus on urban learners of Mansi. In order to help such efforts, we implemented apPILcation, an Android application for Mansi language acquisition, freely available for anyone. In this paper, we present the application and its main functionalities. As far as we know, ours is the first attempt to offer an online tool for smartphones for Mansi language learning.

The paper is structured as follows. First we give an overview of some language teaching mobile applications, then we briefly discuss the current sociolinguistic background of the speakers and learners of Mansi. Next, apPILcation is presented in detail, together with its main functionalities. The paper ends with listing some of the possible uses of the application.

2 Background

Mobile-assisted language learning (MALL) has been an emerging field in language teaching recently due to the widespread usage of smartphones in all over the world (Chinnery, 2006). Traxler (2005) defines mobile learning as "Any educational delivery where the sole or prevailing technologies are handheld or palmtop devices". MALL enables

language learners to study whenever and wherever they are, without the need for desktop computers (Miangah and Nezarat, 2012). However, most surveys on MALL investigated only the institutional use of mobile-based learning while only a few analysed their use outside the classroom (Godwin-Jones, 2017). For instance, Stockwell and Hubbard (2013) define ten principles for MALL, e.g. accommodating language learner differences and keeping activities short.

There are several smartphone applications for language learners from bilingual dictionaries to tools offering grammar exercises. However, the number of languages these tools offer courses for is rather limited: solutions are mostly available for widely spoken languages. Just to name a few such applications, Babbel[1], Duolingo[2], Memrise[3] and Busuu[4] are among the most well-known applications for language learning. Table 1 shows the languages with available courses in the above applications (as of November 2019). Data on the number of the native speakers for each language come from the English Wikipedia.

As can be seen, it is primarily world languages, in addition, smaller languages mostly spoken in Europe that are taught in these applications, not to mention extinct languages like Latin or constructed and fictional languages (beside Esperanto, Klingon and High Valyrian are also available in Duolingo, the two latter owe their popularity to certain television series).

Concerning the mobile-assisted language learning for minority Uralic languages, we are aware of only few applications, e.g. Laring[5], developed at the University of Tromsø for teaching Southern Saami.[6] The system teaches words belonging to different word groups to the user: it reads out the Southern Saami equivalent of Norwegian words, while in another task, listening comprehension can also be practised – a picture corresponding to the heard Saami word should be chosen. Beside this, we are aware of some user-generated Memrise courses for Uralic and Siberian languages, e.g. Ingrian[7], Livo-

nian[8], Kven[9] and Yakut[10]. Also, there are some courses available that contain only some tens of words for languages such as Ulch[11] and Enets[12]. The Mansi version consists of 11 elements (one of which is translated incorrectly), using an inconsistent spelling. Thus our apPILcation is proved not to be the first MALL application for Mansi, but it has very good chance to incorporate more material and attract more users than its predecessor.

3 Mansi language, its speakers and learners

Mansi is an endangered language spoken in Western Siberia. Mansi plays limited role in its Russian-dominated, multiethnic and multilingual environment, its usage is heavily affected by the loss of the traditional way of life and rapid urbanisation as well. While the Mansi have been (and in some respect still are) regarded as followers of traditional, semi-nomadic lifestyles, and are expected to live in rural conditions, the majority of the Mansi live in a multi-ethnic urban environment.

The principles of Soviet language policy according to which the Mansi literary language and written standard have been designed kept changing from time to time. The first, Latin alphabet for Mansi was created in 1931 at the Institute of the Peoples of the North. It was in use for a short period, in 1937 the Mansi language planners had to switch to Cyrillic transcription. This writing system is in use since then, and underwent only minor changes. The marking of vowel length and special characters absent from the Russian alphabet started to appear in printed materials in the 1980s. Since the 1990s two parallel spellings are in use (differing only in one element), one used by the leading specialists (mainly following the Soviet academic policy, publishing a small amount of Mansi texts) and the journalists (using and promoting the language on a daily basis, with the largest active number of followers). Taking into consideration the history and the status of the Mansi language, in our application we use the colloquial literary written Mansi standard, that is, the

---

[1] https://www.babbel.com/

[2] https://www.duolingo.com/

[3] https://www.memrise.com/

[4] https://www.busuu.com/

[5] http://divvun.no/laring/laring.html

[6] When writing this paper, we could have access only to the iPhone version of the application, the Android version being unavailable for download.

[7] https://decks.memrise.com/course/2107565/ingrian/

[8] https://decks.memrise.com/course/5603933/livonian/

[9] https://decks.memrise.com/course/5596403/kven/

[10] https://decks.memrise.com/course/362501/basic-yakut/

[11] https://decks.memrise.com/course/1064732/ulch-language/

[12] https://decks.memrise.com/course/1843983/family-words-in-enets/

Language	Number of native speakers	Babbel	Duolingo	Memrise	Busuu
Chinese	1500M		●	●	●
Spanish	400M	●	●	●	●
English	332M		●		●
Hindi	370M		●		
Arabic	300M		●	●	●
Portuguese	230M	●	●	●	●
French	220M	●	●	●	●
Russian	145M	●	●	●	●
Japanese	126M		●	●	●
German	90M	●	●	●	●
Korean	78M		●	●	
Vietnamese	70M		●		
Italian	63M	●	●	●	●
Turkish	60M	●	●	●	●
Polish	50M	●	●	●	●
Indonesian	43M	●	●		
Ukranian	35M		●		
Romanian	24M		●		
Dutch	22M	●	●	●	
Greek	20M		●		
Hungarian	15M		●		
Czech	12M		●		
Catalan	10M		●		
Swedish	9M	●	●	●	
Hebrew	6M		●		
Danish	5,5M	●	●	●	
Guarani	4,8M		●		
Norwegian	4,6M	●	●	●	
Mongolian	3,6M			●	
Slovenian	2,5M			●	
Swahili	2M		●		
Welsh	610K		●		
Icelandic	310K			●	
Irish	260K		●		
Navajo	170K		●		
Hawaiian	2K		●		
Esperanto	0		●		
Klingon	0		●		
Latin	0		●		
High Valyrian	0		●		

Table 1: Languages with language courses available on smartphones.

Cyrillic-based form of spelling used in the press.

Although the prestige of Mansi language and culture is rising, the number of Mansi speakers is critically low. The Mansi speakers' community is traditionally divided into three major age groups (cf. Skribnik and Koshkaryova (2006)). The eldest speakers were born and raised in monolingual Mansi families and remained more or less monolingual Mansi themselves, with only a limited command of Russian. The middle-aged speakers were born and raised in monolingual Mansi families and speak Mansi as their mother tongue, in line with becoming Mansi-Russian bilinguals through education, and generally they live in Russian-dominated multilingual environment. The youngest generation of Mansi speakers consists of considerably less people than the former two, and with the exception of those being raised in a few peripherical Mansi villages to be found outside the Khanty-Mansi Autonomous Okrug, none of them can be considered as Mansi monolingual even until their school years. Thus the level of the speakers' proficiency in Mansi is typically related to their age: the older the speakers are, the more likely they are to have native competence in Mansi. This general tendency is often counterbalanced by the speaker's place of birth and residence: younger speakers born and raised in smaller, monolingual Mansi settlements often have good command of the Mansi language.

The majority of Mansi children are born outside the Mansi-speaking settlements. They usually reside in multiethnic, multicultural towns and cities, and live in families with Russian as the language of communication. Their Mansi parents usually have not taught them the Mansi language, hence these children cannot acquire it while listening to their parents' conversations either (since the parents tend to use Russian between themselves, as well), which leaves education as the only possible domain available of Mansi language acquisition for children. A small number of alternative institutions were founded in larger, urbanised settlements with a large Mansi population to complement Mansi children's knowledge of their heritage, culture and language, which they could not completely acquire within their family, but they do not serve as stable domains for language use either (cf. Horváth (2015, 2016)).

The group of middle-aged and especially young language learners, who had no ties with Mansi-speaking families or other domains, or for some reason were unable to acquire the Mansi language in their family, but who are still interested and motivated to attend Mansi language courses at school or to study the language on their own, form our main target group.

4 apPILcation / PILozhenie

Our Android-based tool is called apPILcation in English, apPILkáció in Hungarian and PILozhenie in Russian, *pil* meaning "berry" in Mansi. Our idea behind the name was that by using the application, it is as easy to pick up words in Mansi as picking up berries in the forest.

4.1 The Underlying Dictionary

Researchers of the Mansi language already compiled some dictionaries of the language about one hundred years ago, which were only lately published (Munkácsi and Kálmán, 1986; Kannisto, 2013). These dictionaries contain words from all the dialects, also from those that are now extinct. There are also some modern dictionaries of the Northern Mansi dialect available (Rombandeeva, 2005; Rombandeeva and Kuzakova, 1982). These dictionaries form the base for the vocabulary used in our application, all of which come from the Northern dialect of Mansi.

The vocabulary used in the application is built on an online Mansi dictionary that contains approximately 20,000 entries (Horváth et al., 2017). The Mansi forms were retrieved from the PDF versions of Rombandeeva's and Kuzakova's, as well as Rombandeeva's dictionaries (Rombandeeva, 2005; Rombandeeva and Kuzakova, 1982) by means of optical character recognition, then lexical entries from different sources were merged. The Mansi lexemes are supplemented with the Russian translation given by the dictionaries, and Hungarian and English translations were provided by linguists. Thus, the potential learners of Mansi can choose whether they want to learn Mansi with Russian, English or Hungarian as the source language.

Table 2 contains the number of vocabulary items for each language in the dictionary.

4.2 Functionalities

The main functionalities of the application are as follows. First, a randomized Mansi word is shown to the learner, together with its translation in the given source language, so that he or she can check whether he or she already knows the word. If not, he or she

Language	Entry
Mansi	13,948
Russian	14,344
Hungarian	2,334
English	458

Table 2: Statistics on languages.

can now memorize the word. Second, there is a possibility for playing some quizzes, for instance, the learner can choose certain topics (e.g. colors, family terms, berries or animals) and he or she has to match Mansi words that belong to these semantic categories with their other language equivalents. In this way, words with similar meaning can be learnt and practised together. Third, some information on Mansi grammar, Mansi geography and Mansi culture is also available in the application, in order to deepen cultural knowledge on Mansi as well.

4.2.1 Vocabulary Learning

First, the user can select which language pair s/he wants to work with: Mansi–Hungarian, Mansi–Russian, Mansi–English, Hungarian–Mansi, Russian–Mansi or English–Mansi (see Figure 1[13]). Next, by tapping the button *Give me a new word* a word is shown to him or her in the selected source language and he or she can decide whether sh/he knows the meaning in the target language. By tapping the button *Show the target language equivalent*, he or she can check the meaning of the word as shown in Figure 2.

4.2.2 Thematic Word-guessing Games

In the case of thematic word-guessing games, the user can select which semantic group of words he or she wants to practice. Then a word is shown to him or her from the given semantic field, together with four words from the target language, out of which one is correct while the other three are incorrect. By clicking on the correct equivalent, it turns green, indicating that it is the correct answer (see Figure 3). On the other hand, incorrect words are highlighted with red when clicking on them (see Figures 4 and 5). By clicking on the button *New quiz*, a new word is offered to the user for practice.

Currently, apPILcation contains four semantic groups, namely, colors, family terms, berries and animals, which we are planning to extend with other

Figure 1: Dictionaries available in apPILcation.

---

[13]In the screenshots, we use the Hungarian version of the application as this is another Uralic language.

semantic groups of words in the near future.

Figure 3: Correct selection of the meaning of the word.

5 Applicability

Our apPILcation may serve various purposes in its present state. It may be used by any user who can read Russian, English or Hungarian, and Mansi written in Cyrillic. We expect attention from the experts and university students specialised on Ob-Ugric languages from all over the world, but first and foremost from pupils, students and Mansi language teachers living on the territory of the Khanty-Mansi Autonomous Okrug.

The word matching module of the application is targeted for beginning language learners, especially for pupils on 1-5 classes. In the beta version of the application, the module contains four word sets, which can be extended according the users' feedback.

The randomised Mansi word learning module is targeted for more advanced language learners, especially for pupils of senior classes and students, who spend regular, but short periods using the applica-

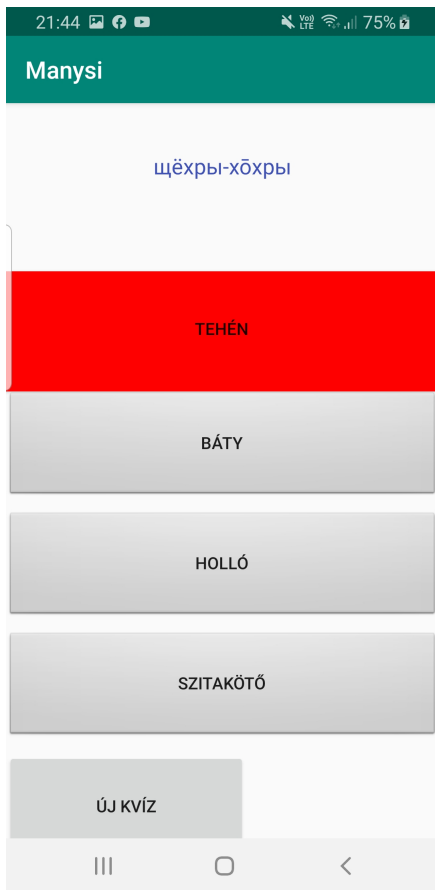Figure 2: A Mansi–Hungarian entry.

Figure 4: Incorrect selection of the meaning of the word.

Figure 5: Incorrect, then correct selection of the meaning of the word.

tion.

Both modules are ideal for independent language learning. The word matching module may be used to complement teacher assisted language learning as well.

apPILcation will be open access and it is going to be available for use without charge. The advertisement of the program seems to be unproblematic due to the creators' connection with European specialists possibly interested in the application on the one hand, and both offline and online Mansi speaker groups on the other hand. The promotion is planned to take place on the different pages and in chat threads of the most popular Russian social media site, as well as in reports or advertisements to be published in the only Mansi newspaper and the only Mansi journal for children. Creators expect to get direct feedback via social media pages created on two social media sites, via the email address of the application, while indirect feedback with the help of the Mansi intermediators and distributors of the application, first and foremost from specialists working in press and educational institutions.

6   Conclusions

In this paper, we presented our Android application for Mansi language learning, called apPILcation / PILozhenie / apPILkáció. Learners can select Hungarian, English or Russian as the source language while learning Mansi. Currently, the application offers a general vocabulary practising session as well as a thematic word guessing game for specific groups of words (e.g. colors). apPILcation is primarily supposed to be used by learners of the Mansi language but it may serve useful for language teachers and linguists as well.

The application can be freely used for anyone interested, and will be soon made available for download.

As future work, we would like to extend the vocabulary of apPILcation, besides, we would like to implement other modules for assisting vocabulary learning. Moreover, we would like to add some grammar-based drills and tasks to the modules of the application. Lastly, we would like to create AP-PLEcation (priLOMTzhenie in Russian and AL-MAlkazás in Hungarian), the iPhone version of ap-PILcation.

References

George M. Chinnery. 2006. Going to the MALL: Mobile Assisted Language Learning. *Language Learning & Technology*, 10(1):9–16.

Robert Godwin-Jones. 2017. Smartphones and language learning. *Language Learning & Technology*, 21(2):3–17.

Csilla Horváth. 2015. Beading and language class. Introducing the Lylyng Soyum Children Education Centre's attempt to revitalise Ob-Ugric languages and cultures. *Zeszyty Łużyckie*, 48:115–127.

Csilla Horváth. 2016. A manysi örökségnyelv oktatási kísérletei és eredményei. *Általános Nyelvészeti Tanulmányok*, 28:295–306.

Csilla Horváth, Norbert Szilágyi, Ágoston Nagy, and Veronika Vincze. 2017. Language technology resources and tools for Mansi: an overview. In *Proceedings of the Third International Workshop on Computational Linguistics for Uralic Languages*, St. Petersburg, Russia.

Artturi Kannisto. 2013. *Wogulisches Wörterbuch*. Kotimaisten Kielten Keskuksen Julkaisuja, Helsinki.

Tayebeh Mosavi Miangah and Amin Nezarat. 2012. Mobile-Assisted Language Learning. *International Journal of Distributed and Parallel Systems*, 3(1):309–319.

Bernát Munkácsi and Béla Kálmán. 1986. *Wogulisches Wörterbuch*. Akadémiai Kiadó, Budapest.

Evdokija Ivanova Rombandeeva. 2005. *Russko-mansijskij slovar'*. Mirall, Sankt-Peterburg.

Evdokija Ivanova Rombandeeva and Evdokija Aleksandrova Kuzakova. 1982. *Slovar' mansijsko-russkij i russko-mansijskij*. Prosvešenije, Leningrad.

Elena Skribnik and Natalya Koshkaryova. 2006. Khanty and Mansi: the contemporary linguistic situation. In *Shamanism and northern ecology*, pages 207–218, The Hague. Mouton de Gruyter.

Glen Stockwell and Philip Hubbard. 2013. Some emerging principles for mobile-assisted language learning. Technical report, The International Research Foundation for English Language Education, Monterey, CA.

John Traxler. 2005. Defining Mobile Learning. In *IADIS International Conference on Mobile Learning*, pages 261–266, Malta.