LANGUAGE CONTACT AND SOUND ARCHIVES IN RUSSIA

Liya Vasilyevna Bondarko and Tjeerd de Graaf

University of Saint-Petersburg, Russia and University of Groningen, The Netherlands

ABSTRACT

In the Russian Federation, Russian is used for interethnic and official communication, and until recently this language served the same purpose throughout the former USSR. This situation resulted in contacts between Russian and other languages, such as Uralic, Altaic, Paleo-Asiatic and Baltic as well as other Slavic languages. These contacts are interesting from a phonetic point of view: We can study the influence that Russian and the other languages exert on each other in a situation of language contact. The main subject of this presentation is the versatility of problems we face in creating sound archives, problems which determine the structure of such archives. We consider the preservation of old sound recordings, the development of representative databases for the modern language situation, the analysis of the interrelation between different sound systems, the registration of the speech sounds for endangered languages and their revitalisation.

1. INTRODUCTION

After the phonograph of Thomas Edison was introduced at the end of the previous century, for the first time in history people were able to store and listen to acoustic data, in particular speech, songs and music. A central facility was created for the preservation of the valuable material which had been collected. This led to the establishment of several archives, and in Russia. In section 2 we describe the use of these sound archives in an international research project where we investigate the languages of Russia both diachronically and synchronically by creating an acoustic database. We discuss the structure of databases for the present day sound system of Russian in section 3. Both kinds of databases are used for the study of language contrast (section 4) and for the description and revitalisation of the endangered languages of Russia (section 5).

2. SOUND ARCHIVES IN RUSSIA

In the beginning of the 20th century the first Russian sound recordings were made, which were later stored in the sound archives of the Russian Academy of Sciences (Pushkinsky Dom) in Saint-Petersburg. These sound archives contain about 10,000 wax cylinders of the Edison phonograph and more than 500 old wax discs. In addition, there is an extensive fund of grammophone cylinders of the Edison phonograph and more than 500 old wax discs. The sound archives in Saint-Petersburg also contain unique data on Yiddish, the language of the Jews in Eastern Europe which at the beginning of this century was used by millions of speakers. In the archives we found an unpublished manuscript by Sofia Magid on 'The Ballad in Jewish Folklore', together with many corresponding wax cylinders. We are preparing a further exploration of the acoustic data in the sound archive and an edition of the book. This takes place in the framework of a project with the title "Voices from the Shtetl: the Past and Present of the Yiddish Language in Russia", for which we obtained financial support from the Dutch National Science Foundation (NWO).

An important part of the linguistic database project is related to the analysis of recordings of Russian dialects and minority languages in the Russian Federation, such as Nenets, Komi, Nivkh, Evenki and Yakut. We have already prepared collections of Russian fairy tales, Tales from the Russian North and Ritual Poetry of the Russian North, based on historical sound recordings and material from the archives, such as written texts and comments by ethnomusicologists, folklorists, phoneticians and other specialists.

Part of the recordings in the Museum of Literature of the Russian Academy of Literature (Pushkin Dom) have been reconstructed and made available in a phonetic database of the languages of Russia. This database has many scientific, cultural and technical applications and the research conducted within this project provides us with valuable insights in the history and culture of the languages concerned and their speakers. The material will also be used to study the process of language change during the last century as well as the phonetic systems of the languages of Russia from a typological point of view.

3. THE RUSSIAN SOUND SYSTEM

Our second task consists of constructing sound archives for the registration of the present day language situation. There are numerous acoustics databases for various languages which are used for a number of tasks, in the first place for projects with an applied
representation. This is required in order to obtain statistical data on the automatic grapheme-phoneme conversion, which allows the structure of wordforms are obtained.

(3) An important component of the database system used is the automatic grapheme-phoneme conversion, which allows the transition from an orthographic text to its phonemic or allomorphic representation. This is required in order to obtain statistical data on the functional load of certain units of the sound system [4].

A number of investigations have recently been carried out in these directions. The Phonic Fund of the Russian language has been created on the basis of recordings made from 4 speakers of Russian representing the Moscow and Saint-Petersburg pronunciation variant. The input sound material which contains 180 syllables of the CV type and a phonetically representative text segmented into open syllables (about 1000 syllables in the realisation of each speaker), gives a complete picture about changes of the acoustic characteristics of the speech sounds, depending on the type and size of the sound sequence (isolated syllables or syllables excised from a text). The experiment showed not only a considerable dependence of the acoustic realisations of the segmental units on their position in the phrase, but also on their left and right context [3,4]. The perceptual study of these syllables has made it possible to evaluate the degree of adequacy of the existing transcription systems for real properties of speech sounds in connected speech [5]. In view of the development of Russian text-to-speech synthesis systems new principles of automatic transcription have been developed [6].

It is important that the acoustical component of the system is organized on the basis of data which allow the selection and classification of elements with given features [7]. This facilitates the further development of the system and allows to create sound archives which will include both standard speech realisations and various deviations from the speech norm. The acoustic characteristics of the pronunciation of one and the same text by various speakers can be used for a model to describe the variability of the Russian sound system. This model can be applied both to the development of a linguistic theory and to the construction of practical devices in speech recognition and speech synthesis systems.

In the former Soviet Union, Russian was an official language as well as the language used for interethnic communication. Thus Russian was used not only by native speakers, but also by people whose native language was other than Russian. The Russian speech of these non-native speakers can be analysed in order to establish the phonetic interference from a given speaker's native language. For this purpose, a detailed study of the acoustic properties of such interference was made, including among other things the realisation of vowel phonemes after soft consonants. On the basis of a thorough analysis of the differences between the standard pronunciation and a pronunciation which shows interference from other languages, we are now able to develop a reliable method of measuring the amount of such interference. This method can then also be used to measure the degree of stability of various Russian speech segments in the speech of bilinguals [8].

4. APPLICATIONS TO THE STUDY OF LANGUAGE CONTRASTS AND MINORITY LANGUAGES

A striking feature of the present day situation in the Russian Federation is the increasing national self-awareness of numerous native peoples and the growing interest in their own languages. It is our goal to prepare a Phonetic Database of the Languages of Russia, which consists of compact and complete information on the phonetic (phonological) systems of these languages. This database can be used by scientists working on linguistic problems but it should also be made available to the growing number of speakers of minority languages who are interested in preserving their mother tongue. The change in attitude toward the minority languages and the subsequent increase of the number of domains in which these languages are being used, have led to various activities, most importantly the development of teaching material and - in some cases - the revision of orthographies which do not correctly represent the phoneme system of a given language.

Data on the sound systems of minority languages can also be used to study the influence of these languages on the Russian speech of non-native speakers of Russian. An example is our program "A Phonetic Database of the Finno-Ugric Languages of Russia", comparable to the project on the Phonetic Database of Russian. Data on the Mari language have been used in order to establish the basic acoustic parameters for the sound elements of this language, and an automatic transcriptor is designed which enables us to obtain statistic data on its various speech sounds. So we can study the Russian speech of native speakers of Mari and determine the amount and kind of interference of Mari in this particular variety of Russian. Minor languages who are interested in preserving their mother tongue. The change in attitude toward the minority languages and the subsequent increase of the number of domains in which these languages are being used, have led to various activities, most importantly the development of teaching material and - in some cases - the revision of orthographies which do not correctly represent the phoneme system of a given language.

Data on the sound systems of minority languages can also be used to study the influence of these languages on the Russian speech of non-native speakers of Russian. An example is our program "A Phonetic Database of the Finno-Ugric Languages of Russia", comparable to the project on the Phonetic Database of Russian. Data on the Mari language have been used in order to establish the basic acoustic parameters for the sound elements of this language, and an automatic transcriptor is designed which enables us to obtain statistic data on its various speech sounds. So we can study the Russian speech of native speakers of Mari and determine the amount and kind of interference of Mari in this particular variety of Russian. Minor languages who are interested in preserving their mother tongue. The change in attitude toward the minority languages and the subsequent increase of the number of domains in which these languages are being used, have led to various activities, most importantly the development of teaching material and - in some cases - the revision of orthographies which do not correctly represent the phoneme system of a given language.

5. THE REVIVAL OF ENDANGERED LANGUAGES IN RUSSIA

The extinction of languages is a process which takes place everywhere in the world. In our INTAS report of 1996 [1] it was stated that numerous languages in the territory of the Russian Federation are under threat of total extinction and that measures should be taken to end this process of language death. Linguists and ethnologists should work together with speakers of endangered languages in order to find solutions to these problems.

We consider it our task to create sound archives for the endangered languages of minor ethnic groups. It is well known that...
these languages are used less and less and that many of them are on the brink of extinction. For the purpose of conserving these languages we collect a variety of texts, and we create a sound archive with a wide range of material which will represent the entire (grammatical) structure of the languages. Such an archive can be used for the revitalisation of endangered or even completely extinct languages. As an example we mention the creation of a computerized database for the sound dictionary of the Nenets language at the Phonetic Department of Saint-Petersburg University.

During fieldwork expeditions to Sakhalin and Northern Yakutia we were able to observe the ongoing process of Language Shift and Language Death. It is the urgent task of linguists to record the speech of the last speakers of these languages, making use of modern equipment. The results of present day fieldwork, combined with the reconstructed data from sound archives, provide important data for language descriptions, grammars, dictionaries and edited collections of oral and written literature. The results can also be used to develop teaching methods, in particular for the younger members of certain ethnic groups who do not have a sufficient knowledge of their native language and to make them aware of their heritage. In this way, the Russian Federation can develop a basis for multi-ethnic co-existence of Russians and the many other ethnic groups in this country. In certain cases the revival of endangered languages and cultures will be possible and for this purpose the work of linguists and ethnologists is of inestimable value.

6. CONCLUSIONS

The various sound archives described in this paper present a wealth of information on the languages of the Russian Federation. They are not only important for phonetics, but also for a number of related disciplines like language description, ethnography, ethnomusicology, folklore and the study of literature. The preparation of these archives is based mainly on detailed phonetic work by specialists, which consists of the segmentation of recorded sound samples, the perceptive analysis of recorded texts in order to provide a transcription, resulting in a database which is exhaustive with respect to phonetic specifications of the languages concerned. The solution of the problems involved in creating this database will contribute to the further progress of the phonetic sciences.

REFERENCES

5. Bondarko, L.V., Kuznetsov, V.I., Stepanova, S.B. Teoreticheskie i prakticheskie problemy transkripcii teksta, Problemy fonetiki, I Moskva, 1993, pp. 8-20