Phonetic Decomposition for Speech Recognition of Lesser-Studied Languages

Vijay John
Department of Linguistics
University of Texas
Austin, Texas
vijayjohn@mail.utexas.edu

ABSTRACT

This paper deals with voice transcription systems for lesser-studied languages. In particular, it deals with creating phonetic decompositions for words in these languages, an important step in creating a voice transcription system. The two languages cited here as examples of lesser-studied languages are the San Juan Quiahije variety of Chatino and Vlax Romani. For these languages, an English-based phonemic decomposition is inadequate, because lesser-studied languages are not written with the orthographic rules used for English. The phonemic decomposition proposed here is composed of two stages: separation of words into sounds and expansion of sounds into phonemes.

Keywords
Phonetics, Sphinx, Chatino, Romani, Speech

INTRODUCTION

Lesser-studied languages, such as Native American languages, are typically spoken in environments where a more dominant language, such as English or Spanish is the main language used for communication. This limits the ability of speakers of the lesser-studied language to participate in social and economic activity. For these and other reasons, use of the lesser-studied language is typically discouraged. This can contribute to the disappearance of some languages. However, creating voice transcription systems for these languages could facilitate intercultural communication between speakers of lesser-studied languages and speakers of more dominant languages. Voice transcriptions can also help in preserving oral accounts in many lesser-studied languages are in danger of extinction. Even for widely spoken lesser-studied languages, political weakness of speakers of the language can be remedied to some extent by converting oral information into written text through voice transcription.

The University of Texas at Austin is one institution with linguists working on a wide variety of lesser-studied languages, most notably Native American languages but also Romani and lesser-studied languages of Asia. This paper concerns the development of voice transcription systems for two languages. The first is San Juan Quiahije Chatino (henceforward SJQ). Chatino is a tonal Indigenous ("Native American") language within the Zapotecan language family and is spoken in Oaxaca in southern Mexico. There are many varieties of Chatino spoken in different towns, and they are not necessarily mutually intelligible; SJQ is one variety of Chatino. The second is Vlax Romani. Romani is the language spoken by the Romani (or "Gypsy") people, especially (and originally) in Europe. Vlax Romani is the most widespread dialect of Romani and originates in Romania (but is also spoken elsewhere, including outside of Europe).

We use the Sphinx speech recognition system from Carnegie-Mellon University. Although Sphinx is generally used for recognizing English, it has been used in recognizing speech in other languages, e.g. Mexican Spanish [1], Mandarin Chinese [2], and even some Indigenous American languages [3]. In developing recognizers for SJQ and Romani, we need to develop phonetic decomposition models for words in these languages.

Phonetic terminology used in this paper may be found in most phonetics textbooks, for example [4].

Sphinx is trained using acoustic models developed from transcribed audio corpora. Part of the acoustic model is a dictionary that decomposes words to be recognized by the speech recognizer into phones. Sphinx recognizes these phones in new speech by relating each phone with some probability; this is used along with more information about the language to recognize this speech. Carnegie Mellon has developed a tool [5], called lmtool, for creating, among other things, a phonetic decompositions of words. However, this tool seems to use English-based

pronunciations from a large pronunciation dictionary (cmudict). (Incidentally, our conclusion is based on publicly accessible versions of the lmtool script, such as Simple_LM available through SourceForge.) This does not result in the proper phonemic decomposition for words in SJQ, Romani or other related languages. A further complication is that SJQ as well as other languages use tones that are relevant in proper recognition of the spoken words.

While it is possible to develop pronunciation dictionaries for each language, this is a time-consuming process. Instead we have developed a method that allows us to realize phonetic decomposition through a two-stage process. In the first stage, we use rules developed based on linguistic properties of the relevant language to create a decomposition of written words into sound units. The next stage decomposes these sound units into sets of phones. By breaking down words into sound units, the problem of phonetic decomposition is reduced to a problem of mapping a finite number of sounds into phones. This list is considerably smaller than a pronunciation dictionary such as cmudict.

PHONETIC DECOMPOSITIONS FOR ROMANI AND SJQ

The phonemic decompositions for Romani and SJQ are done in two stages. We use information from [6] for SJQ and from [7] for Vlax Romani. Words from these languages are written through transliteration into Romanized letters. In the first stage of decomposition, this transliteration is decomposed into sounds. In the second stage the sounds are later decomposed into phones.

Examples

One example of a word-to-sound decomposition in Romani would be the following: The Romani verb-stem *ker-* means "to do" or "to make." The verb *kerel* means "does, makes" (i.e. the third person form of the verb "to do" or "to make," as in "he does," "she makes," etc.). It would be decomposed into three parts:

kerel -> ker, e, l

In the IPA or International Phonetic Alphabet [4], each one of these parts would be [kəɾ], [e], and [ɬ].

In this example, the 'e' within the verb-stem 'ker' is phonetically different from the second 'e' in the word. After the above decomposition, we can map the last two parts into Romani phonemes (/e/ and /ɬ/). However, the first part is made up of three sounds: [k], [ə], and [ɾ]. Breaking up ker- into these three phonemic components would be an example of a sound-to-phone decomposition.

In tonal languages, phonetic decomposition may involve consideration of tones as well. This is the case for SJQ. An example of word-to-sound and sound-to-phone decomposition would be the following case in SJQ. (Note that here, the transliteration involve several numbers indicating phonological properties of the spoken word.) For example, the word *n7a3ki732* means "kitchen" in Chatino. It is a word composed of two syllables (*n7a3* and *ki732*), where the number at the end of each syllable (3 or 32) identifies the tone of the syllable and the number "7" represents a glottal stop (based on the IPA symbol [ʔ]) [4][6]. The transliteration-to-sound decomposition of this word would be as follows:

n7a3ki732 -> n7, a, 3, k, i, 7, 32

In IPA, the pronunciation of these parts would be represented as [nʔ], [a], [k], [i], and [ʔ]. Note that /nʔ/ is a phoneme in Chatino that is contrastive with /n/ and /ʔ/. Note that "3" and "32" are tone numbers, which are not represented in the main set of IPA symbols.

The sound-to-phone decomposition would be:

n7, a, 3, k, i, 7, 32 -> n, 7, a, 3, k, i, 7, 32

In this case, the combination n7 was broken up into two parts: [n] and a glottal stop [ʔ].

However, in Chatino (as in many other tonal languages), tones often change in the presence of other tones. This phenomenon (change in tone) is known as *tone sandhi*. To account for tone sandhi, it is necessary to implement additional rules in the program to do the phonetic decomposition. In the example presented above, no tone sandhi takes place. However, the following case based on an example from [6] is another example of word-to-sound and sound-to-phone decomposition in Chatino, this time including tone sandhi: The word for "you ground" (where "ground" is the past tense of "grind") in Chatino is *yo1*, and the word for "tortilla" is *yja4*. But when the two words are combined to mean "you ground tortillas," they are pronounced *yo1 yja24*, with the tone "4" in *yja4* changing to another tone "24."

To determine the underlying tones of *yo1 yja24*, it is necessary to include a rule that tone "4" changes to tone "24" after tone "1." With this rule, the transliteration-to-sound decomposition of this sentence would be:

yo1 yja24 -> y, o, 1, y, j, a, 4

In IPA, these parts would be [j], [o], (tone number 1), [j], [h], [a], (tone number 4).

Phonetic decomposition method

For each language, we create tables consisting of two columns. The first column contains sound patterns in that language. The second column gives a phonetic decomposition of that sound pattern. The actual tables used for the decomposition are somewhat large (of the order of hundreds of patterns). We show a part of the table here for each language.

In Romani, many of the patterns are single letters, but a few patterns involve multiple letters.

č	[tʃ]
čh	[ʂ]
dž	[ʐ]
kh	[k^h]
ph	[p^h]
th	[$t̪^h$]
z	[z]
ž	[ʒ]
ker	[k], [ə], [ɾ]
ljumja	[l^j], [ɨ], [m], [j], [a]
muca	[m], [ɨ], [ts], [a]

In SJQ, the tables include combinations of glottal stops (which are represented by the number "7") and other letters. SJQ words may contain other numbers, which represent tones; these numbers are not included in the patterns below.

tykwa	[t^j], [k^w], [a]
7	[ʔ]
7n	[ʔ], [n]
7ny	[ʔ], [n^j]
ly	[l^j]
7w	[ʔ], [w]
7y	[ʔ], [j]
7nyo	[ʔ], [n^j], [o]

The method for decomposing words into sounds is based on the one in [8]. We find it convenient to work from the end of the string. Generally speaking, single letters do represent sounds, unless they occur in combination with something else. The method is a finite state machine, but it avoids order dependence that is frequently a problem in linguistic finite state machine implementations. Effectively, we try to match the string with the longest pattern we can find in the table of sound patterns.

CONCLUSION

Using this method, we can use the tables of patterns to decompose words. The following shows one decomposition in Romani:

ljumjatar [l^j], [ɨ], [m], [j], [a], [$t̪$], [a], [ɾ]

Similarly, the following is an example produced from the SJQ table:

7nyo24=tykwa20 [ʔ], [n^j], [o], [t^j], [k^w], [a]

In the implementation for Sphinx, the IPA symbols and tones (such as the tones "24" and "20" in the above example) are represented using ASCII letters based on the ARPAbet.

Even though the methods here were applied to just one dialect of Chatino and one (albeit main) dialect of Romani, the results are applicable elsewhere. To model other dialects or languages, one needs to only change the table of sound patterns and phones.

Note that the standard lmtool will produce entirely different and incorrect results on these words. This, however, is not unexpected since the standard lmtool is based on the English pronunciation dictionary.

While it appears that most words are decomposed correctly, there are some incorrect phonetic decompositions with our method. Exceptions can be added to the table as new rows. Thus, the accuracy of these methods is being continually improved.

FURTHER WORK

While the rules developed to produce phonetic transcriptions are specifically for the SJQ and Vlax Romani, they are in a form that can be modified to suit other related languages. For example, SJQ is just one of the varieties of Chatino. Different varieties of Chatino differ slightly in terms of their phonemic inventory and tonal characteristics, but rules for a variety of Chatino other than SJQ can be produced quickly by altering the tables we use for phonetic transcription. Similarly, there are different versions of Romani that differ from the Vlax Romani considered here. For example, the word for "house" in Serbian Romani differs phonetically from the word in Vlax Romani. However, it is easy to adapt the phonetic transcription rules for Vlax Romani to create one for Serbian Romani.

The methods here also apply to other commonly studied languages such as Mandarin Chinese and Hindustani (a.k.a. Hindi/Urdu). For Mandarin Chinese, tone sandhi is relevant, though tones change less often in Chinese. However, in Hindustani (as in Romani), there are no tonal variations to be considered because Hindustani is not a tonal language.

The rules indicated here are adequate for phonetic decomposition of most words in the respective languages. There are, however, some exceptions, e.g. the phonetic decomposition of the word *ljumja* in Vlax Romani is a special case. There are other special cases that need to be considered and incorporated into the phonetic decomposition system.

ACKNOWLEDGMENTS

I wish to thank Tony Woodbury, Emiliana Cruz and Hilaria Cruz for providing me with materials on SJQ, for rerecording some audio files, and for enthusiastic support for using Sphinx. I also wish to thank Ian Hancock for providing me with materials on, and introducing me to and familiarizing me with, Romani. I would also like to thank him for access to the Romani Archives.

REFERENCES

1. A. Varela, H. Cuayáhuitl and J.A. Nolazco-Flores, "Creating a Mexican Spanish Version of the CMU Sphinx-III Speech Recognition System", Progress in Pattern Recognition, Speech and Image Analysis, 251-258, Springer 2003
2. Hsiao-Wuen Hon; Baosheng Yuan; Yen-Lu Chow; Narayan, S.; Kai-Fu Lee, "Towards large vocabulary Mandarin Chinese speech recognition," Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., vol.1, 19-22 Apr 1994
3. Juan A. Nolazco-Flores , Luis R. Salgado-Garza and Marco Peña-Díaz, "Speaker Dependent ASRs for Huastec and Western-Huastec Náhuatl Languages," Lecture Notes in Computer Science: Pattern Recognition and Image Analysis, 595-602, Springer, 2005
4. Peter Ladefoged, A Course in Phonetics, 4th Edition, Heinle & Heinle, 2001.
5. Carnegie Mellon Speech Group, Sphinx Knowledge Base Tool, http://www.speech.cs.cmu.edu/tools/lmtool.html and Simple LM from http://sourceforge.net/project/showfiles.php?group_id=1904
6. Emiliana Cruz and Tony Woodbury, "El sandhi de los tonos en el Chatino de Quiahije," Las memorias del Congreso de Idiomas Indígenas de Latinoamérica-II. Archive of the Indigenous Languages of Latin America, 2006. http://www.ailla.utexas.org/site/cilla2/ECruzWoodbury_CILLA2_sandhi.pdf.
7. Ian Hancock, A Handbook of Vlax Romani, Slavica, 1995.
8. Vijay John. A Method for Enhancing Search Using Transliteration of Mandarin Chinese. Texas Linguistics Society 10, 2006 University of Texas. http://uts.cc.utexas.edu/~tls/2006tls/papers/john_tlsx.pdf.