ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken
Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

GRAPHEME-TO-PHONEME CONVERSION IN CHINESE TTS SYSTEM

[†]*Honghui Dong*, [‡]*Jianhua Tao*, [§]*Bo Xu*

[†‡]National Laboratory of Pattern Recognition, [§]High Technology Innovation Center
Institute of Automation, Chinese Academy of Science
{[†]hhdong, [‡]jhtao}@nlpr.ia.ac.cn, [§]xubo@hitic.ia.ac.cn

ABSTRACT

Phonetization is an important component in Chinese TTS system. However the polyphonic characters make this problem more complex. This paper reports the study on the relation between the Chinese characters and their pronunciation, proposes the solution to the disambiguation of polyphonic characters, dictionary-based method, and rules-based method. In the rule-based method, we used the statistic decision list method. The phonetization plan is proved effective in the experiment. Most of the improvements on the accuracy of polyphone phonetization are beyond 10%.

1. INTRODUCTION

Grapheme-to-phoneme conversion plays an important role in speech synthesis. Normally, we get the pronunciation of non-polyphonic characters from the lexicon directly. However, it is more difficult for the prediction of polyphonic characters or words, in that we don’t know if the pronunciation is correct or not without any help from context information

Previously, the polyphone is resolved as a homograph disambiguation. Yarowsky[1] proposed a efficient method for English homograph disambiguation. Decision lists were used to decide the context of the polyphonic word. However, Chinese is much different than other languages. Word boundaries and Part of speech are the major reasons for pronouncing ambiguity. Some previous works were more focused on the rule-based method [2][3]. The pronunciation correspondence to the Chinese characters has itself distinguishing characteristics.

The paper makes the statistic analysis of the relationship between Chinese transcription and their pronunciation with the mode of pinyin, and presents the plan of the grapheme-to-phoneme conversion. In the analysis, we focus on the polyphonic characters, since the polyphonic words have the same characteristic instead of word segmentation feature.

The whole paper is organized as following. Section 2 analyzes the correspondence relationship between the Chinese characters and their pronunciations within different linguistic features. Section 3 proposed the methods to phonetize Chinese characters, and to disambiguate the polyphonic characters and words. The experiment results are provided in section 4. In section 5, we discuss the results and give the conclusion.

2. STUDY ON CHINESE CHARACTERS AND THEIR PRONUNCIATIONS

Generally there are some special relationship between the Chinese characters and their pronunciations, between the linguistic features of the characters or words and their pronunciations. The mapping relationship between them, maybe bring out some ideas on how to phonetize Chinese characters and how to resolve the problem of polyphonic characters. The distribution of polyphonic characters/ words and the frequency of the pronunciations in the polyphonic characters/words must also be helpful for polyphone disambiguation.

2.1. Mapping between the character and the pinyin

In Chinese, there is a correspondence between the characters and their pronunciations, sometimes one-to-one, and sometimes one-to-more relation, in which the characters are called polyphonic characters. We also studied the mapping relationship between the polyphonic characters/ words’ pronunciation and their morphological features. Due to the high accuracy of the morphological analysis, we only consider two linguistic characteristics, word segmentation and POS tagging.

In this study, we discovered that the pronunciations of some polyphonic characters are highly relative to the word segmentation and POS tagging. Figure 1 is an example which describes the correspondence relationship between the pronunciation and word segmentation. The pronunciations of character ‘的’ are directly relative to the words composed of it. When ‘的’ is used as a word on its

Supported by Knowledge Innovation Programs of Chinese Academy of Science

own, it is pronounced ‘de5’ . But it is pronounced ‘di2’ in word ‘的确’, ‘di4’ in word ‘目的’. Another example is “阿” which is usually pronounced as “a1”. It has another pronunciation ‘e1’ only appearing as the words, such as “阿谀奉承”, “阿弥陀佛”, etc.. There are many characters like this, such as “否大咖…”. This is a one-to-one relation. It’s to say, we can decide their pronunciation through the word composed of it.

Part of speech also has the same characteristic, with different POS corresponded to different pronunciation. Like “长”, it is pronounced “zhang3” as a verb word, and “chang2” as a adjective word as described in figure 2. On the other hand, there are still 334 polyphonic characters and 27 polyphonic words which can not be distinguished by their POS.

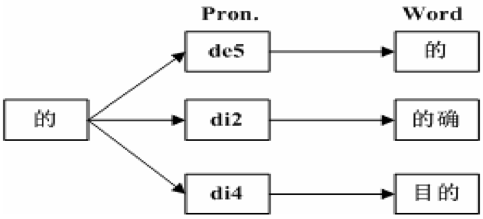

Figure 1. 的’s pronunciations and the words with it

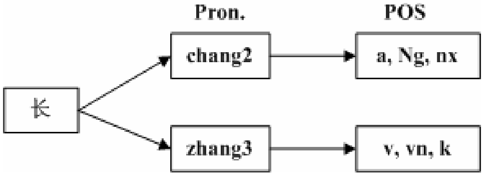

Figure 2. 长’s pronunciation and its POS

2.2. The distribution of polyphones

In our lexicon, there are 851 polyphonic characters and 54 polyphonic words. Nevertheless, in our corpus these polyphonic characters or words are over 10%, most of which are polyphonic characters. There are two kinds of distribution of polyphonic characters/words needing to study, the distribution of every character/word’s frequency and the distribution of pronunciations in each polyphonic characters or words, which are very useful for polyphonic characters/words process.

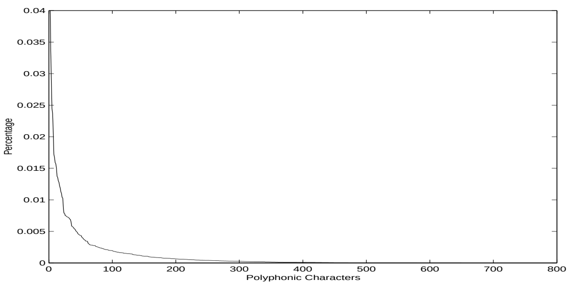

Figure 3. The distribution of the polyphonic characters.

We counted the 20 years people daily, and discovered that the frequencies of the polyphonic characters are very different between each other. The distribution of all polyphonic characters’ frequency is shown in figure 3. We found that all polyphone focus on some characters. About the first 220 polyphonic characters take the 95% in all polyphonic characters. And the first 30 polyphonic words take the 98% in all polyphonic words. These polyphonic characters and words are the emphasis we should pay attention to in the polyphonic characters and words process.

Table 1 describes the pronunciations’ distribution in the polyphonic characters. In some polyphonic character the high-frequent pronunciation take more than 98% in all its pronunciation, like “说”.

	default (%)	others (%)
了	93.4 {le5}	6.6
上	80.8 {shang4}	19.2
着	85.3 {zhe5}	14.6
为	65.1 {wei2}	34.9
得	76.1 {de5}	23.9
说	98.5{shuo1}	1.5
……	……	……

Table 1. Distributions in the polyphonic characters

2.3. Extracting more complex polyphones

From the analysis above, it can be concluded that there are so many polyphonic characters and words that we can not analysis every polyphonic character at the same time. We need consider the most necessary Chinese polyphonic characters which take the most percentage, and are more difficult to process. Basing on section 2.1 and 2.2, we make the principles for extracting the polyphonic characters and words as follows.

1. It must be in the first 98% of all polyphonic characters or words.

2. The high frequency pronunciation must be less than 98%. If the percentage of the high frequency pronunciation is very high, we can consider this character can only pronounce one case. Why not? It pronounces one pronunciation almost all the time.

3. Its pronunciations can not be distinguished by word segmenting.

Why we don’t consider whether the word POS can classify the pronunciations as a principle of extracting the polyphonic characters/words? Although the accuracy of word segmentation is very high, the POS tagging is relative not satisfied. If we used the POS as a distinguishing measure, the wrong POS tagging may bring some new false results.

According to these principles, we extracted 45 Chinese polyphonic characters and 24 Chinese words as the analysis emphases. These Chinese polyphonic characters are "了上为着得过还看长种…". And the Chinese polyphonic words are "转向，好事，东西，倒车，播种，地方，调配…". We will study these characters and words to build the suitable rules, and use other methods to phonetize the correct pronunciation.

3. METHODS OF GRAPHEME-TO-PHONEME CONVERSION

From the analysis in section 2, we have known word segmentation and POS tagging information play the important roles in disambiguation of the polyphonic characters. We used the two methods for Grapheme-to-phoneme conversion. (1) dictionary-based method. (2) rule-based method.

3.1. Dictionary-based method

Most of the Chinese characters are non-polyphonic characters. So the pronunciation of these words can be phonetized directly through the dictionary.

And this method can also resolve the problem of the polyphonic characters which can be distinguished by word segmentation. This demands that the non-high frequency pronunciation words are at the most included in the dictionary. We can get the correct pronunciation for the polyphonic word through the mapping between the pronunciation and the words in the dictionary. And the lexical analysis system should recognize the names, which can resolve the surnames' pronunciation, like "曾，区，仇，查，单……".

3.2. Rule-based method

```
if(word == "待" && the behind word == "在" or "了")
{
        待's pronunciation = "dai1";
}
else if(word == "待" && the behind POS == "verb" or "nr")
{
        待's pronunciation = "dai4";
}
if(word == "得了" && the previous POS == "r" or "nr")
{
        了's pronunciation = "le5";
}
else if(word == "得了" && the previous POS == "verb" or "ad.")
{
        了's pronunciation = "liao3";
}
```

Figure 4. Examples of hand-crafted rules

3.2.1. Hand-Crafted rules

For the polyphonic characters and words extracted, we built some rules according to the context information of each special pronunciation. These hand-crafted rules are based on the keywords and the POS information, which can represent the special context.

Figure 4 shows some rules on the polyphonic character '待' and '了'. In our system, we built about two hundred rules like this for the polyphonic characters and words.

3.2.2. Statistical Decision Lists

Although the hand-crafted rule method is an easy and efficient way, it has some disadvantages. First, the hand-crafted rules construction is costly. Second, rule interactions are hard to manage. Third, the hand-crafted rule systems usually output all possible results without associated weights [6].

Due to the disadvantages of hand-crafted, we use statistical decision lists in our system. Decision lists method is efficient and flexible in using data, easy to interpret and modify, and can handle both wide and narrow context information [1]. Homograph disambiguation using decision lists method and decision lists learning had been described in detail in [1] and [5]. This method includes the following steps:

- Corpus collection and labeling: For each polyphonic character or word, we collected all instances from a large text corpus, 20 years' People Daily. Then we label each example of the target polyphonic characters/words with its correct pronunciation in that context. Now we have got enough labeled corpus of some polyphonic characters for statistic decision lists analysis. These polyphonic characters are: "了长为还过". The corpus labeling of other polyphonic characters and words is being continued.

Position	Collocation	zhang3	chang2
Word to the left (+1 w)	最长 太长 米长	0 0 0	43 20 24
Word to the right(+1 w)	长大 长时间	249 0	0 84

Within ±k words	高(in +5 words)
	路(in ±5 words)
		107
		0
			0
			85

Table 2. 长's collocation distribution

- Measuring collocation distribution: We classified decision lists into three: the left word to the polyphonic character, the right word to the polyphonic character, and widely context of the polyphonic. We counted the most likely collocation words of the polyphonic character. Table 2 is an example how to measure the collocation distribution of the polyphonic character ‘长’.

- Likelihood Ratio Computation and Using the Decision Lists:
  We use the following formula to compute the likelihood ratio of the evidence in the decision list [1].
  $$LogL(i) = Abs(Log(\frac{P(pronunciation_1 \mid collocation_i)}{P(pronunciation_2 \mid collocation_i)})$$
  Smoothing method we used is plus a very small number $\alpha = 0.1$ to every collocation count. The more the likelihood ratio of the evidence in the list is, the more reliably does it disambiguate the target character/word. We arrange the evidence in the likelihood ratio order from high to low. When a word in a new context is to be assigned a pronunciation, the evidence’s likelihood ratio should be the highest in the list. Only the single most reliable evidence matched in the target context is used.

Decision List for “长”		
LogL	Evidence	pron.
7.82	长+大	zhang3
6.98	高 in +5 words	zhang3
6.75	路 in ±5 words	chang2
6.73	长+时间	chang2
6.07	最+长	chang2
5.48	米+长	chang2
5.30	太+长	chang2

Table 3. The decision list for “长”

4. EVALUATION

Word	Pron1	Pron2	Sample Size	Prior Prob.	% Correct
了	le5	liao3	5000	93	99.2
长	zhang3	chang2	400	58	93.1
为	wei2	wei4	900	57	83.5
还	hai2	huan2	1700	92	96.5
过	guo5	guo4	2000	85	95.3

Table 4 Experiment results of some polyphonic characters

Table 4 gives some cases of our system’s performance. We have had enough corpuses for the polyphonic characters. We take one part of these corpuses for testing and the left for training the decision lists. Because the size of corpuses is different between the polyphonic characters, the sizes of testing corpus are very different. The prior probability represents the ratio of the high frequency pronunciation. The accuracy of the pronunciation of the polyphonic characters is improved relative to the probability of the high frequency pronunciation. Most of the improvements are over 10%. Especially, the accuracy of ‘长’, one of the most hard-handling polyphonic characters, reaches 93.1%.

5. DISCUSSION AND CONCLUSIONS

Through analysis on the relationship between the Chinese characters and the pronunciation, we designed the method of grapheme-to-phoneme conversion: word segmentation, hand-crafted rules and statistical decision lists method.

From the result of experiment provided in section 4, our plan for the Chinese character phonetization is proved successful, though we only got several polyphonic Chinese characters’ corpus. We still have a lot of work to do, building enough corpuses for all polyphonic Chinese characters and words to be disambiguated, and building the decision list for them. With the rise of the corpuses’ size, the phonetization accuracy will be improved further. The method of disambiguation used in this paper can also be extended to other classification problems in TTS text analysis, such as distinguishing between fractions and dates, deciding the pronunciation mode of numbers.

Although our phonetization system can make a satisfying result, the main method in our plan, decision list, only uses some partial cues, like key words. The combination of the context information must be more useful for the polyphonic characters disambiguation. This is the direction we should do in the future.

6. REFERENCES

[1] Yarowsky, D., Homograph disambiguation in speech synthesis. In J. van santen, R. Sproat, J. Olive and J. Hirschberg (eds.), Progress in Speech Synthesis, Springer-Verlag, 1997, pp. 159-175.

[2] Zhang Zirong, Chu Min, A Statistical Approach for Grapheme-to-Phoneme Conversion in Chinese, Journal of Chinese Information Processing.

[3] Zhang Hong, Yu Jiangsheng, Zhan Weidong, Yu Shiwen,, Disambiguation of Chinese Polyphonic Characters, The First International Workshop on MultiMedia Annotation (MMA2001), 2001. 1. 30 - 1.31, Tokyo

[4] Liu Kaiying, Automatic Word Segmentation and Tagging of Chinese Texts, the Commercial Press. Beijing, May 2000.

[5] Rivest, R., Learning decision lists, Machine Learning, 2 (1987), 229-246.

[6] Michael Riley, Richard Sproat, Text Analysis Tools in Spoken-Language Processing, Tutorial at ACL 1994.