ON PHONEME-TO-CHARACTER CONVERSION SYSTEMS IN CHINESE PROCESSING
中文語言處理中的音轉字問題

Wen-Lian Hsu and Yi-Shiou Chen
Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.
許聞廉，陳逸修， 中央研究院資訊所

Keywords: phoneme-to-character, conversion, natural language understanding
關鍵詞：音轉字，轉換，自然語言理解

Abstract

We propose a new goal for constructing a Chinese phoneme-to-character automatic conversion system. Instead of trying to get the largest number of correct characters converted, this goal strives for extracting the correct events from a given sentence. We believe this objective is more fruitful in that the extracted events can be better utilized in other applications such as information retrieval, extraction and dialogue systems. Furthermore, it would motivate researchers to work on the more challenging problem of natural language understanding.

An information map is proposed for knowledge representation. These maps give very concise summary of different aspects of an object. They also provide more convenient ways to implement semantic template matching.

摘要

我們對於「音轉字」系統提出一個新的目標。就是希望從一個音節所構成的句子中擷取其「事件」，而非計算字的轉換正確率。我們認為這個新目標更有價值，因為擷取出來的「事件」還可供其他應用系統（如：資訊檢索，抽取以及對話系統）使用。同時，這樣的目標也較容易讓這個領域的學者的研究轉向自然語言「理解」，而不再沈醉於純統計式的「轉換正確率」競賽。

我們提出以「資訊地圖」作為知識表達的方法。「資訊地圖」能夠對事物的不同特性歸納，並且在實際操作時加快語意模版的對應。

1. Introduction

In Hsu (1995), a context sensitive method was developed for constructing a Chinese phoneme-to-character (*PTC*) automatic conversion system called "*GOING*" with a conversion rate close to 96%. There have been a number of approaches for the

PTC system [1,2,3,5,6,7,9,10,13,15,17,19]. Besides a few attempts on using grammar or semantic analysis [6,9,10,13,15,17], most of these methods are based on the manipulation (for example, using dynamic programming, Hidden Markov Model) of word frequency and k-gram. We believe that it is time to rethink about the goal of the *PTC* system. The traditional measure of the word accuracy rate of the conversion seems to be too myopic since the ultimate goal of the *PTC* system is to understand what the user is trying to convey in the message. A system that gets the best word conversion rate is not guaranteed to know anything about user's intention. The latter is particularly important in most speech recognition system because, eventually, speech is the best communication mode for human-computer interaction. In this paper we shall discuss a new goal for the *PTC* system.

Before delving into the detailed discussion of the *PTC* systems, let us consider several other application systems that are similar in nature to the *PTC* system. These are Chinese text-to-speech (*TTS*), English style checker (*ESC*) and English-to-Chinese machine translation (*MT*). There is a common feature among all of these systems: it is crucial for the system to resolve ambiguity based on the context. In the following we briefly describe these systems and their similar features. In a Chinese *TTS* system, one needs to determine the correct pronunciation for each character in a sentence since many Chinese words have more than one pronunciation (such as "為", "行", "得"). Furthermore, the system has to adjust the pitch and duration of each syllable based on the context. In an English style checker, each word or phrase has to be checked against a set of alternative word or phrase candidates to determine which should be the best choice. In an English-to-Chinese *MT* system, each English word has around 5 different meanings (or Chinese translations), the *MT* software must choose the most appropriate Chinese translation based on the context. Plus, the system should arrange for the necessary word permutation to conform to "regular" Chinese word order.

The ambiguity resolution in these systems can be divided into several layers: sentence layer, phrase layer, and word layer. Our approach for the *PTC* system is based on the consideration of the disambiguation issues related to all of the above applications. In fact, we think any approach that can simultaneously handle all of the problems in the above systems is really worth investigating. An approach like that is yet to be discovered. However, we try to set a goal for our PTC system that would potentially lead to such an approach.

2. Comparison among Various Methods for the *PTC* system

There are basically three methods for the *PTC* system, namely, statistical methods based on *k*-gram, rule-based methods and probabilistic template-based

methods. Below, we briefly describe these methods.

Most statistical methods employ the Markov language model. Given a word sequence $W_{1,n} = w_1,w_2,\ldots,w_n$, the language model probability for this sequence is

$$\Pr(W_{1,n}) = \prod_{i=1}^{n} \Pr(w_i|W_{1,i-1})$$

To reduce the number of parameters, *k*-gram model assume the probability for the current word w_i depends only on its previous *k*-1 words $W_{i-k+1,i-1}$ and equation (1) becomes

$$\Pr(W_{1,n}) = \prod_{i=1}^{n} \Pr(w_i|W_{i-k+1,i-1})$$

When *k* is chosen to be 2 or 3, the *k*-gram model is referred to as *bigram* or *trigram*. *K*-gram statistics are collected for word association relationships in a large corpus. These statistics can be used quite effectively for the analysis of speech-input as shown in Wang et al. (1997) and the techniques are largely language independent. Furthermore, statistical methods require much less human effort and can be quickly adopted to attain an acceptable accuracy in laboratory experiments. Therefore, these methods have received wide attention in computational linguistics. They have also been widely used in existing commercial products. An extension of k-gram to long-distance modeling for the *PTC* system has also been explored in Ho et al. (1997). However, human users are very difficult to be pleased with and there are a lot of common sense not captured by statistics, which lead to the following problems:

1. Since statistical methods cannot make use of detailed semantic information, it is difficult to obtain a highly accurate, satisfactory system. Basically, certain errors committed by the computer are considered excusable but others are considered ridiculous. It is difficult for statistical methods to avoid making ridiculous errors. Furthermore, user adaptation is sometimes difficult since it is hard to weigh the little statistics gathered from user correction against those from the large corpus.
2. In the selection of homophones, statistical methods normally use a scoring mechanism to obtain a word sequence with the largest overall aggregate probability. Without much semantic guidance, such a sequence could be highly erroneous at times. Furthermore, it is difficult to guarantee that any specific character or word is correct.
3. The statistics gathered depends too much on the training corpus. The recognition result would degrade drastically when applied to an untrained domain. Nevertheless, it is impractical to store too many different statistics for different domains. Moreover, the statistics gathered is very much application-dependent.
4. An ideal *PTC* system has to perform the following tasks well: unknown word recognition, proper name identification, word segmentation and automatic

learning. Different statistical methods are appropriate for tackling different tasks. It would be awfully difficult to combine these various statistical methods to simultaneously overcome all of the above problems.

On the other hand, the traditional rule-based method, seems to be more appealing to a human designer. Linguistic rules are constructed by specialists, which cover a wider domain than the training corpus. These rules and their resulting *PTC* systems tend to make more sense for human beings and can be adjusted quickly. However, they suffer from the following drawbacks:

1. Rule-based methods are too labor-intensive and rule construction requires extensive linguistic training.
2. Rule consistency is difficult to maintain even for the same designer. Furthermore, common sense is often difficult to encode.
3. The rules designed by different experts could sometimes contradict with each other and thus, affect the overall system performance.

A third alternative is the *probabilistic template-based method* (*PTM*). An example is the approach used in the *GOING* system of Hsu (1995). *PTM* is based on the pattern matching of a large collection of semantic "templates". Each template is a description of the occurrence of a word or a collection of words. The structure of a template is a hybrid combination of syntactic and semantic features. Such a description could be either local or global. In *PTM*, linguistic information as well as common sense are modeled as patterns in much the same way as those used in image processing. Templates are different from rules in that many of them are quite independent. They are divided into layers those within the same layer can be processed in parallel. It is their aggregate weight that affects the homophone selection rather than their individual logical implication.

Some examples of templates are given below for the *PTM*. In Chinese, different measures are used in accordance with the nouns they specified. Consider the following sentences, "A very lovely cat", "A very lovely flower" and "A very lovely candle". The same quantifier "a" is used with all three nouns in three English sentences, but their counterparts in Chinese are all different.

a	very	lovely	( )
一 隻	很	可 愛 的	貓
yi4 zhi1	heng3	ke3 ai4 de5	mou1 (cat)

{ determinative | 隻 (zhi1) | adv. | adj. | 的 | noun[animal] }

一 枝 很 可 愛 的 花
yi4 zhi1 heng3 ke3 ai4 de5 hua1 (flower)
{ determinative | 枝 (zhi1) | adv. | adj. | 的 | noun[plant] }

一 支 很 可 愛 的 蠟 燭
yi4 zhi1 heng3 ke3 ai4 de5 la4 chu2 (candle)
{ determinative |支 (zhi1) | adv. | adj. | 的 | noun[artifact] }

There are several advantages of the *PTM*:

1. Each template when matched successfully will usually have a very local effect. Unlike a rule-based system that adopts rules which cover a wide scope, the *PTM* used detailed local templates whose probabilistic weights are properly (often statistically) controlled in such a way that the effect of an incorrectly matched template can be minimized. Similarly, the emergence of a correct selection will often rely on more than one local template be matched successfully (so that the aggregate weight is large enough to win).
2. Templates are divided into layers. These template layers form a pyramid shape with only a few at the top layer. All templates in the same layer are processed in parallel. A lower level conflict could be resolved in an upper layer provided there are enough structural templates in upper layers supporting the winner. Such a hierarchical structure conforms well with that of the sentential analysis from phonemes to characters, words, phrases, and then to the whole sentence. Thus, semantic information can be embedded into the template structures to enhance the analysis (for example, to carry out partial parsing) in the phoneme level.
3. The following important tasks such as unknown word recognition, proper name identification, word segmentation and automatic learning can all be accommodated by the template system in a cohesive way. Since semantic information is used pervasively in the *PTM*, the system is less domain-dependent.
4. The semantic template matching system adopted in *PTM* intends to simulate human understanding. Similar template-construction strategy can be applied to Chinese text-to-speech, Chinese parsing and English-to-Chinese machine translation.

Although the *PTM* seems to be a good solution for the *PTC* problem and its idea is general enough to be applied to other related applications, it still suffers from the following drawbacks. Template construction is time-consuming and their utility is application-dependent. For example, templates constructed for the *PTC* system in

general would not be useful for the *TTS* system nor the machine translation system. Even though the construction schemes are similar, one virtually has to start from scratch in rebuilding a different template system for a different application.

We believe that the main problem surrounding the statistical methods and rule-based approaches is that they do not attempt to "understand" the sentences they are processing. On the plus side, the *PTM* is different in this respect. But, even though the semantic templates of the *PTM* can capture certain semantic structure of the processed sentence, it would take an enormous amount of effort for the *PTM* template system to incorporate world knowledge into its analysis. A different way to put it is that, there has to be a more efficient way to implement the template system in *PTM*. Thus, in our new approach for dealing with such ambiguity resolution problems, we shall propose a new goal for the *PTC* system. Instead of getting the largest number of correct character converted, this goal strives for extracting the correct events from a given sentence. Our new approach is based on a model for concept understanding in Hsu et al. (1995, 1998), which is briefly described in the next section.

3. A Context Sensitive Model for Concept Understanding

We shall temporarily step away from ordinary linguistic issues and probe into the more philosophical aspect of Chinese language processing. People often wonder what can constitute "understanding" on the part of a computer software system. Although "Turing test" provides a paradigm for the "intelligence test", we believe a more pragmatic goal should be set for each software system that is currently being constructed to exhibit some sort of self-disambiguation ability. In Hsu et al. (1998), we have developed a context-sensitive model for concept understanding, which is based on query answering. We now briefly describe the basic framework for concept understanding in Hsu et al. (1998). It should be emphasized that, since this framework is to be implemented in a computer to build a simulation model, our definition has a strong "problem-solving" flavor. In other words, this approach is not meant to be a cognitive study of the true nature of human understanding, it merely represents a system designer's viewpoint for a plausible simulation.

Since a concept is not considered as an isolated notion, we shall refer to it as a *contextual concept*. Each contextual concept will be denoted by a pair, (concept name, query set), where the concept name *C* will be associated with a query set *Q*, in which every query has an answer (let us put aside the "correctness" issue for the time being). This *CQ*-pair serves as a description of the contextual concept. For example, the *frequently asked questions* (*FAQ*) about a concept could serve as a good example of the query set *Q*, though in practice the query set may be much larger. If the query set

Q is obtained from a particular person's view, we say that the *CQ*-pair is a description of the contextual concept *in the mind of* that person. Thus, it is appropriate to think that the query set provides the context for this concept.

Such a model is quite popular for testing human understanding. Most tests given to human beings can be regarded as sampled query sets for the understanding of certain topics. For example, each elementary school can conduct its own mathematics test for grade 3 kids to judge if they pass or fail (in the mind of the school). Next, consider the example of understanding a specific "person". For his colleagues, the queries could be related to his character, hobbies, or abilities. For his doctor, the queries could be related to his medical record. Thus, query set could contain actions, attributes and so forth. Different types of query sets are associated with different facets of the concept and could be best represented by different schemes or combinations thereof.

In *CSM*, we regard the ability to understand a sentence the same as that to extract the correct event about the sentence. Hence, the goal of our *PTC* system is not to calculate the portion of characters correctly converted, but to count the number of events correctly extracted by the system. The reason for adopting such an evaluation criterion is as follows. First of all, the correctness of other indicators in the *PTC* system depends heavily on the extracted events. Secondly, the final goal of the conversion system is to utilize the extracted events in other applications such as information retrieval, extraction and dialogue systems, where events, rather than characters, play a major role.

4. Knowledge Representation – The Information Maps

Our *CSM* is implemented based on concepts. A concept can be a word, a phrase, a sentence or even a paragraph. To represent a concept, we adopt the *frame* system of Minsky (1975) as well as the *script* system of Shank (1992). A frame typically is used to describe a class of objects. It consists of a collection of slots that describe aspects of the objects. Procedural information can be associated with the slots. We implement the procedural information using natural language scripts. Such scripts can refer to a collection of events and describe their relationships in more detail. In this aspect, a collection of event scripts is more powerful than an event net.

Each Chinese word is considered a word concept. We adopt the idea of How-net Dong (1999) in constructing our basic concepts. A phrase concept can be a frame consisting of several words satisfying certain semantic and syntactic structure. We developed time frame, name frame, location frame, NP frame and VP frame for a sentence. The VP frame is essentially the corresponding event of the sentence.

Furthermore, to assist common sense reasoning, we have also developed a large network of *knowledge frames*. These frames are instances of a general knowledge representation scheme called information maps. An *information map* is an abstract description about the relationships among several objects, both abstract concepts and real world entities. In the information map each node denotes a "concept". The relationships among these nodes are described through tree-like charts, natural language scripts as well as constraint satisfaction rules. Similar to a geographical map, an information map will guide us (and hopefully, the computer) to traverse through information nodes to perform many tasks such as proper name identification, *NP* identification, relation identification and anaphora disambiguation. An example for the event frame is as follows. There are several roles for each verb. The semantic category of each role is listed after "=".

[買(buy)]

1. agent = human, organization
2. possession = physical
3. source = human, place, organization
4. cost = wealth
5. beneficiary = human, animal, organization

An example of the time frame is as follows.

[年月日次序]

1.[幾年]
2.[幾月]
3.[幾日]
4.[日夜時段]
5.[幾時]
6.[幾分]
7.[幾秒]

(In English:

[the order of date, month and year]

1. [the hour]
2. [the minute]
3. [the second]
4. [the month]
5. [the date]
6. [the year])

An example of the address frame is as follows.

[門牌號碼次序]

1.{[某市]（院轄市、省轄市)、[某縣]}

2.{[某區]、[某市]、[某鎮]、[某鄉]}

3.{[某里]、[某村]}

4.[幾鄰]

5.{[某路]、[某街]}

6.[幾巷]

7.[幾弄]

8.[幾號]

9.[之幾]

10.[幾樓]

(In English:

[description of address]

1. [the number]
2. [the Section]
3. [the street]
4. [the city]
5. [the state]
6. [the zip code]
7. [the country])

An example of the NP frame is as follows.

[接量詞 NP]

1.ne

2.量詞

3.副詞

4.形容詞

5.的

6.{名詞、複合名詞}

(In English:

[NP with quantifier]

1. "a" or "the"
2. adverb
3. adjective
4. noun)

An example of the (partial) knowledge frame about the concept [汽車] is as follows.

[汽車]

1.使用方式

買車、賣車、修車、洗車、開車、賽車

2.汽車結構

汽車零件

3.汽車種類

以用途分類

自用、出租、計程、運貨
以型態分類
小客車、卡車、貨櫃車
以廠牌分類
福特，通用、裕隆、日產
以汽缸大小分類
(In English:
[the description of a car]
1. the utilization
buy, sell, fix, wash, drive, race
2. the structure
various parts of a car
3. different types
classified by usage
passenger car, rental car, tour bus
classified by outlook
sedan, coupe, truck
classified by production company
Ford, GM, Toyota, Nissan, Honda
classified by engine horse power
1500cc, 1600cc, 2000cc, 3000cc

REFERENCES

1. Chang, J.S., S.D. Chern and C.D. Chen, 1991, "Conversion of Phonemic-Input to Chinese Text Through Constraint Satisfaction," *Proceedings of ICCPOL'91*, 30-36.
2. Chen, K.J., 1993, "A Mathematical Model for Chinese Input," *Computer Processing of Chinese and Oriental Languages* 7, 75-84.
3. Chen, S.I, C.T Chang, J.J. Kuo and M.S. Hsieh, 1987, "The Continuous Conversion Algorithm of Chinese Character's Phonetic Symbols to Chinese Characters," *Proceedings of National Computer Symposium*, Taipei, 337-342.
4. Dong, Z.D., 1999, "How-net," http://www.how-net.com.
5. Fan, C.K. and W.H. Tsai, 1988, "Disambiguation of Phonetic Chinese Input by Relaxation-based Word Identification," *Proceedings of ROCLING I*, 145-160.
6. Gie, T.H., 1991, "A Phonetic Input System for Chinese Characters Using a Word Dictionary and Statistics," Master Thesis, National Taiwan University.
7. Gu, H.Y., C.Y. Tseng and L.S. Lee, 1989, "Markov Modeling of Chinese Language for Linguistically Decoding the Mandarin Phonetic Input, *Proceedings of National Computer Symposium*, Taipei, 759-767.
8. Ho, T.H., K.C. Yang, J.S. Lin and L.S. Lee, 1997, "Integrating Long-Distance Language Modeling to Phoneme-to-Text Conversion," *ROCLING X*, 287-299.
9. Hsieh, M.L., T.T. Lo and C.H. Lin, 1989, "Grammatical Approach to Converting Phonetic Symbols into Characters," *Proceedings of National Computer Symposium*, Taipei, 453-461.
10. Hsu, W.L., 1995, "Chinese parsing in a phoneme-to-character conversion system based on semantic pattern matching," *International Journal on Computer Processing of Chinese and Oriental Languages* 40, 227-236.
11. Hsu, W.L., Y.S. Chen and Y.K. Wang, 1998, "A Context Sensitive Model for Concept Understanding" Proceedings of ITALLC'98, 161-169.
12. Hsu, W.L., W.K. Shih and P.H. Yeh, 1995, "Object Oriented Concept Representation," *Proceedings of ICCPOL'95*.
13. Hsu, W.L., K.J. Chen, 1993, "The Semantic Analysis in GOING - An Intelligent Chinese Input System," Proceedings of the Second Joint Conference of Computational Linguistics, 338-343. (In Chinese: 許聞廉、陳克健, 1993, "「國音」智慧型輸入系統的語意分析——「脈絡會意法」," 第二屆計算語言學聯合學術會議論文集, 廈門, 338-343)
14. Kuo, J.J., et al., 1986, "The Development of New Chinese Input Method - Chinese Word String Input Method," Proceedings of International Computer Symposium, Taipei.

15. Lua, K.T. and K.W. Gan, 1992, "A Touch-Typing Pinyin Input System," Computer Processing of Chinese and Oriental Languages 6, 85-94.
16. Minsky, M., 1975, "A Framework for Representing Kcnowledge", in *Psychology of Computer Vision*, P. Winston (Ed.), McGraw-ill, New York.
17. Pong M.C. and Y.G. Zhang, 1993, "An Input-Method Independent Model for Chinese Input and Exploiting the Knowledge of Chinese Natural Language for Postprocessing of Chinese Input," Proceedings of the Second Joint Conference of Computational Linguistics, 344-349. (In Chinese: 龐民治、張永光, 1993, "獨立於輸入法的漢字輸入模型及利用中文自然語言知識後處理的漢字輸入," 第二屆計算語言學聯合學術會議論文集, 廈門, 344-349)
18. Schank, R.C., 1972, Conceptual Dependency: A Theory of Natural Language Understanding, *Cognitive Psychology*, Vol. 3, no. 4.
19. Sproat, R., 1990, "An Application of Statistical Optimization with Dynamic Programming to Phonemic-Input-to-Character Conversion for Chinese," *Proceedings of ROCLING III*, 379-390.
20. Wang, H.M. T.H. Ho, R.C. Yang, J.L. Shen, B.R. Bai, J.C. Hong, W.P. Chen, T.L. Yu, and L.S. Lee, 1997, "Complete Recognition of Continuous Mandarin Speech for Chines Language with Very Large Vocabulary Using Limited Training Data," *IEEE trans. on Speech and Audio Processing* 5, 195-200.