Applying Repair Processing in Chinese Homophone Disambiguation

Yue-Shi Lee and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan, R.O.C.

E-Mail: {leeys, hh_chen}@csie.ntu.edu.tw

Abstract

Repair processing plays an important role in spoken language processing systems. This paper proposes a method for correcting Chinese repetition repairs and demonstrates the effects of repair processing in Chinese homophone disambiguation. The experimental results show that the precision rate of 93.87% and the recall rate of 90.65% can be achieved for the repair processing. At the same time, 50% of errors in the repairing segments can be reduced for Chinese homophone disambiguation.

1 Introduction

Repair is a very common phenomenon in spoken languages. Speakers usually repeat, add, replace, or even abandon some constructions in the utterances for some mental reasons. Typical repairs in Chinese spoken data[1] are shown in (1) and (2).

(1) ...我 喝 酒 也 是%--
     I drink wine too SHI

   ...有時候 .. 就是%,-
   sometimes   SHI

   ...(1.3)不 ^太--
     not too

 ..不太 節制 啊.\
 not too constrain-oneself particle

(I am sometimes addicted to drinking.)

(2) [[因爲 眼-]]--
    because eye

 ..眼睛 一 閉.\
   eye once close

(Because eye .. Once (you) close your eyes, ...)

Heeman and Allen (1994) describe that 25% of turns contain at least one speech repair in their corpus. In our study, 17% of turns contain at least one speech repair. Thus, the repair processing cannot be negligible and has influences to a certain extent.

Recently, text-first approach and speech-first approach have been proposed to touch on repairs in English. The text-first approach assumes the speech recognizer could provide a correct transcription. That is, it tries to detect and correct speech repairs automatically using text alone. Hindle (1983) adds rules to a deterministic parser to tackle the problem of correcting speech repairs. His parsing strategy depends on the successful disambiguation of the syntactic categories. Although syntactic categories can be determined well by local context, Hindle admits that speech repair disrupts the local context. Bear, *et al.* (1992) firstly try to parse the input sentence and then invoke a repair processing when the parsing fails. For repair processing, a simple pattern matcher finds the candidates based on the lexical cues at the first stage. Then the syntactic and semantic processing filters out the impossible candidates. Heeman and Allen (1994) present an algorithm that detects and corrects modification and abridged repairs. The algorithm uses some repair patterns to capture potential repairs. These patterns are built based on the identification of word fragments, editing terms[2], and word correspondences between the repaired segment

---

[1] The transcription system proposed by Bois, *et al.* (1992) is used to transcribe the spoken data. The three symbols ...(N), ... and .. denote an unfilled pause (silence) is long, medium and short, respectively. The symbol % denotes the glottal stop.

[2] The editing terms can either be filled pauses (e.g., um, un, er) or cue phrases (e.g., I mean, I guess, well).

and the repairing segment[3]. The resulting potential repairs are then passed to a statistical filter that judges the proposal as either fluent speech or an actual repair.

The speech-first approach tries to identify speech repairs using acoustic and prosodic cues. Nakatani and Hirschberg (1993a; 1993b) investigate the detection of the interruption point of speech repairs based on this line. The cues that they found are the occurrence of a filled pause, the presence of a word fragment, the energy peaks in each word and other features such as accent. However, they do not address the problem of correcting speech repairs. In other words, they do not determine which words are undesired.

These approaches cannot be adopted to deal with Chinese speech repairs for the following reasons. First, a Chinese sentence is composed of a string of characters without any word boundaries. In other words, it is necessary to segment Chinese sentence before tagging and parsing (Chen and Liu, 1992; Sproat, *et al.*, 1994; Chen and Lee, 1996). Repairs make segmentation and text-first approach more difficult. Second, Chinese repairs may not always have an editing terms between a repaired segment and a repairing segment. In other words, editing terms do not have much effect in Chinese repair processing. Third, duplicate constructions (e.g., 幫幫忙 (bang1 bang1 mang2, help), 研究研究 (yan2 jiu4 yan2 jiu4, study)) generated by repeating words or phrases in Chinese utterances are used very often, but they do not always initiate a repair. That is, a simple pattern matching mechanism cannot be workable. Forth, the Chinese speech repairs may be initiated at various syntactic environment (Chui, 1995), e.g., before the subject, during the subject, after the subject and before the verb, during the verb, during a direct object, during a prepositional phrase, during subordination, and so on. The variety makes the identification of Chinese speech repairs more troublesome.

Because the repairs introduce much noise in language processing, we cannot defer the task of repair processing to the latter stages. This paper employs acoustic and prosodic information to correct Chinese repetition repairs. The results are applied to Chinese homophone disambiguation. Section 2 defines four major types of speech repairs. Section 3 introduces the spoken corpus. Sections 4 and 5 describe the

baseline and the advanced models for repetition repairs, respectively. Section 6 demonstrates the effects of repair processing in Chinese homophone disambiguation. Section 7 concludes the remarks.

2 Types of Chinese Speech Repairs

Heeman and Allen (1994) divide English speech repairs into three types: fresh start, modification repair and abridged repair. For Chinese speech repairs, Chui (1995) classify them into eleven patterns. In this section, we map these eleven patterns into four major types according to their surface forms.

Let A, B, C, D, X and Y be character strings and # be interruption point[4]. The four major types of speech repairs are described below. The repair, the repaired segment and the repairing segment are in underline, *italic* and **boldface**, respectively. They appear within an utterance or between two consecutive utterances.

(a) Repetition Repair
The repetition repair can be represented as follows:

$$X\underline{A}\#\mathbf{A}Y$$

The repetition can range from a portion of a word up to several words. After being repaired, the utterances become XAY. (1) and (2) are two examples.

(b) Addition Repair
There are two types of addition repairs.
  (i) The type I addition repair can be represented as follows:

$$X\underline{AB}\#\mathbf{ACB}Y$$

After being repaired, the utterances become XACBY. (3) shows an example.

(3) [^他 說,-
      He say

..他 今天] 才 說,-
He today until say

...^最 不 喜歡 喝酒 的 是 他=.\
most not like drink DE SHI he

(He said .. Until today he said that he is the one who dislikes drinking most.)

---

[3] A repair is composed of a repaired segment and a repairing segment which immediately follows the repaired segment. A repaired segment denotes the portion of the utterance which is being repaired, and a repairing segment denotes the portion which is accomplishing the repair (Fox and Jasperson, forthcoming). That is, the repaired segment is replaced by the repairing segment.

[4] The end of the repaired segment is called the interruption point. It is often accompanied by a disruption in the intonation contour.

(ii) The type II addition repair can be represented as follows:

$$X A\#BAY$$

After being repaired, the utterances become **XBAY**.

(c) Replacement Repair
There are five types of replacement repairs.
(i) The type I replacement repair can be represented as follows:

$$X AB\#ACY$$

After being repaired, the utterances become **XACY**. (4) shows an example.

(4) ..我們 等%-
　　We wait

..[我們] 聊天.\
　　We chat

(We wait .. We chat.)

(ii) The type II replacement repair can be represented as follows:

$$X AB\#CBY$$

After being repaired, the utterances become **XCBY**.

(iii) The type III replacement repair can be represented as follows:

$$X ABC\#ADCY$$

After being repaired, the utterances become **XADCY**.

(iv) The type IV replacement repair can be represented as follows:

$$X ABC\#ACY$$

After being repaired, the utterances become **XACY**.

(v) Different from the above replacement repairs, the repaired segment and the repairing segment in this type do not match any characters. It can be represented as follows:

$$X A\#BY$$

After being repaired, the utterances become **XBY**. (5) shows an example.

(5) ..他們--
　　They
...她 唸 的,-
　　She majors-in DE
..^不是 <L2 Computer Science L2>.\
　　not Computer Science
(They .. Her major is not Computer Science.)

(d) Abandon Repair
The original utterance is discarded and a new utterance is initiated. (6) shows an example.
(6) ...(.9)那就 不要 一堆--
　　Then don't together

..大家 坐 在一起 幹嘛 啊?/
all of us sit together for what particle

(Then don't together .. For what do all of us sit (here) together?)

3 Spoken Corpus

The spoken corpus used in this paper consists of two commonplace, everyday conversations among friends. Each is about forty-minute long. There are four and five speakers in these two conversations, respectively. In total, this corpus contains 5,395 utterances, 22,409 words and 2,602 turns. There are totally 440 self-repairs[5]. On the average, 17% of turns contain at least one repair. Table 1 lists the frequency distribution of each type of repairs in two conversations.

Table 1. Frequency Distribution of Repairs in Two Conversations

Conv.	Repeat	Add	Replace	Abandon
1	122	23	16	7
2	199	26	22	25
Total	321	49	38	32

In Table 1, the repetition repairs form the majority

---

[5] The speech repairs discussed in this paper are all self-repairs. That is, only the repairs accomplished by the same speaker are considered. This is because this kind of repairs is the most common form of repairs. Nevertheless, the present study includes repairs placed across different turns.

(72.62% in conversation 1 and 73.16% in conversation 2) of the repairs. Addition (Replacement) repairs have 13.69% (9.52%) and 9.56% (8.09%) in conversations 1 and 2, respectively. The rest (4.17% in conversation 1 and 9.19% in conversation 2) are the most complex type of repairs, i.e., Abandon. Because this paper corrects repairs based on acoustic and prosodic cues, the Chinese characters in the spoken corpus are converted into the corresponding syllables manually[6].

4 Baseline Model

4.1 Simple Pattern Matching

Because the repetition repairs form the majority, we focus on the repetition repairs in this paper. Although the repetition repairs have the simple surface form, correcting such a kind of speech repairs is not trivial. That is, a simple pattern matching mechanism cannot work perfectly. Table 2 explains this point. A repair is proposed when a string of syllables repeats within an utterance or between two consecutive utterances.

Table 2. The Experimental Results Using Simple Pattern Matching

Conversation	Total	Proposed	Correct
1	122	243	118
2	199	412	196
Total	321	655	314

Columns 2, 3 and 4 denote the total repetition repairs, the number of repairs proposed by the simple pattern matcher and the number of correct proposed repairs, respectively. For example, 243 repairs are proposed by the simple pattern matcher in conversation 1, but only 118 of them are correct. That is, there are 125 false alarms. Since there are 122 repetition repairs in conversation 1, 4 repetition repairs are not captured. They are all English repairs. Because only Chinese repairs are considered, English repairs are lost. Although this technique can achieve recall rate of 97.82%, it has a relatively low precision rate, i.e., 47.94%.

Since the simple pattern matching mechanism cannot solve this problem properly, two additional cues are firstly considered in the baseline model: the length of the repeated syllable string and the number of inter-utterances.

---

[6] Because we focus our efforts on correcting speech repairs, the identification of acoustic and prosodic cues does not discuss in this paper.

4.2 The Length of the Repeated Syllable String

How many syllables are repeated in the repetition repairs is an interesting problem in cognition. Table 3 lists the distribution of length of the repeated syllable strings in the repetition repairs.

Table 3. The Distribution of Length of the Repeated Syllable Strings

Conversation\Length	1	2	3	4
1	71	40	6	1
2	107	72	15	2
Total	178	112	21	3

The length ranges from 1 to 4. Thus, when a string of syllables repeats and the length of this string is greater than 4, we do not regard it as a repetition repair.

4.3 The Number of Inter-Utterances

In human conversation, most of the repetition repairs occur within an utterance or between two consecutive utterances of one speaker without interrupting by other speakers. That is, if many utterances issued by other speakers are inserted between two utterances of the same speaker, the repetition repairs usually do not occur. The spoken corpus shows this point.

- Total 13.69% of repetition repairs occur in the same utterance.

- Total 71.66% of repetition repairs occur between two consecutive utterances without interrupting by other speakers.

- Only 0.32% of repetition repairs occur across more than 3 utterances issued by other speakers.

According to the heuristic rule, when more than 3 utterances pronounced by other speakers interrupt the speech of a speaker, we do not check whether there is a repetition repair or not.

5 Advanced Model

5.1 Unfilled Pause (...)

In spontaneous or conversational speech, we find that there is a significant unfilled pause (silence) between a repaired segment and a repairing segment for repetition

repairs[7], whereas actual or intended repeated characters (syllables) usually do not have any unfilled pauses between them. (1) and (2) are examples. After the unfilled pause information is added to the baseline model, the experimental results for two conversations are listed below.

Table 4. The Experimental Results Using Unfilled Pause

Conversation	Total	Proposed	Correct
1	122	99	86
2	199	191	158
Total	321	290	244

The experimental results show that the precision rate is increased to 84.14%, and the recall rate is decreased to 76.01%.

5.2 Glottal Stop (%)

Glottal stop has the similar functions to unfilled pause. That is, a glottal stop may occur between the repaired segment and the repairing segment for the repetition repairs, whereas actual repeated characters usually do not have such a marker between them. (1) is an example. Table 5 shows the results when the glottal stop information is used to enhance the baseline model.

Table 5. The Experimental Results Using Glottal Stop

Conversation	Total	Proposed	Correct
1	122	31	31
2	199	85	82
Total	321	116	113

From Table 5, we find that glottal stop is a more reliable cue than unfilled pause, but it does not occur as frequently as unfilled pause. These points are verified by the high precision rate (97.41%) and the low recall rate (35.20%). When the unfilled pause information and the glottal stop information are all applied to the baseline model, the experimental results for two conversations are listed in Table 6. Both the precision rate (84.71%) and the recall rate (82.87%) are all better than those in the former models.

---

[7] Because the filled pauses such as um, un and er do not occur frequently in the spoken corpus, the effects of filled pauses are not demonstrated in this paper.

Table 6. The Experimental Results Using Unfilled Pause and Glottal Stop

Conversation	Total	Proposed	Correct
1	122	110	97
2	199	204	169
Total	321	314	266

5.3 Two Consecutive Equal Utterances

If two consecutive utterances are equal, repetition repairs usually do not occur within and between them when the length of the utterances is long enough. This is because the matched string usually denotes an emphasis when it is long enough. This cue can eliminate some implausible repairs, so that the precision rate can be increased.

5.4 Cue Patterns

In Chinese conversation, some words or phrases are frequently repeated, but they are not repairs. Typical examples are interjections (e.g., 哦 (o2, oh)) and phrase-final particles (e.g., 啊 (a5, a)). These patterns called type I cue patterns are used to increase the precision rate. That is, a repair is proposed when a string of syllables repeats, satisfies the criteria of baseline model, unfilled pause and glottal stop, and the first syllable of the string does not belong to type I cue patterns.

In contrast to type I cue patterns, another kind of patterns, type II cue patterns, are also considered to increase the recall rate. That is, some repeated syllable strings that do not satisfy the criteria of unfilled pause and glottal stop, but they are usually repetition repairs. Typical examples are pronouns such as 我 (wo3, I) and 你 (ni3, you). Based on type II cue patterns, some additional repairs can be proposed when a string of syllables repeats, it does not satisfy the criteria of unfilled pause and glottal stop, but the first syllable of the string belongs to type II cue patterns.

When all the cues proposed in the previous subsections are all applied to the baseline model, the final experimental results are listed in Table 7.

Table 7. The Final Experimental Results

Conversation	Total	Proposed	Correct
1	122	120	111
2	199	190	180
Total	321	310	291

The experimental results show that the precision rate of 93.87% and the recall rate of 90.65% can be achieved.

6 Repair Processing in Chinese Homophone Disambiguation

Mandarin Chinese has approximately 1,300 syllables, 13,094 commonly used characters, and more than 100,000 words. Each character is pronounced as a syllable and many syllables are shared by several characters. Some syllables correspond to even more than 100 characters. Thus, Chinese homophone disambiguation is difficult but important in a Chinese phonetic input method and a Chinese speech recognition system.

The problem of Chinese homophone disambiguation is defined as how to convert a sequence of syllables S into the corresponding sequence of characters $\tilde{C}$ correctly. Thus, Chinese homophone disambiguation can be regarded as a process of conversion of syllable-to-character. Let $S=<s_1, s_2, s_3, ..., s_n>$ be a syllable string and $C=<c_1, c_2, c_3, ..., c_n>$ be one corresponding character string. Here, s_i denotes one of 1,300 Chinese syllables and c_i denotes one of 13,094 Chinese characters. The conversion can be formulated as follows.

$$\tilde{C} \equiv \underset{c}{argmax}\ P(C|S)$$

$$\equiv \underset{c}{argmax}\ \frac{P(S|C)*P(C)}{P(S)}$$

The denominator part does not effect the maximization and it merely serves as a constant multiplier. The above formula therefore becomes as follows.

$$\tilde{C} \equiv \underset{c}{argmax}\ P(S|C)*P(C)$$

As most Chinese characters are unambiguous in their pronunciation (Sproat, 1990), we assume that $P(S|C)$ is one in general case. Finally, this formula is simplified as a Markov character bigram model shown below.

$$\tilde{C} \cong \underset{c}{argmax}\ P(C)$$

$$\cong \underset{c}{argmax}\ P(c_1)*\prod_{i=2}^{n} P(c_i|c_{i-1})$$

The language model is usually trained on fluent text

only. Because repairs introduce much noise, direct application of this method without repair processing is expected to have worse performance[8].

For evaluating the effects of repair processing in this application, we count how many syllables in the repairing segments are wrongly converted and how many wrongly converted syllables are recovered after the repair processing. The experimental results are listed below[9].

Table 8. The Experimental Results for Homophone Disambiguation

Conversation	Wrong	WC	CW	Net
1	45	29	2	27
2	81	41	5	36
Total	126	70	7	63

Column 2 (Wrong) denotes the number of wrongly converted syllables before the repair processing. Columns 3 and 4 then indicate the performance changes. They are classified into two types: Wrong-to-Correct (WC) and Correct-to-Wrong (CW). In the WC type, a wrongly converted syllable is changed to the correct one by the repair processing. In the CW type, a syllable which is correctly converted before repair processing, is changed to a wrong one after the repair processing. The performance of the repair processing can be evaluated as the net gain shown as follows.

Net Gain = # of WC - # of CW

In Table 8, the number of the original errors is 126. After the repair processing, the number of the errors is reduced to 63. That is, 63 (50%) errors are recovered by the repair processing. It reveals that the repair processing has much effect in these experiments.

7 Concluding Remarks

Any spoken language systems will not perform well without treating speech repairs. Correcting speech repairs make more reliable environments for the subsequent processing. This paper employs acoustic and prosodic cues to correct the repetition repairs. The experimental results show that our method can

---

[8] Stolcke and Shriberg (1996) described that "cleaning up" disfluencies reduces perplexity.
[9] The Academia Sinica Balance Corpus (1995) is adopted as the training corpus in this experiment. It contains text of several categories and includes approximately 360,000 sentences comprising of about 3,300,000 characters.

achieve the precision rate of 93.87% and the recall rate of 90.65%. At the same time, 50% of errors in the repairing segment can be reduced for the Chinese homophone disambiguation.

O'Shaughnessy (1992) claims that most speech repairs do not have lengthening prior to the hesitation pause. If this cue is used in our model, it can slightly increase the precision rate (95.37%), but the recall rate (76.95%) is greatly decreased.

Although our method can perform well in repetition repairs, other kinds of repairs such as addition, replacement and abandon repairs are not addressed in this paper. They have more complex surface forms and should be investigated further.

8 Acknowledgments

References

J. Bear, J. Dowding and E. Shriberg (1992) "Integrating Multiple Knowledge Sources for Detection and Correction of Repairs in Human-Computer Dialog," In *30th Annual Meeting of the Association for Computational Linguistics*, pages 56-63.

D. Bois, *et al.* (1992) "Discourse Transcription," *Santa Barbara Papers in Linguistics*, Vol. 4.

H.H. Chen and J.C. Lee (1996) "Identification and Classification of Proper Nouns in Chinese Texts," In *16th International Conference on Computational Linguistics*, pages 222-229.

K.J. Chen and S.H. Liu (1992) "Word Identification for Mandarin Chinese Sentences," In *14th International Conference on Computational Linguistics*, pages 101-107.

K. Chui (1995) "Repair in Chinese Conversation," In *2th International Symposium on Language in Taiwan*, pages 75-96.

B.A. Fox and R. Jasperson (forthcoming) "A Syntactic Exploration of Repair in English Conversation," In P. Davis, editor, *Descriptive and Theoretical Models in the Alternative Linguistics*, forthcoming.

P. Heeman and J. Allen (1994) "Detecting and Correcting Speech Repairs," In *32th Annual Meeting of the Association for Computational Linguistics*, pages 295-302.

D. Hindle (1983) "Deterministic Parsing of Syntactic Nonfluencies," In *21th Annual Meeting of the Association for Computational Linguistics*, pages 123-128.

C.R. Huang, *et al.* (1995) "An Introduction to Academia Sinica Balance Corpus," In *8th R.O.C. Computational Linguistics Conference*, pages 81-99.

C. Nakatani and J. Hirschberg (1993a) "A Speech-First Model for Repair Detection and Correction," In *3th European Conference on Speech Communication and Technology*, pages 1173-1176.

C. Nakatani and J. Hirschberg (1993b) "A Speech-First Model for Repair Detection and Correction," In *31th Annual Meeting of the Association for Computational Linguistics*, pages 46-53.

D. O'Shaughnessy (1992) "Recognition of hesitation in Spontaneous Speech," In *2th International Conference on Spoken Language Processing*, pages 521-524.

R. Sproat, *et al.* (1994) "A Stochastic Finite-State Word-Segmentation Algorithm for Chinese," In *32th Annual Meeting of the Association for Computational Linguistics*, pages 66-73.

R. Sproat (1990) "An Application of Statistical Optimization with Dynamic Programming to Phonemic-Input-to-Character Conversion for Chinese," In *3th R.O.C. Computational Linguistics Conference*, pages 377-390.

A. Stolcke and E.E. Shriberg (1996) "Statistical Language Modeling for Speech Disfluencies," In *IEEE International Conference on Acoustic, Speech, and Signal Processing*, pages 405-408.