Using Structural Information to Detect Speech Repairs[*]

Peter A. Heeman and Kyung-ho Loken-Kim
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02 JAPAN
{heeman,kyungho}@itl.atr.co.jp

Abstract

Previous approaches to detecting and correcting speech repairs have for the most part separated these two problems. In this paper, we present a statistical model of speech repairs that uses information about the postulated repair structure (correction) to help decide whether a speech repair actually occurred. By better modeling the interactions between detection and correction, we are able to improve our detection results.

1 Introduction

Interactive spoken dialog provides many new challenges for spoken language systems. One of the most critical is the prevalence of speech repairs. Speech repairs are dysfluencies where some of the words that the speaker utters need to be removed in order to correctly understand the speaker's meaning.

Fortunately for the hearer, speech repairs tend to have a fairly standard form. As illustrated in the example below from the Trains corpus (d92a-5.2 utt34), they can be divided into three intervals, or stretches of speech: the *reparadum*, the *editing terms*, and the *alteration*.[1]

$$\text{we'll pick up } \underbrace{\text{a tank of}}_{\textit{reparadum}} \underset{\textit{interruption point}}{\uparrow} \underbrace{\text{uh}}_{\textit{editing terms}} \ \underbrace{\text{the tanker of}}_{\textit{alteration}} \text{ oranges}$$

The reparadum is the stretch of speech that the speaker intends to replace, and this could end with a *word fragment*, where the speaker interrupts herself during the middle of the current word. The end of the reparadum is called the *interruption point* and is often accompanied by a disruption in the intonational contour. This is then followed by the editing terms, which can either be a filled pause, such as "um" or "uh" or a cue phrase, such as "I mean", "well", or "let's see". The last part is the alteration, which is the speech that the speaker intends as the replacement for the reparadum. In order to *correct* a speech repair, the reparadum and the editing terms need to be deleted in order to determine what the speaker intends to say.[2]

In the Trains corpus of human-human problem solving dialogs, speech repairs abound. Just considering repairs that consist of something more than editing terms (i.e. the reparadum is not empty), we find that they occur in 20% of all speaker turns. As the length of a turn increases, so does the chance of finding such a repair. For turns of at least ten words in length, 48% of them have at least one repair, and for turns of at least thirty words, 76% have at least one. In terms of number of words, 9.4% of the words in the Trains corpus are in the reparadum or are an editing term of a speech repair.

Psycholinguistic work in understanding speech repairs and the implications that they pose for theories of speech production have come up with a number of classification systems. A lot of these are based on comparing the content of the reparadum with the alteration, such as whether the material was repeated, material was inserted, or made more appropriate, or if it was a production error or a planning error. For work in computational detecting and correcting speech repairs, we follow Hindle's approach (1983) and use a much simpler classification scheme, based on what the hearer needs to do to correct a speech repair. We divide speech repairs into three types: *fresh starts*, *modification repairs*, and *abridged repairs*. A fresh start is where the speaker abandons the current utterance and starts again, where the abandonment seems accoustically signaled (d93-12.1 utt30).

$$\underbrace{\text{so it'll take}}_{\textit{reparadum}} \underset{\textit{interruption point}}{\uparrow} \underbrace{\text{um}}_{\textit{editing term}} \ \underbrace{\text{so you want to do what}}_{\textit{alteration}}$$

The second type of repairs are the modification repairs.

[*] Presented at a workshop held by the Institute of Electronics, Information and Communication Engineers (IEICE), Japan, TR SP95-91.

[1] Notation adapted from Levelt (1983). Following Shriberg (1994) and Nakatani and Hirschberg (1993), we use *reparadum* to refer to the entire interval being replaced, rather than just the non repeated words. We have made the same change in definition for *alteration*.

[2] The reparadum and the editing terms might still contain pragmatic information, as the following contrived example displays, "Peter was ... well ... he was fired."

These include all other repairs in which the reparadum is not empty (d92a-1.3 utt65).

$$\text{so that } \underbrace{\text{will total}}_{\textit{reparadum}} \underset{\textit{interruption point}}{\uparrow} \underbrace{\text{will take}}_{\textit{alteration}} \text{ seven hours to do that}$$

The third type of repairs are the abridged repairs, which consist solely of an editing term (d93-14.3 utt42).

$$\text{we need to } \underset{\textit{interruption point}}{\uparrow} \underbrace{\text{um}}_{\textit{editing term}} \text{ manage to get the bananas}$$

Problematic to the above classification scheme are the repairs whose reparadum consists solely of a word fragment. Under this scheme, these will either be fresh starts or modification repairs. However, our input is the word transcription (as would be provided by an ideal speech recognizer), with the word fragments marked. Given this, a word fragment is much like a filled pause in that we know we must always remove them. Hence, for this paper, we will treat such repairs in the same class as the abridged repairs.

The strategies that a hearer can use for correcting speech repairs depends on the type of repair. For fresh starts, the hearer must determine the beginning of the current utterance, and takes this as being the onset of the reparadum. For modification repairs, the hearer can make use of the *repair structure*, the parallel structure that often exists between the reparadum and alteration, to determine the extent of the reparadum. For abridged repairs, there is no reparadum, and so simply knowing that it is abridged automatically gives the correction.

Previous work in correcting speech repairs (Levelt, 1983; Hindle, 1983; Kikui and Morimoto, 1994) has assumed that speech repairs are accompanied by an acoustic editing signal (Labov, 1966). Given the interruption point, the type of repair, and the syntactic categories of the words involved, correction rates of around 95% can be achieved.

However, a reliable accoustic signal has yet to be found (Bear, Dowding, and Shriberg, 1992). Rather, detection of speech repairs probably relies on the combination of a number clues, both accoustic and lexical. Furthermore, the assumption that detection and correction can be done as separate processes might not be appropriate. Although experiments by Lickley and Bard (1992) have found that hearers were able to recognize a disfluency by the end of the first word of the alteration in 85.4% of the cases, this still leaves 16.6% of the repairs in their test set unaccounted for. In order to detect these, the hearer must need more context. Part of this context might be the presence of a suitable correction. Hence, strategies for speech repair detection and correction that separate these two tasks will be unable to account for a significant number of repairs. The only solution is to use information about the likely correction for a potential interruption point as a clue for deciding if it is in fact a repair.

In this paper, we will focus on modification repairs and show how our existing statistical model for detecting modification repairs (Heeman and Allen, 1994) can be augmented to use information about the proposed correction.[3] We have categorized potential corrections into a set of ten different groups, which differ in terms of how strongly they signal a modification repair. When detecting speech repairs, we use the correction algorithm to determine the proposed correction, and then use the category of the proposed correction as part of the context for deciding whether a speech repair has actually occurred. This approach of interleaving detection and correction lets us better model the interdependencies that exist between these two tasks.

2 Previous Work

Most previous research work has separated the problems of detecting speech repair and of correcting them. One of the first computational approaches to correction was undertaken by Hindle (1983). In this work, Hindle assumed that the input was marked with the interruption points of speech repairs as well as the syntactic category of the input words. If he found a word sequence repetition across the interruption point, he took the first sequence to be the reparadum. For the rest of the repairs, he used a deterministic parser to look for constituent correspondences across the interruption point, preferring correspondences where both are complete over those in which the first is incomplete. With this parsing based approach, Hindle was able to achieve a correction recall rate of 97%.

Kikui and Morimoto (1994) also worked from the premise that the interruption points of speech repairs have already been detected and part-of-speech assignment already assigned. Working with a Japanese spoken language corpus, they employed two techniques to determine the onset of the reparadums. First, they find all possible onsets for the reparadum that cause the resulting correction to be well-formed, according to an adjacency matrix that lists syntactically well-formed POS transitions. Second, they used a similarity-based analyzer (Kurohashi and Nagao, 1992) that finds the best path through all possible repair structures using dynamic programming. Each type of word correspondence has been given a different weight. The best path was then altered to take into account the well-formedness information from the first step. Using this approach, they were able to achieve a correction recall rate of 92% on a test corpus of 300 utterances.

One of the few works on computational detecting speech repairs was done by Nakatani and Hirschberg (1993). Using hand-transcribed prosodic annotations,

[3] Our reason for excluding fresh starts is that we are using a corpus of spoken dialogs, rather than isolated utterances. Hence this corpus poses the additional problem of determining the onset of the reparadum of fresh starts.

they trained a classifier on a 172 utterance training set to identify the interruption point (each utterance contained at least one repair). On a test set of 186 utterances, also each containing at least one repair, they obtained a recall rate of 83.4% and a precision of 93.9% in detecting speech repairs. The clues that they found relevant were duration of pause between words, presence of fragments, and lexical matching within a window of three words. However, speech repairs occurred in only 5.2% of the utterances in their corpus, so it is difficult to say how this would impact their precision rate if they tested over a representative sample of utterances.

In contrast to the above approaches, the SRI group (Bear et al., 1992) concentrated on detecting and correcting speech repairs. They employed simple pattern matching techniques for detecting and correcting modification repairs and fresh starts. For detection, they were able to achieve a recall rate of 76%, and a precision of 62%, and they were able to find the correct repair 57% of the time, leading to an overall correction recall of 43% and correction precision of 50%. In later work (Dowding et al., 1993), they also tried combining syntactic and semantic knowledge in a "parser-first" approach—first try to parse the input and if that fails, invoke repair strategies based on word patterns in the input. In a test set containing 26 repairs, they obtained a detection recall rate of 42% and a precision of 84.6%; for correction, they obtained a recall rate of 30% and a recall rate of 62%.

In earlier work, Heeman and Allen (1994) also looked at detecting and correcting speech repairs. In this work, structural analysis of the word correspondences was used to propose potential modification and abridged repairs. Acting in conjunction with a part-of-speech tagger trained on spoken dialogue, it achieved a correction recall rate of 86% and a precision of 43% across modification and abridged repairs (with fresh starts removed from the corpus). To counteract the low precision rate, the hypothesized modification repairs were filtered by a statistical model of speech repairs. This resulted in an overall correction recall rate of 80% and precision of 86% over abridged and modification repairs. The statistical model however suffers from not using the proposed repair structure as a clue in deciding if a repair actually occurs, but simply relies on the yes/no decision of the correction module.

3 The Trains Corpus

As part of the TRAINS project (Allen et al., 1995), which is a long term research project to build a conversationally proficient planning assistant, we have collected a corpus of problem solving dialogs (Heeman and Allen, 1995). The dialogs involve two human participants, one who is playing the role of a user and has a certain task to accomplish, and another who is playing the role of the system by acting as a planning assistant. The collection methodology was designed to make the setting as close to human-computer interaction as possible, but was not a *wizard* scenario, where one person pretends to be a computer. Rather, the user knows that he is talking to another person.

The corpus consists of 98 dialogs totaling six and a half hours in length and containing about 55,000 words, 5900 speaker turns, and 34 different speakers. These dialogs have been segmented into single speaker utterance files and word annotated using the Waves software (Ent, 1993). The corpus is available from the Linguistics Data Consortium on CD-ROM (Heeman and Allen, 1995).

The speech repairs in the dialog corpus have been hand-annotated. There is typically a correspondence between the removed text and the resumed text, and following Bear, Dowding and Shriberg (1992), we annotate this using the labels **m** for word matching and **r** for word replacements (words of the same syntactic category). Each pair is given a unique index. Other words in the reparadum and alteration are annotated with an **x**. Also, editing terms (filled pauses and clue words) are labeled with **et**, and the interruption point with **ip**, which will occur before any editing terms associated with the repair, and after the fragment, if present. The interruption point can also be marked as to whether the repair is a fresh start or a modification repair, in which cases, we use **ip:can** and **ip:mod**, respectively. The example below illustrates how a repair is annotated in this scheme.

engine	two	from	Elmi-	or	engine	three	from	Elmira
m1	**r2**	**m3**	**m4**	**et**	**m1**	**r2**	**m3**	**m4**
			ip:mod					

Further details of this scheme can be found in (Heeman and Allen, 1996a).

4 Statistical Model

For detecting speech repairs, we use a statistical model based on a part-of-speech tagger. For modification repairs, the category transition probabilities from the last word of the reparadum to the first word of the alteration have a different distribution than category transitions for fluent speech. So, by giving these distributions to the part-of-speech tagger, the tagger can decide if a transition signals a modification repair or not. In fact, in our general model (Heeman and Allen, 1996b), we feel that these different distributions can be used as the basis for detecting fresh starts and intonational phrase boundaries as well.

Part-of-speech tagging is the process of assigning to a word the category that is most probable given the sentential context (Church, 1988). The sentential context is typically approximated by only a set number of previous categories, usually one or two. Good part-of-speech results can be obtained using only the preceding category (Weischedel et al., 1993), which is what we will be using. In this case, the number of states of the Markov model will be N, where N is the number of tags. By using the Viterbi algorithm, the part-of-speech tags that

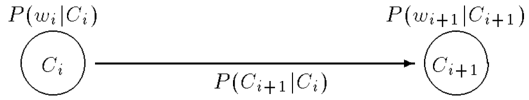

Figure 1: Markov Model of Part-of-Speech Tagging

lead to the maximum probability path can be found in linear time.

Figure 1 gives a simplified view of a Markov model for part-of-speech tagging, where C_i is a possible category for the ith word, w_i, and C_{i+1} is a possible category for word w_{i+1}. The category transition probability is simply the probability of category C_{i+1} following category C_i, which is written as $P(C_{i+1}|C_i)$. The probability of word w_{i+1} given category C_{i+1} is $P(w_{i+1}|C_{i+1})$. The category assignment that maximizes the product of these probabilities is taken to be the best category assignment.

To incorporate knowledge about modification repairs, we let T_i be a variable that indicates whether the transition from word w_i to w_{i+1} contains the interruption point of a modification repair. Rather than tag each word, w_i, with just a category, C_i, we tag it with $T_{i-1}C_i$, the category and the presence of a modification repair. So, we need the following probabilities, $P(T_iC_{i+1}|T_{i-1}C_i)$ and $P(w_i|T_{i-1}C_i)$. To keep the model simple, and ease problems with sparse data, we make several independence assumptions. By assuming that T_{i-1} and T_iC_{i+1} are independent, given C_i, we can simplify the first probability to $P(T_i|C_i)*P(C_{i+1}|C_iT_i)$; and by assuming that T_{i-1} and w_i are independent, given C_i, we can simplify the second one to $P(w_i|C_i)$. The model that results from this is given in Figure 2. As can be seen, these manipulations allow us to view the problem as tagging null tokens between words as either the interruption point of a modification repair, $T_i = \mathcal{M}$, or as fluent speech, $T_i = \mathcal{P}$. For completeness, we also show the transitions for fresh starts, $T_i = \mathcal{C}$, and for intonational phrase boundaries, $T_i = \mathcal{T}$.

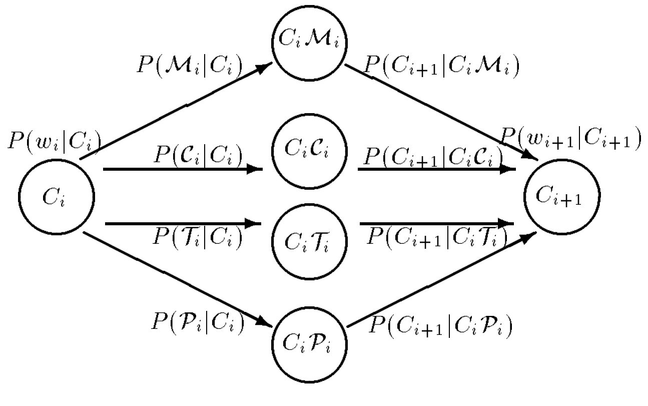

Figure 2: Statistical Model of Speech Repairs

Modification repairs can be signaled by other indicators than just syntactic anomalies. For instance, the presence of word fragments and filled pauses, editing terms, silence duration and word matches also indicate their presence. This information can be added in by viewing the presence of these clues as part of the context to be used in computing the probabilities of the transition type. So, we replace $P(T_i|C_i)$ by $P(T_i|C_iF_iE_iS_iM_i)$, where F_i indicates the presence of a word fragment, E_i indicates the presence of an editing term, S_i indicates the precence of a pause, and M_i indicates the precence of a word matching. If we make independence assumptions about the occurrence of these clues, we can rewrite this as the following.

$$P(T_i|C_i) \cdot P(T_i|F_i)/P(T_i) \cdot P(E_i|F_i)/P(T_i) \cdot \ldots$$

5 Structural Analysis

To determine the correction of a modification repair, we use the well-formedness constraints for repair structures defined by Heeman and Allen (1994). Here however we assume that the statistical model will determine the presence of a modification repair, rather than expecting the structural analysis to do this.

The well-formedness constraints make use of word correspondences to find the parallel structure that often exists between the reparadum and the alteration. These word correspondences consist of both word matches and word replacements based on part-of-speech labels, given by the statistical model. The repair structure is built by using constraints that limit what can be added to the hypothesized repair structure.

The first constraint is that correspondences must be between a word in the reparadum and a word in the alteration; in order words, they must cross the interruption point.

(1) All correspondences must cross the interruption point and editing terms if present.

The next constraints are used to start the application of word correspondences when no correspondences are yet in the repair structure.

(2) A word matching can be added to the repair structure if there are at most 3 intervening words, excluding fragments and editing terms, between the first part and the second part of the correspondence.

(3) An adjacent pair of word matches can be added to the repair structure if there is at most 6 intervening words, excluding fragments and editing terms, between them.

(4) A word replacement can be added to the repair structure if there are no intervening words between the two words.

The rest of the constraints are used to restrict the word correspondences that are added to an existing repair structure. But first we need to introduce some notation. Figure 3 shows two word correspondences $\mathbf{m_i}$

and $\mathbf{m_j}$. We denote the part of each correspondence that is in the reparadum with a superscript $\mathbf{r}$, and the part that is in the alteration with an $\mathbf{a}$. The interval $w^r_{i,j}$ refers to the sequence of words between $\mathbf{m^r_i}$ and $\mathbf{m^r_j}$; and the interval $w^a_{i,j}$ refers to the sequence of words between $\mathbf{m^a_i}$ and $\mathbf{m^a_j}$. Word correspondences i and j are *adjacent* if there are no words labeled with a word correspondence in the $w^r_{i,j}$ and $w^a_{i,j}$ intervals.

$$\mathbf{m^r_i} \underbrace{\cdots}_{w^r_{i,j}} \mathbf{m^r_j} \cdots \mathbf{ip} \cdots \mathbf{m^a_i} \underbrace{\cdots}_{w^a_{i,j}} \mathbf{m^a_j}$$

Figure 3: Distance between correspondences

Constraint (5) restricts word correspondences so that they are cross serial. This reflects the tendency of speakers not to change the order of words between the reparadum and alteration.

(5) Word correspondences must be cross-serial; for two correspondences indexed by i and j in the repair structure, if $\mathbf{m^r_i}$ precedes $\mathbf{m^r_j}$, then $\mathbf{m^a_i}$ must precede $\mathbf{m^a_j}$.

For two adjacent word correspondences, Constraint (6) ensures that there is at most 4 intervening words in the reparadum, and Constraint (7) ensures that there are at most 4 intervening words in the resumed text.

(6) In the reparadum, two adjacent matches can have at most 4 intervening words ($|w^r_{i,j}| \leq 4$).

(7) In the alteration, two adjacent matches can have at most 4 intervening words ($|w^a_{i,j}| \leq 4$).

The next constraint is used to capture the regularity that words are rarely dropped from the reparadum, instead they tend to be replaced.

(8) For two adjacent matches, the number of intervening words in the reparadum can be at most one more than the number of intervening words in the alteration ($|w^r_{i,j}| \leq |w^a_{i,j}| + 1$).

The last constraint is used to restrict word replacements. From an analysis of our corpus, we found that word replacement correspondences are rarely isolated from other word correspondences.

(9) For a word replacement (except those added by constraint 4), there must be a word correspondence in which there are no intervening words in either the reparadum or the resumed text ($w^r_{i,j} = w^a_{i,j} = 0$).

5.1 Results

In Table 1, we give the performance of the well-formedness constraints in determining the correction for modification repairs. These rates compare favorably to those reported by Hindle (1983) and Kikui and Morimoto (1994).

	Training Set	Test Set
Number	427/445	436/456
Recall	96.0	95.6

Table 1: Result of structural analysis

6 Using Structural Analysis as a Detection Clue

From Table 1, it is clear that good results can be achieved in correcting modification repairs given the correct part-of-speech assignment and given the interruption points. So, it would be beneficial if this source of knowledge could be deployed in detecting speech repairs. Unlike the approach taken by Heeman and Allen (1994), we want to integrate this source of information into the statistical model, so that it can be combined with the other sources of information. Rather than trying to devise an ad-hoc scoring scheme, our approach is to categorize the potential repair structures into categories that reflect their likelihood of actually being a repair. These categories can then be used as part of the context for deciding whether a transition is a modification repair or not.

6.1 Inconsistent Matches

The statistical model already employs the presence of word matches as one of its clues. But this is a very rough indicator of the phenomena of repair structure, and is no help in locating the interruption point. Consider the following example (d92a-3.2 utt45).

which engine are we are we taking
m1 $\uparrow$ **m2** $\uparrow$ **m1** $\uparrow$ **m2**
ip? ip:mod ip?

There are three transition points that have word matches across them: the actual one labeled with **ip:mod**, and the two neighboring ones labeled with **ip?**. Only using the presence of a word matching will distinguish the actual interruption point over its two competitors, even though the actual interruption point is the only one that properly accounts for all of the word matches, while the proposed repair structure for the other two would posit an inserted or deleted word for part of the other matching, as illustrated below.

which engine are we are we taking
m1 $\uparrow$ **x** **m1**
ip?

To disprefer proposed interruption points whose repair structure cannot properly account for all word matches, we need to let the structure analysis algorithm check for word matches that are inconsistent with the proposed interruption point, but are consistent with the other matches that are found. These would be word matches that conform to the constraints given in the previous section with the exception that they don't cross the proposed interruption point. Rather, there is some other transition point(s) that all of the word matches do cross, and this transition point would be a much more likely candidate for the interruption point. In the example above, after the structure analysis routine has found the matching on the words "are", it would be free to add the correspondence on the word "we", since there is still a transition point, namely the actual transition point, that both correspondences cross. This would lead to the following proposed repair structure, with the inconsistent matching labeled with **o**.

```
which engine are    we  are we taking
             m1  ↑  o2  m1  o2
                ip?
```

Such inconsistent matches are very rare for actual interruption points of modification repairs, but are common for fluent speech that is close to an interruption point. Hence this helps discredit neighboring transitions from taking credit for the word correspondences that are due to a modification repair.

So, we restrict the usage of Constraint (1) so that it is used only when finding the initial matching. For all other matches that are added, we use the following weaker constraint in its place.

(1a) There must exist a word transition that all word correspondences cross.

With this revised set of constraints, we will be able to categorize repair structures as to whether they account for all word matches in their vicinity.

6.2 Amount of Changed Material

Another way to categorize potential repair structures is by the amount of changed material between the reparadum and alteration. For modification repairs, there is typically one sequence of words that have been changed in the alteration, or a sequence of words that have been inserted, and the rest of the alteration simply repeats the reparadum. Proposed repair structures for transition points that are not interruption points of a modification repair, however, do not necessarily obey this regularity. Consider the following example (d93-14.1 utt10).

```
it could either take you 8 hours  or it could take you 6 hours
m   m     x     m    m   r   m  ↑ x  m   m    m    m  r   m
                               ip?
```

Here, the proposed alteration deletes the word "either", replaces "eight" by "six", and inserts "or", giving 3 regions of changed words.

From a study of the speech repairs in the Trains corpus, we found that for modification repairs, there was a maximum of two such regions. Anymore than this indicated that the proposed repair structure was spurious or due to a fresh start. We also found that if a deleted, inserted, or replaced region consisted of more than four words (as could result from Constraint (3)), it was also not indicative of a modification repair. So, we classify potential repair structures as to whether there are more than two regions of changed words or whether a changed region consists of more than four words.

6.3 Simple Patterns

Now that the above irregular categories have been dealt with, we are left with patterns that only differ by the amount of support that they offer. Four of these patterns are very distinctive and occur often enough that they can form categories on their own. These categories are: single word replacements (**r.r**), single word repetitions (**m.m**), multiple word repetitions (**mm+.mm+**), and cases in which the structural analysis does not find a suitable repair structure, which we take as a single word deletion (**x.**).

6.4 Interruption Point

To distinguish the remaining potential repairs, we looked at how the proposed repair structures constrains the interruption point. There are four possibilities. First, there might be word matchings on both the word before the interruption point and on the word after, and so the pattern constrains the interruption point. Second, there might only be a word matching on the word before (or to the left of) the interruption point, and so the interruption point is only constrained on the left. Third, the interruption point could be constrained on the right side. Or lastly, their might not be word matchings either to the immediate left or right of the proposed interruption point. What we found was that the more constrained the proposed interruption point, the greater the likelihood that it was of a modification repair.[4]

6.5 Repair Structure Categories

In all, we categorized potential repair structures into ten classes. Next, we ran the structure analysis on every transition point, which is the set of potential interruption points. Table 2 gives the number of occurrences of each of the ten categories, and their distribution by actual transition type, be it a plain transition (fluent speech or abridged repair), intonational phrase ending, modification repair, or fresh start. As can be seen, very few modification repairs fall into the categories of *too much changed*, *other matches*, and *ip not constrained.*

Given the category of a potential repair structure, we need to determine how this information can be used as a clue by the statistical model. From Table 2, we can estimate the probability of the transition type given the structural analysis category. As with the other sources of evidence, we use the preference factor $P(T_i|S_i)/P(T_i)$

[4] Other alternatives can be used, such as using any word correspondence, rather than just matches, or looking at the percentage or words involved in the repair are marked with a word correspondence.

Category	P	T	M	C
x. pattern	35624	4192	81	300
r.r pattern	566	135	118	16
too much changed	337	50	0	12
other matches	1110	95	4	3
ip not constrained	919	125	7	21
m.m pattern	44	53	309	47
repetition	5	13	159	40
ip constrained	42	17	50	37
ip constrained on left	787	89	75	71
ip constrained on right	792	210	110	81

Table 2: Table of Repair Structure Categories

to adjust the scores used by the statistical analysis. These preference factors are given in Table 3. In this table, we can see that if a transition point has a proposed repair structure that has *other matches*, we adjust the probability that this point is the interruption point of a modification repair by a factor of 0.17; whereas if the proposed repairs structure was a repetition, we would multiply the probability by a factor of 37.52. We can also see that proposed interruption points that are constrained on the right are more preferred over those constrained on the left. This captures the tendency of alterations to start with word matches and the changed material to be at the end of the reparadum.

Category	P	T	M	C
x. pattern	1.03	0.98	0.10	0.56
r.r pattern	0.79	1.52	7.24	1.43
too much changed	0.98	1.18	0.00	2.24
other matches	1.06	0.74	0.17	0.18
ip unconstrained	1.00	1.09	0.33	1.46
m.m pattern	0.11	1.10	34.92	7.72
repetition	0.03	0.56	37.52	13.72
ip constrained	0.33	1.09	17.53	18.86
ip constrained on left	0.89	0.82	3.76	5.17
ip constrained on right	0.77	1.65	4.72	5.05

Table 3: Transition preferences

Before incorporating the preferences into the statistical model, we need to verify that the structural analysis categories are sufficiently independent from the other context clues that are used. However, two strong dependencies suffice. First, the single word replacement category is very dependent on the category choice for the two words that are involved in the replacement. In fact, this type of repair structure is already modeled by the probability $P(C_{i+1}|T_iC_i)$. So, we do not use this category.

Second, there is a strong dependence between the presence of a word match and the structural analysis category. So we combine these two sources of information, using $P(T_i|S_iM_i)/P(T_i)$ as the preference factor. To cope with the limited amount of training data, we use a back-off model (Katz, 1987), which first back-offs on the identify of the matching word, and then on the distance between the word matching. Next we back-off on the structural analysis score, combining categories that make similar predictions about the occurrences of a modification repair. Next we back-off the POS tag of the matching word to a more general tag.

6.6 Results

Table 4 compares the results achieved without using structural analysis with the results from using it. A cross validation method was used in obtaining the results, in which the corpus is divided into 6 parts. For obtaining results for each part, the other five parts are used for gathering training data. So, we find that by using the above structural analysis categories, the recall rate improves by 4.1% and the precision by 11.8%. This decreases the recall error rate by 10% and decreases the precision error rate by 25.9%.

7 Conclusions

In this paper, we tried to illustrate how the two problems of detecting speech repairs and correcting them are not separable. First, the detection of speech repairs can not just detect the occurrence of a repair, but should classify the repair based on the correction strategy. Second, the correction strategy should be able to categorize the potential repair structure based on how likely it is to in fact be a repair. This will help the detection model to skip over transitions that should be ruled out by the lack of a convincing repair structure. By using structural information as a clue for detecting speech repairs, our recall rate for detecting modification repairs increased by 4.1%, and the precision increased by 11.8%.

8 Acknowledgments

We wish to thank James Allen of the University of Rochester, and Tsuyoshi Morimoto of ATR Interpreting Telecommunications Research Laboratories.

References

Allen, James F., Lenhart K. Schubert, George Ferguson, Peter Heeman, Chung Hee Hwang, Tsuneaki Kato, Marc Light, Nathaniel Martin, Bradford Miller, Massimo Poesio, and David R. Traum. 1995. The Trains project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7:7–48. Also published as Trains TN 94-3 and TR 532, Computer Science Dept., U. Rochester, September 1994.

Model	Recall	Precision
Without Structural Knowledge	72.4%	68.7%
With Structural Knowledge	75.4%	76.8%

Table 4: Results from using repair structure in detecting modification repairs

Bear, John, John Dowding, and Elizabeth Shriberg. 1992. Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 56–63.

Church, K. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the 2nd Conference on Applied Natural Language Processing*, pages 136–143, Febuary.

Dowding, John, Jean Mark Gawron, Doug Appelt, John Bear, Lynn Cherny, Robert Moore, and Douglas Moran. 1993. Gemini: A natural language system for spoken-language understanding. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pages 54–61.

Entropic Research Laboratory, Inc., 1993. *WAVES+ Reference Manual.* Version 5.0.

Heeman, Peter and James Allen. 1994. Detecting and correcting speech repairs. In *Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics*, pages 295–302, Las Cruces, New Mexico, June.

Heeman, Peter A. and James Allen. 1995. The Trains 93 dialogues. Trains Technical Note 94-2, Department of Computer Science, University of Rochester, March.

Heeman, Peter A. and James Allen. 1996a. Annotating speech repairs. Unpublished manuscript.

Heeman, Peter A. and James Allen. 1996b. Using local context to detect and correct speech repairs. In preparation.

Heeman, Peter A. and James F. Allen. 1995. The Trains spoken dialog corpus. CD-ROM, Linguistics Data Consortium, April.

Hindle, Donald. 1983. Deterministic parsing of syntactic non-fluencies. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 123–128.

Katz, Slava M. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pages 400–401, March.

Kikui, Gen-ichiro and Tsuyoshi Morimoto. 1994. Similarity-based identification of repairs in Japanese spoken language. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP-94)*, pages 915–918.

Kurohashi, Sadao and Makoto Nagao. 1992. Dynamic programming method for analyzing conjunctive structures in Japanese. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING '92).*

Labov, William. 1966. On the grammaticality of every-day speech. Paper presented at the Linguistic Society of America Annual Meeting.

Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition*, 14:41–104.

Lickley, R. J. and E. G. Bard. 1992. Processing disfluent speech: Recognizing disfluency before lexical access. In *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP-92)*, pages 935–938, October.

Nakatani, Christine and Julia Hirschberg. 1993. A speech-first model for repair detection and correction. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pages 46–53.

Shriberg, Elizabeth Ellen. 1994. Preliminaries to a theory of speech disfluencies. Doctoral dissertion, University of California at Berkeley.

Weischedel, Ralph, Marie Meteer, Richard Schwartz, Lance Ramshaw, and Jeff Palmucci. 1993. Coping with ambiguity and unknown words through probabilistic models. *Computational Linguistics*, 19(2):359–382.