TR-IT-0107

Linguistic Analysis of Speech Disfluency in the ATR Spoken Language Database

Kyung-ho Loken-Kim Yoshihiro Kitagawa*
Yoko Ohta**

December, 1995

ABSTRACT

This is the first of two technical reports on speech disfluencies found in the ATR spoken language database. This report presents the statistical analysis of the orthographic transcriptions, and provides the following: 1) probability of fluent speech at various sentence lengths, 2) the structure of reparandums (see section 2 in this report), 3) disfluency patterns and occurrences, and 4) word fragments and their significance in speech disfluency. The results of the acoustical analysis will be reported in a subsequent report (TR-IT-0108).

ATR Interpreting Telecommunications Research Laboratories
*Advanced Telecommunications Research Laboratories International
**The Telecommunications Association

CONTENTS

1. INTRODUCTION

Speech disfluencies, such as false starts, repeats, and hesitations (filled and unfilled pauses) are prevalent in spontaneous speech, and their presence has been one of the major impediments to progress in speech recognition and understanding research. Spontaneous speech translation - the ATR-Interpreting Telecommunications Research Laboratories' (ITL) goal - will be a difficult goal to attain without understanding these abnormal (normal ?) speech phenomena, and devising an effective means to deal with them.

In the past, several ATR-ITL's researchers attempted to understand speech disfluencies [Murakami, Takezawa] and to deal with them [Kikui]. For example, Murakami analyzed 11,054 spontaneously spoken sentences extracted from the ATR speech database [Shinozaki], and classified speech disfluencies into seven different patterns. He, nevertheless, failed to provide any statistically significant evidence that could lead us to further investigation. Takezawa gave examples of "ill-formed sentences" found in the ATR Spontaneous Speech Database [Uratani-1], but again offered no statistically meaningful data. Kikui and Morimoto developed a similarity based speech repair identification algorithm exploiting the fact that words in the reparandum and the repair of a sentence are syntactically somewhat similar. Although their algorithm works well for some cases, the similarity scores were assigned heuristically, rather than based on any empirical evidence.

Although it is questionable whether the ATR Spontaneous Speech Database truly represents a spontaneous speech[1)], it is our hope that its analysis will help us understand the nature of speech disfluency. The goal of this study is to provide a statistical basis for building a language model that will be used for detecting and correcting speech disfluencies. This is the first of the two technical reports on the speech disfluencies found in the ATR spoken language database. In this report, we will present the linguistic aspects of the speech disfluencies found in the 750 Japanese orthographic transcription files. Specifics are: probability of fluent speech in relation to the sentence length, the structure of reparandum (see section 2 in this report), disfluency patterns and occurrences, and word fragments and their significance in speech disfluency. The acoustical characteristics of speech disfluencies will appear in a subsequent report [TR-IT- 0108].

[1)] See Uratani- 1 page 2. Recording conditions are stated as follows:

. One utterance should be less than 10 seconds.

. Interruptions should be avoided.

. Avoid extremely long filled pauses and speech repairs.

2. SPEECH DISFLUENCY

As was mentioned before, speech disfluencies include many kinds of linguistic phenomena such as false starts, repeats, and hesitations. In this report, the authors liberally use the term speech disfluency (SD) to indicate self-repairs [Levelt], fresh starts, modification repairs, abridged repairs [Heeman], and repairs [Kikui]. In this sense, SDs are in line with DFs in Shrieberg [Shrieberg-1].

Notwithstanding confound terminologies, there seems to be a general agreement on the structure of SD: the speech interval that will be replaced, an optional editing phrase, and the replacing interval (Figure 1).

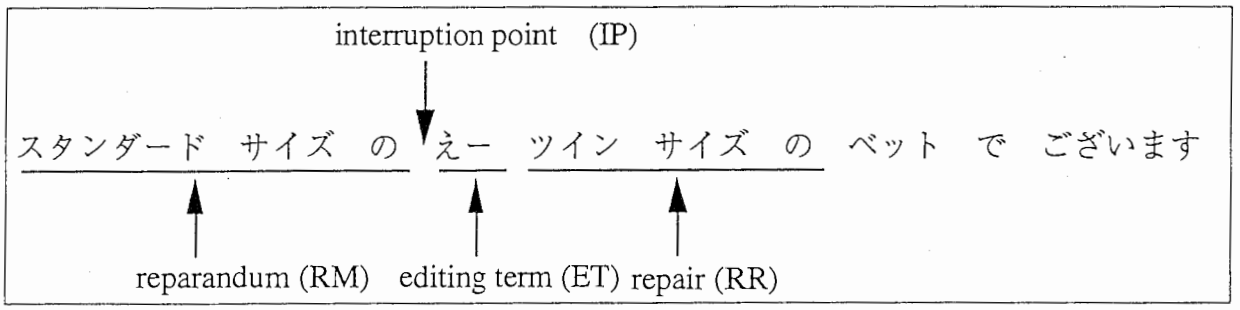

Figure 1. The structure of speech disfluency

The term reparandum (RM), originally used by Levelt, refers the speech that will be replaced (e.g., スタンダードサイズの). The interruption point (IP) is the moment when the flow of fluent speech is interrupted. It is normally followed by a silence or an editing phrase. The editing term (ET) is the same as the interregnum of Shrieberg and editing phase of Levelt. Many times it is a short interjection such as 'えー' or an explicit editing term such as 'すいません'. The repair (RR) region corresponds to the region that corrects RM (ツインサイズの)

3. ATR SPOKEN LANGUAGE DATABASE (ASLD)

As of November 8th, 1995, there were a total of 961 travel conversation files in the ATR Spoken Language Database. Part-of-speech (POS) labeling of each file is a laborious and time consuming job (this work is being done by an outside vendor), and was impeding the progress of the authors' work in collecting the SD statistics. It was in the authors' best interest to stop waiting for the completion of POS labeling of the entire 961 files and start analyzing what they had at the time - 750 files.

Table 1 is the list of the task domains and topics of the 750 files. The majority of the

conversations are recordings of dialogues between two participants, one mimicking a hotel guest and the other a hotel front desk staff, discussing a variety of sub-topics such as room reservations, direction findings, and ticket reservations. Out of the 750 files, 633 are of spontaneous speech (albeit with some restrictions) and 117 files are of the conversations using some form of check sheet. 132 files are of Japanese-Japanese conversations, while the rest are interpreted conversations either from Japanese to English, or English to Japanese.

Table 2 and 3 summarize the contents of the 750 files.

Domain	T 750 :travel information services
Topics	155 :hotel room reservation (dialogue with front desk staff)
	24 :consulting hotel rooms (dialogue with front desk staff)
	339 :hotel services (dialogue with front desk staff)
	36 :consulting hotel conference rooms・reservation
	22 :air reservation・change・cancellation
	22 :inquiring regular bus・train tickets
	28 :inquiring transportation
	8 :direction finding
	20 :inquiring car rentals, reservations
	50 :inquiring tour bus・package tour, reservation, travel
	14 :play・concert ticket reservation・purchase
	12 :restaurant reservation
	8 :food ordering
	8 :troubles・lost and found
	4 :shopping
Degree of Spontaneity	633 :spontaneous
	117 :check sheet
Language (s)	132 :Japanese→Japanese
	272 :Japanese→English
	346 :English→Japanese

numbers indicate the number of files

Table 1. Content of 750 spoken language database files

total 750 files	26,046 sentences	108,448 bunsetsu	304,131words
average per file	34.73 sentences	144.60 bunsetsu	405.51 words
average per sentence		4.16 bunsetsu	11.68 words
average per bunsetsu			2.80 words

Table 2. Total number of files, sentences, bunsetsu phrases, and words

total 750 files	657 disfluencies	1,150 words	2,186 letters
per file	0.88 disfluencies	1.53 words	2.91 letters
per disfluency		1.75 words	3.33 letters
per word			1.90 letters

Table 3. Disfluency statistics

4. DISFLUENCY PATTERN LABELING SYSTEM (PLS)

To represent the range of SD, the authors have adopted the Pattern Labeling System (PLS) which was originally developed at SRI [Shriberg-2, Bear], and later modified by Shriberg [Shriberg-1]. The authors closely followed the PLS introduced by Shriberg with only minor modifications: coindexing was revived, and one extra label was introduced.

Coindexing is described in Bear et al. [Bear] "it involves assigning an index to each word symbol to associate the word with the correct corresponding symbol on the other side of the IP." Shriberg eliminated coindexing in her work to simplify the system and reduce labeling time. The authors, however, feel that coindexing is necessary in order to index reordered words and variable-length substituted/repeated strings. In the following examples, the PLS labels alone will not provide enough details on how the reparandums are altered in the repairs unless coindexing is used.

e.g., reordering case

（御	希望	は	なに	か）	あ	あ	えー	なに	か	御	希望	は	お	あ	り	で	す	か。
r1	r2	r3	r4	r5.	f	f	f	r4	r5	r1	r2	r3						

In case of variable-length substituted strings,

タクシーを（お 呼 び さ せ）呼 ば し ていただ き ま す。
d1 r1 s1 s2 s2. r1 s1 s2

In case of variable-length repeated strings,

e.g., 。。。フロントを（とおれ） お 通 り にな/る/際/に。。。
r1~ . i1 r1 r1

Second, in ASLD, there were several cases where a word is fragmented by another fragment. In this case, we used the symbol 't' to mark the fragment but did not label fragmented word.

e.g., かしこまりま （っひ） した。
Part-of-speech Labeling その他 その他 その他
Pattern Labeling t1- .

The entire PLS (Table 4) labeled files (Table 5) are under the directory of /data/as46/kyungho/DB/DISFLUENCY (see section 8 for details)

symbol	explanation
.	interruption point (IP)
r	repeated word
s	word in substituted string
i	inserted word
d	deleted word
f	filled pause
e	explicit editing term
-	word fragment
*	same word with different POS label
~	misarticulated word
t	an inserted word causing a fragmention of another word

Table 4. Pattern Labeling System (PLS, Shriberg, 1994)

```
90¦0110¦630¦2070¦けさ¦ケサ¦今朝¦普通名詞¦¦¦¦
90¦0110¦640¦2080¦一¦ヒト¦一¦数詞¦¦¦¦
90¦0110¦640¦2090¦つ¦ツ¦つ¦接尾辞¦¦¦¦
90¦0110¦650¦2100¦（¦（¦記号¦¦¦¦
90¦0110¦650¦2110¦飲¦ノ¦飲¦本動詞¦五段マ¦語幹¦¦r1
90¦0110¦650¦2120¦ん¦ン¦む¦語尾¦五段マ¦連用¦ん音便¦r2
90¦0110¦650¦2130¦だら¦ダラ¦だら¦接続助詞¦¦¦¦s1.
90¦0110¦650¦2140¦）¦）¦記号¦¦¦¦
90¦0110¦650¦2150¦飲¦ノ¦飲¦本動詞¦五段マ¦語幹¦¦r1
90¦0110¦650¦2160¦ん¦ン¦む¦語尾¦五段マ¦連用¦ん音便¦r2
90¦0110¦650¦2170¦だ¦ダ¦た¦助動詞¦特殊¦連体¦¦s1
90¦0110¦660¦2180¦後¦アト¦後¦普通名詞¦¦¦¦
90¦0110¦670¦2190¦切¦キ¦切¦本動詞¦五段ラ¦語幹¦¦
90¦0110¦670¦2200¦ら¦ラ¦る¦語尾¦五段ラ¦未然¦¦
90¦0110¦670¦2210¦せ¦セ¦せ¦助動詞¦一段¦語幹¦¦
90¦0110¦670¦2220¦てしま¦テシマ¦てしま¦助動詞¦五段ワ¦語幹¦¦
90¦0110¦670¦2230¦っ¦ッ¦う¦語尾¦五段ワ¦連用¦つ音便¦
90¦0110¦670¦2240¦た¦タ¦た¦助動詞¦特殊¦連体¦¦
90¦0110¦670¦2250¦の¦ノ¦の¦準体助詞¦¦¦¦
90¦0110¦670¦2260¦で¦デ¦で¦助動詞¦特殊サ¦語幹¦¦
90¦0110¦670¦2270¦す¦ス¦す¦語尾¦特殊サ¦終止¦¦
90¦0110¦670¦2280¦。¦。¦記号¦¦¦¦
```

Table 5. An example of a PLS labeled file

4.1. Interruption Point (.)

An interruption point is the moment when the normal flow of fluent speech is interrupted by the speaker who just realized the fact that there was something wrong about what he said. An interruption is usually followed by an unfilled/filled pause, or an explicit editing term (see 4.7). An IP is denoted by a '.'.

e.g., じゃあ、そのツイン ツインでお願いします。
r1 . r1

4.2. Repeated Word (r)

Exactly the same word(s) repeated before and after the IP are indicated by the symbol 'r'. Repetitions may be one or more words.

e.g., exact repetition
そんなに容体も悪 い と は 思 悪 い と は 思わないんですけれども。
r1 r2 r3 r4 r5 . r1 r2 r3 r4 r5

e.g., variable-length repetition

。。。フロントを（とおれ） お 通 り になる/際/に。。。

r1~ . i1 r1 r1

4.3. Word in Substituted String (s)

Each substituted word is marked by the symbol 's'. A substitution does not always match up word for word (variable-length substitution), therefore, they are judged based on their syntactic and semantic correspondences.

e.g., 二日分以上の しゅく えー お 部屋 代を入れていただく。。。。

s1- . f s1 s1

In this case, the authors judged 'しゅく' as a fragmentation of '宿泊' (lodging), which was substituted by two words 'お' and '部屋' (room). The words '宿泊' and 'お部屋' have syntactic/semantic similarities in the travel information service task domain.

4.4. Inserted Word (i)

Inserted words are marked with the symbol 'i'. Inserted words are notably surrounded by repeated words. It is sometimes difficult to decide whether an insertion is an emphatic expression or a disfluency, but we, nevertheless, treated them as disfluencies.

e.g., 朝六時から 十一時 夜 の 十一時まで。。。

r1 . i1 i2 r1

4.5. Deleted Word (d)

Words in the RM that have no syntactically/semantically corresponding words in the RR are labeled with d. In the following example, 'て' is replaced by the semantically unrelated word 'あげます', therefore, judged as a deleted (and fragmented) word.

e.g., お電話さして あげます。

d1-.

4.6. Filled Pause (f)

The ATR part-of-speech labeling convention [Uratani-1, p. 30] defines a filled pause (間投詞) as " A content word. A word that can constitute a bunsetsu. Non- conjugative. Removing them does not alter the structure and meaning of the sentence." Filled pauses occurring in otherwise fluent speech were not labeled, but were labeled if they occurred as an ET inside of an SD.

e.g., スタンダードサイズ のえー ツインサイズ のベットでございます。
s1 r1 . f s1 r1

Table 6 lists all the filled pauses and their occurrences in the 750 files, and Table 7 filled pauses function as an ET.

4.7. Explicit Editing Term (e)

In Japanese, '感動詞's (idiomatic expressions) [Uratani-1, p. 29] such as 'すいません' or 'しつれいしました' are commonly used to retract what was just said. Mostly, they appear right after the RM, and they are marked by 'e'.

e.g., 弘子が エイチ アイ エル あ すいません エイチ アイ アール ケー オー です。
r1 r2 s1. f e r1 r2 s1

Table 8 lists all the idiomatic expressions and their occurrences in the 750 files, and Table 9 idiomatic expressions function as an explicit editing term.

4.8. Word Fragment (-)

Word fragments are represented by '-'. Word fragments are so prevalent in speech disfluency that they warrant special attention. A detailed explanation is provided in section 7.

e.g., かしこまりま っひ した。
t1- .

4.9. Same Word with Different POS/Conjugation (*)

An abrupt ending of RMs occasionally causes the same word with different POS labels and/or conjugation labels. In the following example, two words 'う' have the same POS labels but one was assigned as a conjugation label predicative form (終止形) and the other as an adnomial form (連体形).

e.g., 日本酒を 味わ う 味わ う の を 楽しみにしているんです

pattern labeling		r1	r2*.	r1	r2
POS			word final	word final	
conjugation			predicative	adnominal	

4.10. Mispronounced Word (~)

Mispronounced words are indicated by '~'. A word is judged mispronounced if the word is fragmented and there is no syllabic matchings in the RR.

e.g., そのようにうけ<u>とも</u> たまわっております

d~.

4.11. An Inserted Word Causing a Word Fragmentation (t)

If there is a word or a word fragment causing the fragmentation of another word, the prior is marked with 't'. In the following example, an auxiliary very ' でいただけ' is divided into two by 'あ'.

e.g., 変え <u>ていただ</u> <u>あ</u> <u>か</u> な くですむよに

t-.

えー	2214	うん	12	さようでございますねえ	3
え	1332	そうでございますね	12	そうでございますねえ	3
あ	769	いや	11	ちょっと	3
あの	555	ふーん	11	まっ	3
あのー	378	えーっ	9	んと	3
えーと	310	えーとー	9	あえー	2
そうですね	262	さようでございますね	9	あのね	2
あー	188	でー	9	えーっとー	2
あっ	155	へえー	9	えーとね	2
えーっと	86	うーんと	7	そうなんですか	2
うーん	78	そうですか	7	そうね	2
で	78	っとー	7	なんか	2
そうですねー	60	はー	7	ふん	2
ああ	59	んとー	7	んーとー	2
えっと	58	うー	6	あーっと	1
えと	57	そうそう	6	あう	1
おー	54	とー	6	あうー	1
そうですねえ	41	わー	6	あそう	1
ま	40	んーと	6	あそうですね	1
お	38	あのう	5	あと	1
まあ	38	あのですね	5	あのーあ	1
ん	37	まー	5	あのーえー	1
ええ	34	もしもし	5	あのーですね	1
と	33	あとー	4	あら	1
えっ	32	い	4	あれですか	1
んー	38	えーそうですね	4	いえね	1
えとー	22	へー	4	うーと	1
っと	19	もう	4	うーんあ	1
いやー	16	あのですねー	3	うーんとー	1
えっとー	16	いー	3	うんとー	1
えーとですね	14	うんと	3	えーあ	1
はい	14	えーっとですね	3	えーあの	1
そう	13	えーとですねー	3	えーっとね	1
う	12	こう	3	えーっとや	1

continued

えーで	1	そうでございますねー	1
えあーと	1	そうです	1
えあのー	1	そうですよね	1
ええーと	1	その	1
ええっ	1	それあのー	1
ええと	1	それでですねー	1
えく	1	っとえ	1
えっーとー	1	ていうと	1
えっこ	1	でえ	1
えっとですね	1	ですかー	1
えっとですねー	1	ではあのー	1
えでー	1	ど	1
えの	1	どうかな	1
おう	1	どうしましょう	1
く	1	ねえ	1
ぐ	1	の	1
こうー	1	は	1
さー	1	ひー	1
さあ	1	ふぅーん	1
しー	1	ふう	1
じゃーあの	1	やー	1
じゃあ	1	わあ	1
じゃのー	1	んーとあの	1
すると	1	んーとね	1
そうそうそう	1	んっと	1
そうだなー	1		

==========

Total 7455 (153 different filled pauses)

Table 6. Filled Pauses (間投詞) and their Occurrences

えー	52	あのう	1
あ	15	う	1
あの	11	えーっと	1
え	10	えと	1
えーと	4	えの	1
あっ	3	お	1
あのー	3	と	1
ん	4	ま	1
あー	2	まあ	1
あうー	1		

Total 114 (19 different filled pauses)

Table 7. Occurrences of Filled Pauses Appeared in Disfluencies

はい	4341	もうしわけありません	35	しょうちしました	8
かしこまりました	1115	いらっしゃいませ	33	すみませんでした	8
ありがとうございました	577	どうぞ	32	おめでとうございます	5
ありがとうございます	418	しつれいします	31	こんばんは	5
ええ	384	どうも	29	しつれいしました	5
おねがいします	267	おはようございます	27	うん	4
ありがとう	245	いや	26	おだいじに	4
もしもし	173	いってらっしゃいませ	23	おねがいしますね	4
すみません	172	しつれいいたしました	23	おまたせしました	4
しつれいいたします	149	はいっ	23	なるほど	4
それは	148	え	18	また	4
もうしわけございません	114	もうしわけございませんでした	17	ああ	3
いいえ	109	おまたせいたしました	16	ありがと	3
いえ	104	さようなら	12	おだいじにどうぞ	3
どういたしまして	104	よろしく	11	のちほど	3
おねがいいたします	102	いえいえ	10	いやー	2
しょうちいたしました	99	とんでもございません	9	いやいや	2
そうですね	67	おせわさまでした	8	おきをつけて	2
おそれいります	44	ごめんなさい	8	おっけー	2
こんにちは	39	さよなら	8	おやすみなさいませ	2

continued

こんちは	2	おせわさまです	1
しつれい	2	おやすみなさい	1
じゃあ	2	これは	1
そう	2	ごめんください	1
なるほどね	2	すみませんね	1
はっぴーばーすでー	2	そうですねえ	1
ほら	2	それでは	1
ようこそ	2	ども	1
わあ	2	なるほどお	1
あ	1	はじめまして	1
あっ	1	はっぴばーすでー	1
あら	1	もうしわけない	1
いえっ	1	やれやれ	1
えー	1	よろしくね	1
えーと	1		
おせわさま	1		

==========

Total 9286 (90 different idiomatic expressions)

Table 8. Idiomatic Expressions (感動詞） and Their Occurrences

ごめんなさい	5
しつれいいたしました	3
しつれい	2
すみません	1

==========

Total 11 (4 different idiomatic expressions)

Table 9. Idiomatic Expressions as Explicit Editing Terms

5. DISFLUENCY, SENTENCE LENGTH AND REPARANDUM

5.1. Disfluency and Sentence Length

Shriberg [Shriberg-1] developed the concept of "efficient words" as a measure of sentence length. According to Shriberg the sentence length should be counted as the total words minus the reparandum and the editing term. Shriberg justifies this system by noting that "because there is an inherent correlation between the presence of a disfluency and sentence length. Efficient length avoids this confound by omitting words deleted in disfluency from the word count for the sentence."

Figure 2 is the frequency polygon by sentence length. The majority of sentences (over 72 %) in the corpus are rather short (less than 15 words), and one third of them (27.6 %) are even shorter (less than 5 words) reflecting the fact that many of the sentences are comprised of simple acknowledgements (e.g., yes, I understand). Another reason for this lopsided distribution may be the 10 second limitation - speakers were encouraged to utter sentences shorter than 10 seconds [Uratani-2] which seems to be the limit that current natural language understanding technology can handle.

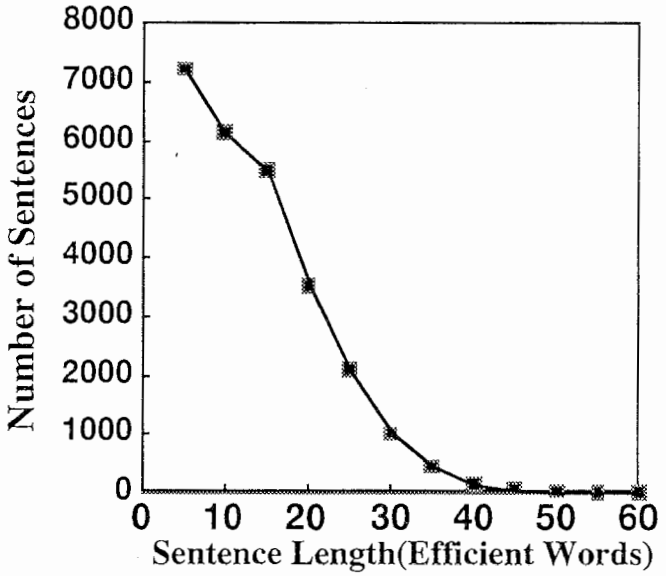

Figure 2. Number of Sentence at Each Sentence Length

By separating fluent sentences from disfluent sentences (Figure 3, 4), we know that most of the short sentences (less than 5 words) are fluent, and the distribution of disfluent sentences is bell shaped peaking at 20 words. We don't think 20 word sentences are particularly prone to

disfluency, but rather it is an artifact of the 10 second rule that caused a smaller number of longer sentences. The simple fact is that longer sentences are more prone to disfluency. In Figure 5, the sentence length is inversely related to the probability of fluent speech, and it shows no particular fluctuations at 20 words.

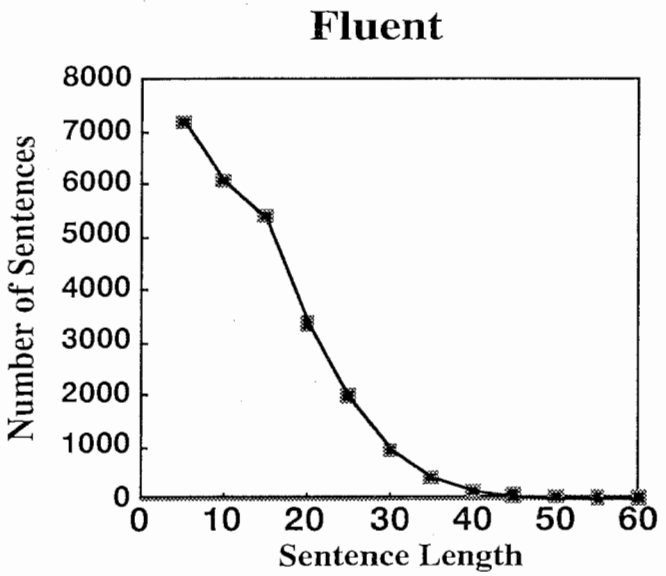

Figure 3. Number of Fluent Sentences at Each Sentence Length

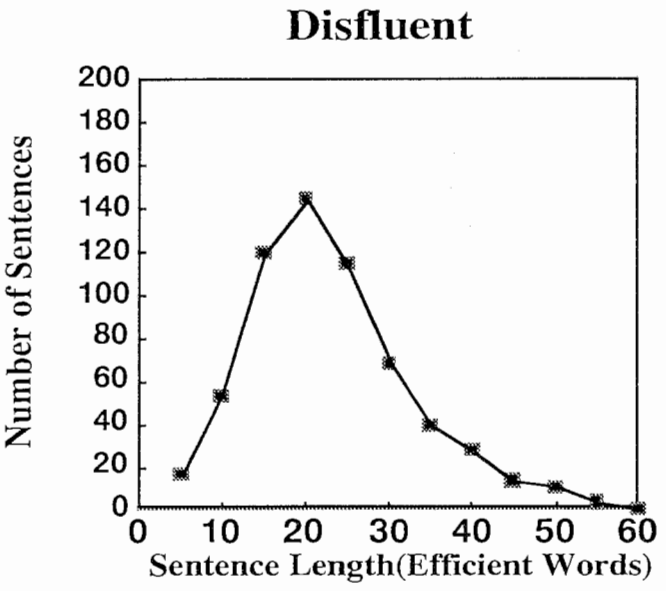

Figure 4. Number of Disfluent Sentences at Each Sentence Length

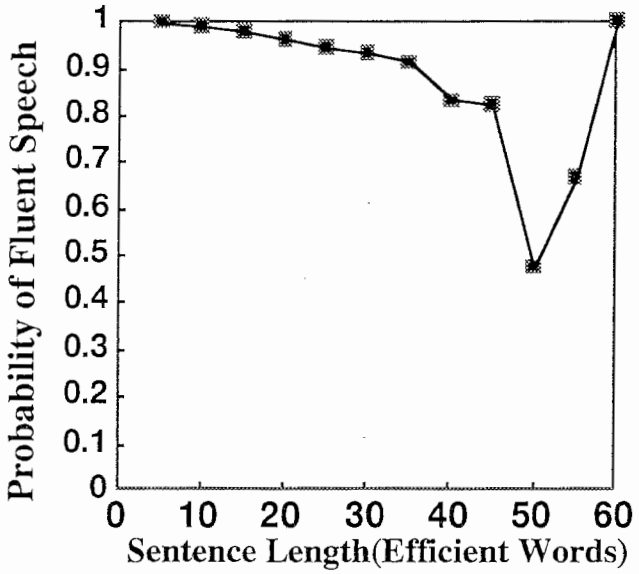

Figure 5. Probability of Fluent Speech at Each Sentence Length

5.2. Reparandum (RM)

Our analysis shows that RMs are mostly short and many of them end in a fragment word. Figures 6 and 7 show that 96.7 % of the RMs are less than five words and 84.9 % are less than five hiragana characters. 68.3 % of the RMs are one word long, and 30.1 % are one character long, such as,

それから お [えー] 主人と私には。。。。
s1-.
鈴木さんで す いらっしゃいますか。
d1-.

Tables 10 and 11 also show that the majority of RMs are single fragmented words or words ending with a その他 (see section 7.1 for definition).

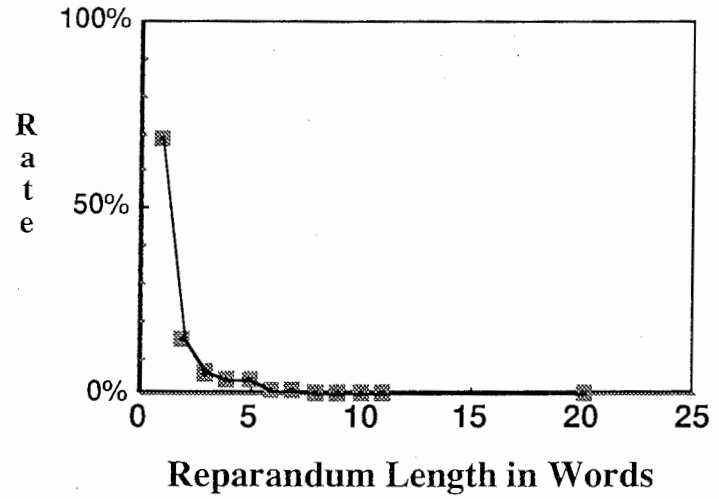

Figure 6. Reparandum Length in Words

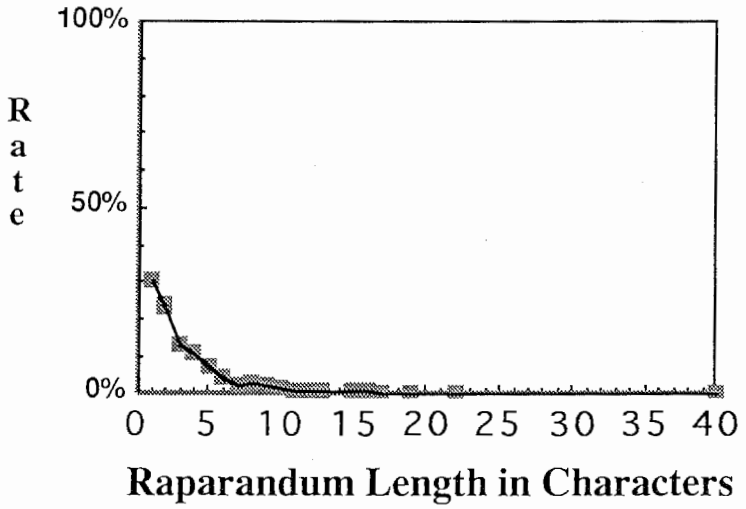

Figure 7. Reparandum Length in Characters

POS	frequency	POS	frequency	POS	frequency
その他	372	補助動詞	10	準体助詞	4
格助詞	46	サ変名詞	9	人名	3
普通名詞	43	数詞	9	形容詞	2
語尾	33	接頭辞	8	並立助詞	2
本動詞	21	日時	6	接続詞	2
連体助詞	19	係助詞	6	感動詞	2
接尾辞	16	代名詞	6	副助詞	2
接続助詞	14	連体詞	5	終助詞	2
助動詞	11	固有名詞	4	Total	657

Table 10. Part-of-Speech Occurrences in Reparandum Endings

6. SPEECH REPAIR PATTERNS AND OCCURRENCES

We next investigated the speech repair patterns in the 609[1)] disfluencies in the annotated corpus. We found a total of 162 disfluency patterns (see Appendix). Table 11 presents the 14 patterns with more than 10 occurrences these accounted for 59.6 percent of the disfluencies.

Noticeably, 10 out of the 14 patterns are the result of removing either word fragments or mispronounced words. This reaffirms our belief that if the speech recognizer can locate these words it will enable us to reliably detect the presence of speech repairs. At the top of the list in this table is the simple deletion of word fragments, d1-. 11.2 %. Next three patterns are one word repetitions: r1.r1, r1-.r1, and r1~.r1; in total they amount to 20.1 % of the disfluencies (127 cases).

[1)] Out of a total of 657 disfluencies, 48 of them with 1) numbers, 2) alphabets, and 3) complex disfluencies are not annotated.

patterns	frequency	%
d1-.	68	11.2
r1.r1	48	7.9
r1-.r1	42	6.9
r1~.r1	37	6.1
s1-.s1	31	5.1
s1.s1	20	3.3
t1-.	18	3.0
r1r2.r1r2	17	2.8
r1~.r1r1	16	2.6
d1-.f	15	2.5
r1s1.r1s1	14	2.3
r1-.fr1	13	2.1
r1-.i1r1	13	2.1
r1-.i1i2r1	11	1.8
total	363	59.6

Table 11. Speech Repair Patterns and Occurrences

We wrote a program that identifies two words in tandem with the same part-of-speech label. We used the entire 750 files, and found out that 780 r1.r1 patterns were detected by the program. This means that, although all 48 repairs in the Table 11 were correctly detected, the precision rate[1)] was too low (6%) due to the high rate of false positives. Most of the false positives were caused by consecutive numbers and alphabets.

---

1)

Recall Rate: percentage of correct answers that were found

Precision Rate: percentage of answers that are correct

e.g., total 100 and the system found 80 and out of 80 60 are correct

Recall Rate = (60/100) X 100 = 60 %

Precision Rate = (60/80) X 100 = 75 %

7. その他 and FRAGMENTS

7.1. その他

'その他' in Japanese means *the rest* or *the others* (Lighthouse). The following ATR's part-of-speech labeling guideline [Uratani-1] was used to mark a word with 'その他' :

1) Out of vocabulary word: The ATR Dictionary, as of March, 27, 1995, contained 54,676 words, and any word that was not in the dictionary was labeled as 'その他'.

2) A word divided by an interjection (間投詞) or other 'その他': In the following example, the word 'かしこまりました' (I understand) is divided into two 'かしこまりま' and 'した' by the word fragment 'っひ'; and all three of them are labeled with 'その他'.

e.g., かしこまりま (っひ) した。

Part-of-speech Labeling その他 その他 その他

3) A word considered as a fragment judged by contextual information: in the following case, the word '東京' in '東京ター' is not labeled as a proper noun (固有名詞) - rather '東京ター' as a whole is viewed as a fragment of 東京タワー, thus it is labeled as a その他.

東京ター 東京タワー

Part-of-speech Labeling その他

4) Word fragments and mispronounced word: Both of them are marked with the 'その他'. label (see section 7.2 Fragments).

7.2. Fragments

'その他' includes several categories of words including fragments. Fragments are mostly one or two hiragana characters long (Table 12), but they can play an important role in detecting SD (Table 13) because they almost always indicate disfluency, and they almost always mark the right edge of the RM (Table 14, p. 124 Shriberg-1). For example, out of a total of 657 disfluencies, 372 (56.6 %) of them involve word fragments (almost equivalent to the ATIS corpus, p. 123 in Shriberg), and out of 422 word fragments (Table 12), 372 of them are found at the ends of reparandums (Table 14). Fragments occur with higher frequency in the monolingual dialogues than in the interpreted dialogues (Table 15). The implication of this is that if a speech recognizer can identify the location of the word fragment, then we can say that 56.6 percent of the interruption points can be identified - automatic recognition of word fragments, we know however, is an unsolved problem.

hiragana characters	tokens	%
1	201	47.6
2	118	28.0
3	57	13.5
4	30	7.1
5	7	1.7
6	5	1.2
7<=	4	0.9
total 422		

Table 12. Lengths of Word Fragments

	# of files	# of disfluency	# of disfluency with word fragments	fragment probability (%)
total Data	750	657	372	56.6

Table 13. Word Fragments in Speech Disfluency

reparandum ending	372
in reparandum	6
others	44

Table 14. Word Fragments in Reparandum

	# of files	# of disfluency	# of disfluency with word fragments	fragment probability (%)
Japanese-Japanese	132	220	163	74.1
Japanese-English	618	437	209	47.8

Table 15. Conversation Mode

8. FILES

The SD information used in this study is stored under the directory /data/as46/kyungho/DB.

1) DISFLUENCY Directory

This directory contains the names of the 307 files (extracted from the 750 files under /SDB/ALL/LNG/JMOR/) each containing a sentence with disfluencies and disfluency pattern labels.

2) DISFLUENCY.FULL Directory

This directory contains the names of the 750 files each containing a sentence with disfluencies and disfluency pattern labels.

3) DISFLUENCY.SPH.LBL Directory

This directory contains the names of the disfluency labeled acoustic files.

4) RESULT Directory

This directory contains several files on the statistics of speech disfluency.

5) SDB Directory

Same as the ATR Spoken Language Database (under /DB/SDB/). There are portions that have been altered to deal with disfluencies (see modify.log).

6) SLDB Directory

Same as the ATR Spoken Language Database (under /DB/SLDB/). There are portions that have been altered to deal with disfluencies (see modify.log).

7) modify.log File

Lists the differences among files in SDB, SLDB and the original ATR Spoken Language Database.

Acknowledgement

The authors would like to express their gratitude to Nick Campbell for his cooperation in labeling speech disfluencies.

REFERENCES

[Bear] J. Bear, J. Dowding, Elizabeth Shriberg, Patty Price, "A System for Labeling Self-Repairs in Speech," Technical Note 522, SRI International in [Shriberg-2], 1992

[Heeman] Peter Heeman, Kyung-ho Loken-Kim, "Using Structural Information to Detect Speech Repairs," 電子情報通信学会研究会、1995

[Kikui] Gen-ichiro Kikui, Tsuyoshi Morimoto, "Similarity-based Identification of Repairs in Japanese Spoken Language," ICSLP'94

[Levelt] Willem Levelt, "Monitoring and Self-repair in Speech," Cognition, 14, pp. 41-104, 1983

[Murakami] 村上　仁一、嵯峨山　茂樹、自然発話音声認識における音響的および言語的な問題点の検討, 電子情報通信学会論文集、 NLC 91 - 57, 1991

[Shinozaki] 篠崎、小倉、森元、”言語データベース作成のためのシミュレーション会話、”弟３７回情報処理全国大会、pp. 1000-1001, 1988 in [Murakami]

[Shriberg-1] Elizabeth Shriberg, "Preliminaries to a Theory of Speech Disfluencies," Ph.D. Thesis, UC Berkeley, 1994

[Shriberg-2] Elizabeth Shriberg, Patty Price, "User Behaviors Affecting Speech Recognition." Proceedings of the International Conference on Spoken Language Processing, pp. 995-998 in [Shriberg-1]

[Takezawa] 竹沢　寿幸、田代　敏久、森元　逞、”音声言語データベースを用いた自然発話現象の調査、”人工知能学会研究会資料, SIG-SLUD-9403-3(02/10), pp. 13-20

[Uratani-1] Noriyoshi Uratani, Toshiyuki Takezawa, Hidehiko Matsuo, Chiho Morita, "ATR Integrated Speech and Language Database," ATR Technical Report TR-IT-0056, May 1994

[Uratani-2] Noriyoshi Uratani, Toshihisa Tashiro, Hisako Yamata, Kaori Matsumoto, "Users Manual for Japanese Morphological Analysis in the ATR Spoken Language Database," ATR Technical Report TR-IT-0009, September 1993

APPENDIX

DISFLUENCY PATTERNS AND OCCURRENCES

d1-.	69	r1r2.i1i2r1r2	4	r1r2s1r3r4.r1r2s1r3r4	2
r1.r1	46	r1r2r3r4r5.r1r2r3r4r5	4	r1r2.r1i1r2	2
r1-.r1	42	r1r2r3s1.r1r2r3s1	4	r1s1r2r3r4r5.r1s1r2r3r4r5	2
r1~.r1	38	r1.i1r1	4	s1s2r1.s1s2r1	1
s1-.s1	31	s1-.s1s1s1	3	r1r2s1.fer1r2s1	1
s1.s1	20	s1r1.fs1r1	3	r1r2.r1i1i2r2	1
t1-.	18	r1d1.r1	3	r1r2s1-.er1r2s1s1	1
r1r2.r1r2	17	r1r2r3.r1r2r3	3	r1~.fr1r1	1
r1~.r1r1	15	r1r2r3~.r1r2r3	3	s1-.fi1s1	1
d1-.f	15	s1r1.s1r1	3	s1r1r2-.fs1r1r2	1
r1s1.r1s1	13	r1r2s1.r1r2s1	3	r1~.fr1	1
r1-.fr1	13	s1-.fs1s1	2	d1d2d3.f	1
r1-.i1r1	13	r1r2r3r4.i1i2r1r2r3r4	2	s1.efs1	1
r1-.i1i2r1	11	r1-.r1r1	2	r1r2r3r4r5.i1i2r1r2r3r4r5	1
s1-.fs1	8	r1r2-.r1r2	2	s1r1*.fs1r1	1
s1-.s1s1	8	r1s1.fr1s1	3	r1r2r3r4r5s1d1d2.fr1r2r3r4r5s1	1
r1~.r1r1r1	7	d1r1.r1	2	r1.i1i2i3r1	1
r1.i1i2r1	7	s1~.s1	1	s1r1r2.s1i1i2r1r2	1
r1r2.i1r1r2	7	s1-.fs1s1s1	2	r1r2r3r4r5.i1r1r2r3r4r5	1
r1.fr1	6	r1-.fi1r1	2	s1s2.s1s2	1
s1.s1s1	6	r1-.r1r1r1	2	r1s1~.r1s1	1
s1.fs1	6	d1.	2	s1s1s1r1.s1r1	1
r1r2~.r1r2	4	r1r2r3r4.r1r2r3r4	2	r1r2s1s1-.r1r2s1	1

continued

s1s1.s1	1
r1s1s1s1s1s1.r1s1s1	1
d1d2d3.	1
r1r2s1d1-.r1r2s1	1
r1r2r3r4r5r6r7.fr1r2r3r4r5r6r7	1
r1r2r3*d1.r1r2i1r3	1
r1-d1-.i1i2r1r1r1	1
r1r2r3r4r5r6r7r8r9r10s1.r1r2r3r4r5r6r7r8r9r10s1	1
s1r1.i1s1r1	1
r1r2d1-.fi1i2r1r2	1
r1*.i1r1	1
r1r2r3d1-.r1r2r3	1
s1d1.fs1	1
r1r2r3r4r5s1r6s2.r1r2r3r4r5i1s1r6s2	1
s1.es1s1	1
r1~.i1i2r1r1r1	1
s1d1.s1	1
d1r1s1s2s2.r1s1s2	1
r1r2d1-.r1r2	1
r1r2r3s1-.r1r2r3s1s1s1	1
r1r2r3s1r4r5.fr1r2r3s1i1i2i3i4i5i6i7r4r5	1
r1r2r3r4.i1i2i3i4i5r1r2r3r4	1
r1r2r3r4r5r6r7s1.r1r2r3r4r5r6r7s1	1
r1r2r3.i1r1r2r3	1
r1r2s1.fr1r2s1	1
r1r2d1r3r4.r1r2r3r4	1
r1d1s1.r1s1s1	1
r1r2r3.i1r1i2r2r3	1
r1r2r3r4s1.r1r2r3r4s1	1
r1r2r3r4s1.fr1r2r3i1i2i3r4s1	1
t1-t1-.	1
r1d1r2r3.r1r2r3	1
s1s1s1.fs1s1s1	1
r1d1d2r2.r1r2	1
r1s1r2.r1s1r2	1
s1s1s1r1r2r3r4.s1r1r2r3r4	1
r1r2r3~.r1r2r3r3	1
r1r2*.r1r2	1
d1r1-.fr1	1
r1.fi1r1	1
s1s1r1.s1r1	1
d1.f	1
r1s1-.fr1s1	1
d1-.ff	1
d1-r1-.r1	1
r1s1r2s2-.fr1s1r2s2	1
r1~.i1r1	1
s1-.ffs1s1	1
r1r2d1d2.fr1r2	1
r1-.fi1i2i3i4r1	1
d1~.	1
r1s1s1.r1s1	1
r1r2fr3r4.fr1r2i1i2r3r4	1
d1d2d3d4d5.	1
r1~.i1r1r1	1
d1~.f	1
r1d1d2d3.r1	1
r1~.fi1i2r1r1r1	1
r1r2~.i1i2i3i4r1r2	1
d1-d2-.	1
r1r2r3.fr1r2i1r3	1
r1r2.fr1r2	1
r1r2s1s1.r1r2s1	1
r1r2d1-.fr1r2	1
r1r2r3r4r5.ffffr4r5r1r2r3	1
s1.fefs1	1
d1d2.	1
r1r2d1d2d3.r1r2	1
r1r2r3r4r5r6r7.i1i2r1r2r3r4r5r6r7	1
r1r2fr3r4r5r6r7s1.i1i2i3i4i5r1r2r3r4r5r6r7s1	1
s1r1r2r3.s1r1r2r3	1

continued

r1s1s2r2r3r4r5r6r7r8.fr1s1s2r2r3r4r5r6r7r8	1
r1r2r3.er1ili2i3i4i5r2r3	1
s1r1.fili2s1r1	1
r1r2r3.fi1r1r2r3	1
r1r2r3d1d2d3d4.er1r2r3	1
r1r2.fr1i1r2	1
s1r1r2s2.s1r1r2s2	1
r1r2r3.ili2i3r1r2r3	1
s1r1.s1s1r1	1
r1r2r3r4r5r6r7r8r9s1r10r11r12r13r14r15r16r17r18-.r1r2r3r4r5r6r7r8r9s1r10r11r12r13r14r15r16r17r18r18	1
s1s1.fs1	1
s1r1.fs1s1r1	1
r1r2r3r4r5.fer1r2r3r4r5	1
r1s1.ei1r1s1	1
d1d2d3d4d5-.ff	1
s1.ili2s1	1
r1r2*d1d2.r1i1r2	1
r1r2r3r4r5s1.r1r2r5s1r3r4	1
r1.ili2i3i4r1	1
s1.s1s1s1s1s1s1s1s1s1s1s1s1	1
r1~.i1r1r1r1	1
s1s1~.es1	1
r1s1-.r1s1	1

s1.fili2s1s1s1	1
r1r2r3r4s1.fr1r2r3r4s1	1
r1~.fr1r1r1	1
r1r2r3d1r4r5.r1r2r3r4r5	1
d1d2.ff	1
r1s1.ei1i2r1s1	1
TOTAL	609