

CONSTRUCTION OF CORPUS-BASED SYNTACTIC RULES FOR ACCURATE SPEECH RECOGNITION

JUNKO HOSAKA

TOSHIYUKI TAKEZAWA

ATR Interpreting Telephony Research Laboratories
Hikaridai 2-2, Seika-cho, Soraku-gun
Kyoto 619-02, Japan
hosaka@atr-la.atr.co.jp
takezawa@atr-la.atr.co.jp

Abstract

This paper describes the syntactic rules which are applied in the Japanese speech recognition module of a speech-to-speech translation system. Japanese is considered to be a free word/phrase order language. Since syntactic rules are applied as constraints to reduce the search space in speech recognition, applying rules which take into account all possible phrase orders can have almost the same effect as using no constraints. Instead, we take into consideration the recognition weaknesses of certain syntactic categories and treat them precisely, so that a minimal number of rules can work most effectively. In this paper we first examine which syntactic categories are easily misrecognized. Second, we consult our dialogue corpus, in order to provide the rules with great generality. Based on both studies, we refine the rules. Finally, we verify the validity of the refinement through speech recognition experiments.

1 Introduction

We are developing the Spoken Language TRANSLATION system (SL-TRANS)[1], in which both speech recognition processing and natural language processing are integrated. Currently we are studying automatic speech translation from Japanese into English in the domain of dialogues with the reception service of an international conference office. In this framework we are constructing syntactic rules for recognition of Japanese speech.

In speech recognition, the most significant concern is raising the recognition accuracy. For that purpose, applying linguistic information turns out to be promising. Various approaches have been taken, such as using stochastic models[2], syntactic rules[3], semantic information[4] and discourse plans[5]. Among stochastic models, the bigram and trigram succeeded in achieving a high recognition accuracy in languages that have a strong tendency toward a standard word order, such as English. On the contrary, Japanese

belongs to free word order languages[6]. For such a language, semantic information is more adequate as a constraint. However, building semantic constraints for a large vocabulary needs a tremendous amount of data. Currently, our data consist of dialogues between the conference registration office and prospective conference participants with approximately 199,000 words in telephone conversations and approximately 72,000 words in keyboard conversations. But our data are still not sufficient to build appropriate semantic constraints for sentences with 700 distinct words. Processing a discourse plan requires excessive calculation and the study of discourse itself must be further developed to be applicable to speech recognition. On the other hand, syntax has been studied in more detail and makes increasing the vocabulary easier.

As we are working on spoken language, we try to reflect real language usage. For this purpose, a stochastic approach beyond trigrams, namely stochastic sentence parsing[7], seems most promising. Ideally, syntactic rules should be generated automatically from a large dialogue corpus and probabilities should also be automatically assigned to each node. But to do so, we need underlying rules. Moreover, coping with phoneme perplexity, which is crucial to speech recognition, with rules created from a dialogue corpus, requires additional research[8].

In this paper we propose taking into account the weaknesses of the speech recognition system in the earliest stage, namely when we construct underlying syntactic rules. First, we examined the speech recognition results to determine which syntactic categories tend to be recognized erroneously. Second, we utilized our dialogue corpus[9] to support the refinement of rules concerning those categories. As examples, we discuss *formal nouns*¹ and *conjunctive postpositions*². Finally, we carried out a speech recognition experiment with the refined rules to verify the validity of our approach.

¹ *Formal nouns*: keishiki-meishi in Japanese.

² *Conjunctive postpositions*: setsuzoku-joshi in Japanese.

2 Issues in HMM-LR Speech Recognition

In the Japanese speech recognition module of our experimental system the combination of generalized LR parsing and Hidden Markov Model (HMM) is realized as HMM-LR [10]. The system predicts phonemes by using an LR parsing table and drives HMM phoneme verifiers to detect/verify them without any intervening structure, such as a phoneme lattice.

The speech recognition unit is a Japanese *bunsetsu*, which roughly corresponds to a phrase and is the next largest unit after the word. The ending of the *bunsetsu* (phrase) is usually marked by a breath point. This justifies its treatment as a distinct unit. A Japanese phrase consists of one independent word (e.g. noun, adverb, verb) and zero, one or more than one dependent words (e.g. postposition, auxiliary verb). The number of words in a phrase ranges from 1 to 14, and the mean number is about 3, according to our dialogue corpus.

We will clarify the weaknesses of HMM-LR speech recognition both in phrases and in sentences.

2.1 Phrase Recognition Errors

We examined which syntactic categories tend to be erroneously recognized, when using HMM-LR phrase speech recognition. For this purpose, we applied syntactic rules containing no constraints on word sequences³. This means that any word can follow any word.

Examples (1) and (2) show the results of HMM-LR Japanese speech recognition⁴. The uttered phoneme strings are enclosed in | |.

(1) |sochirawa| (this, that)

-
- > 1: sochira-wa
 - 2: sochira-wa-hu
 - 3: sochira-hu-wa
 - 4: sochira-hu-wa-hu
 - 5: sochira-wa-hu-hu
-

(2) |aringatougozaimasu| (thank you)

-
- 1: ari-nga-to-wa-eN-hu-su-su-su
 - 2: ari-nga-to-wa-eN-hu-su-su
 - 3: ari-nga-to-wa-eN-hu-su-su-u
 - 4: ari-nga-to-wa-eN-su-su
 - 5: ari-nga-to-wa-eN-hu-su-su-su-u
-

³Japanese verbs, adjectives, etc. are always inflected when used. In syntactic rules containing no word sequence constraints, inflected verbs, inflected adjectives, etc. are considered to be words.

⁴The maximal amount of whole beam width, the global beam width, is set for 16 and the maximal beam width of each branch, the local beam width, 10.

In the examples, the symbols >, -, ng and N have special meaning:

- A correctly recognized phrase is marked with >.
- A word boundary is marked with -.
- A nasalized /g/ is transcribed ng.
- A syllabic nasal is transcribed N.

In (1), after recognizing the first word, the system selected subsequent words solely to produce a phoneme string similar to the original utterance.

(2) is an example of phrase recognition which failed. In this example *to* was erroneously recognized as *to*. Subsequently, no further correct words were selected.

Examples (1) and (2) both show that HMM-LR tends to select words consisting of extremely few phonemes when it fails in word recognition. To avoid this problem, precise rules should be written for sequences of words with small numbers of phonemes. In Japanese, postpositions (e.g. *ga*, *o*, *ni*), *wh*-pronouns (e.g. *itsu*, *nani*, *dare*) [11], numerals (e.g. *ichi*, *ni*, *san*) and certain nouns (e.g. *kata*, *mono*) particularly fit this description.

2.2 Sentence Recognition Errors

To examine the error tendency of sentence speech recognition we applied a two-step method [12]. First, we applied phrase rules to the HMM-LR speech recognition⁵. Second, we applied phrase-based sentence rules to the phrase candidates as a post-filter, in order to obtain sentence candidates, while filtering out unacceptable candidates. We experimented with the 353 phrases making up 137 sentences. The recognition rate for the top candidates was 68.3 % by exact string matching, and for the top 5 candidates 95.5 %.

Based on the top 5 phrase candidates, we conducted a sentence experiment. In this experiment we applied loosely constrained sentence rules. With these rules, approximately 80 % of all the possible combinations of phrase candidates were accepted. Following are examples which did not exactly match the uttered sentences⁶. Notice that misrecognized words consist of a relatively small number of phonemes, as we have seen in section 2.1.

(3) |kaingi-ni moushiko-mi-tai-no-desu-nga|
(? would like to register for
the conference.)

3a: kaingi-ni moushiko-mi-tai-N-desu-nga
3b: kaingi-ni moushiko-mi-tai-no-desu-ka

(4) |kochira-wa kaingizimukyoku-desu|

⁵The global beam width is set for 100 and the local beam width 10.

⁶Since the phrase candidates are obtained by the HMM-LR speech recognition, word boundaries are already marked by -.

(This is the conference office.)

4a: kata-wa kaingizimukyoku-desu

(5) |doumo aringat-ou-gozaima-shi-ta|
(Thank you very much.)

5a: go-o aringat-ou-gozaima-shi-ta
5b: go-mo aringat-ou-gozaima-shi-ta
5c: mono aringat-ou-gozaima-shi-ta

(6) |gozyuusho-to onamae-o onengai-shi-masu|
(Can I have your name and address?)

6a: gozyuusho-to onamae-o
onengai-shi-masu-shi

Though the phoneme string in 3a is different from the uttered phoneme string, the difference between *no* and *Ń* in meaning is minor, and has no effect on translation with the current technique. While (3) is affirmative, 3b is interrogative, which is indicated by the sentence final postposition *ka*. This cannot be treated with sentence rules. To handle this problem, we need dialogue management.

The uttered phrase *kochira-wa* in (4), meaning "this," was recognized erroneously as *kata-wa* in 4a, meaning "person." The word *kata* belongs to the formal noun group, a kind of noun which should be modified by a verbal phrase [13]. Sentence 4a is acceptable, if modified by a verbal phrase, as in 4a':

4a': *midori-no seihuku-o kiteiru kata-wa*
kaigizimukyoku-desu

(The person who is wearing a green uniform is [with] the conference office.)

This is also true of the phrase *mono* in 5c meaning "thing," which was erroneously recognized instead of *doumo* meaning "very much":

5c': *kouka-na mono aringat-ou-gozaima-shi-ta*

(Thank you for the expensive thing.)

In sentence candidates 5a and 5b, the numeral *go*, meaning "five," is used. These sentences may seem strange at first glance, but in a situation such as playing cards, these sentences are quite natural. If someone plays a 5 when you need one, you would say: "Thanks for the five." Similarly, when you need a 3 and a 5, and someone plays a 3 and after that someone else plays a 5, you would say: "Thanks for the five, too."

In the sentence candidate 6a, the *conjunctive-postposition* (*conj-pp*) *shi* is used sentence finally. In principle, a *conj-pp* combines two sentences, functioning like a conjunction, such as "while" and "though," and is used in the middle of a sentence.

Erroneous sentence recognition such as in the case of 3a-b cannot be treated by sentence rules. Therefore, we are trying to cope with erroneous recognition, as seen in sentence candidates 4a, 5a-c and 6a, with sentence rules.

3 Dealing with Speech Recognition Errors

We are going to deal with sentences containing the following phrases:

- Phrases with formal nouns
- Phrases with numerals
- Phrases with *conj-pps* used in the sentence final position

In order to decide how to cope with the above problems, we used our dialogue corpus. Currently we have 177 keyboard conversations consisting of approximately 72,000 words and 181 telephone conversations consisting of approximately 199,000 words⁷. We regard keyboard conversations as representing written Japanese and telephone conversations as representing spoken Japanese. When retrieving the dialogue corpus, we always compare written and spoken Japanese, in order to clarify the features of the latter. We examined the actual usage of formal nouns as well as that of *conj-pps*.

3.1 Formal Nouns

We examined the behavior of formal nouns, such as *koto* and *mono*. Formal nouns are considered to be a kind of noun which lacks the content usually found in common nouns such as "sky" or "apple." They function similarly to relative pronouns and therefore are used with a verbal modifier [13], as in examples 7 and 8:

7: *kinou itta koto-wa torikeshitai.*

(I would like to take back what I said yesterday.)

8: *nedan-ga takai mono-ga shitsu-ga ii wakedenai.*

(It is not always true that an expensive thing has good quality.)

In examples 7 and 8, the formal nouns, *koto* and *mono*, are modified by *kinou itta* (yesterday said) and *nedan-ga takai* (price expensive), respectively. But it is also true that these nouns behave like common nouns and can be used without any verbal modifier, as in examples 9 and 10:

9: *sore-wa koto desu ne.*

⁷The dialogue corpus is growing constantly. When we retrieved formal nouns, we had 113 keyboard conversations and 96 telephone conversations.

(It is a grave matter.)

10 : mono-wa tashika-da.

(This stuff is trustworthy.)

Considering the examples 7-10, we could define two kinds of usage for formal nouns. This distinction is applicable to sentence analysis, but is meaningless from the standpoint of applying syntactic rules as constraints.

3.1.1 Formal Nouns in the Corpus

In our dialogue corpus, *koto*, *mono*, *hou* and *kata* are the most frequently used formal nouns. Table 1 shows how often the formal nouns are used with a verbal modifier. We have also retrieved formal nouns used in the sentence initial position, as in example 10.

Table 1: Formal Nouns

	Keyboard		Telephone	
	Freq.	%	Freq.	%
With Verb. Mod.	358	63	774	40
Without Verb. Mod.	214	37	1,145	60
Sent. Initial	0	0	0	0
Total	572	100	1,919	100

Table 1 indicates that the coverage reaches 63 % in written Japanese, when we allow only formal nouns preceded by a verbal modifier in the syntactic rules. However, the coverage remains at 40 %, which is less than half, in the spoken Japanese we are dealing with. We have further examined those sentences in which formal nouns are not modified by verbals. Most of them are modified by phrases consisting of a noun and postposition *no*, which approximately corresponds to "of." Further, some are modified by phrases consisting of a verb followed by postpositions *to* and *no*. Others are modified by words which can be used exclusively as nominal modifiers such as *donna* (what kind of) and *sono* (that). We found only one example in the keyboard conversation in which a formal noun is not modified at all:

11 : osoraku kyouju-ni koto-no shidai-o tsutaeru koto-ga ii-to omoimasu.

(It might be good if you tell the professor how the thing is going.)

In our dialogue corpus we found 2,491 phrases containing the formal nouns *koto*, *mono*, *hou* and *kata*. Out of 2,491 examples, there is only one which is not modified at all. If we define formal nouns as those which are always modified in some manner, i.e. even if we do not allow formal nouns to be used alone, the coverage still exceeds 99 %. Since the occurrence rate of formal nouns without any modifier is very low, we can treat the usage of formal nouns (as in examples 9-11) as semi-frozen expressions.

3.2 Conjunctive Postpositions

Japanese postpositions such as *ga*, *o* and *ni*, which function as case markers, are usually attached to nominals. Different from this kind of postposition, conj-pps such as *ga*, *te* and *ba* are used after verbals. Conj-pps combine two clauses, functioning similarly to conjunctions such as "because" and "while," and are thus often used in the middle of a sentence, as in example 12. But they can also be used in the sentence final position, as in example 13.

12: *kaigi-ni mousikomi-tai-no-desu-ga,*
 DAT
mousikomiyousi-o ookurikudasai.
 AKK

Because I would like to apply for the conference, please send me a registration form.

13: *kaigi-ni mousikomi-tai-no-desu-ga.*
 DAT

I would like to apply for the conference, ...

Example 13 sounds vague, if uttered in isolation. There should follow some additional words to express the complete meaning. Sentences finishing with a conj-pp leave the interpretation to the hearer. And, in general, the hearer can correctly interpret the sentence from the context. Understanding conj-pps, therefore, plays an important role in treating spoken Japanese.

3.2.1 Sentence Final Conj-pps in the Corpus

In the dialogue corpus the following conj-pps are used:

ga (because, while), *node* and *nde* (because), *te* and *and* (and), *kara* (because, after), *keredomo*, *keredo*, *kedo* and *kedomo* (though, but), *shi* (and, and then), *monode* (because), *tara* (if), *to* (if, when), *ba* (if) and *nagara* (while).

Table 2 shows conj-pps used sentence finally.

According to Table 2, the conj-pp *ga* is the one most used in keyboard conversations. While the usage of conj-pps in keyboard conversations is heavily concentrated on *ga* with an occurrence rate of 85%, it is more balanced in telephone conversations. In addition to *ga* (38%), *keredomo* (30%) and conj-pps which carry a similar meaning such as *keredo*, *kedo* and *kedomo* are frequently used. In telephone conversations, *node* (13%) is also frequently used. Treating only the six conj-pps in sentence final position, the coverage reaches 91% for spoken Japanese. Differentiating conj-pps which can be used in sentence final position from those which can be used only in the middle of a sentence is also supported by the speech recognition results[14]. The conj-pps *shi* and *cha* are especially subject to erroneous recognition.

Table 2: Sentence Final Conj-pps

Conj-pp	Keyboard		Telephone	
	Frequency	%	Frequency	%
<i>ga</i>	197	85	274	38
<i>node</i>	11	5	96	13
<i>nde</i>	0	0	5	1
<i>te</i>	8	3	23	3
<i>de</i>	0	0	1	0
<i>kara</i>	6	3	14	2
<i>keredomo</i>	5	2	212	30
<i>keredo</i>	1	0	18	3
<i>kedo</i>	1	0	12	2
<i>kedomo</i>	0	0	37	5
<i>shi</i>	2	1	10	1
<i>monode</i>	1	0	0	0
<i>tara</i>	0	0	5	1
<i>to</i>	0	0	2	0
<i>ba</i>	0	0	2	0
<i>nagara</i>	0	0	1	0

3.3 Syntactic Rules for Speech Recognition

Based on the corpus retrieval we decided to deal with formal nouns and conj-pps as described below. And we decided to treat numerals only in a restricted environment, because they are significant noise factors in speech recognition⁸:

- Phrases with formal nouns must be modified.
- Phrases with numerals can be used only in certain environments. Numerals are allowed in addresses, telephone numbers, dates and prices. Japanese numerals consist of an extremely small number of phonemes, e.g. *ichi*, *ni*, *san* (1, 2, 3) and are therefore especially easy to misrecognize⁹. Thus, they should be strongly constrained. The domain we have chosen is limited to dialogues between an international conference receptionist and prospective participants and we are going to deal only with the anticipated usage in the domain. Another condition, such as playing cards, will be treated when speech recognition is further improved.
- We classify conj-pps into two groups: conj-pps which can be used in the sentence final position as well as in the middle of a sentence, and conj-pps which can be used only in the middle of a sentence.

We refined the loosely constrained syntactic rules introduced in section 2.2. In the new version of the sentence rules, formal nouns, numerals and conj-pps are more precisely treated. In the following, we explain the rules for formal nouns and conj-pps.

⁸See Figure 2.

⁹Numbers greater than ten are in principle the combination of basic numbers.

The format for syntactic rules is as follows:

(**<CAT1>** **<->** (**<CAT2>** **<CAT3>**))

Nonterminals are surrounded by **<>**¹⁰. The above rule indicates that **CAT1** consists of **CAT2** and **CAT3**. To make the distinction between phrase categories which are terminals in phrase-based sentence rules and those which are not, we will write the former all in lower-case.

In the process of sentence construction, phrases containing a formal noun **np-formal** are treated as follows¹¹:

(**<M-NN>** **<->** (**<NN>**)
 (**<M-NN>** **<->** (**<MOD-N>** **<NN-FORM>**))
 (**<M-NN>** **<->** (**<MOD-N>** **<M-NN>**))

 (**<NN>** **<->** (**<np>**))
 (**<NN-FORM>** **<->** (**<np-formal>**))

The above rules say that noun phrases **M-NN** can, in principle, be modified by some modifier **MOD-N**. In the case of a common noun **NN**, the phrase can be modified but need not be. But in the case of a formal noun **NN-FORM**, the phrase must be modified.

Phrases with a conj-pp which is exclusively used in the middle of a sentence **vaux-s**, those with a conj-pp which is used both in the middle of a sentence and in the sentence final **vaux-s+f**, and verb phrases without any conj-pps **vaux**, are treated as follows:

(**<SS>** **<->** (**<NVS>**)
 (**<NVS>** **<->** (**<VS>**))
 (**<VS>** **<->** (**<VC>**))
 (**<VS>** **<->** (**<ADVPH>** **<VS>**))

 (**<ADVPH>** **<->** (**<ADV-s>**)
 (**<ADV-s>** **<->** (**<ADV1>**))
 (**<ADV-s>** **<->** (**<ADV1>** **<ADV-s>**))
 (**<ADVPH>** **<->** (**<ADV-c>**)
 (**<ADV-c>** **<->** (**<VADVS>**))
 (**<VADVS>** **<->** (**<VADV>**))
 (**<VADVS>** **<->** (**<ADV-s>** **<VADVS>**))

 (**<VADV>** **<->** (**<vaux-s>**))
 (**<VADV>** **<->** (**<vaux-s+f>**))
 (**<VC>** **<->** (**<vaux>**))
 (**<VC>** **<->** (**<vaux-s+f>**))

A sentence **SS** does not always need a noun phrase. A sentence **SS** can consist of only one verb phrase **VC**, or can be preceded by adverbial phrases **ADVPH**. A sentence **SS** can end either with a verb phrase without a conj-pps **vaux** or with a verb phrase with a certain kind of conj-pps **vaux-s+f**. An adverbial phrase **ADVPH** can consist of only adverbs **ADV1** and can also consist of verbal phrases **VADVS**. The verbal phrases

¹⁰For terminals we have a different notation. Terminals in phrase rules are phoneme strings, whose transcription is defined by the HMM-LR phoneme model.

¹¹For the sake of explanation, the rules are simplified.

VADVS can contain any conj-pps, which means both *vaux-s* and *vaux-s+f*.

Compared with the first version, which accepts approximately 80 % of the sentence candidates consisting of all the possible combinations of phrase candidates, the refined version only accepts approximately 30 %. Table 3 shows the size and the perplexity of the phrase rules and phrase-based sentence rules.

Table 3: Size and Perplexity of Syntactic Rules

	Phrase Rules	Sentence Rules
No. of Rules	1,973	471
No. of Terminals	744	133
Perplexity	3.57/Phoneme	99.7/Phrase

4 Validity of Rule Refinements

We tested the improvement in two ways: speech recognition accuracy and the acceptance rate[12]. To estimate the latter we checked how many sentence candidates were filtered out by applying phrase-based sentence rules as a post-filter. We verified the rule refinements through comparison of results gained by five different rule sets: the refined version of sentence rules which contain all three refinements (**New Grammar**); the refined version without conj-pp treatment (**No Sentence Final Conj-pp**), without formal noun treatment (**No Formal Noun Treating**), and without numeral treatment (**No Numeral Treating**); and rules which allow all combinations of phrase candidates (**No Grammar**). For the first four of these rule sets we determined ranks based on the probabilities of phoneme strings predicted by syntactic rules. But in the **No Grammar** case we determined the rank solely based on phoneme probability. We experimented with the same 353 phrases which make up 137 sentences as in section 2.2. The phrase recognition rate for the top 5 candidates was again 95.5% by exact string matching.

4.1 Speech Recognition Accuracy

We conducted speech recognition experiments. Figure 1 shows the constraint effectiveness of the phrase-based sentence rules given the five conditions examined. These five conditions are compared in the graph, based on their abilities to correctly recognize the spoken sentences among the top-ranked 20 candidates.

While the sentence recognition rate for the top candidates remains 37.2 % when probability is the only factor in determining the candidates, the recognition rate rises to 70.1 % when the refined syntactic rules are applied as constraints. Differentiating conj-pps is highly effective. Without this treatment, the recognition rate remains 48.2%. Formal noun and numeral treatments are not as effective. Figure 1 indicates

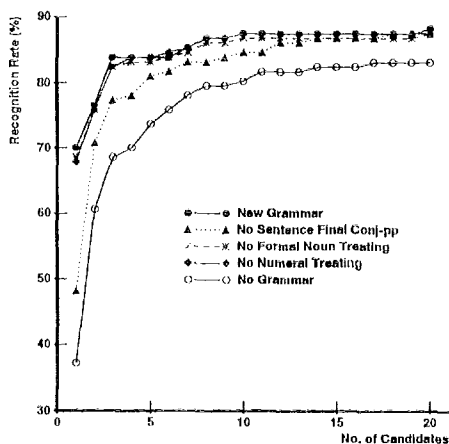


Figure 1: Comparison of Recognition Rates

that the effect according to each syntactic constraint is especially distinct up to rank 5, and that the recognition rates saturate when we take into account sentence candidates up to rank 10.

4.2 Acceptance Rate

We also verified the validity of sentence rules through the acceptance rate. We examined how many sentence candidates were filtered out. Table 4 shows the frequencies of sentences consisting of different numbers of phrases in our test corpus:

Table 4: Phrase Number and Frequency

Phrase Number	1	2	3	4	5	6	7	8
Frequency	45	37	23	11	11	5	4	1

Figure 2 shows the acceptance rates when applying four different syntactic rules. When applying rules which allow all combinations of phrase candidates, the acceptance rate remains 100 %.

The effect of constraints is especially clear for sentences with a small number of phrases. In sentences with one phrase, the acceptance rate for the revised version is 41%, and for the version without conj-pp constraints 70%. In comparison with Figure 1, treating numerals contributes toward filtering out sentence candidates rather than raising speech recognition accuracy. Independent of the constraint strength, the more phrases there are in a sentence, the more effectively the rules work. The value for a sentence with 8 phrases is unreliable, as we have only one example.

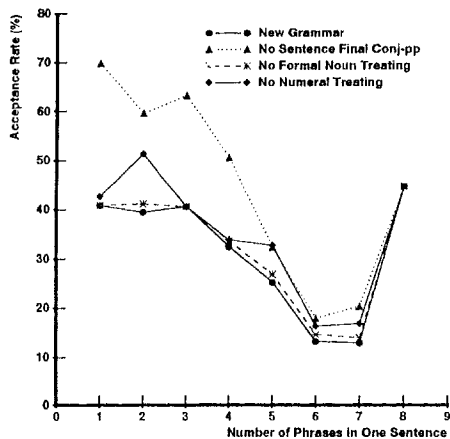


Figure 2: Acceptance Rate

5 Conclusion

We have described phrase-based syntactic rules which are used as constraints in the Japanese speech recognition module of our experimental speech-to-speech translation system. For constructing rules we took into account the error tendency in speech recognition. We treated precisely those syntactic categories which tend to be recognized erroneously. To increase the efficacy of each rule, the rule construction is strongly motivated by our dialogue corpus. By applying the refined phrase-based syntactic rules, the speech recognition rate for the top candidates improved from 37.2 % to 70.1 % and for the top 5 candidates from 73.7 % to 83.9 %.

The implementation of syntactic rules based on our dialogue corpus is continuing in order to increase coverage. Currently we are studying postposition deletion in nominal phrases, which is one of the features of spoken Japanese. When adding rules and enlarging vocabulary, we cannot avoid decreasing speech recognition accuracy, but our further experiments showed that careful rule construction filtered out unacceptable sentence candidates much more effectively. Though we believe that our dialogue corpus for the current domain provides enough expressions of spoken Japanese, we are going to apply the same method to other domains to establish the generality of the rules.

Acknowledgements

The authors wish to thank Dr.A.Kurematsu, President of ATR Interpreting Telephony Research Labs for his continued support, Mr.T.Morimoto for discussion of the various

stages of this work, Mr.K.Inoue for his help in database retrieval and Dr.S.Luperfoy and Dr.L.Fais for reading an earlier draft.

References

- [1] Morimoto, T., Shikano, K., Iida, H., Kurematsu, A.(1990): "Integration of Speech Recognition and Language Processing in Spoken Language Translation System(SL-TRANS)," Proc. of ICSLP-90, pp.921-924.
- [2] Lee, K.-F. and Hon, H.-W.(1988): "Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM," Proc. of ICASSP-88, pp.123-126.
- [3] Ney, H.(1987): "Dynamic Programming Speech Recognition Using a Context-Free Grammar," Proc. of ICASSP-87, pp.69-72.
- [4] Matsunaga, S., Sagayama, S., Homma, S. and Furui, S.(1990): "A Continuous Speech Recognition System Based on a Two-Level Grammar Approach," Proc. of ICASSP-90, pp.589-592.
- [5] Yamaoka, T. and Iida, H.(1990): "A Method to Predict the Next Utterance Using a Four-layered Plan Recognition Model," Proc. of ECAI-90, pp.726-731.
- [6] Kuno, S. (1973): *The Structure of the Japanese Language*, The MIT Press, Cambridge, Massachusetts and London.
- [7] Fujisaki, T. (1984): "A Stochastic Approach to Sentence Parsing," Proc. of COLING-84, pp.16-19.
- [8] Ferretti, M., Maltese, G., Scarci, S. (1990): "Measuring Information Provided by Language Model and Acoustic Model in Probabilistic Speech Recognition: Theory and Experimental Results," *Speech Communication* 9, pp.531-539.
- [9] Ehara, T., Ogura, K., Morimoto, T. (1990): "ATR Dialogue Database," Proc. of ICSLP-90, pp. 1093-1096.
- [10] Kita, K., Kawabata, T., Saito, H. (1989): "HMM Continuous Speech Recognition Using Predictive LR Parsing," Proc. of ICASSP-89, pp.703-706.
- [11] Hosaka, J., Ogura, K., Kogure, K. (1990): "Word Sequence Constraints for Japanese Speech Recognition," Proc. of ECAI-90, pp. 363-365.
- [12] Takezawa, T., Kita, K., Hosaka, J., Morimoto, T. (1991): "Linguistic Constraints for Continuous Speech Recognition in Goal-Directed Dialogue," Proc. of ICASSP 91, pp.801-804.
- [13] Ogawa, Y., Hayashi, H., et al. (1982, 1988): *Nihongo Kyouniku Jiten*, Taishukan, Tokyo, (In Japanese).
- [14] Hosaka, J., Takezawa, T., Ehara, T. (1991): "Utilizing Empirical Data for Postposition Classification toward Spoken Japanese Speech Recognition," Proc. of ESCA-91, pp. 573-576.