SPEECH RECOGNITION SYSTEM FOR SPOKEN JAPANESE SENTENCES

Minoru Shigenaga, Yoshihiro Sekiguchi and Chia-horng Lai

Faculty of Engineering, Yamanashi University
Takeda-4, Kofu 400, Japan

Summary: A speech recognition system for continuously spoken Japanese simple sentences is described. The acoustic analyser based on a psychological assumption for phoneme identification can represent the speech sound by a phoneme string in an expanded sense which contains acoustic features such as buzz and silence as well as ordinary phonemes. Each item of the word dictionary is written in Roman letters of Hepburn system, and the reference phoneme string and the reference characteristic phoneme string necessary for matching procedure of input phoneme sequences are obtained from the word dictionary using a translating routine. In syntax analysis, inflexion of verbs and adjectives and those of some main auxiliary verbs are taken into account. The syntax analyser uses a network dealing with state transition among parts of speech, predicts following words and outputs their syntactic interpretation of the input phoneme string. The semantic knowledge system deals with semantic definition of each verb, semantic nature of each word and the schema of the sentence, and conconstructs a semantic network. The semantic analyser examines semantic validity of the recognized sentence as to whether each word in the sentence meets the definition of the recognized verb or others. The present object of recognition is a Japanese fairy tale composed of simple sentences alone. The syntactic and semantic analysers work well and can recognize simple sentences provided that the acoustic analyser outputs correct phoneme strings. For real speech, though the level of semantic processing is yet low, it can recognize 25 blocks out of 33 blocks (A block means a part of speech sound uttered in a breath.), and 9 sentences out of 16 sentences uttered by an adult male.

1. Introduction

Intensive studies of speech recognition or speech understanding are being carried out [1-3], but there are some fundamental problems to be solved both in acoustic analysis and linguistic processing. The authors think there must exist some fundamental procedures to be applicable to any task in speech recognition, and are trying to solve the problems through the behavior of two recognition systems which deal with Japanese sentences [4] and FORTRAN programs [5] spoken without interruption.

Both the recognition systems consist of two parts: an acoustic analyser and a linguistic processor. In the acoustic analysis, recognition model based on a psychological assumption is introduced for phoneme identification. As a result, speech sound has come to easily be expressed in a phoneme string in an expanded sense that contains some acoustic features such as buzz and silence as well as ordinary phonemes. The systems require a process of learning a small number of training samples [6] for identification of the speaker's vowels, nasals and buzz. In the linguistic processor, using major acoustic features as well as linguistic information has made it possible to effectively reduce the number of candidate words. For sequences of phonemes with erroneous ones has also been devised a graphic matching method [7] more suitable for matching than the one using dynamic programming.

In the previous system for Japanese sentences, sentences were narrowly limited in a pre-decided style. In the new system, as shown in Fig. 1.1, the knowledge system is much reinforced. That is, in the syntax analysis, inflexion of verbs and adjectives and those of some main auxiliary verbs can be referred; thus the syntax analyser may be able to deal with various kinds of simple sentences. A simulation has confirmed the ability of syntax analyser for simple sentences which have been offered in terms of Roman letters without any partition between words. In the semantic knowledge source, semantic definition of verbs, natures of nouns, a simple schema for a topic are stored, and semantic network will be constructed as a recognition process goes on. This semantic knowledge is used to yield, at the end of spoken sentence, the most semantically probable sentence as an output and occasionally to reduce the number of candidate words in co-operation with the syntax analyser.

2. Acoustic Analyser and Matching Method

A psychology based model is used to obtain neat phoneme string from speech wave using the following feature parameters determined every ten milli-seconds [5].

(i) Maximum value of amplitudes,
(ii) Number of zero-crossing,
(iii) Normalized prediction error,
(iv) Parcor-coefficients,
(v) Variation of Parcor-coefficients between successive frames,
(vi) Frequency spectrum,
(vii) Formant frequencies.

The output phonemes and their decision methods are given in Table 2.1. The obtained output phoneme strings contain 5 Japanese vowels, a nasal group, an unvoiced stop consonant group, /s/, /h/, /r/, buzz parts and silence. Discrimination of each stop consonant [8] and that of each nasal consonant are not yet embodied in this system.

Vowels and /s/ having long duration and silent parts are used as characteristic phonemes.

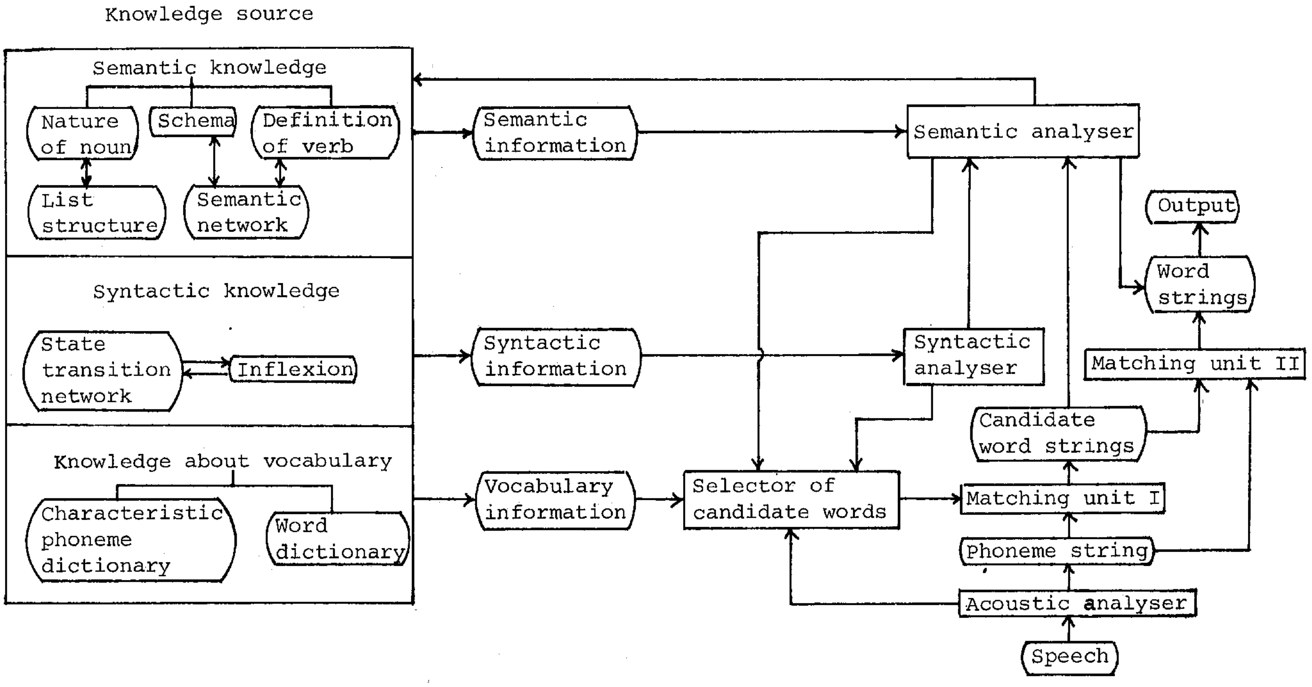


Fig. 1.1 Speech recognition system.

Besides an ordinary word dictionary, a characteristic phoneme dictionary (This dictionary exists only implicitly and is automatically composed from the word dictionary which is written in Roman letters.) is prepared and presents major acoustic features of each word. These major features are used for reduction of the number of candidate words.

For matching between a phoneme string with erroneous phonemes and items of the word or characteristic phoneme dictionaries, a new matching method using graph theory is devised [7].

These acoustic and matching processings are the same as the ones in the previous systems.

3. Knowledge Representation

3.1. Syntactic Knowledge

3.1.1. Classification of Japanese words for machine recognition

In order to automatically recognize continuously spoken natural languages, it is necessary to use syntactic rules. However using the original form of Japanese grammar written by grammarians is not necessarily suitable for mechanical recognition. Moreover it is very difficult to reduce the number of predicted words only by syntactic information because of the nature of Japanese language which does not require to keep the word order so rigorously. Taking account of these conditions, Japanese words are classified as described in the following article and the syntax may preferably be represented by state transition networks as shown in section 3.1.3.

3.1.1.1. Classification of words by parts of speech

Each word is classified grammatically as given in Table 3.1. In Japanese nouns, pronouns, numerals and quasi-nouns (KEISHIKI-MEISHI in Japanese) are called substantives (inflexionless parts of speech in Japanese grammar, TAIGEN in Japanese), and verbs, auxiliary verbs and adjectives are called inflexional words (inflexional parts of speech, YOGEN in Japanese). Meanwhile the words No. 1 - No. 11 in Table 3.1 are inflexionless words and the words No. 12 - No. 15 are

Table 2.1 Output phonemes and their decision methods.

Class	Output Phoneme	Decision Method
Vowel	i,e,a,o,u	Parcor-coefficients k, using Bayes decision theory
Nasal	m,n,ŋ,N	
Buzz	denoted by B	
Fricative	s	Number of zero-crossings
	h	Variations of amplitude and spectrum, Number of zero-crossings, and Unsimilarity to vowels and nasals
Liquid	r	Variations of amplitude and first formant frequency, Number of zero-crossings
Unvoiced stop	p,t,k	Following after silence and Having high frequency components
Silence	•	Small amplitude

inflexional words. In No. 16 the inflexion rules necessary for each inflexional word are written in appropriate forms. The additional word "carriage return" in No. 17 is a special symbol. We ask each speaker to utter the word "carriage return" at the end of each sentence in order to inform the recognizer of the end of a sentence.

Japanese verbs, adjectives and auxiliary verbs are inflexional. The verb's inflexion has been classified traditionally into 5 kinds of inflexion types: GODAN-KATSUYO (inflexion), KAMI-ITCHIDAN-KATSUYO, SHIMO-ICHIDAN-KATSUYO, SAGYO-HENKAKU-KATSUYO and KAGYO-HENKAKU-KATSUYO. But we classify them into 14 types as given in Table 3.2 taking into account the combination of the stem, a consonant following the stem and the inflexional ending of each word. Examples are shown in Fig. 3.1. By so doing the number of inflexion tables becomes smaller.

The adjectives and verbal-adjectives(KEIYO-DOSHI in Japanese) have we classified into 3 types according to their inflexion. Two types of them are shown in Fig. 3.2.

The inflexion of auxiliary verbs is the same as the traditional one. Some examples are shown in Fig. 3.3.

Table 3.1 Classification of words by parts of speech. No.16 and 17 are exceptional.

No.	part of speech
1	noun
2	pronoun
3	numeral
4	quasi-noun
5	prefix
6	suffix
7	part modifying substantives
8	adverb
9	conjunction
10	exclamation
11	particle
12	verb
13	adjective
14	auxiliary verb
15	subsidiary verb
16*	inflexion
17*	carriage return

Table 3.2 Classification of verbs.

No.	Inflexion	Example
1	GODAN-KATSUYO 1	IKU
2	" 2	KATSU
3	" 3	NORU
4	" 4	KAU
5	" 5	SHINU
6	" 6	YOMU
7	" 7	YOBU
8	" 8	SAKU
9	" 9	OSU
10	" 10	OYOGU
11	KAMI-ICHIDAN-KATSUYO, SHIMO-ICHIDAN-KATSUYO	OKIRU NAGERU
12	SAGYO-HENKAKU-KATSUYO	SURU
13	KAGYO-HENKAKU-KATSUYO	KURU
14	Verb: ARU (be)	ARU

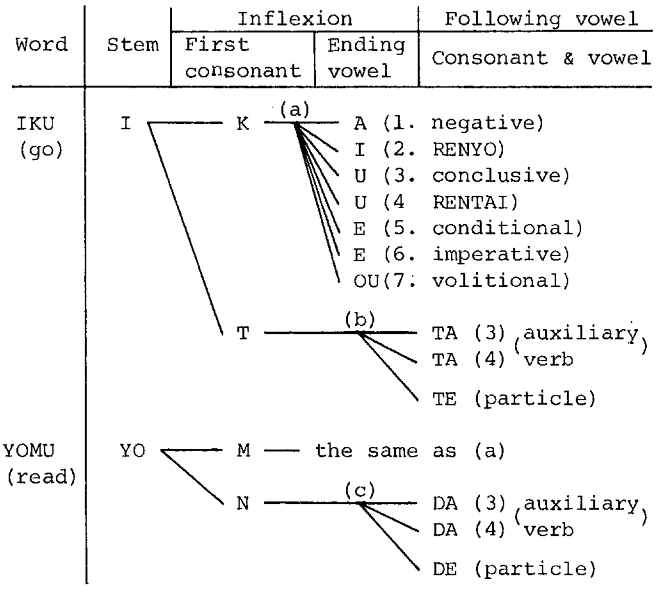


Fig. 3.1 Inflexion of verbs: IKU (go)(No.1 in Table 3.2) and YOMU (read)(No.6 in Table 3.2). RENYO or RENTAI means that the following word must be inflexional or substantive respectively. The following words TA and DA are auxiliary verbs and TE and DE are particles.

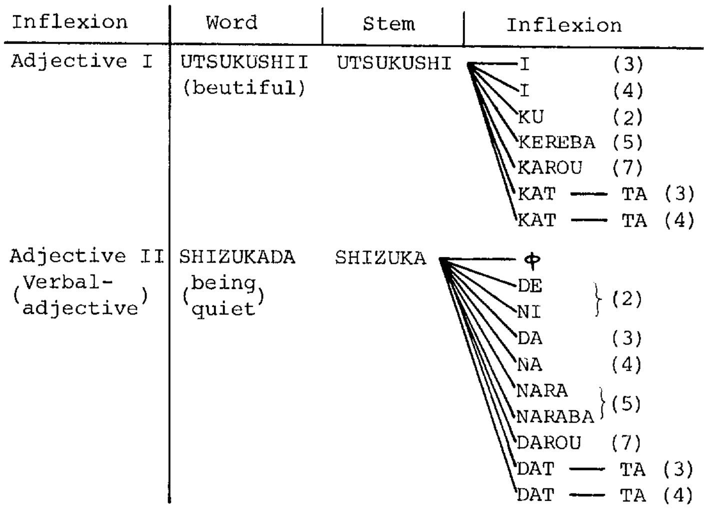


Fig. 3.2 Examples of inflexion of an adjective and a verbal-adjective. The numbers in parentheses are identified with the ones in Fig. 3.1.

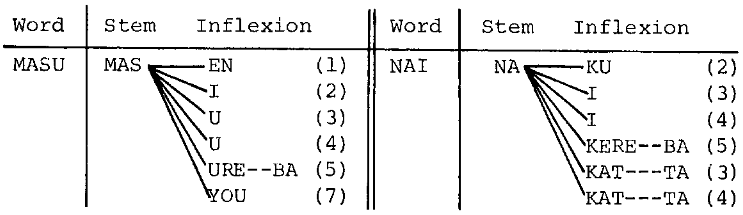


Fig. 3.3 Examples of inflexion of auxiliary verbs. The numbers in parentheses are identified with the ones in Fig. 3.1.

3.1.1.2. Classification of words by syntactic functions

In a Japanese sentence some words express material (noēma) such as substantives and verbs, and the others express syntactic function (noēsis) such as particles and auxiliary verbs [9]. The latter controls the syntactic function of the former, or, in other words, gives a material word or phrase a modifying function and these two words usually appear in a pair in sentences. The pair is called a phrase, and some modifying relation is established between phrases. And those modifying relations between phrases compose a sentence. In some cases a phrase consists of only a word such as an adjective, an adverb and some inflexional word, without being accompanied by any word that expresses a syntactic function, and itself carries a syntactic function. Some examples are shown here.

```
(i)WATASHI (pronoun)  NO (particle) HON (noun)
           (   I   )                    (book)
   ----------------------
          phrase
                   modifying relation
                       (my books)

   SHIROI (adjective)  HANA (noun  )
   ------ (  white  )       (flower)
   phrase
       modifying relation
         (white flowers)

   ISHI (noun )  NO (particle)  IE (noun )
        (stone)                    (house)
   ---------------
      phrase
           modifying relation
             (stone houses)

   KONO (7 in Table 3.1) HON (noun) GA(particle)..
   ---- (     this     )     (book)
  phrase
     modifying relation
   ----------------------------------
           phrase     (This book ...)

(ii) TOKYO (noun )  E (particle)  IKU (verb)
           (TOKYO)    (   to   )      ( go )
     ---------------
        phrase
             modifying relation
               (go to TOKYO)

     HON (noun)  UO (particle)  KAU (verb)
         (book)                     (buy )
     -------------
        phrase
             modifying relation
               (buy a book)
```

The syntactic relation is classified into three categories:

(a) Modification of a substantive word or phrase
Some examples are shown in above (i).

(b) Modification of an inflexional word or phrase
Some examples are shown in above (ii).

(c) Termination (the end of a sentence).

3.1.3. Syntactic state transition network

A syntactic state transition network is a network which represents the Japanese syntax[10]. The standard form is shown in Fig. 3.4, where each S represents a syntactic state, an arrow a transition path to the next state, C a part of speech, and I syntactic information. Therefore, if a state S_0 is followed by the part of speech C_0 then the state transits context-freely to S_1 outputting syntactic information I_0.

To an inflexional word a transition network is also applied and represents the inflexion. In speech recognition it is necessary to pursue the whole transition from the stem of an inflexional word to the end of inflexion, in other words, to predict the stem of an inflexional word with its inflexional ending and to output the syntactic information comprehensively for the whole words including their inflexions. In Fig. 3.5 is shown an example of transition network and accompanying syntactic information for two verbs "IKU(go)"

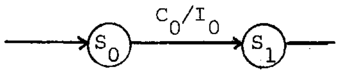


Fig. 3.4 Standard form of syntactic state transition network. S_0, S_1: states, C_0: part of speech or inflection, I_0: syntactic information.

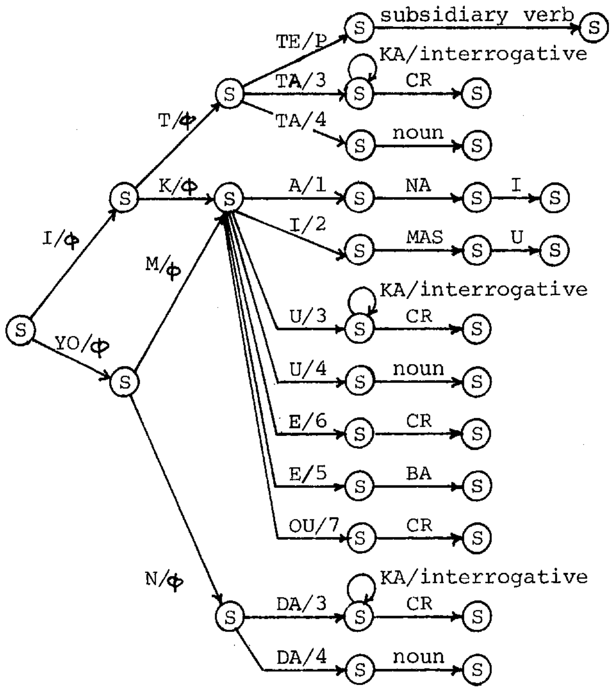


Fig. 3.5 Transition network for verbs: "IKU (go) and YOMU (read)" with their inflexion and syntactic information. X/Z means that X is output letters and Z is the syntactic information. φ: empty, CR: carriage return, P: particle, and the numbers are identified with the ones in Fig. 3.1.

and "YOMU (read)". This procedure corresponds to predicting all possible combinations of a verb with auxiliary verbs. For example, for a word "go", it may be better to predict probable combinations: go, goes, will go, will have gone, went and so on, though the number of probable combinations will be restricted.

The syntactic state transition network can not only predicts combinable words but also outputs syntactic information about modifying relation between phrases.

3.2. Knowledge about Vocabulary

3.2.1. Word dictionary

Each word is entered in a word dictionary in group according to part of speech as shown in Fig. 3.6. Each entry and its inflexion table are represented in Roman letters together with semantic information. If a part of speech is predicted using the syntactic state transition network, a word group of the predicted part of speech is picked out from the dictionary.

3.2.2. Automatic translating routine for Roman letter strings and inflexion tables

This routine translates a word written in Roman letters into a phoneme string using a table [11]. A translated phoneme string of a predicted word is used as a reference for matching an input phoneme string. This routine can also extract the characteristic phoneme string of a word. A characteristic phoneme string of a word contains only phonemes to be surely extracted from the speech wave. It is composed of vowels, /s/ and silence, and represents major acoustic information of a word. Some examples of the phoneme strings are shown in Table 3.3.

For matching procedure between an input phoneme string and a predicted word are used both phoneme and characteristic phoneme strings of the word. Here, these phoneme strings are not stored in the word dictionary. The system has only one word dictionary written in Roman letters and phoneme strings necessary for matching are produced each time from the word dictionary using the translating routine. This fact makes it very easy to enrich the entry of vocabulary.

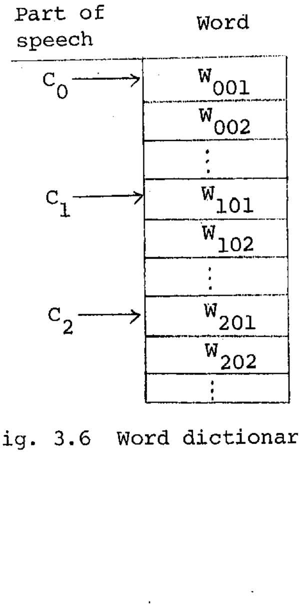


Fig. 3.6 Word dictionary.

Table 3.3 Examples of phoneme and characteristic phoneme strings of words. P: unvoiced stop, N: nasal, B: buzz, .: silence.

Word (Pronunciation)	Phoneme string	Characteristic phoneme string
OZIISAN	OBSIISAN	OISA
YAMA	IEAMA	AA
SENTAKU	SEN.PA.PU	SE.A.U
OOKII	OO.PSI	O.SI

3.3. Semantic Knowledge

Semantic information is used for the following purposes.

(i) Elimination of semantically inconsistent sentences which have been recognized using only acoustic and syntactic information.

(ii) Future development to semantic understanding of natural language by forming semantic networks.

(iii) Control of transition on the syntactic state transition network through the syntax analyser.

3.3.1. Semantic information

One of the semantic information dealt with is "knowledge about meaning". This knowledge involves (i) what each word means, (ii) verb-centered semantic structure, and (iii) schema of a story [10]. The other information is, so called, "remembrance of episode" which means the remembrance of a topic of conversation. In the present system, meaning of a word is represented by a list structure, and the others are represented by networks.

In the system the knowledge about meaning must be given from outside and can not yet be increased or updated by itself, but remembrance of episode can be increased or updated whenever new information comes in. While, if a schema has been already formed for a topic to be talked from now on, the knowledge of the topic will help recognition of the spoken topic. In the following sections how semantic information works in the recognition system will be explained.

3.3.1.1. Meaning of a word

Denote a word by n, its characteristic features by $f_i (i=1,\ldots,m;$ m is the number of features). Then, the meaning of a word may be expressed as follows:

$$n(f_1, f_2, \ldots, f_m),$$

where

$f_i = 1$ when the word has the characteristic feature f_i,

$f_i = 0$ when the word has not the feature f_i.

For example, if f_1= concrete, f_2= creature, f_3= animal, ..., then

hill (1, 0, 0, ...), dog (1, 1, 1, ...).

3.3.1.2. Definition of a verb

A verb plays very important semantic role in a simple sentence. A semantic representation of meaning of a verb is shown in Fig. 3.7, where n_0, n_1, ..., n_i are nodes, and Ar_1, Ar_2, .., Ar_i attatched to each arc are the natures of each arc. The nature of a node n_p is determined by a nature Ar_p attatched to the arc directing to the node n_p. Thus,

Structure = (V, Ar_1, Ar_2, ..., Ar_i),

$$\text{Restriction}\begin{cases} n_1 = \text{a word or node qualified by a nature } Ar_1, \\ \vdots \\ n_i = \text{a word or node qualified by a nature } Ar_i. \end{cases}$$

For example, a verb "IKU (go)" is defined by Fig. 3.8.

3.3.1.3. Schema

The form of a schema can not be determined uniquely. Dealing with a story, we may be able to represent the schema, for example, as shown in Table 3.4 and Table 3.5.

3.3.1.4. Remembrance of an episode --- Formation of a semantic network

Refering to the results of syntactic analysis and the relation between the nature of an arc and a case particle (partly involving another particle), the system forms a semantic network for a simple sentence centering a recognized verb. For instance, if a word sequence

OZIISAN WA YAMA E SHIBAKARI NI IKIMASHITA.
(An old man went to a hill for gathering firewoods.)

with syntactic information is given, a network shown in Fig. 3.9 will be formed. In Fig. 3.9 a process constructing a sentence is also shown.

3.3.2. Linking a semantic network for a sentence with a semantic network for an episode

After a network for a sentence has been formed, the network must be linked up with the already constructed network for the current episode. For this purpose a new node must be identified with the same node in the episode network.

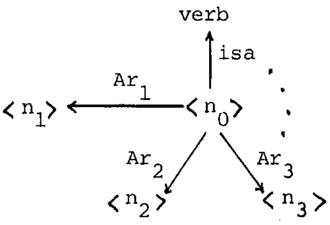


Fig. 3.7 Definition of a verb.
n: node, Ar: nature of an arc,
isa: is an instance of.

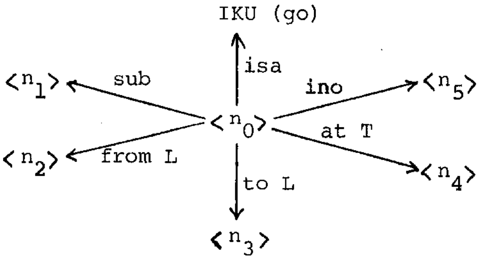


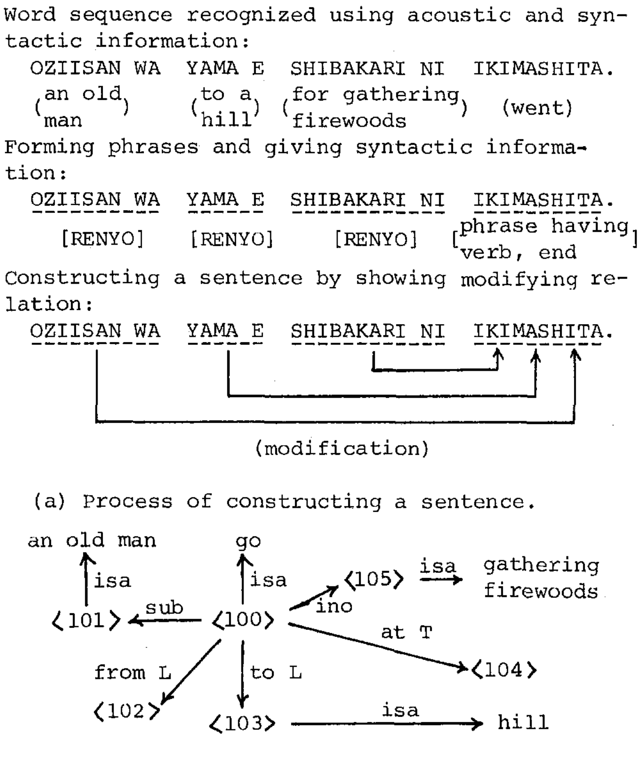
Fig. 3.8 Definition of a verb "IKU (go)".
sub: subject, L: location, T: time, isa: is an instance of, ino: in order to.

Table 3.4 A schema of a story.

Story	Title				
Scenes	Opening scene	Episode			
		event 1	event 2		event n
Characters	A, B, C, D	A, B, E, F	A, C		X, Y, Z
Other key words	m, n, o	k, l, m	m, n		a, b, c

Table 3.5 A schema for a tale "MOMOTARO (a brave boy born out of a peach)".

Story	MOMOTARO			
Scenes	Opening scene	Episode		
		event 1		event 5
Characters	an old man an old woman	an old man		Momotaro, dog, monkey, pheasant
Other key words	once upon a time, live	hill, firewoods, go		treasure, bring


Word sequence recognized using acoustic and syntactic information:
OZIISAN WA YAMA E SHIBAKARI NI IKIMASHITA.
(an old man) (to a hill) (for gathering firewoods) (went)
Forming phrases and giving syntactic information:
OZIISAN WA YAMA E SHIBAKARI NI IKIMASHITA.
[RENYO] [RENYO] [RENYO] [phrase having verb, end]
Constructing a sentence by showing modifying relation:
OZIISAN WA YAMA E SHIBAKARI NI IKIMASHITA.
(modification)
an old man
go
isa
isa
<105>
isa
gathering firewoods
<101>
sub
<100>
ino
at T
<104>
from L
to L
<102>
<103>
isa
hill

(a) Process of constructing a sentence.

(b) Semantic network.

Fig. 3.9 Process of constructing a sentence (a) and its semantic network (b) for "An old man went to a hill for gathering firewoods.". --- shows a phrase, └─↑ shows modification and RENYO in [ ] means this phrase modifies an inflexional word or phrase. ino: in order to.

In the present system all relations explicitly appearing in sentences and nodes expressing location are examined whether they have already appeared or not. Time relation is not handled unless it appears explicitly in sentences. Deeper structures of meaning such as causality or reasoning are not yet able to be dealt with. Fig. 3.10 illustrates a network for the episode, which has been constructed after the system has processed several sentences at the beginning of the tale of "MOMOTARO" shown below.

There lived an old man and an old woman.
The old man went to a hill for gathering firewoods.
The old woman went to a brook for washing.
She was washing on a brookside.

3.3.3. Word prediction by a conjunction "TO (and)"

When the syntax analyser has found a conjunction "TO (and)" which is used to enumerate some nouns, the system can predict a following noun group. For instance, for the input "MOMOTARO WA INU TO ... (MOMOTARO was accompanied by a dog and ... ", the system picks up as a following noun a noun group having similar natures to those a dog has.

3.3.4. Application of semantic knowledge to speech recognition

Using semantic knowledge the system advances recognition process as follows:

(i) Using acoustic and syntactic information, and sometimes semantic information, the system processes an input sentence and outputs several word sequences. The syntax analyser gives to each word sequence necessary syntactic information such as part of speech of each component word, phrase and modifying relation between

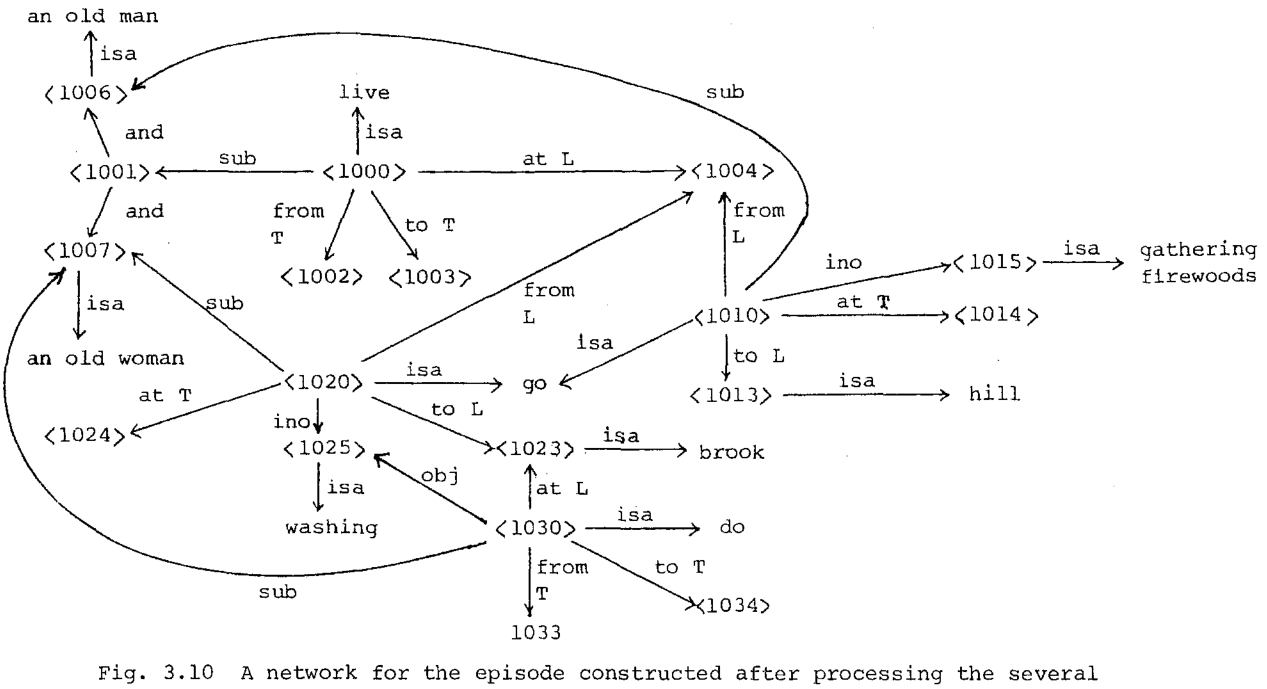


Fig. 3.10 A network for the episode constructed after processing the several sentences at the beginning of the tale of "MOMOTARO".

phrases.

(ii) The semantic processor, using this syntactic information, forms a semantic network for each word sequence.

(iii) A word sequence for which a semantic network failed to be formed satisfactorily is rejected because of semantic inconsistency. For instance, for an input sentence: "OZIISAN WA YAMA E SHIBAKARI NI IKIMASHITA.(An old man went to a hill for gathering firewoods.)", an output word sequence: "OZIISAN WA HANA (flower) E SHIBAKARI NI IKIMASHITA." is rejected, because the verb "IKU (go)" has an arc "to Location" but the output word sequence has no word meaning location and also the word "HANA (flower)" has no appropriate arc in the network.

(iv) Taking into account the result of syntax analysis and reliability of acoustic matching, the most reliable word sequence is output.

(v) Finally, the semantic network of the output sentence is linked with the semantic network of the episode formed by this process stage.

4. Results

We have been dealing with a Japanese fairy tale, "MOMOTARO" consisting of simple sentences and are now improving the system performance. The system's vocabulary is 99 words in total excepting inflexion of verbs, auxiliary verbs and adjectives. For simple sentences, the syntactic and semantic analysers work well. Furthermore the syntactic analyser alone can exactly recognize simple sentences with correct phoneme strings which would be provided from an ideal acoustic analyser. Though the level of semantic analysis is in its first stage, for simple sentences the semantic analyser can reject semantically inconsistent word sequences.

Therefore the acoustic analyser must be improved first of all. Its performance is as follows: The total number of output phonemes expected for an ideal acoustic analyser is 826 for the whole 16 test sentences from the tale, while the number of correct phonemes obtained from the analyser is 741 (89.7 %), and that of erroneous phonemes is 125 (15.1 %), in which the numbers of mis-identified phonemes, missing phonemes and superfluous phonemes are 25, 60 and 40 respectively.

The system can successfully recognize 25 blocks (a part of a sentence uttered in a breath) out of 33 blocks, and 9 sentences out of 16 sentences.

5. Conclusion

We have just started to construct a speech recognition system which can deal with semantic information and inflexion of words and have many problems to be solved. However, from this experiment it may be able to say as follows:

(i) The acoustic analyser gives pretty neat phoneme strings, if only a learning process using Bayes decision theory for a group of vowels, nasals and buzz is executed for each speaker.

ii) Use of global acoustic features is effective to reduce the number of predicted candidate words, though its effectiveness is not so much as in case of our isolatedly spoken word recognition system [12].

(iii) In Japanese, inflexion of inflexional words are complicated, and the number of Roman letters involved in the stem and inflexional ending of each verb or each auxiliary verb is usually very small. Especially the number of letters which very important particles have is much smaller. These aspects are very unfavorable for speech recognition in which ideal acoustic processing can not be expected. But the syntactic and matching processors can, to some extent, process input phoneme strings with erroneous phonemes satisfactorily.

(iv) Developing the vocabulary is very easy.

Of course we must improve the capability of the syntactic and semantic analysers and also develop the vocabulary.

References

1. Reddy, D.R.: "Speech recognition", Invited papers presented at the 1974 IEEE symposium, Academic Press (1975).
2. Sakai, T. and Nakagawa, S.: "A speech understanding system of simple Japanese sentences in a task domain", Trans. IECE Japan, Vol. E60, No. 1, p.13 (1977).
3. Koda, M., Nakatsu, R., Shikano, K. and Itoh, K.: "On line question answering system by conversational speech", J. Acoust. Soc. Japan, Vol. 34, No. 3, p.194 (1978).
4. Sekiguchi, Y. and Shigenaga, M.: "Speech recognition system for Japanese Sentences", J. Acoust. Soc. Japan, Vol. 34, No. 3, p.204 (1978).
5. Shigenaga, M. and Sekiguchi, Y.: "Speech recognition of connectedly spoken FORTRAN programs", Trans. IECE Japan, Vol. E62, No. 7, p.466 (1979).
6. Sekiguchi, Y. and Shigenaga, M.: "A method of phoneme identification among vowels and nasals using small training samples", Acous. Soc. Japan Tech. Rep., S78-17 (1978).
7. Sekiguchi, Y. and Shigenaga, M.: "A method of classification of symbol strings with some errors by using graph theory and its application to speech recognition", Information Processing, Vol. 19, No. 9, p.831 (1978).
8. Shigenaga, M. and Sekiguchi, Y.: "Recognition of stop consonants", 10th ICA, (1980).
9. Suzuki, K.:"NIPPON BUNPO HONSHITSURON"(Fundamental study on Japanese grammar), Meiji-shoin (1976).
10. Norman, D.A. and Rumelhart, D.E.: "Explorations in cognition", W.H. Freeman and Company, (1975).
11. Sekiguchi, Y. and Shigenaga, M.: "On the word dictionary in speech recognition system", Reports of Faculty of Eng., Yamanashi Univ., No. 28, p.122 (1977).
12. Sekiguchi, Y., Oowa, H., Aoki, K. and Shigenaga, M.: "Speech recognition system for FORTRAN programs", Information Processing, Vol.18 No. 5, p.445 (1977).