

国立国語研究所学術情報リポジトリ

コンピュータによる言語資源の研究

メタデータ	言語: English 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 林, 四郎, HAYASHI, Shiro メールアドレス: 所属:
URL	https://doi.org/10.15084/00001025

COMPUTER-BASED LINGUISTIC STUDY OF LITERARY TEXT

General Report on the Study of
Computational Linguistics
in
The National Language Research Institute
and the Author's Personal Study on C.L.
Shiro Hayashi

The present paper is the enlarged edition of the paper prepared for the 1st USA-Japan Computer Conference. The meeting of the Conference was held from October 3rd to 5th, 1972, at Nihon Toshi Center of Tokyo. The speech of S. Hayashi was performed in the Session of "Computers and the Humanities" (Oct. 3rd) under the title of "Computer-based Linguistic Research of Literary Text" in Japanese.

Contents

1. Computational Linguistics
2. Word Count by Computer
 - 2.1 Two Kinds of Word Unit
 - 2.2 Word Count and Character Count
 - 2.3 Four Principles of Data Classification
3. Problems of Computational Linguistics Brought out from the Word Count
 - 3.1 Discrimination of Kanji Homographs
 - 3.2 Input/Output of Kanji Data
 - 3.3 Automatic Word Segmentation
 - 3.4 Automatic Reading of Kanji
4. Linguistic Studies Based on the KWIC Concordance
 - 4.1 Study of Phrase-formation Rules
 - 4.2 Quantitative Analysis of Honorific Expressions in our Daily Life
 - 4.3 Examination of Activities of Demonstrative Adjectives in a Discourse
 - 4.4 Study of Hypersyntactic Elements in Japanese and English

Sentences

1. Computational Linguistics

It is a well-known fact that a division of technology called Computational Linguistics is established on the common field between computer science and linguistics. Strictly speaking, the computational linguistics contains only linguistic studies by computer. However, as a matter of fact, the computational linguistics must contain linguistic studies for computer, because we have many difficult problems on the processing of language data at the present stage, especially in the case of Japanese. While the linguistic study by computer is the study proceeded using computer, the linguistic study for computer is the preliminary study for searching the method of mechanical processing of language data. The present paper reports the outline of studies on computational linguistics in both meanings in the National Language Research Institute.

2. Word Count by Computer

Our Institute gained the computer HITAC 3010 and Kanzi-teletype-writers in 1966. Using these machines, we are engaged in the word count research of newspaper texts. Our materials are Asahi-sinbun, Mainiti-sinbun and Yomiuri-sinbun which are representative newspapers in present Japan. The corpus of our research is the whole language data in them in 1966. The number of words in the universe is 180 millions and the number of sample words is 3 millions, which is one sixtyth of the universe. The research program started in 1966 and will be finished in 1973. By the end of 1972, one of the last reports of this research is to be published. We can mention three remarkable points about this research as follows.

2.1 Two Kinds of Word Unit

In segmenting each sentence into words, we adopted two kinds of word unit, long unit and short unit. These two units are necessary for analysis of Japanese language data. Look at the following sentence.

田中総理大臣の中国訪問によつて、日中国交回復の道が開かれた。

TANAKASŌRIDAIZINnoTYŪGOKUHŌMONniYOtte,

NITTYŪKOKKŌKAIHUKUUnMITIgaHIRAKeta.

(A way to the restoration of diplomatic relations between Japan and China has been prepared by the prime minister TANAKA's visiting China.)

This sentence is segmented into words as constituents of the sentence as follows.

田中総理大臣/の/中国訪問/に/よっ/て/, /日中国交回復/の/道/が/開かれ/た。

TANAKASÔRIDAIZIN/no/TYÛGOKUHÔMON/ni/YOt/te/,
/NITTYÛKOKKÔKAIHUKU/no/MITI/ga/HIRAKe/ta/.

Each character string separated by slant lines is a long unit word. In the next step, they are segmented into short words, as follows.

田中総理大臣	TANAKASÔRIDAIZIN
田中/総理/大臣	TANAKA/SÔRI/DAIZIN
中国訪問	TYÛGOKUHÔMON
中国/訪問	TYÛGOKU/HÔMON
日中国交回復	NITTYÛKOKKÔKAIHUKU
日中/国交/回復	NITTYÛ/KOKKÔ/KAIHUKU

Short words are nearly equivalent to morphemes.

MITI, HIRAKe, YOt, *no*, *ni*, *te*, *ga* and *wa* are short words, and at the same time, long words. Three millions is the number of short words contained in the corpus. When the number of words is counted by long unit, that will be about two millions.

2.2 Word Count and Character Count

Our word count research is joined to the character count research. Therefore, when both researches are synthesized, following two points will become clear.

1) How is each Kanzi (Chinese character) used in short words, and how are those short words used in long words?

(character)	(short word)	(long word)
	中国	中国訪問
中	日中	日中国交回復
	田中	田中総理大臣

2) How has each word the variety of writing forms?

(word) (variety of writing forms)

MITI 道, 路, みち, ミチ

2.3 Four Principles of Data Classification

Before data processing, all texts in the corpus were examined from four points of view.

- 1) What kind of topics is the content of each article concerned to?
- 2) What kind of literary genre does each article belong to?
- 3) What position does each sentence occupy in the article?
- 4) In what form is the writer of each article shown?

Each word comes to be accompanied by informations concerning features of the article or the sentence of which the word is a constituent. Those features are found through observations due to the following:

classification lists.

1) The classification of topics

1. politics 2. diplomacy 3. economics 4. labour 5. human interests
6. foreign affairs 7. culture 8. local news 9. sports 10. domestic
life 11. amusement 12. advertisement

2) The classification of literary genres

1. news 2. news review 3. editorial or column 4. readings related
to current topics 5. special readings 6. criticism 7. practical read-
ings 8. documentary report 9. periodical summary 10. business
announcement 11. interview 12. reader's composition 13. corres-
pondence between readers and editor

3) The classification of positions

1. headline 2. title 3. lead 4. body 5. signature 6. table 7. sen-
tences accompanying to illustration

4) The classification of forms showing the name of writers

1. unsigned in general 2. news of agency 3. top-signed 4. end-
signed by outside writer 5. end-signed by inside writer 6. end-
signed with abridged signature 7. top-signed foreign telegram 8.
unsigned by editor 9. unsigned by outside writer

The corpus was stratified by these classification principles. Through
the use of these stratifications, two kinds of meaningful word group
are obtained, one is the group of prevailing words in whole corpus
and the other is the group of words being dominant in any particular
stratum of the corpus. Fig. 1 shows the process of obtaining these
word groups from the stratified text.

The chief planners of the word count research are A. Tanaka and
T. Ishiwata. The main programs of word count were made by H. Saito
and S. Kimura, many other programs were made by H. Nakano, K.
Egawa, S. Muraki and A. Turuoka. The planners and programmers
of the character count research are S. Tsuchiya and M. Nomura.
S. Hayashi is the promoter of both researches.

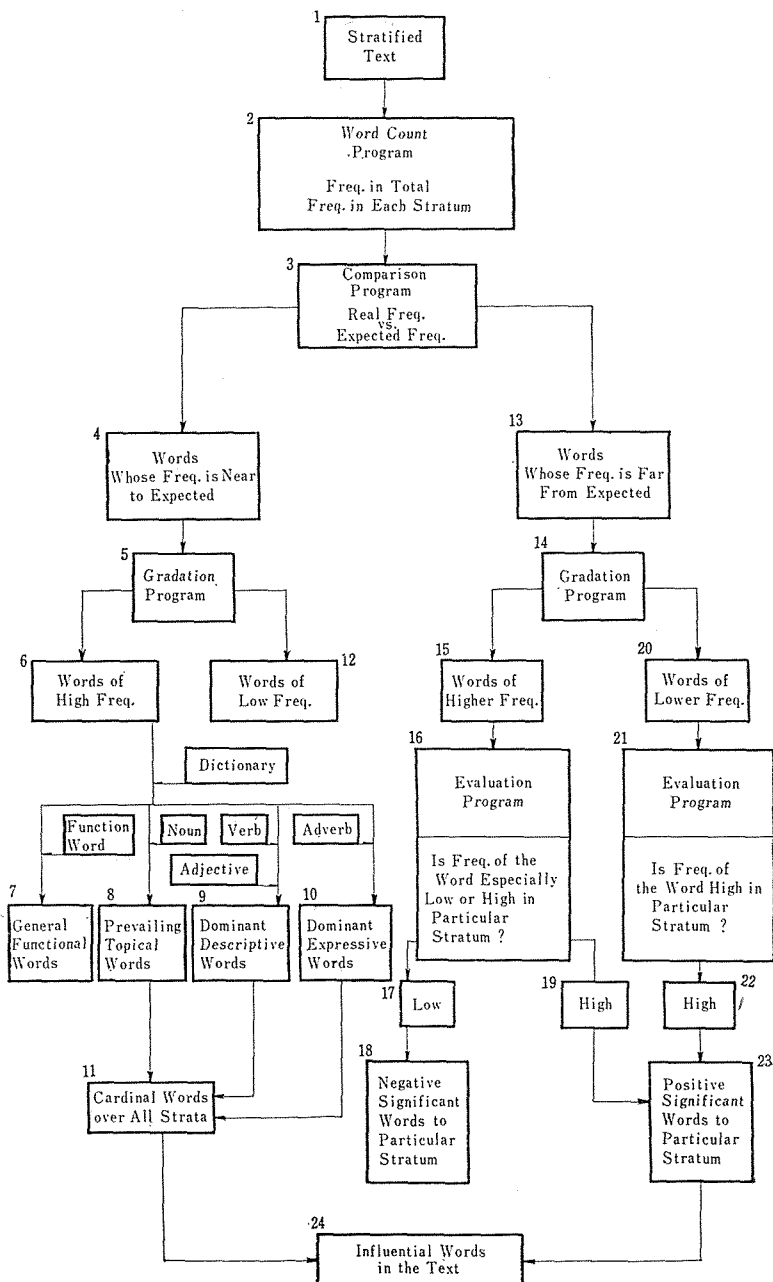
3. Problems of Computational Linguistics Brought out from the Word Count

3.1 Discrimination of Kanzi Homographs

The vocabulary of Japanese contains many words of Kanzi homo-
graph as follows.

KÔHU	}	工夫	NINKI	}	人氣	TAIKA	}	大家
KUHÔ			ZINKI			TAIKE		
			HITOE			ÔYA		
		ÔIE						

Fig. 1 Program to find influential words in a text



In actual sentences, it will be easy for normal adult readers to find a relevant reading of Kanzi in the following phrases.

人気のある人 NINKI no ARu HITO
人気のいい土地 JINKI no Ii TOTI
人気のない場所 HITOKE no NAI BASYO

But it is very difficult for a computer to do such things.

In order to prepare a program for discrimination of Kanzi homographs, many semantic researches must be performed hereafter.

3.2 Input/output of Kanzi Data

We are using Kanzi-teletypewriters for the input of Kanzi data. Better ways than this cannot be found at the present stage. For the output also, we are using Kanzi-teleprinter which can print only 120 characters a minute. Its slowness is a great obstacle for the effective procedure of the job. Recently, some high speed online Kanzi-printers have come to be produced, which have the capacity of printing more than 100 thousands characters per minute. When those high speed Kanzi-printers become popular, the jobs of Kanzi output will be in good condition.

3.3 Automatic Word Segmentation

In usual Japanese passages of Kanzi-Kana mode, no sentence has any space between words. Accordingly, in order to make the computer identify words in those sentences, it is necessary to put any segmental symbol between words before computer processing. In our word count research, these pre-editing of word segmentation was executed by men. Besides the normal job, we have tried to find the way of automatic word segmentation. Through several experiments, we have found two ways of processing. In one way, words are recognized according to the pattern of Kanzi-Kana arrangement; as a rule, a string of Kanzi characters is recognized to be a word. In the other way, it is repeated throughout each sentence to recognize a string of characters to be a word by matching with a dictionary, and, hereon, syntactic relations between words are examined with provided rules. The latter way is extremely sophisticated but faultless. The experiment by the former way was programmed by K. Egawa and the latter by T. Ishiwata.

3.4 Automatic Reading of Kanzi

In order to arrange all counted words into any alphabetical order every Kanzi must be transformed into Kana or Roman character. Automatic transformation, Kanji to Kana, is also our important subject. In this subject, too, two ways of approach were tried. In one

approach, the reading of each Kanzi is determined due to the primary principle that singular Kanzi surrounded by Kanas is, as a rule, used with the mode of KUN (pronunciation as pure Japanese word), and double Kanzis are both used with the mode of ON (Chinese style pronunciation). In the other approach, the same determination is made word by word according to the instruction in the dictionary. The experiment by the former way brought out adequate determinations in 87 cases among 100, and the experiment by the latter 98 cases among 100. Both methods have their own advantages and disadvantages. By the former way, we shall have no case of impossibility though in danger of misreading. By the latter, we shall have few cases of misreading, but unless a great size of dictionary is provided, we shall have some cases of impossibility. In both ways, automatic discrimination of Kanzi homographs is difficult at the present stage. The experiment by the former way was programmed by A. Tanaka, and the latter by T. Ishiwata.

4. Linguistic Studies Based on KWIC Concordance

Computer is now very useful tool for the study of our language. For instance, using computer, we can obtain a literary text's concordance of KWIC form by comparatively simple program. KWIC concordance of a text supplies us very useful materials for the linguistic study of the text, and moreover, such study by actual mass data leads us to the study for automatic processing of language data. Japanese Kanzi-Kana texts have above-mentioned handicaps compared to the texts of European languages written by Roman characters, nevertheless, such production of KWIC concordance by computer have the same merits with the case of European texts. Hereafter we can expect the steady progress of linguistic science through the study of actual data owing to the plentiful materials supplied by KWIC concordances. The remaining part of this paper will be assigned to some reports of the study of computational linguistic based on KWIC concordances of various texts.

4.1 Study of Phrase-formation Rules

T. Ishiwata is engaged in the descriptive study of Japanese phrase patterns. The first subject of his study is to describe the combination pattern of verb with noun intermediated by a class of Zyosi (postpositional particle). In the beginning, Zyosi NI was observed. Verb phrases which have the form of "Noun+NI+Verb" were analyzed according to semantic categories of verbs and nouns, and their patterns of combination were described. In this study, the semantic category of each word was identified according to the system of *The*

Word List by Semantic Principles. (editor, The National Language Research Institute, published in 1964)

The outline of his description is as follows.

grp. 1 Verb phrases combining two nouns in a particular relation
[Nouns in general]+NI+[Verbs expressing the abstract
relation of things]

KOKUSAI-KEIZAI ni KANsuru HÔKOKU

(the report concerning the international economics)

grp. 2-1 Verb phrases expressing special locomotion

[Nouns connoting certain
locations]+NI+[Verbs expressing
thing's locomotion]

BYÔIN ni YUKu (go to the hospital)

grp. 2-2 Verb phrases expressing some purposeful behaviors

[Nouns denoting
human conducts]+NI+[Verbs expressing
thig's locomotion]

SYUZYUTU ni YUKu (go to the hospital in order to undergo an operation)

grp. 3-1 Verb phrases expressing actual existence or physical actions

[Nouns connoting
locations or directions]+NI+[Verbs expressing thing's
existence or actions]

KOEN ni ARU SINRIN (the wood in the park)

KAO o YOKO ni MUKeru (turn one's face aside)

grp. 3-2 Verb phrases expressing actions towards certain objects

[Nouns denoting
physical objects]+NI+[Verbs expressing thing's
actions]

TE ni MOTu (hold in one's hand)

DENSYA ni NORu (take a train)

grp. 4 Verb phrases expressing progressions, development or transference of matters

[Nouns implying the
concept of time]+NI+[Verbs expresisng some intangible
state or bodyless motions]

SYÔRAI ni SONAeru (prepare for future)

ENSOKU ga GOGATU TUITATI ni NOBi-ta. (Our excursion has been prolonged to the May 1st)

grp. 5 Verb phrases expressing one's activity towards another being

[Nouns denoting human beings
or human organizations]+NI+[Verbs expressing
human activities]

OYA ni AMAeru (take advantage of one's parents' kindness)

NAKAMA ni SIRASeru (give information to one's colleagues)

grp. 6 Verb phrases that express human activities and are almost equivalent to solitary verbs

[Nouns denoting
human activities]+NI+[Verbs in general]

SÔDAN ni Ôzuru (advise)

SIPPAI ni OWARu (fail)

After these major classification, minor classifications were carried out, and many patterns of [a class of Nouns]+NI+[a class of Verbs] were described. The study of this kind will be continued henceforth about the verb phrases assuming the form of [Noun+O+Verb] or the form of [Noun+GA+Verb] and the noun phrases of the form of [Noun+NO+Noun]. In accordance with the progression of these studies, grammatical rules for the automatic analysis of sentence structure will be provided in the better condition.

4.2 The Quantitative Analysis of Honorific Expressions in Our Daily Life

In our daily use of language, we can observe considerably sophisticated system of honorific expressions. This system is not yet described exactly. Provided well collected mass data are analyzed by well designed computer programs, the exact description will be possible about the honorific expression system in any language. A research was tried with this aim. F. Minami analyzed the colloquial data of Matue dialect that had been recorded in a family dwelling in Matue city, the capital of Simane-prefecture. This study was consistently pursued from the point of view of discourse analysis. All discourses in the materials were classified into several categories according to their functions in verbal communication (greetings, conversation for business, chatting etc.), tone of speech (neutral, joking, ironical, complaining etc.) and topics (matters in daily life, business gossip etc.). And every actual use of the honorific expressions was examined in each category of the discourse. For example, a definite difference among the morphemes used in exalted expressions was found in examining the discourses in which they appeared: -RARE- appeared, in most cases, in the discourses of gossip and was used in the third person 194 times out of a total of 201. On the other hand, -NASAR- mainly appeared in the discourses of topics related to daily life or business matters and was used for the addressee 54 times out of a total of 60. The program for analysis was provided by A. Matumoto.

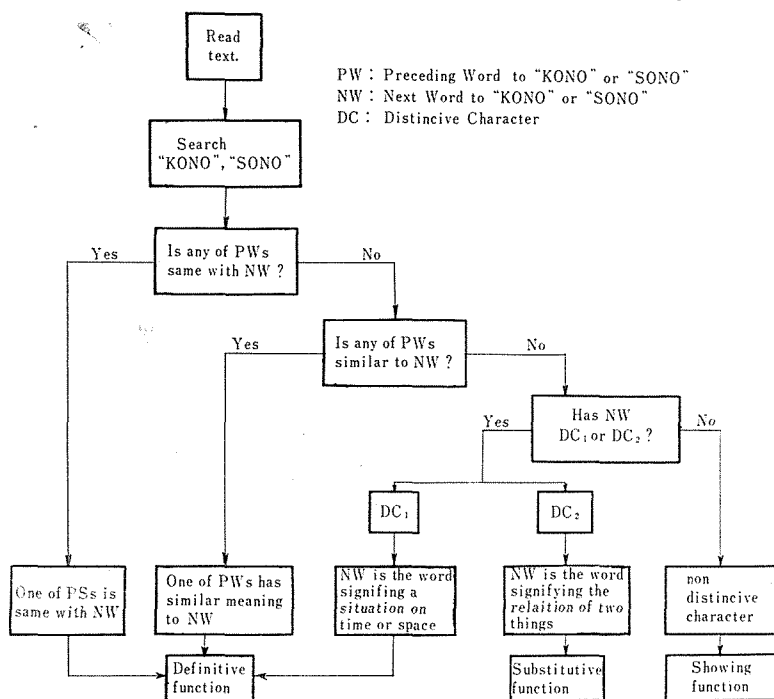
4.3 Examination of Activities of Demonstrative Adjectives in a Discourse

In many languages several classes of words are recognized to have the grammatical role as connective links between sentences. They are found in word groups of so-called conjunction, pronoun, adjective, adverb and some others. In Japanese, a group of words which have

the form of KO- SO- or A- in their heads are called demonstrative. Of these demonstratives, KORE, SORE, ARE, KOKO, SOKO and ASOKO have the nature of noun (demonstrative pronoun), KONO, SONO, ANO, KONNA, SONNA and ANNA, of adjective (demonstrative adjective), KÔ, SÔ and Â, of adverb (demonstrative adverb). It is necessary for the analysis of any discourse to make clear what the demonstratives in the text indicate. Therefore, it comes to have great value to describe explicitly the general types of relations between demonstratives and objects indicated by them.

S. Hayashi, the author of this paper, had a small research on the

Fig. 2 Program to discriminate function of demonstrative adjective



activity of demonstrative adjectives as the first step of the study of demonstratives in general. He examined every context of KONO and SONO in a short novel, Mori Ogai's *Takasebune*, using the KWIC concordance of the novel.

Through the examination of whole 45 cases (KONO 20, SONO 25), he found some systematic relations between the words defined by KONO or SONO and the words or things indicated by KONO or SONO. When these relation are generalized, a hypothetical rule is

gained for the identification of these words' contextual functions as follows.

1. When a word is found to be of the same form with the word in the later context succeeding to KONO or SONO, the content of that word is pointed out by KONO or SONO. (The meaning of the word is transferred to the denotation from the connotation by KONO or SONO.)

2. When no word is found to be of the same form with the word in the later context succeeding to KONO or SONO;

2-1 In the case that a word is found to be nearly related to the word in question, the situation is regarded to be the same with the case 1.

2-2 In the case that the word succeeding to KONO or SONO belongs to noun expressing any kind of relation, KO or SO of KONO or SONO acts as the substitute for a certain antecedent.

2-3 In the case that the word succeeding to KONO or SONO belongs to a group of empty noun expressing loosely a condition of time or place, whole matter described in the preceding sentence is put into the condition.

2-4 In the other cases from above three cases, KONO or SONO points out something existing in the speech situation.

The flow chart of Fig. 2 shows the process to discriminate five types of activities of KONO or SONO in a Japanese text.

4.4 Study of Hypersyntactic Elements in Japanese and English Sentences

S. Hayashi performed the research on positional types of Japanese sentences through the examination of a thousand sentences in the elementary school text books, and tried a precise description of those types. Sentences are classified into three major types; starting type, following type and turning type. The sentence of starting type sets a starting point of a discourse and arises the stream of thought in minds of the speaker and hearers. The sentence of following type expresses something related to the content of the preceding sentence, and it makes the mental stream go on smoothly. The sentence of turning type gives a turning point to the discourse and changes the direction of the stream.

Each sentence in a discourse is shaped into any of these types by three kinds of type-making factor, symbol, agent and position. When a word in a sentence is clearly recognized to be acting as a type-making factor according to the grammatical meaning of the word, the word may be called type-making symbol. When any type-making factor is observed not so clearly in the sentence which has no type-making symbol, the factor observed is called type-making agent. When

a sentence which has neither symbol nor agent shows any positional type owing to the position of the sentence, that position itself is recognized to be the type-making factor.

Kinds of positional sentence type and kinds of type-making factor are as follows.

(positional sentence type)	(type-making factor)
starting type with symbol	starting symbol
starting type with agent	starting agent
starting type by position	starting position
following type with symbol	following symbol
following type with agent	following agent
following type by position	following position
turning type with symbol	turning symbol
turning type with agent	turning agent
turning type by position	turning position

In the composition of a typical form, the starting type will be found only in the first sentence of the whole composition, and the turning type will be found in the first sentences of some paragraphs, other sentences will be of the following type. Accordingly, in the ordinary discourses the number of following type sentences is far greater than the number of sentences of other two types. Every sentence in a discourse is related by these hypersyntactic elements from the beginning to the end.

Based on the assumption that these sentence types and their activities are of the nature of universality, Hayashi examined hypersyntactic relations in English sentences and contrasted them with those of Japanese, limiting the subject to following symbols and following agents. Examples of English sentences are taken from James Joyce's short novel *Eveline* (143 sentences in total) in his *Dubliners*. For the contrastive study, KWIC concordances of English text (programmed by H. Saitô) and of Japanese text (programmed by S. Tsuchiya) were used.

Following Symbol A conjunctions expressing development from a preceding event to a following event

Following Symbol A1 conjunctions for expected development

Jap. **SOSITE, SOREKARA**

Eng. then, and then

Japanese conjunction **SOSITE** is considered by many Japanese people to have almost the same meaning as English conjunction *and*. But their natures are different from each other. **SOSITE** is in many cases used at the head of sentence with the meaning of *afterwards*,

but, English *and* is not so much used at the head of sentences. The meaning of *and* is near to that of *as well as* or *added to*, not *afterwards*. Japanese conjunction SOREKARA expresses the simple development in sequential matters. English *then* and *and then* are nearly equivalent to Japanese SOREKARA.

Following Symbol A2 conjunction for unexpected development
Jap. SURUTO

SURUTO, as well as SOSITE, leads the sentence describing an event that occurred just after the event described in the last sentence. But, while SOSITE is used for the expected event, SURUTO introduces the succeeding event as an unexpected event. SURUTO is used especially much in the sentence of the narrative story. It seems difficult to find any English conjunction which is equivalent to Japanese SURUTO.

Following Symbol B conjunctions expressing logical or psychological development.

Following Symbol B1 illative particles
Jap. SOREDE, SOKODE, DAKARA
Eng. therefore, then, so, and

The meaning of *and* concerns to the logical or psychological process of thinking, not to the chronological order.

Following Symbol B2 adversative conjunctions
Jap. DAGA, SIKASI, KEREDOMO, SOREDEMO
Eng. but, yet, and yet, however

The considerable equivalency is found between both languages in conjunctions of this group.

Following Symbol B3 miscellaneous conjunctions
Jap. SORENI, TOKOROGA, SOREDEWA, DEWA,
SORENISITEMO
Eng. beside, and beside

Following Symbol C words for substitution

Following Symbol C1 demonstratives indicating something close at hand
Jap. KORE, KONO
Eng. this

KORE or KONO is used as the following symbol in Japanese sentences more often than *this* in English sentences.

Following Symbol C2 demonstratives indicating something existing far from the speaker and near to the hearer, or personal pronoun of third person

Jap. SORE, SONO

Eng. the, that, it, he, she, they

Japanese SORE belongs to the pronoun and SONO to the adjective, but English *that* belongs to both. In usual sentences, SONO is equivalent to *the* rather than to *that*. *It, he, she* and *they* all belong to the pronoun. They are usually considered to be equivalent to KARE, KANOZYU, and KARERA, but these Japanese words are not so often used for the following symbol, while *he, she,* and *they* are often used for the following symbol.

Following Symbol D responsives

The words called responsives show that the sentence including them was uttered for the purpose of a response to the preceding question.

Following Symbol D1 assenting responsives

Jap. HAI, EE, SÔDESU

Eng. yes, yea, aye

Following Symbol D2 negative responsives

Jap. IIE

Eng. no

Following Agent A lexical-semantic following factor

When the meaning of a word in a sentence is related to the meaning of a word in the preceding sentence, and the relation of them causes the meaning of each sentence to accord with the other, the very relation will be regarded as a kind of following agent. The remarkable equivalency is observed between both languages in every phase of the Following Agent A.

Following Agent A1 words expressing relations to other beings

Jap. SOBA, TUGI, KADO, TONARI, HOKA, MAE, MUKÔ, USIRO

Eng. still, rather, another

EX. 1 *Still* they seemed to have been *rather* happy then.

EX. 2 She was about to explore *another* life with Frank.

Following Agent A2 words implying occurrence of preceding events

Jap. MATA, MAMONAKU, ITUNOMANIKU, SUGU, SIBARAKU, TÔTÔ

Eng. again, latterly, in the end

EX. 1 Perhaps she would never see *again* those familiar objects...

EX. 2 ...but *latterly* he had begun to threaten her...

EX. 3 *In the end* he would give her the money...

Following Agent A3 relation of synonyms

EX. 1 ...the invariable *squabble* for money on Saturday nights had begun to weary her unspeakably. She gave her entire wages..., but the *trouble* was to get any money from her father.

EX. 2 In the end he would give her the money and ask her had she any intention of *buying* Sunday's dinner. Then she has to rush out as quickly as she could and so her *marketing*,...

Following Agent A5 co-occurrence relation

When two words are apt to co-occur in the same context of ordinary discourse, the relation of both words is called co-occurrence relation. This relation has very strong power to connect two sentences.

EX. 1 Her head was *leaned* against the window curtains,... She was *tired*.

EX. 2 The man out of the last house *passed* on his way home; she heard his *footsteps* clacking along...

EX. 3 She sometimes felt herself in danger of her father's *violence*. She knew it was that that had given her the *palpitations*.

EX. 4 Down far in the avenue she could hear a street *organ* playing. She knew the *air*.

EX. 5 All the *seas* of the world tumbled about her heart. ...he would *drown* her.

Following Agent A6 relation of antonyms

EX. 1 Her *father* was not so bad then; and besides, her *mother* was alive.

EX. 2 ...he had begun to *threaten* her... And now she had nobody to *protect* her.

EX. 3 Why should she be *unhappy*; she had a right to *happiness*.

Following Agent A7 relation between general recognition and individual recognition

1. from individual to general

EX. 1 How well *she* remembered the first time she had seen *him*;... Then *they* had come to know each other.

EX. 2 *She* and *her brothers* and *sisters* were all grown up; *her mother* was dead. *Tizzie Dune* was dead too, and the *Waters* had gone back to England. *Everything* changes.

2. from general to individual

EX. 1 Few *people* passed. The *man* out of the last house passed on

his way home.

- EX. 2 The white of *two letters* in her lap grew indistinct. *One* was to Harry; *the other* was to her father.

Following Agent A8 sequential relation

The sequential chain is apt to be formed by numerical order or time order.

- EX. 1 *When* they were growing up he had never gone for her,...; but *latterly* he had begun to threaten her... And *now* she had nobody to protect her.

Following Agent A9 cause-effect relation

- EX. 1 What would they say of her in the Store when they found out that *she had run away with a fellow?*... Then she would be *married*—she, Eveline.

Following Agent B syntactic following factor

When the syntactic structure of a sentence serves to connect the sentence to the preceding sentence, the very structure is regarded as the following agent. Such an agent is Following Agent B. On the contrary to the equivalency of lexical-semantic following factors in both languages, the syntactic following factors in them are different according to the difference of syntax in both languages.

Following Agent B1 Japanese post-position MO

Every Japanese post-position is not a word which has any substantial meaning but an linguistic form which shows the syntactic role of the preceding word. For instance, post-position GA shows that the noun accompanied by it is the subject for the succeeding predicate.

Post-position MO shows that the word accompanied by it was once used in the preceding sentence. MO is usually translated into *too* or *also*.

WATASI mo SORE ga SUKI da. (I like it, too.)

KARE mo KIta. (He came also.)

Too and *also* belong to the following Agent A2, because they are adverbs possessed of a substantial meaning.

- EX. 1 ...her mother was dead. Tizzie Dune was dead, *too*,...

Following Agent B2 explanatory predicate in Japanese sentence

The Japanese post-position NO has an usage of noun-forming agent which is similar to the usage of the noun KOTO. This kind of NO is called noun-forming post-position (Zyuntai-Zyosi), and KOTO empty noun (Keisiki-meisi).

HANASu no o YAMe-ta. (HANASu=verb. HANASu no=noun phrase)

HANASu KOTO o YAMe-ta. (HANASu KOTO=noun phrase)

When the noun-forming NO is used in the predicate of a sentence combined with the judging form DA or DESU and gets the form of NODA or NODESU, the predicate explains the meaning of the description in the preceding sentence.

In the next sentence succession, the second sentence shows the reason why the first sentence was declared.

KYŪ ni AKARUKu NARi-masi-ta. RESSYA wa TONNERU o Deta-nodesu.

In English, it is difficult to find such a special form of predicate to explain the meaning of the preceding description. The above example of Japanese will be translated into English as follows.

1. It has gotten light suddenly. For our train got out of the tunnel.
 2. It has gotten light suddenly. Our train is now out of the tunnel.
- In the example 1., the conjunction *for* (Following Symbol B3) shows the relation of the second sentence to the first sentence explicitly. But, there is no conjunction in the second sentence of the example 2., and the predicate has no special form.

EX. 1 Every thing changes. Now she was going to away like the others, to leave her home. (SUBETE ga UTURi-KAWARu. IMA, KANOZYU mo UTI o SUTe-te, TA no HITOBITO to ONAzi YŌni De-te YUKô to Si-te Iru-noda.)

Following Agent C omission of a sentence element

In Japanese sentence, the word which is an element of a sentence is often omitted owing to the existence of the same word in the preceding sentence.

AKIRA SAN no IE ni KOINU ga Ki-masi-ta.

KOINU wa MADA UMARE-ta BAKARI no KAWAi INU desi-ta.

“KOINU wa”, the subject of the second sentence, can be omitted owing to the existence of “KOINU ga” in the first sentence.

It seems that no such omission of the sentence element is found in usual English sentences.

Following Agent D repeated use of a word used in the preceding sentence

A certain word in a sentence has the tendency to be used again in the succeeding sentence. This tendency is commonly found in both languages.

OTÔ-SAN ga KAETe Ki-masi-ta. OTÔ-SAN wa TORI-KAGO o SAGe-te I-masi-ta. (*My father has come home. My father has a*

bird-cage in his hand.)

In usual English sentences, *my father* at the head of the second sentence would be expressed by *he*, and the *he* will be used repeatedly in succeeding sentences. While *he*, *she* or *they* is repeatedly used in ordinary English sentences, KARE, KANOZYU, or KARERA will be used rather rarely in ordinary Japanese sentences. On the contrary to the repeated use of these third person pronouns, they are apt to be omitted in ordinary Japanese sentences (see Following Agent C), otherwise, any noun used once in a discourse would be used repeatedly as it was, for instance OTÔ-SAN (my father) in the above example.

EX. 1 She was about to explore another life with *Frank*. *Frank* was very kind, manly, open-hearted. She was to go away with him by the night-boat to be his wife and to live with *him* in Buenos Ayres where *he* had a home waiting for her. How well she remembered the first time she had seen *him*; *he* was lodging in a house on the main road where she used to visit.

Summary

1. The remarkable equivalency is found in the lexical-semantic factors for sentence succession of both languages.
2. Some points are observed to be different in two languages in their grammatical or syntactic factors.
 - i Conjunctions, demonstratives or words for substitution in each language have their own functions.
 - ii The explanatory predicate, which is the special form in the predicate for successiveness, is found abundant in Japanese, but no such form can be found in English.
 - iii In Japanese discourses, a word is often absent in a sentence, because of the presence of the same word in the immediately preceding sentence. In English, no such absence occurs.
 - iv In both languages, some words in sentences have a tendency to be used again in the succeeding sentence. In these cases, the third person pronoun is much more used in English than in Japanese.

REFERENCES

Papers in the following list are all written by Japanese.

Abbreviations

- N.L.R.I. the National Language Research Institute
S.C.L. I Studies in Computational Linguistics I,
N.L.R.I., Syuei Syuppan Co., Tokyo, 1968
S.C.L. II Studies in Computational Linguistics II,

- N.L.R.I., Syuei Syuppan Co., Tokyo, 1969
- S.C.L. III Studies in Computational Linguistics III,
N.L.R.I., Syuei Syuppan Co., Tokyo, 1971
- S.C.L. IV Studies in Computational Linguistics IV,
N.L.R.I., Syuei Syuppan Co., Tokyo, 1972

Word Count Research

(System Design or Programming)

1. HAYASHI, Shiro: Classification of Language Data in Newspapers, S.C.L. II, 1-15
2. ISHIWATA, Toshio: A System Analysis of the Word Count Programs, S.C.L. I, 47-55
3. TANAKA, Akio & H. SAITO: The Routine of Random-sampling by Computer, S.C.L. I, 105-114
4. SAITO, Hidenori: A System of Word Count Program, S.C.L. II, 16-28
5. SAITO, Hidenori: A System of Word Count Program II, S.C.L. III, 77-97
6. TANAKA, Akio: On the Tabulation of "Japanese Word List" by Computer, S.C.L. I, 115-132

(Input/Output)

7. MATUMOTO, Akira: Language Data Processing by the Kanzi-teletypewriter, S.C.L. I, 57-90
8. KIMURA, Sigeru: An Automatic Checking System for Input Data by Kanzi-teletypewriter, S.C.L. I, 133-150

(Word List)

9. N.L.R.I.: Studies on the Vocabulary of Modern Newspaper vol. 1, N.L.R.I., Syuei Syuppan Co., Tokyo, 1970, 342pp
10. N.L.R.I.: Studies on the Vocabulary of Modern Newspapers vol. 2, N.L.R.I., Syuei Syuppan Co., Tokyo, 1971, 314pp
11. N.L.R.I.: Studies on the Vocabulary of Modern Newspapers vol. 3, N.L.R.I., Syuei Syuppan Co., Tokyo, 1972, 159pp

(Analytic Study of Newspaper Vocabulary)

12. HAYASHI, Shiro: Outline of Newspaper Vocabulary Research, S.C.L. I, 1-46
13. HAYASHI, Shiro: Word Counting and Basic Vocabulary, S.C.L. II, 1-35
14. KIMURA, Sigeru: Discrimination of Characteristic Words in Various Fields, S.C.L. II, 29-37
15. TANAKA, Akio: Homophonic Words and Homographic Words in Modern Newspapers, S.C.L. III, 121-145
16. NAKANO, Hiroshi: A Table of the Classified Vocabulary on Research into Newspaper Vocabulary, S.C.L. II, 38-54

17. NAKANO, Hiroshi: Automatic Classification of Parts of Speech, S.C.L. III, 98-120
18. MURAKI, Shinjiro: A Study on the Structure of Ambiguous Expressions, S.C.L. IV, 44-56

Character Count Research

19. N.L.R.I.: Studies on the Frequency of Chinese Characters in Modern Newspapers (Interim Report), N.L.R.I., Syuei Syuppan Co., Tokyo, 1971, 90pp
20. NOMURA, Masaaki: A System Analysis of Kanji Data Count Program, S.C.L. III, 146-164
21. NOMURA, Masaaki: An Experimental Study on Usage of Kanji (Chinese Characters) in Newspaper, S.C.L. II, 89-106
22. NOMURA, Masaaki: Orthographic Patterns in Japanese Kanji-Kana Strings, S.C.L. IV, 1-34

Automatic Word Segmentation

23. EGAWA, Kiyosi: An Inquiry into the "Automatic Segmentation" of Japanese Text, Mathematical Linguistics No. 43/44, 1968, 46-52
24. EGAWA, Kiyosi: A System of Automatic Segmentation for Japanese Text, Mathematical Linguistics No. 51, 1969, 17-22
25. ISHIWATA, Toshio, H. SAITO & S. KIMURA: A Japanese-Sentence-Segmentation Program, Mathematical Linguistics No. 50, 1969, 24-38

Automatic Reading of Kanji

26. TANAKA, Akio: A Program System of Transliteration, from Kanji to Kana, and from Kanji to Romaji, S.C.L. II, 107-138
27. ISHIWATA, Toshio: An Application of KWIC System to the Processing of Newspaper Vocabulary Data, S.C.L. III, 36-76

Programming for KWIC Concordance

28. TSUCHIYA, Shin-ichi: A Word Index System with Kana-Input, S.C.L. IV, 35-43

Linguistic Study Based on KWIC Concordance

29. ISHIWATA, Toshio: Structure of Verb Phrase Containing the Particle "ni", S.C.L. IV, 57-109
30. N.L.R.I.: Some Aspects of Homorific Expressions, N.L.R.I., Syuei Syuppan Co., Tokyo, 1971, 231pp
31. HAYASHI, Shiro: Contextual Functions of Demonstrative "KONO" and "SONO", S.C.L. IV, 110-131
32. HAYASHI, Shiro: Words or Phrases in Sentence as the Connective Links of Ideas in Discourses, Mathematical Linguistics No. 39, 41, 42, 43/44, 45, 1967-1968