
1 Units for the analysis of Japanese written text and spoken discourse¹

SAKUMA Mayumi

Waseda University

Abstract

Written text and spoken discourse, which use letters and sounds, respectively, as the medium for communication, are the largest and most concrete linguistic units. Both, as the sole actual forms of Japanese communication, are complete coherent wholes which dynamically unify linguistic behaviour.

Issues related to the units of written text and of spoken discourse are both old and new. With respect to written text units, the question of which written text constituents (sentences, sentence sequences, paragraphs, *bundan* ‘written grammatico-semantic paragraphs’, etc.) are appropriate has been debated since Tokieda (1950:289). Similarly, with regard to spoken discourse, Minami (1997:295–356) investigated criteria for the identification of “written text (*bunshō*)”, “conversation (*kaiwa*)” and “discourse (*danwa*)” units, and Hayashi (1998:394–396) argued for the necessity of “qualitative units” which he called “communication units”.

Investigation of the similarities and differences between “written text” and “spoken discourse” as effective analytic units for the comprehensive description of linguistic behaviour is an issue that cannot be avoided in “written text/spoken discourse theory” in Japanese linguistics. In this paper, I explore the potential for *bundan* (groups of utterances in written texts) and *wadan* (groups of utterances in spoken discourse) to be effective analytic units of written text/spoken discourse focusing on their “unifying function” and “multiple structure”.

Keywords: Analytic units (*bunseki tan’i*), Japanese discourse analysis (*bunshō-danwaron*), grammatico-semantic written paragraph (*bundan*), grammatico-semantic spoken paragraph (*wadan*), coherency function (*tōkatsu kinō*)

1 This study is extensive elaboration of my original text, Sakuma (2006).

1 Introduction

Written texts and spoken discourses (*bunshō-danwa* 文章・談話) are the largest and most concrete units of written and spoken language, respectively. I will use the term “linguistic unit” (*gen-go tan’i* 言語単位) to refer to a dynamic unit (discrete whole) of verbal and nonverbal action in Japanese communication.

Tokieda, the founder of Japanese discourse analysis (*bunshōron* 文章論), proposed a “qualitative view of units” (*shitsuteki tan’ikan* 質的単位観)² based on his unique theory of “language as process” (*gen-go kateisetsu* 言語過程説) within the framework of Japanese traditional linguistics (*kokugogaku* 国語学) (1950: 15-17; 1960: 9-11). He claims that the “qualitative view of language” is a view of units that presumes that a whole (*ichi zentai* 一全体), understood as qualitatively unified whole (*shitsuteki tōitsutai* 質的統一), is not the ultimate end of analysis, but is rather already given at the beginning of research.

In Tokieda’s (1950: 15-17) view, the “basis for establishing the field of Japanese text study (*bunshō kenkyū* 文章研究) lies in the fact that among the three “units of language,” i.e., “word” (*go* 語), “sentence” (*bun* 文), and “written text” (*bunshō* 文章), the written text differs from the other two units, because it is “a whole with a unified structure”.

The problem of units in Japanese written text and spoken discourse is both old and new. Since Tokieda (1950:289), there has been continuous debate concerning units in Japanese written texts, specifically on which “written text constituents” (*bunshō no seibun* 文章の成分) (e.g., sentence, sequence of sentences (*renbun* 連文), formal (indented) paragraph (*danraku* 段落), grammatico-semantic written paragraph (*bundan* 文段), etc.), are appropriate. Although scholars have differed in their views on language, grammar and the position of text study, they all have agreed that the written text is a unit beyond the sentence.

Regarding the units of Japanese spoken discourse, Minami (1997:295-356) addresses the question of which criteria are applicable for identifying units, such as a written text, a conversation, and discourse. Following Minami’s view of discourse³ as a unit intermediate between sentence and conversation, Hayashi (1998:394-396) pointed out that in addition to units that are intermediate between sentence and written text, i.e., sentence clusters (*bunkai* 文塊) and formal (indented) paragraphs, it is necessary to have “qualitative units” (*shitsuteki tan’i* 質的単位), which he called “communication units” (*komyunikēshon tan’i* コミュニケーション単位), particularly in spoken discourse.

2 Tokieda (1950) contrasts his theory with Saussure’s “structural view of language” and “atomistic view of units”.

3 See Minami (1997: 297) and Minami (1997: 337-355). After 1970, Minami established “a (spoken) discourse” (*danwa* 談話) as a unit of “conversation” (*kaiwa* 会話), comparing it to “something like an indented paragraph of a written text” (Minami 1997: 297). However, he later renamed it as a “coherent unity of a conversation” (*kaiwa no matomari* 会話のまとまり) (Minami 1997: 337-355).

On the other hand, there is also a view that questions the very existence of structure and units in spoken discourse of oral communication⁴. Nonetheless, when positioning written text/spoken discourse analysis in Japanese linguistics (*nihongogaku* 日本語学), there is no question that it is an unavoidable and an important challenge to compare and contrast the units in spoken discourse, which have entered a participants' memory the moment they are uttered, and the units of a written text which are fixed as strings of characters.

Thus, beginning with the premise that written texts and spoken discourse themselves are units of verbal communication, it is possible to refine their analysis and description by establishing the multiple levels (*dankai* 段階) and elements (*yōso* 要素) involved in the process of communicating linguistic information. One could say that it is in fact self-evident that some "part" (*bubun* 部分) and "process" (*katei* 過程) must exist to support the "whole" (*zentai* 全体) / "completion" (*kanketsu* 完結) of the highest level units, i.e., written texts / spoken discourses, which are composed of lower level units such as sentences, words etc.

Based on the assumption that we establish linguistic units according to their usefulness for the analysis and goals of the research, in the remainder of this paper I will examine the constituent elements of written texts and spoken discourses and discuss their similarities and differences. The goal will be to ascertain what units are useful for analyzing the organization of written text and spoken discourse.

2 Sentence and grammatico-semantic written paragraph as the units of Japanese written text

While Tokieda (1950:289) did not consider individual sentences to be constituents of "Japanese written texts", he did consider "paragraphs" (*bunsetsu* 文節, *danraku* 段落, *bundan* 文段), and "chapters" (*shō* 章, *hen* 篇) as constituents. However, he did not provide a detailed definition regarding these constituents.

Nagano (1972/1986), systematizing "grammatical Japanese text analysis" (*bunpōron-teki bunshōron* 文法論的文章論), proposed "sentence" and "formal (indented) paragraph" (*danraku* 段落) as units. In contrast, Ichikawa (1978), taking the view of "general Japanese text analysis" (*hanbunshōron* 汎文章論), proposed "sentence" and "grammatico-semantic written paragraph" (*bundan* 文段) as units. Like Ichikawa's "grammatico-semantic written paragraph", Tsukahara (1966) does not consider "formal (indented) paragraphs" (in his terminology "rhetorical paragraphs" (*shūjiteki danraku* 修辞的段落)), to be a constituent

4 See Nomura (2002: 110). A similar opinion was also expressed at the Symposium of the Society for Japanese Linguistics held in spring in 2006 (Session A ("Japanese) written texts/ spoken discourse").

unit of Japanese written texts, but rather he considers “logical paragraph” (*ronriteki danraku* 論理的段落) to be the constituent unit. Furthermore, by dividing paragraphs into “basic paragraphs” (*kibon danraku* 基本段落) and “paragraph clusters” (*danraku rengō* 段落連合), he admits the possibility that a complex sentence might consist of several “paragraphs”. Nagata (1995), on the basis of “sentence sequence theory” (*renbunron* 連文論) and Japanese written text analysis, proposes “word”, “sentence” and “paragraph” (*danraku* 段落) as units, basing his understanding of paragraph on Tsukahara’s definition.

There is also a view that does not recognize the “formal (indented) paragraph” as a “linguistic unit” based on objective criteria, considering it more as something belonging to the sphere of punctuation rules. Furthermore, regarding grammatico-semantic written paragraphs with multiple structure (*jūsō kōzō* 重層構造) deriving from “coherent organized units based on big and small topics” (*daishō no wadai no matomari* 大小の話題のまとまり), opinions disagree as to which relative division into units, at what level, and with what amount/extent of content should be taken as a basis for constituent units of Japanese written text and discourse. The existence itself of formal indicators/criteria for grammatico-semantic written paragraph has also been questioned.

2.1 Analysis of ‘cohesion between sentences’ (*bun no tsunagari* 文のつながり) based on sentence as a unit

Nagano’s (1972/1986) “theory of cohesion” (*rensetsuron* 連接論), “theory of continuity” (*rensarōn* 連鎖論), and “theory of coherency” (*tōkatsuron* 統括論) share a view that takes sentence as the basic unit for analyzing the structure of Japanese written texts. This point of view considers the structure of a Japanese written text, defined as a unified body consisting of connected sentences (*bun no renzoku tōitsutai* 文の連続統一体), to be relations between individual sentences and between formal (indented) paragraphs. These relations are viewed as “conjunctive relations between sentences” (*bun no rensetsu kankei* 文の連接関係) based on the cohesion between individual sentences or formal (indented) paragraphs, and, furthermore, as “continuity relations between sentences” (*bun no rensa kankei* 文の連鎖関係), based on chains of subjects (*shugo* 主語), predications (*chinjutsu* 陳述)⁵, and principal words and phrases (*shuyōgoku* 主要語句). In my opinion, this view of Japanese written text is being too microscopic and a mere application of the results of grammar research to written texts.

A “sentence sequence” (*renbun* 連文) is a body of semantically connected sentences. The smallest sequence of sentences is a pair of two adjacent sentences, and the largest coincides with the whole written text. In addition, a “complex sentence” (*fukubun* 複文), consisting of several clauses, can be identified as a grammatico-semantic written

5 The usage here follows Heiko Narrog (2009) *Modality in Japanese: The layered structure of the clause and hierarchies of functional categories*, Amsterdam: John Benjamins, translating *chinjutsu* 陳述 as “predication” (translator’s comment).

paragraph, based on the fact that the function of the adverbial predicate forms (*ren'yōkei* 連用形) that make up these sequences corresponds to that of connectives. Thus, complex sentences can be considered to be units akin to a sentence sequence.

“Sentence sequence theory in the narrow sense” (*kyōgi renbunron* 狹義連文論) is concerned with the semantic connections between adjacent sentences, while “sentence sequence theory in the wide sense” (*kōgi renbunron* 広義連文論) takes into account the whole written text. In other words, “cohesion between sentences” is the precondition for a “coherent unity of sentences”. The grammatico-semantic written paragraph is a unit composed of connected sentences which constitute a written text. In contrast, sentence sequences are merely a part of the grammatico-semantic written paragraph, that is, a body of connected sentences expressing a fragment of some topic. This limits the extent to which a sentence sequence can be a constituent unit of a written text.

2.2 Analysis of ‘coherent unity of sentences’ (*bun no matomari* 文のまとまり) based on grammatico-semantic written paragraph as a unit

“Grammatico-semantic written paragraphs”, occupying the middle ground between sentence and written text, can be embedded to larger units, that is, “semantic paragraph sequences” (*rendan* 連段). These “large grammatico-semantic written paragraphs” (*dai-bundan* 大文段), which are made of several grammatico-semantic written paragraphs express a larger coherent unity of semantically related themes. In the final analysis of a written text, grammatico-semantic written paragraphs and semantic paragraph sequences, mutually related through coherency (*tōkatsu* 統括), establish the largest multiple structure of the written text, on the basis of the coherent unity of topics.

In Sakuma (2003:91-119), I made a distinction between “topic/core sentences” (*chūshinbun* 中心文), sentences that express the principal information of a grammatico-semantic written paragraph in the most straightforward way, and “thesis sentences” (*shudaibun* 主題文), sentences that express the theme of the whole written text. It is usually the case that a written text consists of several semantic paragraph sequences with a unified theme (*shudai* 主題) and that a grammatico-semantic written paragraph consists of several sentences with a unified topic (*wadai* 話題). Here, by “topic” I mean an expression of the principal content in a grammatico-semantic written paragraph. Topic sentences and thesis sentences both possess the coherency function (*tōkatsu kinō* 統括機能), imparting coherent unity to the large and small topics in a written text.

Topic sentences and thesis sentences impart a relative strength of coherency corresponding to the scope and frequency of the topics in grammatico-semantic written paragraphs, as well as to the text-developing function (*bunshō tenkai kinō* 文章展開機能) of these topics. Furthermore, the thesis sentence of the theme/core paragraph (*chūshindan* 中心段), the grammatico-semantic written paragraph with the largest strength of

coherency (*tōkatsuryoku* 統括力) in the written text, organizes and completes the whole written text.

The coherency function of the grammatico-semantic written paragraph performs the following roles:

- (i) topic presentation (*wadai teiji* 話題提示)
- (ii) conclusion (*ketsuron hyōmei* 結論表明)
- (iii) issue raising (*mondai teiki* 問題提起)
- (iv) introduction of the problem/ issue to be solved (*kadai dōnyū* 課題導入)
- (v) connection with the preceding context and introduction of the following context (*shōzen kigo* 承前起後)
- (vi) introduction (*maeoki* 前置き)
- (vii) appending (*atozuke* 後付け)

In addition, both topic/core sentences and theme/core paragraphs can appear in a grammatico-semantic written paragraph and in the written texts respectively, in one of the following six positions:

- 1) beginning (*saisho* 最初)
- 2) end (*saigo* 最後)
- 3) beginning and end (*saisho to saigo* 最初と最後)
- 4) middle (*chūkan* 中間)
- 5) several dispersed positions (*fukusū bunsan* 複数分散)
- 6) implicit (*senzai* 潜在)

Furthermore, the position of topic/core sentences and theme/core paragraphs may differ depending on the type of written text, on their position and frequency within grammatico-semantic written paragraphs, and on their text-developing function in the written text. The “topic presentation” (*wadai teiji* 話題提示), the “introduction of a problem/issue to be solved” (*kadai dōnyū* 課題導入) and the “introduction” (*maeoki* 前置き) itself tend to appear at the beginning, while the expression of the “conclusion” (*ketsuron hyōmei* 結論表明), the “issue raising” (*mondai teiki* 問題提起), and the “appending” (*atozuke* 後付け) tend to appear at the end. Finally, the “connection with the preceding context and introduction of the following context” (*shōzen kigo* 承前起後) tend to appear in the middle or in the grammatico-semantic written paragraphs that are constituted by a single sentence.

3 Utterances and grammatico-semantic spoken paragraphs as units of Japanese spoken discourse

Japanese spoken discourse is made up of “utterances” (*batsuwa* 発話), subunits of Japanese spoken discourse, that is, spoken linguistic units of various sizes. It is nonetheless reasonable to think that the structure of Japanese spoken discourse can be described to a considerable degree from the point of view of the analysis of Japanese written text.

3.1 Analysis of ‘topic sequences’ with utterances as units

Sugito (1987: 83) analyzes the “transfer/continuity of utterances” (*batsuwa no uketsugi* 発話のうけつぎ) in data from round table discussions. He defines an “utterance” as follows:

This [an utterance] is a unit, an internally consistent chunk of a continuum of spoken language by a single participant (also including laughter and short back channeling). Each chunk is delimited by continuum of spoken language (same as above) produced by other co-participant[s] and by pauses [gaps in time], and is counted as a separate unit. (Example omitted). This unit, i.e., utterance, may be shorter than what is considered to be a sentence in grammar, [and often] corresponds to a phrase (*bunsetsu* 文節, *ku* 句) or a word or just a part of an interrupted expression. Yet sometimes it may be longer, appearing to be a sequence of two or more sentences. The “utterance” is a unit which may appear variously in long or short form. If one regards units as necessarily uniform/homogeneous, it is true that it is difficult to call such an internally consistent chunk a unit. Nonetheless, for the purpose of following the verbal exchange between participants in Japanese spoken discourse, the utterance provides a rather explicit clue regarding the actual divisions in such verbal exchange.

(Sugito 1987: 83, underline by Sakuma)

If an “utterance” with no fixed length is a unit lacking a “homogeneous” size, then a “grammatico-semantic spoken paragraph”, defined as a content-based relative division, shares a similar characteristic, i.e., a lack of “homogeneity” of its length due to the variable complexity of its content. However, is it not that this property characterizes the communication units of all verbal behaviour? It can be stated that in this respect “grammatico-semantic written paragraphs” in written text are similar communication units.

Szatrowski (1993) identified *wadan* (i.e., grammatico-semantic spoken paragraph, which she reinterpreted as ‘stages’) in invitation conversations based on the differences in the participants’ goals and the interaction between “utterance functions”, and described the overall structure (*zentaiteki kōzō* 全体的構造) of telephone conversations.

Her analysis provided empirical evidence for the limitations of analyzing Japanese conversation structure based only on adjacency pairs and utterance sequences.

3.2 Analysis of ‘coherent unity of topics’ with grammatico-semantic spoken paragraphs as units

Minami (2005: 537), assuming that “linguistic units are set up in linguistic analysis and description methodology as the basic elements constituting language”, proposes units for each of the five realms of language:

(5) [Units] based on spoken discourse (written text) and related to verbal behaviour: sentences (and utterances equivalent to sentences), as well as various coherently organized discourse units belonging to written texts and spoken discourse (i.e., formal [indented] paragraphs, grammatico-semantic spoken paragraphs (*wadan* 話段), etc.), and in addition, also coherently organized units of communication behaviour including both verbal and nonverbal expressions also belong here.

(Minami 2005: 537, underline by Sakuma)

Here, “spoken discourse (written text)” refers to the largest linguistic unit in the same way as “written texts/spoken discourses” does in this paper. However, in particular the units Minami has put in parenthesis (), that is “utterance” and “formal (indented) paragraph”, “grammatico-semantic spoken paragraph” (*wadan*), etc., require special attention. It is particularly relevant for the present research that Minami (1997), who has been consistently investigating “units of Japanese spoken discourse”, at this point for the first time proposes a unit called “grammatico-semantic spoken paragraph”.

Sakuma (1987) was the first to propose the term “grammatico-semantic spoken paragraph” (*wadan*) for a constituent of Japanese spoken discourse, corresponding to “grammatico-semantic written paragraph” (*bundan*), a unit of Japanese written texts. In addition, Sakuma (1992; 2003:91) proposes and defines the “grammatico-semantic paragraph” (*dan*), as a group of one or more sentences or utterances which in principle form a “communicative unit”. Verification of the six to eight criteria proposed by Minami (1997) as clues for the identification of the units of Japanese spoken discourse remains the most urgent task in Japanese discourse analysis (*bunshō danwaron* 文章・談話論).

As units of Japanese spoken discourse, neither “utterances” nor “grammatico-semantic spoken paragraphs” can be included in the framework of particular formal units such as “words”, “phrases” (*ku*), “clauses” (*setsu*), “sentences”, etc. As coherently organized units of spoken/sound/vocalized expression and semantic content, they come into existence fluidly during the communication process of spoken discourse. “Grammatico-semantic spoken paragraphs”, like “grammatico-semantic written paragraphs”, possessing

an internal multiple structure imparted by the coherency function. The coherency function itself originates in coherently organized units based on topics, supports the macro-structure of Japanese spoken discourse and is thus a dynamic unit of linguistic behaviour, that is deeply involved in the process of spoken communication.

Sugito (1984) considers the units of Japanese spoken discourse to be entities which fulfill the conditions of “necessity, validity/effectiveness (*yūkōsei* 有効性) and sufficiency (*jusokusei* 充足性) based on his research goals and the analytic viewpoint”, and proposes the following four “basic characteristics of linguistic units”: parallelism (*heiretsusei* 並列性), possibility of combination (*ketsugōsei* 結合性), “multilayeredness” (*jūsōsei* 重層性) and “exhaustiveness” (*mōrasei* 網羅性). However, it is my claim that among these characteristics, “multilayeredness” in particular seems to reveal the essence of the units of Japanese spoken discourse.

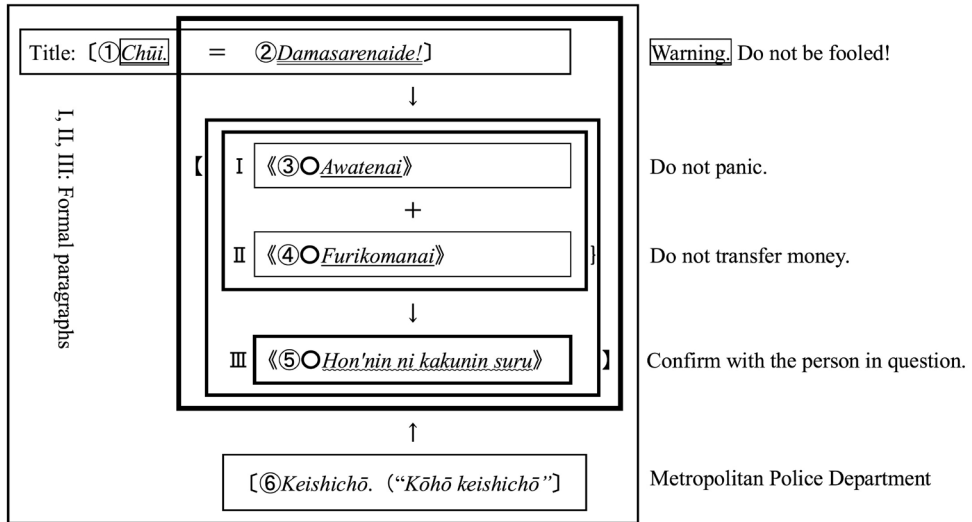
4 “The grammatico-semantic paragraph” *dan* as a unit of Japanese written text and spoken discourse

Following Minami’s (2005; 1997) “coherently organized communicative behaviour units, based on linguistic and nonlinguistic expression”, valid analytic units reflecting various aspects of communication in Japanese possess several clues that constitute criteria for their identification, as well as linguistic expressions, which function as formal markers.

If the “grammatico-semantic paragraph” (*dan* 段) is used as a general term which includes both “grammatico-semantic written paragraphs” of Japanese written text and “semantic spoken paragraphs” of Japanese spoken discourse, then, in addition to topic, there are various other criteria for the identification of units. These include criteria based on factors such as participants, scene, communicative function, attitudes of expression, etc. These criteria reveal elements and levels which are indispensable for the analysis and description of the structure and function of “Japanese written text and spoken discourse”.

Example (1) below is a short public information text which, including the title, consists of 6 sentences and 3 formal (indented) paragraphs (I, II, III). Its text structure consists of an enumeration of specific measures to take in response to a warning. Coherency of the 3 formal (indented) paragraphs, realized in the body of the text as the sentences ③, ④ and ⑤ respectively, is imparted by the thesis sentence ②.

(1)



“Kōhō keishichō” 16 gō, 2005. 1. 16. (Metropolitan Police Department public information No. 16, 16. Jan. 2005.), underline by Sakuma, with *dan* units (in boxes) and conjunctive relations among them. For the explanation of symbols =, ↓, +, ↑, and for the annotated original text, see Appendix 1. Brackets are explained in the footnote 5 below.

Sentences ① and ② represent the Opening section (*kaishibu* 開始部) and ⑥ the Closing section (*shūryōbu* 終了部) of the text, displaying its overall structure. Sentence ② is part of the title and conveys information by addressing the reader with an expression of prohibition (*kinshi hyōgen* 禁止表現). Although the main text ③~⑤ consists of three formal (indented) paragraphs, from the point of view of content, they are all part of a grammatico-semantic written paragraph with sentence ⑤ as the topic/core sentence. It is difficult to recognize formal (indented) paragraphs, employed for visual effect, as intrinsic units of the text. On the other hand, this grammatico-semantic written paragraph goes beyond the limits of the main text ③~⑤. Sentence ② from the title imparts coherency to the whole text, while the predication of sentence ⑤, i.e., an imperative expression, imparts coherency to the expressions of prohibition in sentences ②~④. Based on the properties of ⑤, it is possible to identify the multiple structure of the whole grammatico-semantic large written paragraph ①~⑥.

Example (2) is the Opening section of a spoken discourse, a public lecture, and since it is a monologue, the whole text (including quotations), is one utterance, with a multiple structure (imparted by several grammatico-semantic spoken paragraphs belonging to different dimensions). Furthermore, it contains complex sentences that are

made up of several grammatico-semantic spoken paragraphs. In addition, it also has formal markers that indicate its multiple structures based on the coherency function of big and small grammatico-semantic spoken paragraphs.

Example (2) below is divided into 4 grammatico-semantic spoken paragraphs. The super large grammatico-semantic spoken paragraph A, the semantic paragraph cluster consisting of grammatico-semantic paragraphs I, II, III, corresponds to the Opening section of this lecture discourse.

Grammatico-semantic paragraph IV opens a new topic realized in the text by the repetition of “*ikutsu gurai no kotoba o, kyō—*” (how many words did (you) say today?) in ⑨c and ⑩b. It is part of the super large grammatico-semantic spoken paragraph B (which belongs to the Development section (*tenkaibu* 展開部)). B begins by addressing the audience with the topic expression “*minasama—*” (Dear guests--), and then presenting its content in sentences ⑨ “*batsuon nasutta ndeshō ka*” (did (you) say?) and ⑩ “*o kangae n natta ndeshō ka*” (did (you) think about), further impressing the audience with two repeated interrogative expressions.

In the omitted part after sentence ⑪, the lecturer provides answers using first person topic expressions as in ⑪a “*atakushi wa*” (I) and “*atashi wa*” (I), the lecturer proceeds with the theme of vocabulary in everyday use.

Grammatico-semantic paragraph I consists of the Opening a self-introduction with greetings in the Opening section and grammatico-semantic paragraph II introduces the topic touching upon the theme of the lecture, i.e., “*kotoba ga kowai*” (the words are scary). The large grammatico-semantic spoken paragraph I develops into grammatico-semantic spoken paragraph II. After grammatico-semantic paragraph II, topic expressions concerning ‘*kotoba* (words)’ are repeated. Although first person topic expressions, such as “*watakushi wa*” (I), are ellipted in the grammatico-semantic paragraph I, and sentences ① and ②, they are expressed explicitly in grammatico-semantic paragraph II sentence ④a “*watakushi wa*” (I) and ⑤c “*watashi wa*” (I). The reason they are expressed explicitly is because after a different topic expression has been introduced in sentence ② “*Hotondo no kata ga*” (most of you [present here]) a new topic expression, “*kotoba ga*” (the words) has been introduced.

(2)

A 1 I	<p>① a <i>Tadaima, goshōkai ni azukarimashita,</i> b <i>Mukōda Kuniko de gozaimasu.</i></p> <p>② a <i>Hotondo no kata ga</i> <i>ohatsu ni ome ni kakaru n da to</i> b <i>omoimasu.</i></p> <p>③ a <i>yoroshiku</i> b <i>onegai shimasu.</i></p>	<p>① b (I) am Mukōda Kuniko a who has just now been introduced.</p> <p>② a It's that (this) is the first time (for me) to meet most of you b I think.</p> <p>③ a Kindly b I (humbly) request (that you treat me well).</p>
2 II	<p>④ < a <i>ano-, watakushi wa, kono goro n natte-</i> > b <i>e-, kotoba ga-,</i> <i>totemo kowaku narimashita.</i> ></p> <p>< ⑤ [a <i>ano-, kī ga tsukimasu to,</i> > > b <i>umarete</i> > > c <i>sugu ni-, hantoshi gurai de-,</i> > > d <i>'mamma' tte iu</i> > > e <i>kotoba o, watashi wa shabetta n da sō</i> <i>desu keredomo-</i> >]</p> <p>[< f <i>sorekara gojūichi nen,</i> > > g <i>gojū-, ima, ichi n narimasu keredomo,</i> > > h <i>ano-, sono aida ni-,</i> > > i <i>zūibun takusan no kotoba o, ma,</i> <i>shabetteri,</i> > > j <i>kangae tari,</i> >]</p> <p>[< k <i>kotoba to iu no wa, oto ni</i> > <i>dasanakatemo,</i> > > l <i>kokoro no naka ni aru matomatta</i> > > m <i>koto o omoeba,</i> > > n <i>sore wa, kotoba da, to</i> > > o <i>omou n desu.</i> >]</p> <p>⑥ [< a <i>shabetteri,</i> > > b <i>omottari,</i> > > c <i>ma, saikin wa, kaitari,</i> > > d <i>ma, shite, orimasu.</i> >] ></p> <p>< ⑦ [< a <i>hotondo, zenhan wa,</i> <i>maishiki de, tsukatteta to</i> > > b <i>omou n desu keredomo,</i> >]</p>	<p>④ a U:m, I, at this time, b um, words have become very scary for me.</p> <p>⑤ a U:m, when (I) became aware b after [I was] born c immediately, in about half a year, d "Mamma" e they say it's that I spoke [those] words, but, f since then 51 years, g it has been 51 years, but h um, during that time i I have spoken, well, very many words, j and have thought (about those words), k it's that even if you do not pronounce the words m if you think about things l which make sense in your heart, n those are words o I think.</p> <p>⑥ a Speaking, b thinking, c well, recently, writing d (I) have been doing that.</p> <p>⑦ a It's that (I) have used (words) unconsciously through nearly all of the first half [of my life] b I think but,</p>

<p>III</p>	<p>[< c <i>shokugyō ni shite,</i> >] [< d <i>shikamo-, kotoba de, gohan o taberu yō n natte</i> >] [< e <i>ni jū-nen ni narimasu to,</i> >] [< f <i>mukashi-, maishiki ni tsukatteta</i> >] [< g <i>kotoba ga-, ma, kono goro n natte,</i> >] [< h <i>dondon, dondon kowaku natte kimashite,</i> >] [< i <i>e-, kotoba to iu</i> >] [< j <i>mono o, chotto, jibun nari ni</i> >] [< k <i>kangaete miru yō ni narimashita.</i> >]] > }]</p> <p>[[⑧ < a <i>ma, kyō wa, son'na tokoro o, chotto-, myakuraku naku</i> >] [< b <i>ohanashi shite miyō to</i> >] [< c <i>omoimasu.</i> >]]]</p>	<p>c (I) made it (my) vocation, and d furthermore, (I) reached the point where (I) could earn my living with words, and e (after) 20 years, g the words f (which I) used previously without thinking h have gradually become scary, and k (I) have started thinking [about them] i um so called "words", j (those) things a little bit, on my own terms.</p> <p>⑧ b (I) will try to talk. a well, today (about) those kind of things a bit out of context c I think.</p>
<p>B 3 IV</p>	<p>[< ⑨ [< a <i>minasama-, kesa, ooki n natte,</i> >] [< b <i>koko e irassharu made no aida ni,</i> >] [< [ft-] <i>toiki</i> >] [< c <i>ikatsu gurai no kotoba o, kyō, hatsuon nasatta ndeshō ka.</i> >]] > }]</p> <p>[< ⑩ [< a <i>sorekara-, hatsuon nasaranai made mo,</i> >] [< b <i>ikatsu gurai no kotoba o, kyō, okan^{ga}e n natta ndeshō ka.</i> >]] > }]</p> <p>[< ⑪ [< a <i>atakushi wa, hitorigurashi desu kara,</i> >] [< b <i>koko e karu made no aida,</i> >] [< c <i>denwa ga go, roppon kakatta igai wa,</i> >] [< d <i>ano-, hito to shaberazu ni,</i> >] [< e <i>koko e mairimashita.</i> >]] > }]</p> <p>(The rest is omitted)</p>	<p>⑨ a Dear guests, (from the time) you got up this morning b till when (you) came here [sigh] c how many words did (you) say today?</p> <p>⑩ a And then, even if (you) did not say (them) b about how many words did (you) think about today?</p> <p>⑪ a (I) live alone, so b in the time it took to come here c aside from 5 or 6 phone calls (I) received e I have come here d um without speaking to people.</p>

Mukōda Kuniko (向田邦子) ‘*Kotoba ga kowai*’ (the words are scary) “*Shinchō kasetto kōen* (Shincho cassette lectures)” 1991, Tokyo: Shinchōsha (underline and symbols added by Sakuma). A ~ B, 1 ~ 3 are “large grammatico-semantic spoken paragraphs” (*daiwadan* 大話段), I ~IV are grammatico-semantic spoken paragraphs⁶, ①~⑪ are sentences, and a ~ o are clauses in respective sentences. Annotated original text is shown in Appendix 2.

In this lecture, several formal markers within a single utterance to indicate the multiple structure of grammatico-semantic spoken paragraphs are used. In addition, several formal markers can be observed that hint at the multiple structure created by the coherency function of grammatico-semantic spoken paragraphs, which belong to different dimensions.

Grammatico-semantic paragraph II consists of three small grammatico-semantic paragraphs (*shōwadan* 小話段), i.e., sentence ④, sentences ⑤ and ⑥, and sentence ⑦. In sentences ④ and ⑦, expressions related to the theme of the text (such as “*kono goro n natte--*” (it is in these days...), “*kotoba GA --*” (words), “*kowaku narimashita / natte kimashite*” (became very scary / (have gradually become scary) are repeated.

Grammatico-semantic paragraph II is made of sentences ④ through ⑦. Sentence ④, the topic/core sentence (中心文) in the grammatico-semantic paragraph II, is described in more detail in the sequence of sentences ⑤~⑦.

Because there were different new inserted topic expressions in the second half of the compound sentence ⑤ (i.e., ⑤k “*kotoba TOIUNOWA*” (words), and ⑤n “*sore WA*” (that)), the clause ⑤i “*shabettari*” (talk and...) from the first half of ⑤ is repeated in ⑥a, and reworded in ⑥b “*omottari*” (think and/thinking...), ⑥c “*saikin WA, kaitari*” (well, recently, I write), and ⑥d “*ma, shite, orimasu*” (well, doing).

Grammatico-semantic paragraph III begins with sentence ⑧ which hints in advance about a change in the topic (*wadai tenkan* 話題轉換), with the filler “*ma*” (well). Subsequently, the overall theme of the lecture is presented with the topic expression ⑧a “*kyō WA sonna tokoro o*” (today, **such** points), and with the predications ⑧b “*obanashi shite miyō to*” (would like to talk) and ⑧c “*omoimasu*” (I think).

Grammatico-semantic paragraph III consists of sentence ⑧, the large topic/core sentence (*daichūshinbun* 大中心文). By imparting coherency to the preceding grammatico-semantic spoken paragraph II consisting of four sentences ④~⑦, the large grammatico-semantic spoken paragraph 2. In addition, this large grammatico-semantic spoken paragraph 2 derives coherency from large grammatico-semantic spoken paragraph 3 in the Development section (*tenkaibu* 展開部). The large

6 Expressions in brackets [{ << [< >] >> }] are grammatico-semantic spoken paragraphs, formed by larger and smaller coherently organized units based on theme. Ordering of brackets shows the hierarchy of layers from large to small. Brackets are visualized by enclosures.

grammatico-semantic spoken paragraph 3 includes grammatico-semantic paragraph IV, and forms the complex multiple structure of grammatico-semantic paragraphs over the whole discourse of this lecture. The fact that this multiple structure all results from the coherency relations among several sentences and grammatico-semantic spoken paragraphs within a single utterance is an important characteristic of the units of Japanese spoken discourse in monologue data.

5 Conclusions

If we think about the units of analysis of Japanese written text and spoken discourse from the point of view of the actually realized forms in communication, then the importance of “grammatico-semantic paragraph units beyond the sentence” as units whose coherency function organizes topics, becomes apparent. In other words, it becomes clear that the “grammatico-semantic paragraph” (*dan*) is the very unit for conveying linguistic information.

The medium for conveying linguistic information is not only sound and writing but also nonverbal expressions including gestures, visual images, etc. Thus, means of communication diversify in complex ways, and, when attempting to exchange the information content in a more effective way, communicative behaviour is accomplished by adapting to various stages and elements of the multiple structure of grammatico-semantic paragraphs in Japanese written texts and spoken discourse. In establishing valid units for the analysis and description of the entire process of expressing and understanding Japanese written text and spoken discourse, from its opening to its closing, the task faced by Japanese discourse analysis is to elucidate the dynamics of linguistic performance.

Valid units of verbal or written communication are important as foundation for detailed analysis of linguistic exchange, as is exemplified in the conjunctive relations analysis of example (1) in the Appendix 1.

There is still a need to strive to establish even more valid units of analysis: that will enable our understanding of both the processing of various kinds of nonverbal information and of the differences in the scale of all kinds of communicated expressions. These include “complex written texts and spoken discourses” (*fukubunshō/danwa* 複文章・談話), “hybrid written text and spoken discourse” (*kongōbunshō/danwa* 混合文章・談話), and “simple written text and spoken discourse” (*tanbunshō/danwa* 単文章・談話).

Acknowledgements

I would like to thank Andrej Bekeš for inviting me to join this project, for his useful comments, and for translating my text from Japanese into English. I would also like to thank Polly Szatrowski for her valuable comments during the translation process. Of course, the responsibility for the final text remains mine.

Literature

- Hayashi S. (1998) *Bunshōron no kiso mondai* (Basic issues in text analysis). Tokyo: Sanseidō.
- Ichikawa T. (1978) *Kokugo kyōiku no tame no bunshōron gaisetsu* (Outline of text theory for pedagogy of Japanese as a national language) Tokyo: Kyōiku shuppan.
- Minami F. (1997) *Gendai nihongo kenkyū* (Studies of modern Japanese). Tokyo: Sanseidō.
- Minami F. (2005) 'Gengo no tan'i' (Units of language), Nihongo kyōiku gakkai (ed.), *Shinpan nihongo kyōiku jiten* (Dictionary of Japanese language pedagogy - New edition), 537-538. Tokyo: Taishūkan shoten.
- Nagano M. (1972) *Bunshōron shōsetsu* (Detailed text analysis). Tokyo: Asakura shoten.
- Nagano M. (1986) *Bunshōron sōsetsu* (General remarks on text analysis). Tokyo: Asakura shoten.
- Nomura M. (2000) *Nihongo no tekusuto: kankei, kōka, yōsō* (Japanese text: relations, effects, appearance). Tokyo: Hitsuji shobō.
- Nagata H. (1995) *Kokugo bunshōron* (Text analysis of Japanese). Osaka: Izumi shoin.
- Sakuma M. (1987) 'Bundan nintei no ichi kijun (I): teidai hyōgen no tōkatsu (A criterion for recognition of grammatico-semantic written paragraphs (I): coherency of thematizing expressions', *Bungei, gengo kenkyū - gengo hen* 11: 89-135.
- Sakuma M. (1992) 'Bunshō to bun: dan no bunmyaku no tōkatsu' (Text and sentence: coherency of the paragraph context). *Nihongogaku* 11 - 14: 41-48.
- Sakuma M. (2003) 'Bunshō-danwa ni okeru dan no tōkatsu kinō' (Coherency function of grammatico-semantic paragraphs in text and/or spoken discourse), Kitahara Yasuo (editorial supervision), Sakuma Mayumi (ed.) *Asakura nihongo kōza 7 bunshō-danwa*, 91-119. Tokyo: Asakura shoten.
- Sakuma M. (2006) 'Bunshō-danwa no bunseki tan'i (Units for the analysis of Japanese written texts and/or spoken discourse) *Gekkan Gengo*. Vol.35, No.10: 65-67. Tokyo: Taishūkan shoten.
- Sakuma M. (2012) 'Bunshō-danwa no bunseki tan'i (Units for the analysis of Japanese written texts and/or spoken discourse). *Gengo serekushon* (Selection from 'Gengo') Pt. 1, 93-100. (republication of Sakuma 2006) Tokyo: Taishūkan shoten.

- Sugito S.(1984) 'Danwa no tan'i ni tsuite' (On the units of discourse), *Gengo seikatsu* 393: 34-41.
- Sugito S. (1987) 'Hatsuwa no uketsugi (Inheriting the utterance)' Ch.2.2, National Institute for Japanese Language and Linguistics (ed.) *Danwa kōdō no shosō* (Aspects of discursive actions), 68-106. Tokyo: Sanseidō.
- Szatrowski, P. (1993) *Nihongo no danwa no kōzō bunseki* (Analysis of Japanese conversation structure). Tokyo: Kuroshio shuppan.
- Tokieda M. (1950) *Nihon bunpō kōgo hen* (Japanese grammar: spoken language). Tokyo: Iwanami shoten.
- Tokieda M. (1960) *Bunshō kenkyū josetsu* (Introduction to text study). Tokyo: Yamada shoin.
- Tsukahara T. (1966) 'Ronri-teki danraku to shūji-teki danraku' (Logical paragraphs and rhetorical paragraphs), *Hyōgen kenkyū* 4: 1-9.

Appendix 1

Original text in example (1)

① 注意 ② だまされないで!

I ③ 〇慌てない

II ④ 〇振り込まない

III ⑤ 〇本人に確認する

⑥ 警視庁

(『広報けいしちょう』16号(2005.1.16) (下線、筆者付す。))

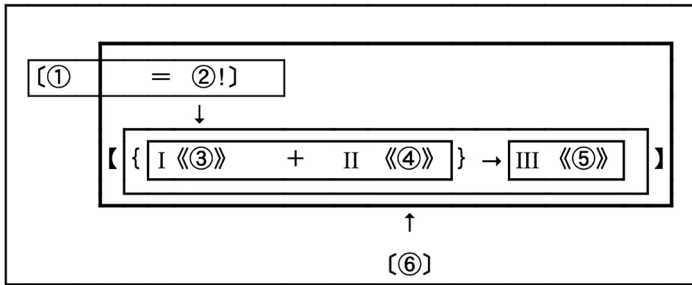


Figure 1. Conjunctive relations between units in Example (1)

Explanation of symbols (cf. Ichikawa 1978: 89-93):

- | | | |
|--|---|---|
| 1. ①→② causal/illative
(<i>junssetsugata</i>) | 2. ①Z② adversative
(<i>gyakusetsugata</i>) | 3. ①+② additive
(<i>tenkagata</i>) |
| 4. ①↔② contrastive
(<i>taihigata</i>) | 5. ①=② illustrative
(<i>dōretsugata</i>) | 6. ①←② causative/supplementing
(<i>hosokugata</i>) |
| 7. ①↓② transitional
(<i>tenkangata</i>) | 8. ①-② sequential
(<i>rensagata</i>) | |

Appendix 2

Original text in example (2)

A 1 I 【{①a ただ今、ご紹介にあずかりました、b 向田邦子でございます。】

②a ほとんどの方がお初にお目にかかるんだとb 思います。】

③a よろしく b お願いします。】】

2 II 【{<④ <a あの一、わたくしは、この頃なって一、> <b え一、言葉が一、とても怖くなり
ました。> <⑤ <a あの一、気が付きますと、> <b 生まれて> <c すぐに一、半年ぐらい
で一> <d 「マンマ」っていう e 言葉を、わたしはしゃべったんだそうですけれども、>]
[<f それから五十一年、> <g 五十一、今、一んなりますけれども、> <h あの一、その間に
二、> <i ずいぶん沢山の言葉を、ま、しゃべったり、> <j 考えたり、>] [<k 言葉というの
は、音に出さなくても、> <l 心の中であるまとまつたmことを思えば、> <n それは、言葉
だ、と思うんです。>] [⑥ <a しゃべったり、> <b 思つたり、> <c ま、最近は、書いたり、>
d ま、して、おります。>] > <⑦ [<a ほとんど、前半は、無意識で、使つてたと b 思うん
ですけれども、>] [<c 職業にして、> <d しかも一、言葉で、ご飯を食べるようなつて、>

〈e二十年になりますと一〉〈f昔□、無意識に使ってたg言葉が一、ま、この頃んなつて、〉〈hどンドン、どンドン怖くなってきてまして、〉〔〈iえー、言葉というjものを□、ちよつと、自分なりに〉〈k考えてみるようになりました。〉〕>〕

III ⑧〔〈aま、今日□は、そんなところを□、ちよつと一、脈絡なく〉〈bお話ししてみよう□とc思います。〉〕〕

B3 IV 【{<< [⑨ 〈a皆様□、今朝、お起きんなって、〉〈bここへいらっしやるまでの間に、〉〔「フッー」吐息〕〈cいくつぐらいの言葉□を、今日、発音なすったんでしょうか。〉〕 ⑩〔〈aそれから一、発音なさないまでも、〉〈bいくつぐらいの言葉□を、今日、お考えんなったんでしょうか。〉〕 >< [⑪ 〈aあたくし□は、独り暮らしですから、〉〈bここへ来るまでの間、〉〈c電話□が五、六本かかった以外□は、〉〈dあの一、人としゃべらずに、〉〈eここへ参りました。〉〕 >〕} (以下、省略)

(向田邦子「言葉が怖い」『新潮カセット講演』1991 新潮社)

(各種の下線・記号は筆者付す。A～B、1～3、I～IVは話段、①～⑪は文、a～oは節の番号を示す。下線_____は時間表現、____に□□囲みは提題表現、____と____は大小の中心文を示す。)

要旨 (Abstract in Japanese)

「日本語の文章・談話における分析単位」

佐久間まゆみ (早稲田大学)

「文章」と「談話」は、それぞれ、文字と音声を伝達媒体とする最大かつ最も具体的な言語単位である。いずれも、日本語のコミュニケーションの唯一の実現形態として、言語行動の動的なまとまりを表す完結統一体とされる。

文章・談話の単位に関する課題は古くて新しい。文章の単位は、時枝(1950:289)以来、「文章の成分」として「文」「連文」「段落」「文段」等のいずれが「文章の成分」として妥当かが論議され、談話の単位も、南(1997:295-356)が「文章」「会話」「談話」の単位認定の手がかりを問い、林(1998:394-396)も「コミュニケーション単位」という「質的単位」が必要だとしている。

日本語学の「文章・談話論」における言語行動を包括的に記述する有効な分析単位として、「文章」と「談話」の異同の検討が不可避の課題である。文章における文のまとまりからなる「文段」、談話における発話のまとまりからなる「話段」の「統括機能」の「多重構造」を中心に、文章・談話の有効な分析単位の可能性を探る。