

5 Modal expressions and verbal interaction type: Suppositional adverbs as discriminators of Japanese corpora according to oral and written discourse varieties¹

Andrej BEKEŠ

University of Ljubljana

Abstract

Modal expressions can be conceived as speaker's/writer's signals for a particular linguistic exchange or as a trace of such a linguistic exchange. Thus, modal expressions are the prime candidate as discriminators of the verbal interaction type, and consequently, of discourse type. In Japanese, the typical means to express modality are sentence-final modality expressions and modal adverbs. Because modal adverbs are easier to identify than sentence-final modal expressions, this chapter will examine the possibility of using a subset, i.e. suppositional adverbs, as discriminators of the Japanese oral and written discourse type. Several public and closed corpora belonging to different genres have been analysed here. Cluster analysis showed that distribution of suppositional adverbs in the examined corpora varied according to the discourse type. The outlier was a corpus of Elementary school *kokugo* textbooks, due to its heterogeneity of genres. The differences within similar corpora (Cluster of Informal conversations, Formal interviews, and CSJ lectures) mainly follow the axis of formality. On the other hand, in the more loosely correlated cluster of Diet speeches and Science textbooks there was a clear distinction between the stress on rhetoric in the former vs. stress on the precision of argument in the latter.

Keywords: verbal interaction type, discourse type, genre, suppositional adverbs, cluster analysis

0 Introduction

Text and discourse. Texts, either written, transcribed or recorded conversation or monologue, Aristotle's *ergon*, are not something static but traces of verbal interaction, of discourse: Aristotle's *energeia*, evolving in a particular social context with particular goals (cf. Coseriu 1973). Halliday (1978, 1991) describes the context of a situation, where communication takes place within the triplet of a *field*, i.e. 'subject matter and social action', *tenor*,

1 An earlier version of this chapter was presented at the International Conference: "Context-based Spoken Japanese Language" at the University Bordeaux Montaigne, April 4th-5th, 2014.

i.e. ‘participants’ relationship’, and *mode*, i.e. ‘medium and participation’. Modality in its broadest sense is concerned with the type of linguistic exchange in a particular situation. Thus, modal expressions can be conceived as a speaker’s/writer’s signals for the nature of a particular linguistic exchange or as a trace of such a linguistic exchange. This view is in line with Bakhtin’s (1981) observation regarding the dialogic nature of text, which is, needless to say, more obvious in spoken interaction than in written.

Modality. The study of modality and related phenomena in Japanese has predominantly focused on the linguistic means which express modal meanings and the meaning of linguistic means, employed to express modality, the scope of investigation typically being a single utterance. With Jespersen as a starting point, Minami (1974, 1993) used Jespersen’s terms *modus* and *dictum*. Teramura (1982), partly to avoid the misinterpretations based on traditional approaches, used *koto* (what is actually happening) and *mūdo* (from English *mood*). More recent approaches (cf. Nitta 1989, Masuoka 1991, Narrog 2009) prefer to view utterance as consisting of two parts: proposition (*meidai*) and modality (*modariti*), where modality reflects speaker’s ‘subjectivity’.

Modality, especially in planned linguistic exchanges such as written communication, tends to be formally expressed in a single utterance. Nonetheless, since modality expressions are not there just to fulfil some syntactic role, it is necessary to go beyond a single utterance in order to adequately understand modality related phenomena, as has already been pointed out by Teramura (1984:278-290).

In Japanese and in many other SOV languages, modality is typically expressed by the modal elements attached to the predicate (cf. Minami 1974). In addition, this can also be expressed by modal adverbs. As Minami (*ibid.*, cf. also Narrog 2009), has shown, this is intimately connected with the layered nature of Japanese sentences. Co-occurrence of both modal adverb and sentence-final modality form, such as in (1) below, results in bracket structures (cf. Bekeš 2008c).

- (1) *tabun* *kiku* *-n* *-darō...* (from NUCC)
 probably hear *focus particle* **probably** (...she is probably going to hear..)

As has already been mentioned, taking verbal interaction / linguistic exchange as a starting point, modality is a signal/trace of the kind of verbal interaction taking place in a particular context of communication. Thus, we can expect that different verbal interactions in different contexts may have a different distribution of modal expressions, depending on the kind of linguistic exchange taking place. Moreover, the reverse is also possible, i.e. that similar types of distribution of modal expressions suggest similar types of verbal exchange.

Goal of this study. The goal of this study is to explore the possibility of using distribution of modal expressions as an analytic tool or indicator for discriminating

different types of verbal interactions represented by a particular text or groups of texts. Depending on the outcome, its usefulness as an indicator of genre will be assessed. To achieve this goal the focus will be on modal adverbs, in particular, on their subset of suppositional adverbs. Because of their formal simplicity, ease of identification and, as it will be shown, intricate relationship with different types of verbal interaction, suppositional adverbs seem to be most expedient for the purpose.

1 Discourse - text - genre - register - corpus

In empirical language research, two basic orientations can be distinguished. One type of research is predominantly concerned with clarifying “what is taking place in a particular linguistic exchange”, i.e. the *ichigo-ichie* (once in a lifetime) type of approach. This is typical of conversation analysis and discourse analysis approaches, where the focus on the micro analysis of individual linguistic exchanges is predominant.

On the other hand, the so-called ‘corpus approach’ is more concerned with what is now called ‘large data’, focusing on the regularities in a mass of discourses produced within a definite period. This approach is typical of corpus lexicography, corpus based and corpus driven language studies and related research. As mentioned earlier, this present study uses a corpus approach.²

1.1 Text type, genre and register

It is expected that different communicative situations engender different goals and thus different types of texts. In situations with generically similar goals, similar solutions regarding the application of linguistic means for the achievement of a given goal emerge, resulting in notions of ‘text type’, ‘genre’ and ‘register’ (cf. Halliday 1991, Trosborg 1997). Definitions of what ‘text type’, ‘genre’ and ‘register’ actually are, and how they are related to each other, differ among authors and schools. This study follows Halliday (1991) and Hasan (2009). Figure 1 below shows the complex relationship between the context (including context of situation) and various aspects of language, i.e. system and instance. On the basis of this, register can be conceived as a toolbox of linguistic means, belonging to a language system, for achieving specific goals in a given (type of) context or situation. Correspondingly, text type is associated with situation type, different situations requiring different text types.

² It is interesting to compare these research orientations with what historian Fernand Braudel (1958) called the opposition between “histoire événementielle”, a non-structural history of events, and the history of “longue durée”, i.e. history covering a longer period of time - which has the potential to highlight emergent structures in historical processes. As in linguistics, both views of historical processes are mutually interdependent.

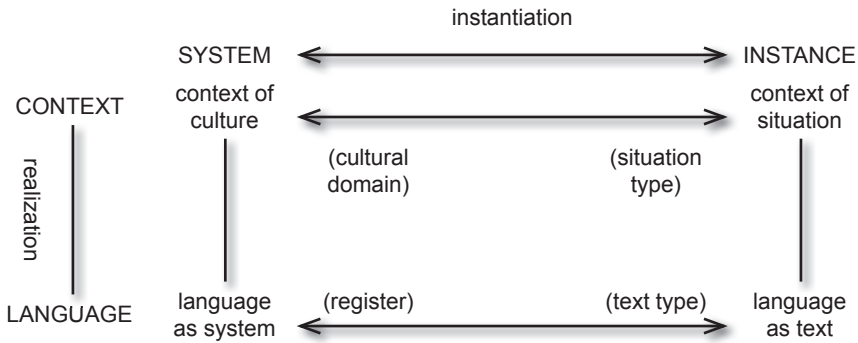


Figure 1. *Language and context: system and instance* (Hasan 2009 after Halliday 1991)

A more detailed relation between language/text, register, genre, and the context of situation is shown in Figure 2 below. Here, the term genre is used instead of the term text type as seen in Figure 1, as it is a more precise term referring to the generic purposes of communication. A type of situation defining generic purpose is also considered what motivates the choice of linguistic means for a particular occasion, i.e. the register. Register is thus closely associated with the features of the context of situation, i.e. with field, tenor and mode.

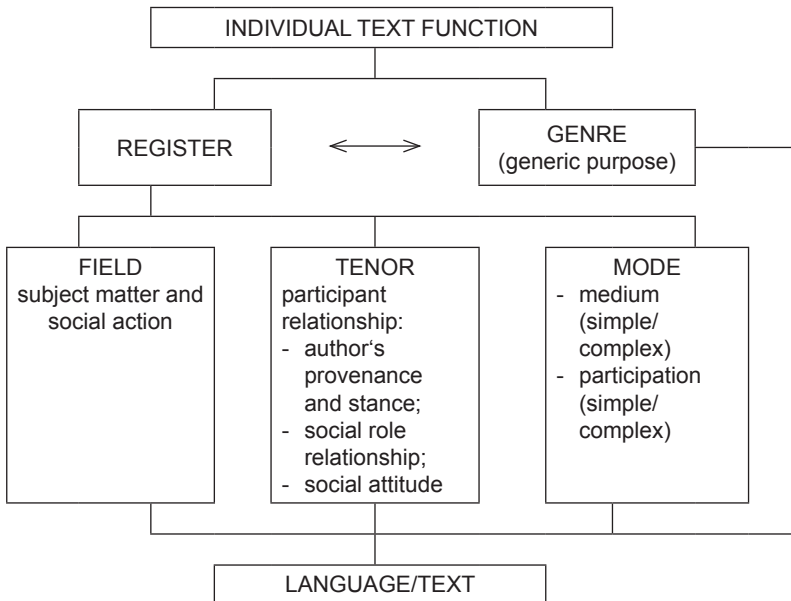


Figure 2. *Relation between register, genre and the context of situation* (cf. House 1997, Halliday 1991)

1.2 Modal adverbs in Japanese

As mentioned earlier, modal adverbs are one of the common means used to express modality in Japanese. Alongside sentence-final modal elements, attached to the predicate stem, they can be conceived as an important signal/trace of the kind of linguistic interaction taking place. Kudō's (2000) work, widely covering different types of adverbs, is a seminal work on this topic. Here, suppositional modal adverbs are of particular importance, existing in a kind of semantically based pseudo agreement with sentence-final modal expressions, as illustrated by example (1) above.

2 Suppositional adverbs as indicators of the type of linguistic exchange

2.0 Data

In this section, I will attempt to further investigate the potential of Japanese suppositional adverbs as indicators of the type of linguistic exchange.

The data used for this research consist of several corpora, some of which are openly accessible and some, due to copyright issues, are used only for research and are not openly accessible. These corpora were chosen because they consist of texts, representing widely different types of linguistic exchange, i.e.:

- (i) Informal conversations: Ohso, Mieko (2003) *Meidai kairwa kōpasu* (Nagoya University Conversation Corpus, NUCC). About 100 informal conversations between familiar participants. Text file size 3.5 MB.
- (ii) Formal interviews: Oikawa Terufumi (1998) *Jinbunkagaku to konpyūtā DATABASE vol. 1*, Sōgō kenkyū daigaku. Transcribed interviews with 50 Japanese native speakers. Unfamiliar participants, systematic status differences. Text file size 0.82 MB.
- (iii) Academic conference lectures: NINJAL CSJ corpus core data (Himawari 1.02 CSJ core data, 2005). Core part of the NINJAL Contemporary spoken Japanese Corpus.
- (iv) Japanese *Kokkai* (the Diet) speeches: NINJAL Chūnagon (1990-1999), Transcribed Diet speeches are part of this corpus.
- (v) Science textbooks: Nishina Kikuko (unpublished material prepared at the Tokyo Institute of Technology) 16 University Science textbooks for basic subjects - corpus of Science texts.
- (vi) Elementary school *kokugo* (national language) textbooks: Nishina Kikuko's (unpublished material prepared at the Tokyo Institute of Technology) 60 elementary school *kokugo* textbooks. Corpus of elementary school texts.

Characteristics of the contextual features of texts from the corpora above are displayed in Table 1 below.

‘Undemanding’ refers to the linguistic interaction regarding common topics, where specialised knowledge of a particular field is not required. On the other hand, ‘demanding’ refers to linguistic interaction requiring specialised knowledge of a particular field.

Table 1. *Contextual features of texts in the analysed corpora*

Contextual features		Informal conversations	Formal interviews	CSJ lectures	<i>Kokkai</i> speeches	Science textbooks	Elem. s. texts
FIELD	<i>subject matter</i>	private topics, undemanding	private topics, undemanding	public topics, science, demanding	public topics, state policy issues, demanding	public topics, science, demanding	fiction, medium demanding
	<i>social action</i>	private conversation	public interview, onesided conversation	public, one-sided imparting of knowledge, expository	public, antagonistic debate, argumentative	public, one-sided imparting of knowledge, expository	public, literary text, - classroom activity,
TENOR	<i>social role relationship</i>	family, friends, colleagues, little or no seniority, high familiarity	interviewer vs(+seniority) interviewee (-seniority), high seniority, low familiarity	lecturer vs. audience (students), high seniority, low familiarity	peer vs peer (Diet members), low familiarity	public, writer-reader, low familiarity	public, writer-reader
	<i>social attitude</i>	informal, familiar	formal	formal	formal	formal	formal
MODE	<i>medium</i>	spoken	spoken	spoken	spoken	written	written
	<i>participation</i>	direct	direct	direct	indirect	indirect	indirect

2.1 Suppositional adverbs

The following is a list of the most common suppositional adverbs, based on Kudō (2000).

Table 2. *Suppositional adverbs* (cf. Kudō 2000)

Modality type	No.	Adverb
NECESSITY (NEC)	1	<i>kitto</i> (surely)
	2	<i>kanarazu</i> (certainly)
	3	<i>zettai ni</i> (absolutely)
EXPECTATION (EXP)	4	<i>osoraku</i> (probably)
	5	<i>tabun</i> (likely)
	6	<i>sazo</i> (surely)
	7	<i>ōkata</i> (probably)
	8	<i>taitai</i> (usually)
	9	<i>taigai</i> (mostly)
CONJECTURE (CON)	10	<i>dōyara</i> (somehow)
	11	<i>yohodo/yoppodo</i> (quite)
POSSIBILITY (POSS)	12	<i>moshika-suruto/-shitara/-sureba</i> (maybe)
	13	<i>hyottosuruto/hyottoshitara</i> (possibly)
	14	<i>kotoniyoreba/kotoniyoruto</i> (possibly)
	15	<i>angai</i> (fairly)
	16	<i>kanarazushimo ... (nai)</i> ([not] necessarily)

According to the modality type they encode, these adverbs are divided into four discrete groups; from the greatest to the least probability: “necessity” (abbreviated as NEC in subsequent tables and graphs), “expectation” (EXP), “conjecture” (CON), and “possibility” (POSS).

2.2 Distribution of suppositional adverbs in different types of corpora

Following Kudō (2000), the distribution of suppositional adverbs, either alone or in association with sentence final modal expressions, has been further examined in various corpora by Srdanović et al. (2008a,b), Srdanović et al. (2009), etc. The purpose was to verify Kudō’s observations, evaluate the importance of suppositional adverb - sentence-final modal expression associations for inclusion in Japanese language curricula, and genre characteristics. The results of present study are presented in Table 3, where some differences between the examined corpora are obvious already on the basis of raw frequencies (**frq** in the table) and relative frequencies (expressed as percentages) of suppositional adverbs in chosen corpora selected; for example adverbs such *tabun* (likely) being frequent in informal conversations, yet less likely to appear in other types of linguistic exchange.

Table 3. *Distribution of suppositional adverbs*

Modality type	No.	Adverb	Informal convers.			Formal interviews			CSJ lectures			Kokkai (Diet) speeches			Science textbooks			Elem. school texts		
			frq	%	dens.	frq	%	dens.	frq	%	dens.	frq	%	dens.	frq	%	dens.	frq	%	dens.
NEC	1	<i>kitto</i>	172	18	184	11	9	52	0	0	0	9	2	73	0	0	0	129	44	168
	2	<i>kanarazu</i>	42	4	45	19	16	90	0	0	0	84	15	684	56	46	86	42	14	55
	3	<i>zettai ni</i>	35	4	37	1	1	5	0	0	0	37	6	301	5	4	8	12	4	16
EXP	4	<i>osoraku</i>	9	1	10	8	7	38	1	33	41	194	33	1579	2	2	3	8	3	10
	5	<i>tabun</i>	580	61	621	50	40	236	2	67	82	72	12	586	1	1	2	10	3	13
	6	<i>sazo</i>	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	15	5	20
	7	<i>ökata</i>	0	0	0	1	1	5	0	0	0	7	1	57	0	0	0	2	1	3
	8	<i>taitei</i>	5	1	5	10	8	47	0	0	0	4	1	33	0	0	0	28	9	36
	9	<i>taigai</i>	11	1	12	1	1	5	0	0	0	2	0	16	0	0	0	0	0	0
CON	10	<i>döyara</i>	11	1	12	0	0	0	0	0	0	5	1	41	0	0	0	11	4	14
	11	<i>yobodo/ yoppado</i>	34	4	36	2	2	9	0	0	0	0	0	0	0	0	0	17	6	22
POSS	12	<i>moshika- suruto/ -shitara/ -sureba</i>	33	3	35	12	10	57	0	0	0	10	2	81	0	0	0	12	4	16
	13	<i>hyottosuruto /hyottoshitara</i>	4	0	4	1	1	5	0	0	0	3	1	24	0	0	0	3	1	4
	14	<i>kotoniyor- eba /kotoniyoruto</i>	0	0	0	0	0	0	0	0	0	0	0	1	1	2	0	0	0	
	15	<i>angai</i>	14	1	15	1	1	5	0	0	0	4	1	33	1	1	2	5	2	7
	16	<i>kanarazu shimo ... (nai)</i>	8	1	9	4	3	19	0	0	0	148	25	1205	54	45	83	1	0	1

The table above and the results of research are further elaborated by adding the density of expressions in each corpus, expressed as the observed frequency per 1.000.000 morphemes (**dens.** in the table). Both, relative frequency and density, make a comparison between different corpora or texts more direct and meaningful than raw frequencies since larger corpora include more tokens than smaller corpora. Corpora can be similar in their distribution of relative frequencies of suppositional adverbs, though their density may differ, or they can be similar in the distribution of the density of suppositional adverbs.

For example, the density of *tabun* per million morphemes is high, not only in informal conversations, but, which must be noted, also in *Kokkai* (the Diet) speeches where raw frequencies were considerably lower due to the smaller size of the corpus.

In order to gain a clearer picture of the general properties of the modal adverb distribution of corpora, the agglomeration of data across the modality type is shown in Table 4.

Table 4a. *Agglomerated distribution of suppositional/evidential adverbs across the modality type (density, i.e. frq/1.000.000 morphemes)*

Modality type	Informal convers.	Formal interviews	CSJ lectures	<i>Kokkai</i> (Diet) speeches	Science textbooks	Elem. school <i>kokugo</i> texts
NEC	266	147	0	1058	94	239
EXP	649	331	123	2271	5	82
CON	48	9	0	41	0	36
POSS	63	86	0	1343	87	28

In Table 4a, different shades of grey reflect the relative prevalence or relative absence of a certain modality type in different corpora, based on density. In informal conversations, suppositional adverbs are employed quite frequently, with prevailing types of suppositional modality being NECESSITY and EXPECTATION. On the other hand, in *Kokkai* (the Diet) speeches, suppositional adverbs are used markedly more frequently. In relation to NECESSITY, the use of EXPECTATION is comparable to that used in Informal conversations, yet in addition to this the very frequent use of POSSIBILITY has also been observed. CSJ academic lectures, as they are a monologues in public/formal settings, display, as expected, a lower frequency of the use of suppositional adverbs, with only one prevailing type of suppositional modality, i.e. EXPECTATION. Science textbooks and Elementary school *kokugo* (national language) textbooks also display complementary distributions. As expected, more formal science textbooks display a markedly lower frequency of modal adverbs than Elementary school *kokugo* textbooks, actually, the second lowest frequency among all corpora. Similar conclusions, though not identical, can be reached also on the basis relative frequencies as shown in Table 4b.

Table 4b. *Agglomerated distribution of suppositional/evidential adverbs across the modality type (relative frequencies)*

Modality type	Informal convers.	Formal interviews	CSJ lectures	<i>Kokkai</i> (Diet) speeches	Science textbooks	Elem. school <i>kokugo</i> texts
NEC	26	26	0	23	50	62
EXP	64	57	100	47	3	21
CON	5	2	0	1	0	10
POSS	5	15	0	29	47	7

These observations have led to the conclusion, that when language is considered, it is difficult to think about language in general, as the use of linguistic means greatly varies across different types of linguistic exchange. The picture that emerges - though based here only on the limited case of modal adverbs - is that, within a language, there might be different sets of “sub-languages”, used for different purposes in different linguistic exchanges. That is, a specific set of the means of expression is mobilised in order to achieve the goals for each specific linguistic exchange; thus, similar exchanges imply similar solutions, hence the similarities of suppositional adverb distributions within each homogeneous corpus. This brings us back to the notion of register in Section 1 (cf. Figures 1 and 2).

2.3 Similarities of text types: clustering of different types of text data based ON WORD-SPACE

The notion of WORD-SPACE (cf. Widdows 2004) will be illustrated here by example of a 2-dimensional modality space, spanned between two suppositional adverbs. Let us make the assumption that there are three hypothetical texts: text A asserting a fact of which the speaker/writer is absolutely sure, with only the adverb *zettai(ni)* used (absolutely); text B, in which the speaker/writer expresses facts as possibilities, with only the adverb *angai* (fairly) used; and text C, which is of a mixed nature, in which the speaker/writer both asserts clear facts - using *zettai(ni)* while stating other facts as possibilities, using *angai* in equal proportions. There is also text D, in which *angai* and *zettai(ni)* are used in proportions as in text C. If the frequencies of adverbs are provided as relative frequencies or densities, this situation can be visualised as texts represented by points in a WORD-SPACE, an abstract space of texts or sets of texts, spanned over word vectors; in our case the words are *zettai(ni)* and *angai*. The position of a text in WORD-SPACE depends on the relative frequency or density of each word vector, as shown in Figure 3 below. Text A includes only *zettai(ni)* and is represented by a point in the upper left corner, text B, including only the word vector *angai* is represented by a point in the lower right corner and texts C and D, including both *zettai(ni)* and *angai* are points in similar proportions located in the upper right corner, between texts A and B.

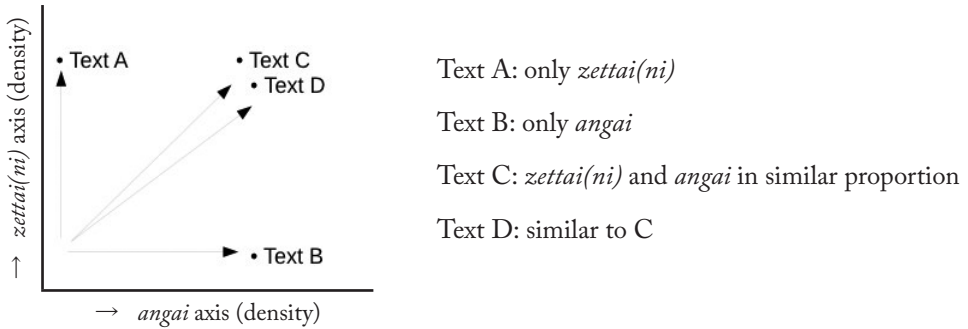


Figure 3. Text/corpus data in a toy modality adverb word space

This idea can also be applied to a more complex abstract space, including more spanning word vectors (in our case 16 suppositional adverbs, as shown in Table 1). Texts with similar proportions of defining words (expressed as word vectors), will be represented as points in close proximity to each other and texts with different proportions as points at a greater distance from each other. The similarity between texts and of course organised groups of texts, such as corpora, can be expressed as the distance between the points representing them or as the angle between text (or corpus) vectors (i.e., lines connecting text points with their origin). On the basis of this, cluster analysis can be systematically applied to group texts or sets of texts.

“Distance” and “angle” as measures of similarity, which highlight the different aspects of similarity and the resulting groupings, are not necessarily the same, as can be seen in Figures 4 and 5.

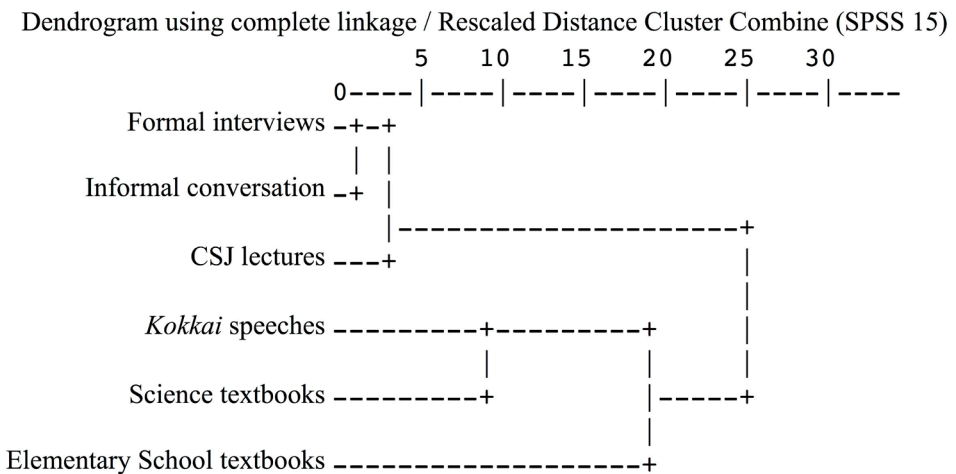


Figure 4. Comparison of corpora 1 - Euclidean “distance”

In Figure 4, the *Kokkai* (the Diet) speeches are shown as a corpus classified in a group of its own, in contrast to the group consisting of all other corpora. With distance as a measure, extremely high normalised frequencies of adverbs of all modality types come out much more strongly than the similarity of their proportions. This results in a dendrogram, where the *Kokkai* (the Diet) speeches are an outlier, at a great distance from the origin and represented with the least similarity to other corpora.

On the other hand, Informal conversations are also a kind of outlier in this cluster analysis, since they also display very high normalised frequency of *tabun*, with the second highest *kitto* being more than three times less frequent.

This similarity of texts and, consequentially, the corpora based on a comparison of the relative frequencies of suppositional adverbs provides us with rather different result, as shown in Figure 5 below.

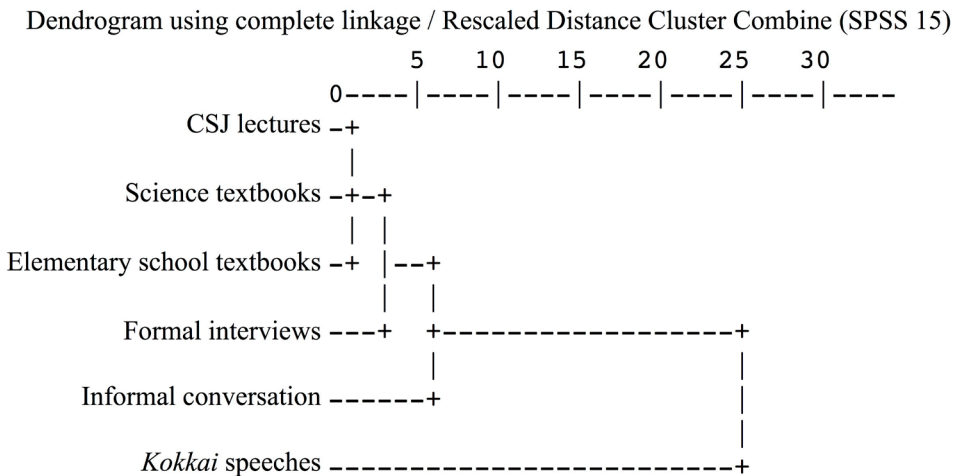


Figure 5. Comparison of corpora 2 – Similarities in relative frequencies – cosine distance

Here the spoken corpora, i.e. Formal Interviews, Informal conversations and CSJ lectures, form quite a distinct group. As can be seen in Table 4b, a similarity based on the proportional predominance of adverbs belonging to the “expected” (*tabun* in particular), are predominant in all three groups over all other modalities. In all the other three corpora *osoraku* is predominant in the category “expected”. Thus, *Kokkai* (the Diet) speeches and Science textbooks, while differing in proportion of the “expected”, share similar proportions in the category of “necessity” and “possibility”. In the second cluster, the outlier consists of a corpus of Elementary school textbooks, where only the proportion of “necessity” is clearly greater than all other modality categories.

Identical results can be obtained without “adverb based word space” interpretation, simply by comparing the correlations (i.e., statistical similarities) between the

distribution of adverbs in corpora, as shown in Table 5 below. The degree of correlation between different corpora is shown with different shades of grey. As in the cluster analysis in Figure 4, Informal conversations and Formal interviews are most closely correlated (0.91), followed by CSJ lectures (0.84 with Informal conversations and 0.83 with Formal interviews). The corpora in this cluster are thus internally closely correlated with each other.

Table 5. *Correlations of adverb distribution between corpora*

	Informal convers.	Formal interviews	CSJ lectures	<i>Kokkai</i> speeches	Science textbooks	Elem. s. texts
Informal convers.	---					
Formal interviews	0.91	---				
CSJ lectures	0.84	0.83	---			
<i>Kokkai</i> speeches	0.13	0.29	0.48	---		
Science textbooks	-0.09	0.13	-0.12	0.55	---	
Elem. s. texts	0.20	0.17	-0.11	-0.10	0.03	---

The other, more loosely correlated cluster is that of the *Kokkai* speeches and Science textbooks (0.55), corresponding to the cluster in the dendrogram in Figure 4. This cluster is loosely correlated to the CSJ lectures in first cluster (0.48). The outlier here is an elementary school *kokugo* texts corpus, which is understandable, since it is an agglomeration of various literary and expository texts and thus has a much lesser internal consistency than other corpora.

Based on the observations of results of both types of cluster analysis, as represented in Figures 4 and 5, Table 5, *Kokkai* (the Diet) speeches display the most complex use, employing all categories of suppositional modality adverbs and, thus, suppositional modality. Indeed, as can be seen in Table 3, *Kokkai* (the Diet) speeches have the highest density (freq./1.000.000 morphemes) of 3 out of 4 suppositional modality categories. It is interesting to note here that, while formal varieties of suppositional adverbs are employed more frequently, informal counterparts also display relatively high densities when compared to other corpora. Indeed, a comparison with Informal conversations shows that *Kokkai* (the Diet) speeches - though public and formal, still display a high frequency of *tabun*, an adverb that is associated with informal contexts of use.

The correlation between the uses of certain forms along the formal-informal axis is obvious, when Informal conversations and Formal interviews are compared.

Table 6a. *Formal - informal axis - osoraku and tabun*

Adverb	Informal conversations	Formal interviews
<i>osoraku</i> (probably)	10	38
<i>tabun</i> (likely)	621	236

Chi-square test: $p < 0.0005$ Table 6b. *Formal - informal axis - kitto and kanarazu*

Adverb	Informal conversations	Formal interviews
<i>kitto</i> (surely)	184	52
<i>kanarazu</i> (certainly)	45	90

Chi-square test: $p < 0.005$

As Tables 6a and 6b show, there is a clear correlation between *osoraku* as a formal choice and *tabun* as an informal one. The same is true also for the pair *kitto* and *kanarazu*: *kitto* prevails in informal conversations and *kanarazu* in the formal ones.

Table 6c. *Proportions of formal vs. informal use (osoraku vs. tabun)*

	<i>Kokkai</i> (Diet) speeches	Style
<i>osoraku</i> (probably)	1579	formal
<i>tabun</i> (likely)	586	informal

Variations between the formal and informal, observed in *Kokkai* (the Diet) speeches (Table 6c above) may be attributed to differences between individual speakers. Yet other corpora with a high degree of formality, such as Science textbooks, do not show this variation to the same extent. What may be the case here is that, though mediated as recorded texts, *Kokkai* (the Diet) speeches still reflect the dynamics of spoken communication. Speakers may be resorting to strategies, such as “characterisation” (cf. Coseriu 1973, *linguaggi d’imitazione*, i.e. imitation languages, Sadanobu 2011, *kyarakuta*), switching between formal and informal registers, in order to achieve certain goals such as attracting greater sympathy with the audience and so on.

This possibility of variation also explains the similarities and dissimilarities between *Kokkai* (the Diet) speeches and Science textbooks. What is similar in the linguistic exchanges of both of these corpora is logical argumentation. On the other hand, where scientific discourse requires a precision of argument, discourse in the Diet preferentially seeks rhetoric effects. Skilful hopping between registers may be a way to achieve this goal.

All corpora, except Elementary school *kokugo* textbooks, seem to be quite homogeneous and are close to what we intuitively understand as genre, i.e. generic purpose (see Figure 2 above). On the other hand, Elementary school *kokugo* textbooks consist of a variety of different texts, fragments of literary works, explanations and comments, and provide a picture of a much dispersed use of various suppositional modalities. Because of this lack of homogeneity, what can be stated here is that this corpus does not represent a genre in the narrow sense of the word, such as academic prose etc.

3 Conclusions

The distribution of suppositional adverbs in the corpora examined here reveals a clearly profiled, distinctive use of suppositional adverbs, according to the type of text. The corpora were internally rather homogeneous, their text type being close to what is also intuitively understood as a *genre*. The outlier *was* the corpus of Elementary school *kokugo* textbooks, because of its heterogeneous nature as a collection of various fragments of literary texts and expository prose, not correlating well with any other corpus.

Differences within similar corpora (Cluster of Informal conversations, Formal interviews, and CSJ lectures are mainly along the axis of formality, i.e. the employment of informal vs. formal items (*tabun* vs. *osoraku*, *kitto* vs. *kanarazu* etc.). On the other hand, in the more loosely correlated cluster of *Kokkai* speeches and Science textbooks there was clear distinction between a stress on rhetoric (use of informal *tabun* when expedient), in the former vs. a stress on precision of argument in the latter.

The results above quite clearly show that elements such as modal adverbs, and the suppositional adverbs among them, intrinsically connected to the type of verbal interaction, are indeed efficient discriminators of various text types. Yet further refinement is a possibility, by using other stable and transparent linguistic elements, also intrinsically connected to various types of verbal interaction. What comes to mind are:

- (i) linguistic elements directly related to types of verbal interaction such as other types of modal adverbs, utterance-final particles, etc.
- (ii) linguistic elements directly related to types of discourse structuring such as connectives (*setsuzoku hyōgen*), thematising elements (*teidai hyōgen*), etc...

On the other hand, elements whose function is predominantly referential, even though they may be relevant in different text types or genres, such as the nouns *bito*, *renchū* and *yatsu*, all referring to a person or human being, are not suitable, as their use depends not only on the type of verbal interaction involved but also on a type of extra-linguistic reality. Since there may be similar text types relating to different kinds of extra-linguistic reality, it can be expected that referential elements are less suitable as efficient discriminators of text types.

Literature

- Bakhtin, M. M. (1986) *The Problem of Speech Genres*, in *Speech Genres and Other Late Essays*. Translated by Vern W. McGee. Austin: University of Texas Press.
- Bakhtin, M. M. (1981) Discourse in the novel. In Michael Holquist (ed.) *The dialogic imagination* (translation by Caryl Emerson and Michael Holquist, of *Voprosy literaturny i estetiki*). Austin: University of Texas Press, Slavic series; No. 1, pp. 259-422 (translation of “Slovo v romane” 1934–1935, also available in original at http://www.gumer.info/bibliotek_Buks/Literat/bahtin/slov_rom.php - last accessed August 2, 2015).
- Bekeš, A. (2008a) Suppositional adverbs as indicators of discourse genre. In: Petrovčič, Mateja (ed.) *Current issues in East Asian languages, (Asian and African studies, ISSN 1408-5429, vol. 12, issue 3)*. Ljubljana: Department of Asian and African Studies, Faculty of Arts, pp. 5-16.
- Bekeš, A. (2008b) *Text and Boundary: A Sideways Glance at Textual Phenomena in Japanese*. Ljubljana: Ljubljana University Press.
- Braudel, F. (1958) Histoire et Sciences sociales: La longue durée. *Annales. Économies, Sociétés, Civilisations*. 13-4, pp. 725-753.
- Coseriu, E. (1973) *Lezioni di linguistica generale*. Editore Boringhieri.
- De Beaugrande, R. A., Dressler, W. (1981), *Introduction to Text Linguistics*. London: Longman.
- Halliday, M. A. K. (1978) *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M. A. K. (1985) *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. A. K. (1990) Some grammatical problems in scientific English. *Annual Review of Applied Linguistics* 6:13-37.
- Halliday, M. A. K. (1991) Language and context: system and instance (in 2002-2007 *Collected Works of M.A.K. Halliday* vol.9, London: Bloomsbury).
- Halliday, M. A. K. (1994) *An Introduction to Functional Grammar*. Arnold.
- Halliday, M. A. K. & Ruqaiya Hasan (1976) *Cohesion in English*. London: Longman.
- Hasan, R. (2009) The place of context in a systemic functional model. In M.A.K. Halliday and Jonathan J. Webster (eds.) *Continuum Companion to Systemic Functional Linguistics*, pp. 166-189. New York: Continuum.
- House, J. (1997). *A Model for Translation Quality Assessment*. Second edition. Tübingen: Günter Narr.
- Kudō, H. (2000) Fukushi to bun no chinjutsu no taipu (Adverbs and the type of sentence-final modality) In Nitta Yoshio and Masuoka Takashi (eds.) *Nihongo no bunpō 3 – modariti* (Japanese grammar 3: modality) pp. 161-234. Tokyo: Iwanami shoten.

- Narrog, H. (2009) *Modality in Japanese: The layered structure of the clause and hierarchies of functional categories*. Amsterdam: John Benjamins.
- Nitta, Y. & Takashi M. (1989) *Nihongo no modariti* (Japanese modality) Tokyo: Kuroshio Shuppan.
- Sadanobu, T. (2011) Kyarakuta wa bunpō o doko made kaeru ka? (How much can a 'character' change the grammar) In Kinsui Satoshi (ed.) *Yakuwari-go kenkyū no hatten* (development of the role language) Tōkyō: Kuroshio shuppan, pp. 17 - 26.
- Sano, M. (2008) Text-classification system for large scale balanced corpus: a systemic functional approach. *Proceedings of the 2008 General Meeting of the MEXT grant-in-aid for Scientific Research Priority Area Program "Japanese Corpus" (in Japanese)*.
- Srdanović Erjavec, I., Bekeš, A., Nishina, K. (2007) Cluster analysis of suppositional adverbs and clause-final modality. *Asian and African Studies*, University of Ljubljana Faculty of arts, pp. 21-31.
- Srdanović Erjavec, I., Bekeš, A., Nishina, K. (2008)a Distant collocations between suppositional adverbs and clause-final modality forms in Japanese language corpora. In: *Large-scale knowledge resources: construction and application* ; Third international conference on large-scale knowledge resources, LKR 2008, Tokyo, Japan, March 3-5, 2008; proceedings, (Lecture Notes in Computer Science, 4938), (Lecture Notes in Artificial Intelligence). Berlin: Springer, pp. 252-266.
- Srdanović Erjavec, I., Bekeš, A., Nishina, K. (2008)b Adverbs and clause-final modality collocations in various corpora. In: *Tokutei ryōiki kenkyū 'Nihongo kōpasu': Heisei 19 nendo kōkai wākushoppu (kenkyūseika hokokukai) Yokōshū : 15.3.-16. 3. Tokyo: Monbukagakusho kagakukenkyūhi tokuteiryōiki kenkyū 'Nihongo kōpasu' Sōkatsuban*, pp. 223-230.
- Srdanović I., Hodošček B., Bekeš A., Nishina K. (2009) Uebukōpasu to kensaku shisutemu o riyō shita suiryō fukushi to modariti keishiki no enkaku kyōki chūshutsu to nihongo kyōiku e no ōyō (Extracting distant collocations of adverbs and modality forms using a web corpus and a query system). *Shizen gengo shori* (Journal of Natural Language Processing) Vol. 16 No. 4.
- Srdanović I., Bekeš A., Nishina K. (2009) Kōpasu ni motozuita goi shirabasu sakusei ni mukete -- suiryō-teki fukushi to bunmatsu modariti no kyōki o chūshin ni shite (Towards the corpus based vocabulary syllabus with special reference to suppositional modality adverbs and sentence-final modality), *Nihongo kyōiku* (Japanese language education), 142 : 69-79.
- Teramura, H. (1982) *Nihongo no shintakusu to imi I*. (Japanese syntax and meaning I) Tokyo: Kurosio.
- Teramura, H. (1984) *Nihongo no shintakusu to imi II* (Japanese syntax and meaning 2). Tokyo: Kuroshio Shuppan.

- Trosborg, A. (1997) Text Typology: Register, Genre and Text Type. In Anna, Trosborg (ed.) *Text typology and translation*. Amsterdam: John Benjamins Publishing Company, pp.3-22.
- Widdows, D. (2004) *Geometry and meaning*. Stanford: CSLI.

Corpora

- NUCC Nagoya University Conversation Corpus
Ohso, Mieko (2003) *Meidai kaiwa kōpasu* (Nagoya University Conversation Corpus).
About 100 informal conversations between familiar participants. Text file size 3.5 MB.
- Formal interviews
Oikawa Terufumi (1998) *Jinbunkagaku to konpyūtā DATABASE* vol. 1, Sōgō kenkyū daigaku. Transcribed interviews by 50 Japanese native speakers. Unfamiliar participants, systematic status differences. Text file size 0.82 MB.
- CSJ conference lectures
NINJAL CSJ corpus core data (Himawari 1.02 CSJ core data, 2005)
- *Kokkai* (Diet) speeches
NINJAL Chūnagon (1990-1999)
- Science textbooks
Nishina Kikuko (unpublished) 16 University Science textbooks for basic subjects - corpus of Science texts
- Elementary school *kokugo* textbooks
Nishina Kikuko (unpublished) 60 elementary school *kokugo* textbooks. Corpus of elementary school Japanese language texts.

概要 (Abstract in Japanese)

「モダリティ表現と言語による相互作用—推量副詞による日本語の会話・文章コーパスのタイプを識別する—可能性について—」

アンドレイ・ベケシュ (リュブリャナ大学)

談話におけるモダリティ表現は、話し手・書き手から聞き手・読み手への、進行中の言葉によるやり取りがどのようなものなのかを示す合図である。そのため、談話分析においてモダリティ表現は、言葉によるやり取りのタイプ、言い換えれば、談話およびテキストのタイプ、を識別する重要な手掛かりと言える。

日本語の場合、モダリティを表す手段として、文末モダリティ表現と、いわゆる陳述副詞がある。陳述副詞は、文末モダリティ表現に比べ同定が容易である。本稿では、日本語の談話およびテキストのタイプを識別するために、陳述副詞のサブタイプである推量的副詞を用いる可能性を検討する。分析の対象として、それぞれが異なったジャンルに属する複数のコーパスを用いる。

分析の結果、推量的副詞の分布がコーパスの種類によって異なることが明らかになった。特に異なったのは小学校の国語教科書コーパスの推量的副詞の分布である。

一方、推量的副詞の分布がある程度類似している親しい仲間たちの会話コーパス、フォーマルなインタビューのコーパスおよびCSJ講義コーパスにおける相違点は、主として、フォーマル・インフォーマルの軸に沿って選択されたレジスタの違いとして現れている。

さらに、国会演説コーパスと自然科学の教科書コーパスのような、若干しか類似していないコーパスの相違点は、主として、レトリックにおける違いおよび論法の精度における違いに現れるようである。