

# Exploring the Microscopic Textual Characteristics of Japanese Prime Ministers' Diet Addresses by Measuring the Quantity and Diversity of Nouns<sup>\*</sup>

Takafumi Suzuki<sup>1</sup> and Kyo Kageura<sup>2</sup>

<sup>1</sup>Graduate School of Interdisciplinary Information Studies, University of Tokyo,  
7-3-1 Hongo, Bunkyo-ku, Japan  
qq16116@iii.u-tokyo.ac.jp

<sup>2</sup>Graduate School of Education, University of Tokyo,  
7-3-1 Hongo, Bunkyo-ku, Japan  
kyo@p.u-tokyo.ac.jp

**Abstract.** This study explores the textual characteristics, more precisely the quantity and diversity of nouns, of Japanese prime ministers' Diet addresses. In the field of stylistics, textual characteristics independent of the content have been examined with the aim on detecting the authors, genres, and chronological variations of texts. This study focuses instead on textual characteristics related to the content of texts, namely the quantity and diversity of nouns, because our aim is to analyze texts to better understand two political phenomena: (a) the difference between the two types of Diet addresses delivered by Japanese prime ministers, and (b) the perceived changes made to these addresses by two powerful prime ministers. It is a case study of the microscopic characterization of texts, which has become more and more important with the expansion in the scope of stylistics and the production of a wide variety of new types of texts following the advent of the Web.

**Keywords:** Japanese prime ministers' speeches, quantity and diversity of nouns, microscopic textual characteristics, Monte Carlo simulation

## 1. Introduction

In the field of stylistics, as computational approaches have been developed and many on-line corpora have been constructed, statistical textual characteristics have been systematically examined. These textual characteristics have traditionally been used for detecting the authors, registers, and chronological variations of texts. Recently, they have also been used for more practical applications such as spam filtering (Argamon, Whitelaw, Chase, Raj Hota, Garg, and Levitan, 2007). With this expansion in the scope of stylistics and the production of a wide variety of new types of texts especially with the growth of the Web (Aitchison and Lewis, 2003), the microscopic characterization of different types of texts has become more and more important.

Turning our eyes to the field of political science, the content of political speeches is regarded as important for analyzing the policies, attitudes and thoughts of political actors (Axelrod, 1976). Among such speeches, prime ministers' Diet addresses are recognized as the most

---

<sup>\*</sup> This study was supported by a Suntory Foundation Research Grant, 2007-2008. We would like to express our gratitude for this support. An earlier version of this study was presented at the 35<sup>th</sup> annual meeting of the Behaviormetric Society of Japan. We would like to thank the participants for their useful comments.

important material for understanding Japanese politics as they reflect Japanese governmental policies, and prime ministerial attitudes and thoughts (Watanabe, 1974; Kusano, 2005). As the role of the media and the performance of politicians increase in importance in contemporary politics (Kusano, 2006), the style of prime ministers' speeches, as well as their content, have attracted more attention (Ahrens, 2005; Azuma, 2006).

Against this background, this study explores the textual characteristics of Japanese prime ministers' Diet addresses, focusing on (a) the difference between the two types of Diet addresses and (b) the perceived changes made to these addresses by two powerful prime ministers, comparing the characteristics of their addresses with those of all prime ministers from 1945 to 2006. In order to clarify these points, we focus on the quantity and diversity of nouns, textual characteristics strongly related to the content of texts, instead of conventional content-independent stylistic characteristics, because the purpose of this study is to analyze these characteristics in order to better understand two political phenomena: the difference and changes in political content. From the point of view of computational linguistics, this study can be seen as a case study for exploring microscopic textual characteristics. The rest of this paper is organized as follows. In Section 2, we review previous work, and in Section 3, explain our two research questions in detail. In Section 4, we describe our data, and in Section 5, discuss the results. In Section 6, we make concluding remarks.

## **2. Previous work**

In the field of stylistics, many textual characteristics have been examined (Grieve, 2007). Those which have been frequently measured are the length of sentences, the length of words and the relative frequency of different parts of speech (Kenny, 1982). Lexical richness measures including those we use in this study have been systematically examined by Tweedie and Baayen (1998). Textual characteristics have been examined for formalizing the mathematical characteristics of texts (e.g., Simpson, 1949; Yule, 1944), and also for authorship attribution (Hoover, 2003) and genre-based text classification (Cortina-Borja and Chappas, 2006). They have proved to be reliable measures for classification because they take the information of function words as well as content words (García and Martín, 2007). But for our purpose, the lexical diversity of specific parts of speech need to be examined, and there are few studies examining these characteristics.

Some studies in political science examine prime ministers' Diet speeches using quantitative methods, but most studies focus only on the frequencies of specific words (e.g., Watanabe, 1974). Though Azuma (2006) examined sentence lengths and other specific functional expressions in the speeches, and Reinem (2005) used comprehensive types of content words for her analysis, the lexical quantity and diversity of Diet speeches have not been examined fully.

## **3. Research questions**

The two issues we address in this study are as follows:

### **3.1 The difference between the two types of Diet addresses**

Japanese prime ministers deliver two types of Diet addresses: *Shisei Hoshin Enzetsu* (Speech to Express Policy: SEP) and *Syoshin Hyomei Enzetsu* (Speech to Express Belief: SEB).<sup>1</sup> The SEP is delivered at the ordinary Diet session in which the budget is compiled. The SEB, on the other hand, is delivered at extraordinary Diet sessions in which the budget is not a major issue. In the SEP, prime ministers deal with many political issues related to the budgets, while in the SEB they highlight the particular political goals they wish to achieve. But some studies insist that the two types of addresses need not be distinguished when analyzing their content because they are

---

<sup>1</sup>Both of these two types of addresses are normally translated as 'general policy speech' in English.

similar (Reinem, 2007). By examining the textual characteristics of the two types of Diet addresses, we will clarify this point. It will help us to determine which (or both) types of addresses should be analyzed when we wish to investigate the political content related to the budget or content related to political goals prime ministers would especially like to achieve.

### 3.2 The perceived changes made to these addresses by two prime ministers

Nakasone, one of the strongest and most powerful prime ministers in Japanese political history, is reported to have increased the length of his Diet addresses by developing his own philosophy and ideas in order to show the concentration of political power in the office of prime minister (Shinoda, 1994). Koizumi, another powerful and strong prime minister, is famous for having used more dramatic and entertaining style of speech (Kabashima and Steel, 2007). His speech style is thought to have been characterized by the repetition of specific topics (typically privatization of the postal system). He succeeded politically partly because he was able to attract people by his way of speaking (Otake, 2006). However, these perceptions about the distinctive styles of speech of these two powerful prime ministers have never been empirically examined. We do so by measuring the textual characteristics of their Diet addresses, thus providing empirical evidences for discussion of their political styles or media strategy as revealed through their speeches.

## 4. Data

The corpora we used in this study consists of 150 Diet addresses covering the 28 tenures in office of 27 Japanese prime ministers<sup>2</sup> from 1945 to 2006.<sup>3</sup> We downloaded the addresses from the on-line database *Sekai to Nihon (The World and Japan)*.<sup>4</sup> We applied morphological analysis to the addresses using ChaSen, a Japanese morphological analysis system (Matsumoto, Kitauchi, Yamashita, Hirano, Matsuda, Takaoka, and Asahara, 2003). We extracted nouns according to part-of-speech tags assigned by ChaSen. Table 1 shows the number of addresses, the number of tokens and the number of types of the SEP and the SEB after 1953.<sup>5</sup> Table 2 sets out the names, the initials, the date of assumption to office, the number of addresses, and the total and mean number of tokens for each prime minister.

**Table 1: Basic data of the two types of Diet addresses**

	Address	Token	Type
SEP	61	120005	6056
SEB	72	91635	5374

## 5. Lexical indices

The three lexical indices we used in this study for analyzing the difference between and the changes in the addresses were number of tokens, type-token ratio and Simpson's  $D$  (Simpson, 1949) of nouns. These simple measures reflect the quantity and the diversity of political content in the Diet addresses, and can explain the two questions we proposed in Section 3.

### Number of tokens

The first index is the number of tokens, which we denote as  $N$ . It shows the number of items

<sup>2</sup> Yoshida was elected prime minister twice. We treat these as different tenures.

<sup>3</sup> The texts of the speeches are written down by secretaries in shorthand.

<sup>4</sup> <http://www.ioc.u-tokyo.ac.jp/~worldjpn/index.html>

<sup>5</sup> We used only the addresses from after 1953 in our analysis, as before that date the two types of addresses were not distinguished.

used in the sample.  $N$  of nouns reflects the quantity of political content mentioned in the addresses.

### Type-token ratio

The second index is a type-token ratio formulated as follows:

$$TTR = \frac{V(N)}{N}$$

where  $V(N)$  represents the number of types. It shows the diversity of items used in the sample.  $TTR$  of nouns reflects the diversity of political content per unit noun-tokens. A high  $TTR$  means that a variety of political content is mentioned, but the average amount of discussion devoted to each topic is limited. A low  $TTR$  indicates the opposite.

**Table 2: Basic data of Japanese prime ministers and their addresses**

Name	Initials	Date	Address	Tokens(total)	Tokens(mean)
HIGASHIKUNI Naruhiko	HN	8/1945	1	5779	5779.0
SHIDEHARA Kijuro	SK	10/1945	1	2981	2981.0
YOSHIDA Shigeru	YS1	5/1946	3	5426	1808.7
KATAYAMA Tetsu	KTe	5/1947	2	9072	4536.0
ASHIDA Hitoshi	AH	3/1948	1	3506	3506.0
YOSHIDA Shigeru	YS2	10/1948	16	26685	1667.8
HATOYAMA Ichiro	HI	12/1954	5	8318	1663.6
ISHIBASHI Tanzan	IT	12/1956	1	2665	2665.0
KISHI Nobusuke	KN	2/1957	9	18089	2009.9
IKEDA Hayato	IH	12/1960	11	35220	3201.8
SATO Eisaku	SE	11/1964	21	56731	2701.5
TANAKA Kakuei	TK	6/1972	4	15097	3774.3
MIKI Takeo	MiT	12/1974	5	18222	3644.4
FUKUDA Takeo	FT	12/1976	5	17030	3406.0
OHIRA Masayoshi	OM	12/1978	4	15629	3907.3
SUZUKI Zenko	SZ	6/1980	4	13878	3469.5
NAKASONE Yasuhiro	NY	11/1982	10	47422	4742.2
TAKESHITA Noboru	TN	11/1987	4	18448	4612.0
UNO Sosuke	US	6/1989	1	3669	3669.0
KAIFU Toshiki	KTo	8/1989	5	23541	4708.2
MIYAZAWA Kiichi	MK	11/1991	4	19110	4777.5
HOSOKAWA Morihiro	HM	8/1993	3	1460	4869.7
HATA Tsutomu	HT	4/1994	1	4185	4185.0
MURAYAMA Tomiichi	MuT	6/1994	4	20454	5113.5
HASHIMOTO Ryutaro	HR	1/1996	5	26809	5361.8
OBUCHI Keizo	OK	6/1998	5	20702	4140.4
MORI Yoshiro	MY	4/2000	4	20149	5037.3
KOIZUMI Junichiro	KJ	4/2001	11	46996	4272.4

Since the number of types and type-token ratio depend heavily on the number of tokens (Tweedie and Baayen, 1998), we adjusted it by Monte Carlo simulations when comparing: the two types of addresses; those before Nakasone, and including and after Nakasone;<sup>6</sup> and Koizumi's and every other prime minister's addresses. The process of simulation was as follows: we merged the addresses to make: the SEPs; the SEBs; those before Nakasone; those including

<sup>6</sup> Nakasone's change of Diet addresses are thought to continue after his resign, we compare the lexical indices before Nakasone, and including and after Nakasone.

and after Nakasone; and those of each prime minister. We then extracted 500 words to 6000 words incremented by 500 words by random sampling. Each extraction was made 1000 times and the mean and 95% confidence intervals were obtained.

### Simpson's $D$

The third index is Simpson's  $D$  which is formulated as follows (Simpson, 1949):

$$D = \sum_{m=1}^{V(N)} \frac{V(m, N)}{N} \frac{m-1}{N-1}$$

where  $V(m, N)$  denotes the number of tokens occurring  $m$  times.<sup>7</sup> It represents the concentration or skewness of distribution of items used in the sample. The  $D$  of nouns reflects the repeat rates of different kinds of political content.  $D$  is an index of both relative concentration, which means that it represents inequality of items in the sample, and absolute concentration, which means that it represents the absolute number of items in the sample (Yoshikane, Kageura, and Tsuji, 2003). This kind of index suits our purposes because the number of times each kind of political content is repeated is important, as is inequality among different kinds of political content when we discuss repetition in political speech. A high  $D$  means that specific political content is repeated. A low  $D$  means that each political content is explained in a balanced way.

## 6. Results and discussion

In Figure 1, the solid line represents the smoothed value of  $N$ ,<sup>8</sup> and the dotted lines represent the smoothed values of  $V(N)$  and  $TTR$ . In Figure 2, the solid line represents the smoothed value of  $D$ .  $N$  and  $D$  in each address is plotted by P (the SEP) and B (the SEB) and the mean value of  $N$  and  $D$  for each prime minister is plotted by the prime ministers' initials (see Table 2) in Figure 1 and Figure 2 respectively.

### 6.1. The different characteristics of the two types of addresses

Table 3 describes the mean values, standard deviations (s.d.) and coefficient of variances (c.v.) of  $N$ ,  $V(N)$ ,  $TTR$  and  $D$  of the SEP and the SEB. The results indicate that the SEP has a significantly larger  $N$  than the SEB, and the two types of addresses are not distinguished by  $D$ .<sup>9</sup> In Figure 3, the solid line represents the mean value of  $TTR$  of the SEP, the dashed line represents the mean value of  $TTR$  of the SEB and dotted lines represent 95% confidence intervals of  $TTR$  of both types of addresses, incremented by 500 tokens constructed by Monte Carlo simulations. The solid line and the dashed line are almost identical in the figure and are not distinguished from each other within 95% confidence intervals. The results show that the two types of addresses are not distinguished by  $TTR$ . The finding that the SEP has a significantly larger  $N$  than the SEB means that the SEP has a greater amount of political content than the SEB.

**Table 3: Comparison of the SEP and the SEB**

		$N$	$V(N)$	$V(N)/N$	$D$
SEP	mean	1738.25*	684.39*	42.57*	41.85
	s.d.	784.83*	235.12**	8.13	10.07

<sup>7</sup> As  $D$  takes all the frequency spectrum into consideration, it is completely constant to the number of tokens.

<sup>8</sup> In this section, we use notations  $N$ ,  $V(N)$ ,  $TTR$  and  $D$  as representing those of nouns.

<sup>9</sup> We used the f-test and t-test. When the  $p$ -value of f-test is less than .05 (variances are not equal), we used the Welch method of the t-test.

	c.v.	0.45	0.34	0.19	0.24
SEB	mean	1261.41*	545.90*	46.56*	42.48
	s.d.	562.82*	178.21**	8.17	8.43
	c.v.	0.45	0.33	0.18	0.20

\* significant ( $p < .01$ )  
 \*\* significant ( $p < .05$ )

The different textual characteristics typically represent the different character of the two types of Diet addresses: the SEP has a greater amount of political content because it relates to the budget compilation and the SEB has more selective political content, namely the goals prime ministers would like to achieve. The finding that the SEP is not distinguished from the SEB regarding *TTR* or *D*, indicates that prime ministers refer to a similar diversity of political content and in a similarly balanced way in the two types of addresses. It means that even if a prime minister talks about the political goals which he would especially like to achieve in the SEB, he does not change the way he speaks at the level of diversity of nouns.

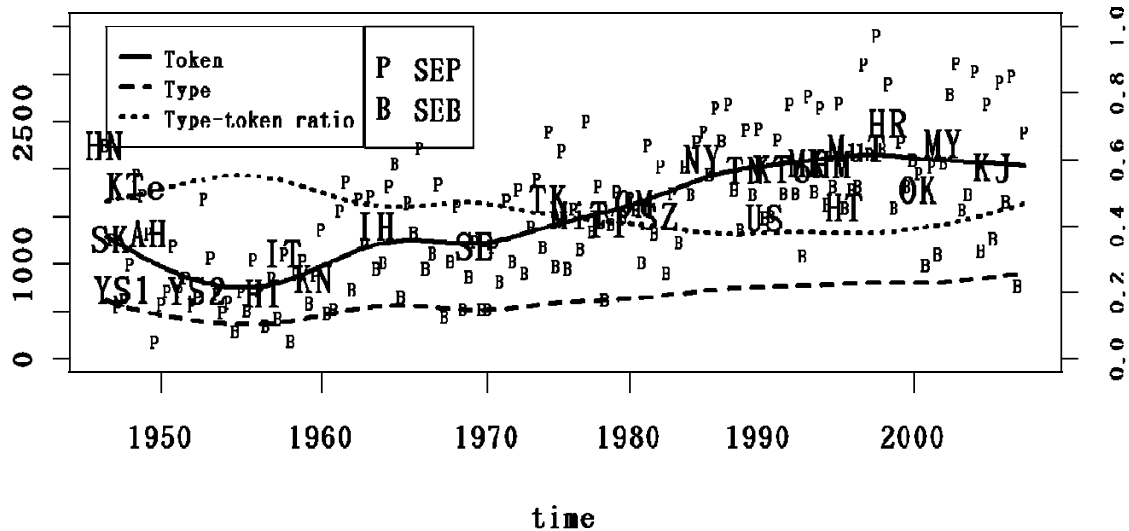


Figure 1: The  $N$  of an address is plotted by P (the SEP) and B (the SEB). The mean value of  $N$  for each prime minister is plotted by the prime minister's initial (see Table 2). The solid line represents the smoothed value of  $N$ , the dashed line represents the smoothed value of  $V(N)$ , and dotted line represents the smoothed values of  $TTR$ .

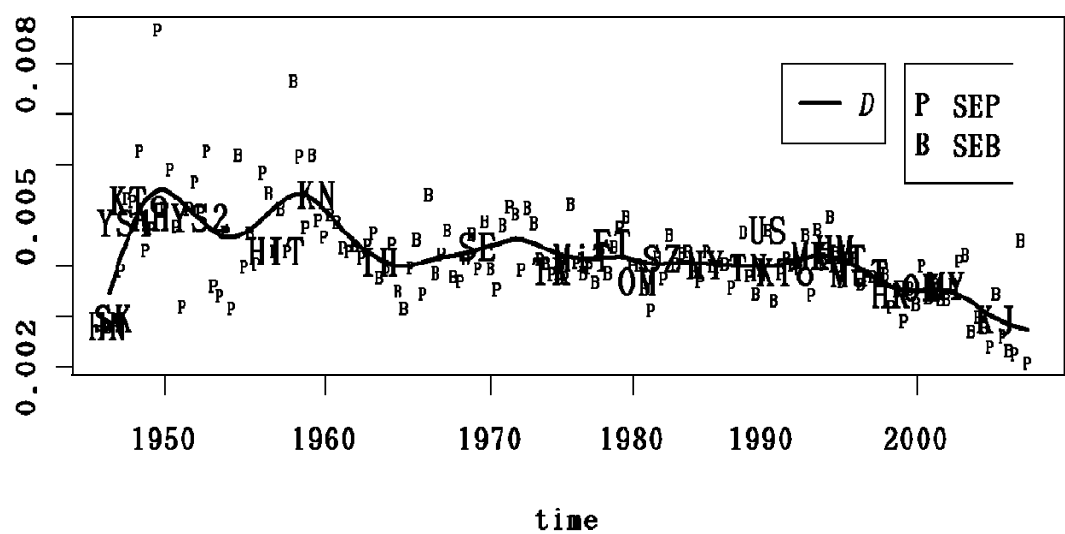


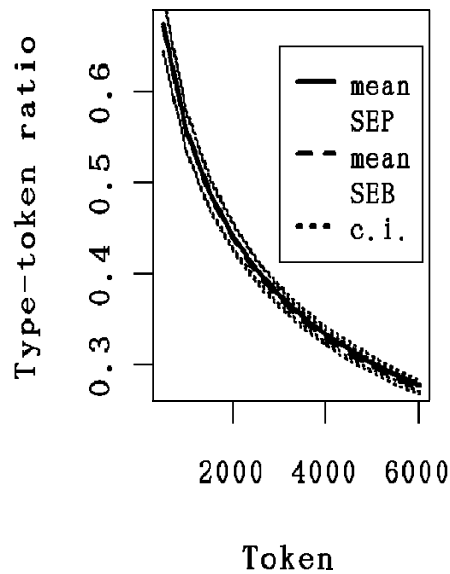
Figure 2: The  $D$  of an address is plotted by P (the SEP) and B (the SEB). The mean value of  $D$  for each prime minister is plotted by the prime minister's initial (see Table 2). The solid line represents the smoothed value of  $D$ .

**6.2. The changes made to these addresses by two prime ministers**

**6.2.1. Before Nakasone, and including and after Nakasone**

Table 4 describes the mean values, standard deviations (s.d.) and coefficients of variances (c.v.) of  $N$ ,  $V(N)$ ,  $TTR$ , and  $D$  before Nakasone, and including and after Nakasone. The results indicate that  $N$  is significantly larger including and after Nakasone than before him.<sup>10</sup> In Figure 4, the solid

<sup>10</sup>It also indicates that  $V(N)$  and  $D$  are also significantly different between the addresses before Nakasone, and including and after Nakasone. These results must be caused by Koizumi (see Section 6.2.2 and also Figure 1 and 2).



**Figure 3: The  $TTR$  per 500, 1000, ... , 6000 tokens constructed by Monte Carlo simulations. The solid line and the dashed line represent the mean values of the SEP and the SEB respectively, and the dotted lines represent the upper and lower limits of the 95% Monte Carlo confidence intervals of the SEP and the SEB.**

line represents the mean value of of  $TTR$  of the addresses before Nakasone, the dashed line represents those including and after Nakasone, and the dotted lines represent 95% confidence intervals of  $TTR$ , incremented by 500 tokens constructed by Monte Carlo simulations. Though the results show that the addresses before Nakasone, and including and after Nakasone start to be distinguished when  $N$  reaches 6000 tokens, they are rather similar to each other and the slight difference must be caused by Koizumi's special value as we discuss in Section 6.2.2. below.

The finding that  $N$  is significantly larger including and after Nakasone than before him corresponds to the argument that Nakasone increased the length of his addresses (Shinoda, 1994). This must have been driven by his intention to concentrate political power on himself. As opposed to that, the diversity of nouns does not change much before Nakasone, and including and after Nakasone. This means that he did not change the way he speaks in the addresses. He may be unconscious of how he speaks at the level of diversity of nouns, as opposed to his interest to the amount of political content in his speech.

### 6.2.2. Koizumi and others

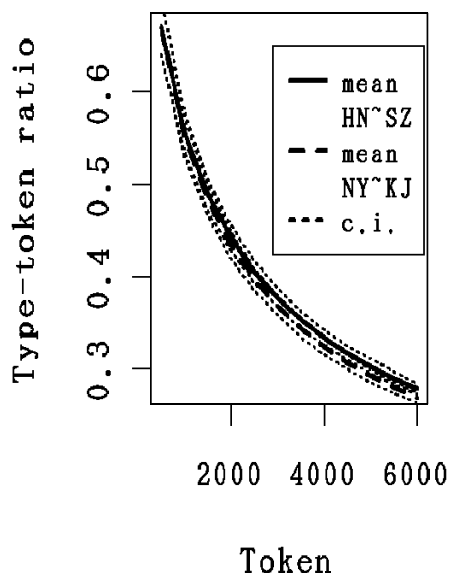
Table 5 describes the mean values, standard deviations (s.d.) and coefficients of variances (c.v.) of  $N$ ,  $V(N)$ ,  $TTR$  and  $D$  for Koizumi and others. The results indicate that Koizumi's  $D$  is significantly lower than the others, though  $N$  did not show significant difference. In Figure 5, the solid line represents Koizumi's mean value of  $TTR$  and the dotted lines represent his 95% confidence intervals of  $TTR$ , incremented by 500 tokens constructed by Monte Carlo simulations. The dashed lines represent the mean values of  $TTR$  of the other prime ministers (more than 6000 tokens) incremented by 500 tokens constructed by Monte Carlo simulations. The results indicate that Koizumi's  $TTR$  is significantly higher than the others and the highest of all prime ministers.



**Table 4: Comparison of the addresses before Nakasone, and including and after Nakasone**

		<i>N</i>	<i>V(N)</i>	<i>V(N)/N</i>	<i>D</i>
Before NY	mean	1170.27*	513.55*	47.53*	45.20*
	s.d.	560.96	177.07	8.38*	9.51*
	c.v.	0.48	0.34	0.18	0.21
Including and after NY	mean	2071.00*	790.63*	39.45*	37.17*
	s.d.	613.80*	171.89	5.44*	6.41*
	c.v.	0.30	0.22	0.14	0.17

\* significant ( $p < .01$ )  
 \*\* significant ( $p < .05$ )



**Figure 4: The *TTR* par 500, 1000, ... , 6000 tokens constructed by Monte Carlo simulations. The solid line represents the mean values of prime ministers before Nakasone. The dashed lines represent those including and after Nakasone. The dotted lines represent the upper and lower limits of the 95% Monte Carlo confidence intervals of prime ministers before Nakasone, and including and after Nakasone.**

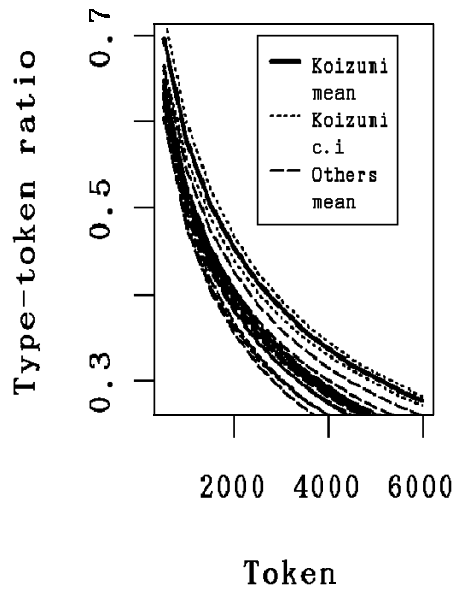
The results demonstrated that Koizumi, though he was not distinguished from the others by the number of tokens of nouns, used more diverse types of nouns and repeated those nouns less. This shows that Koizumi in fact mentioned more diverse political content in a balanced way, despite the common perception that his speaking style was characterized by the repetition of specific political content many times. One possible reason for this may be that the common perception of his speaking style is based on speeches he made outside the Diet. He may have distinguished between Diet speeches and speeches outside the Diet and also distinguished how

he speaks in them strategically.

**Table 5: Comparison of the addresses of Koizumi and others**

		$N$	$V(N)$	$V(N)/N$	$D$
Koizumi	mean	2016.64	863.64*	44.79	29.55*
	s.d.	812.30	266.79	6.02	7.97
	c.v.	0.40	0.31	0.13	0.26
Others	mean	1472.66	599.47*	44.43	43.17*
	s.d.	707.65	205.53	8.53	8.67
	c.v.	0.48	0.34	0.19	0.20

\* significant ( $p < .01$ )  
 \*\* significant ( $p < .05$ )



**Figure 5: The  $TTR$  per 500, 1000, ... , 6000 tokens constructed by Monte Carlo simulations. The upmost solid line represents the mean value of Koizumi. The dotted lines represent the upper and lower limits of the 95% confidence intervals of Koizumi. The dashed lines represent the mean values of other prime ministers.**

## 7. Conclusion

In this study, we explore the microscopic characteristics of Japanese prime ministers' Diet addresses by the quantity and diversity of nouns. Content-dependent textual characteristics tend to be avoided in stylistic text classification studies because they are not good indicators of genres or authors. By examining the quantity and diversity of nouns, we demonstrated these

types of characteristics, instead of conventional stylistic ones, enable us to interpret textual characteristics for questions in political science; namely, clarifying the difference between the two types of Diet addresses and the changes made to them by the two prime ministers.

The distributional characteristics related to the content which this study tried to explore can be called the 'form of content' dimension of textual information. This implies that conventional content analysis studies, which mainly examine the frequency of specific words, and in which the meaning of each word is taken seriously into consideration, focus on the 'meaning of content' dimension of textual information, and as opposed to that, conventional stylistics mainly investigates the 'form of the styles' dimension of textual information; the distributional characteristics of content-independent features. The dimension that has yet to be clarified is the 'meaning of the styles' of textual information. Developing a kind of computational sociolinguistics, which could become a potentially important areas of stylistics (Argamon et al. 2007) would help to shed light on this dimension. We would like to explore this dimension in further studies.

## References

- Ahrens, K. 2005. People in the State of the Union: Viewing Social Change Through the Eyes of Presidents, *Proceedings of PACLIC19, The 19th Asia-Pacific Conference on Language, Information and Computation*, pp. 43-50.
- Aitchison, J. and D. M. Lewis ed. 2003. *New Media Language*. London: Routledge.
- Argamon, S., C. Whitelaw, P. Chase, S. Raj Hota, N. Garg and S. Levitan 2007. Stylistic text classification using functional lexical features, *Journal of the American Society for Information Science and Technology*, 58(6), 802-822.
- Axelrod, R. ed. 1976. *Structure of Decision: the Cognitive Maps of Political Elites*, Princeton, New Jersey: Princeton University Press.
- Azuma, S. 2006. *Rekidai Syusyo no Gengoryoku wo Shindan Suru*. Tokyo: Kenkyu-sya.
- Cortina-Borja, M. and C. Chappas. 2006. A Stylometric Analysis of Newspapers, periodicals and newsscripts, *Journal of Quantitative Linguistics*, 13(23), 285-312.
- García, A. M. and J. C. Martín. 2007. Function words in authorship attribution studies, *Literary and Linguistic Computing*, 22(1), 49-66.
- Grieve, J. 2007. Qualitative authorship attribution: an evaluation of techniques, *Literary and Linguistic Computing*, 22(3), 251-270.
- Hoover, D. L. 2003. Another perspective on vocabulary richness, *Computers and the Humanities*, 37(2), 151-178.
- Kabashima, I. and G. Steel. 2007. How Junichiro Koizumi seized the leadership of Japan's Liberal Democratic Party, *Japanese Journal of Political Science*, 8(1), 95-114.
- Kenny, A. 1982. *The Computation of Style: an Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon Press.
- Kusano, A. 2006. *Terebi wa Seiji wo Ugokasu ka*. Tokyo: NTT publishing.
- Matsumoto, Y., A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, K. Takaoka and M. Asahara. 2003. *Japanese Morphological Analysis System ChaSen ver. 2.2.3* (chasen. naist. jp).
- Otake, H. 2006. *Koizumi Junichiro Popurizumu no Kenkyu: sono Senryaku to Syuho*. Tokyo: Toyo Keizai Shinpo-sya.
- Reinem, M. 2005. Japanese political thought as seen from prime ministers' speech: Koizumi's singularity, *Koukyo Seisaku Kenkyu*, 5, 179-198.
- Reinem, M. 2007. Content analysis of Japanese prime ministers' policy speeches: the political thought of Murayama and Koizumi, *Ronsyu Gendai Bunka / Koukyo Seisaku*, 5, 165-201.
- Shinoda, T. 1994. *Souridaijin no Kenryoku to Shido-ryoku: Yoshida Shigeru kara Murayama Tomiichi made*. Tokyo: Toyo Keizai Shinpo-sya.
- Simpson, E. H. 1949. Measurement of Diversity, *Nature*, 163, 168.
- Tweedie, F. J. and R. H. Baayen 1998. How Variable May a Constant be? Measures of Lexical

- Richness in perspective, *Computers and the Humanities*, 32, 323-352.
- Watanabe, A. 1974. Taigai ishiki ni okeru 'senzen' to 'sengo': syusyo gaisyo no gikai enzetsu no bunseki ni motozuku jakkan no kosatsu, in *Kindai Nihon no Taigai Taido*, ed. by S. Sato, and R. Dingman, pp. 225-274. Tokyo: University of Tokyo Press.
- Yoshikane, F., K. Kageura, and K. Tsuji. 2003. A method for the comparative analysis of concentration of author productivity, giving consideration to the effect of sample size dependency of statistical measures, *Journal of the American Society for Information Science and Technology*, 54(6), 521-528.
- Yule, G. U. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.