

Comparative Analyses of Textual Contents and Styles of Five Major Japanese Newspapers

Takafumi Suzuki

Faculty of Sociology,
Toyo University / 5-28-20,
Hakusan, Bunkyo-ku,
Tokyo, Japan

takafumi_s@toyo.jp

Erina Kanou

Faculty of Sociology,
Toyo University / 5-28-
20, Hakusan, Bunkyo-ku,
Tokyo, Japan

k_69en@yahoo.co.jp

Yui Arakawa

Graduate School of Library
Information and Media Studies,
University of Tsukuba / 1-2 Kasuga,
Tsukuba,
Ibaraki, Japan

s1221577@u.tsukuba.ac.jp

Abstract

Newspapers remain an important media from which people obtain a wide variety of information. In Japan, there are five major newspapers, having their own opinions and ideologies. Although these are readily recognized, they are infrequently investigated from the viewpoint of their textual characteristics. This study analyzes these differences among the five newspaper editorials. We apply morphological analysis and count the frequency of morphemes within the text data. We then apply principal component analysis and random forests classification experiments to examine their similarities and differences. Throughout these statistical analyses, we use function words and content words as features, which enables us to determine which of the two characteristics -styles or content- more powerfully affects the classification types. This study contributes to text classification studies by deliberately comparing the classification performances provided by different feature sets, function words and content words. In addition, this study will provide an empirical basis for understanding the similarities and differences among the five newspapers.

1. Introduction

Newspapers are an important media from which people obtain a wide variety of information, ranging from contemporary political and economic issues to ordinary incidents. Particularly in Japan, where newspapers delivery remains popular, many people read them in their own spaces and have continued to use them as popular information resources, even after the advent and spread of the Web. According to

Nihon Shinbun Kyokai (2012), 87.3 percent people in Japan read newspapers, which is second only to television (98.7 percent) among the five surveyed media including newspaper, television, radio, magazines, and Internet.

There are five major newspapers in Japan: Asahi, Mainichi, Nikkei, Sankei, and Yomiuri, which have publication offices in Tokyo, Osaka, and other areas, and are distributed to almost all regions of Japan. Though all of these newspapers regard correctness, neutrality and unbiased reporting as important, they have their own opinions and ideologies. According to the Hosyu (conservative)-Kakushin (liberal) image survey by Shinbun Tsushin Chosakai (2009), the five major newspapers scored as follows: Yomiuri 5.6, Sankei 5.3, Nikkei 5.2, Mainichi 5.0, and Asahi 4.4, where larger numbers indicate a newspaper perceived as more conservative and smaller numbers indicate a newspaper perceived as more liberal. Excluding Nikkei, which is a specialized newspaper that focuses on economic issues, the survey results show that people see Yomiuri and Sankei as more conservative and Mainichi and Asahi as more liberal. These differences might affect the textual characteristics of newspapers; however, they have not been investigated in a comprehensive and systematic manner.

With the development of natural language processing techniques and the creation of many online text corpora, quantitative text analysis has been expanding in scope. Such methods have begun to be recognized as important tools for solving many theoretical and practical social science research questions.

Particular to newspapers, some studies have applied these quantitative text analysis methods. For example, Newman and Block (2006) determined topics using a probabilistic mixture decomposition method with the Pennsylvania Gazette, a major colonial U.S. newspaper that

was in publication from 1728 to 1800. Higuchi (2011) investigated whether there is significant association between the content of newspaper articles and social consciousness trends by using three Japanese newspapers, Asahi, Yomiuri, and Mainichi. However, these previous newspaper text analyses focused only on the content. Examination of textual characteristics, such as styles of texts, which can reveal attitudes, personalities, psychologies, emotions, text genres, and authors (Argamon et al., 2007; Suzuki, 2009), have rarely been focused despite being intriguing aspects for analysis.

Therefore, this study analyzes the differences among editorials in the five major Japanese newspapers. Among the many types of articles, editorials are one of the most intriguing and colorful, wherein respective viewpoints are expressed (Goto, 1999), and thus are good materials for investigation. We first apply principal component analysis (PCA) to observe the overall distribution of these texts in scatter plots and investigate the factors affecting the textual characteristics. Next, we apply random forests classification experiments using newspapers, editorial dates, and ideology types as classes in order to examine the classification performance and important features of these experiments. Throughout these analyses, we use function words as well as content words as features, which is useful for investigating the similarities and differences of these classes. In addition, these features enable us to clarify which of the two characteristics-styles or content-more powerfully affects these classification types, which is also an interesting text analysis topic. This study contributes to text classification studies by deliberately comparing the classification performance provided by different feature sets, function words and content words. In addition, this study provides empirical findings useful for understanding the characteristics of the five newspapers.

2. Data and methods

2.1. Data

This study focused on the five major Japanese newspapers: Asahi, Mainichi, Nikkei, Sankei, and Yomiuri. We constructed the editorial texts using the following databases.

Yomiuri: Yomidasu Rekishikan (1874-now)

Asahi: Kikuzo II Visual for Libraries (1945-1984, pocket edition 1985-now)

Mainichi: Mainichi News Pack (1987-now)

Nikkei: Nikkei Terekon 21 (1975-now)

Sankei: The Sankei Archives (1992.9-now)

We selected two editorial dates for each newspaper, Jan. 1 and Aug. 15 from 2000 to 2010. As Jan. 1 is New Year's Day, each newspaper runs an editorial reflecting their primary opinions and interests. The Aug. 15 is the anniversary of the end of the Pacific War in Japan, and each newspaper runs an editorial reflecting their view on the war. New Year's Day editorials reflect the general vision of each newspaper and the end-of-war editorials reflect specific visions of the newspapers. In this study, we used 31 editorials from Nikkei and 22 editorials from each of the other newspapers.¹ We removed symbols, lines, and parentheses, i.e., analysis noises, and applied morphological analysis using MeCab.² We divided the morphemes into content words and function words using the tags assigned by MeCab.³ The relative frequencies of morphemes were counted; three types of text-feature matrices (bag-of-words models) were constructed using all morphemes, content words, and function words as features.

2.2. Methods

2.2.1. Principal component analysis

We applied PCA using the variances-covariance matrices constructed from three types of text-feature matrices in order to observe the distribution of the newspaper texts and to examine the factors affecting their textual characteristics.

2.2.2. Random forests

Next, we applied random forests (Breiman, 2001) for classification experiments. Random forests is an improved means of bagging (Breiman, 1996), which is an ensemble-learning method. The basic objective of ensemble learning is to improve the classification performance of previous statistical methods, i.e. decision trees in this case, by repeatedly performing the experiments and calculating the mean or majority votes of the results. However,

¹ Nikkei has two editorials in Aug. 15, and we regard them as separate ones.

² mecab.sourceforge.net

³ We regard noun-dependent, noun-pronominals, adnominals, conjunctions, particles, auxiliary verbs, signs as function words, and others as content words.

the results will always be the same when using exactly the same data. Therefore, ensemble learning methods such as bagging usually use bootstrap samplings from the original data to repeat the experiments. The main improvement in random forests over bagging is the extraction of a random subset from each bootstrapping sample that enlarges the variances in the bootstrapping samples (Breiman, 2001; Jin, 2007). Firstly, we randomly sampled i cases from the original text-feature matrix $M_{i,j}$ with replacements to create a bootstrap sample and extracted square root random subsets of j variables from the bootstrap sample to create a sample for constructing an unpruned decision tree. We used the Gini index, formalized as follows, to split the nodes.

$$Q_{\tau}(T) = \sum_{k=1}^K p_{\tau k}(1 - p_{\tau k})$$

where $p_{\tau k}$ denotes the proportion of data points in region R_{τ} assigned to class k ($k = 1, \dots, K$), which vanishes for $p_{\tau k} = 0$ and $p_{\tau k} = 1$ and has a maximum at $p_{\tau k} = 0.5$ (Bishop, 2006). These sampling, extraction, and tree-constructing processes were repeated 1,000 times, and a new classifier was constructed by a majority vote of the set of trees. When the training set for the current tree was drawn by sampling with replacements, one-third of the cases were omitted from the sample. This is referred to as the out-of-bag (OOB) data, and is used to obtain a running unbiased estimate of the classification errors as trees are added to the forest. It is also used to obtain estimates of variable importance (Breiman and Cutler, 2004). An important characteristic of random forests is that it returns variable importance (VI_{acu}) for classification experiments. To calculate variable importance, we first determined the OOB cases and counted the number of votes cast for the correct class. Next, we randomly permuted the values of the variable m in the OOB cases and placed these cases further down the tree. We subtracted the number of votes for the correct class in the variable- m -permuted OOB data from the number of votes for the correct class in the original untouched OOB data. We calculated the average of this number for all trees in the forest and determined the raw importance score for each variable. Finally, we divided the raw score by the standard error of the variable in the calculation over all trees, which is denoted as VI_{acu} for

each variable (Breiman, 2001; Breiman and Cutler, 2004). The VI_{acu} value represents the degree to which a class loses its specific character when one type of morpheme changes to another type. We used precisions, recall rates, and F1 values for evaluation (Tokunaga, 1999). As random forests use random digits for their experiments, we used the mean values from 100 experiments for these evaluation scores (Jin & Murakami, 2007; Suzuki, 2012).

We used four types of classifications, i.e., five newspaper classes, two editorial date classes, ten editorial classes (two editorial dates from the five newspapers), and three ideology types (Sankei-Yomiuri, Asahi-Mainichi, and Nikkei). We also used three types of features: all morphemes, content words, and function words. We conducted the following 12 types of experiments.

Exp. 1: five newspaper classes, all morpheme features

Exp. 2: ten classes (two editorial dates from five newspaper classes), all morpheme features

Exp. 3: two editorials date classes, all morpheme features

Exp. 4: five newspaper classes, content word features

Exp. 5: ten classes (two editorial dates from five newspaper classes), content word features

Exp. 6: two editorial date classes, content word features

Exp. 7: five newspaper classes, function word features

Exp. 8: ten classes (two editorial dates from five newspaper classes), function word features

Exp. 9: two editorial date classes, function word features

Exp. 10: three ideology type classes, all morpheme features

Exp. 11: three ideology type classes, content word features

Exp. 12: three ideology type classes, function word features

3. Results and discussion

3.1. Basic results

Table 1 shows the mean number of tokens and types of morphemes of two editorials from the five newspapers. It shows that Yomiuri has the longest editorials, which indicates that Yomiuri strives to express their opinions using New Year's editorials more ardently than others. In addition, it shows that all but Mainichi and

Nikkei have longer Jan. 1 editorials than Aug. 15 editorials. This indicates that New Year’s editorials cover more general and diverse content than the end-of-war anniversary editorials.

3.2. Principal component analysis

Figure 1-3 shows PCA scatter plots (x axis: PC1 and y axis: PC2) using all morphemes, content words, and function words, respectively, as features. Each editorial text was plotted using labels representing the five newspapers (a: Asahi, m: Mainichi, n: Nikkei, s: Sankei, Y: Yomiuri) with the editorial date (1: Jan. 1 and 8: Aug. 15). Figure 1 and Figure 3 are similar, which indicates that function words affect the PCA results more strongly than content words when we use all morphemes (simple bag-of-words) are used as features.

The results show that Yomiuri’s Jan. 1 editorials were plotted in a larger area as compared to the other editorials. When we calculated the coefficients of variances using the number of tokens in 11 texts (2000-2010) from each class, Yomiuri’s Jan. 1 editorials had the largest value (.180). The large variance in editorial length explains this PCA result.

Yomiuri, Nikkei, and Sankei’s Jan. 1 and Aug. 15 editorials were grouped respectively. The grouping of editorials from Asahi and Mainichi, for the same dates, was not clearly differentiated, suggesting that Asahi’s and Mainichi’s contents and styles were similar.

Though previous survey results (Shinbun Tsushin Chosakai, 2009) indicated the similarity between Asahi and Mainichi, and Yomiuri and Sankei, the PCA results do not show this point clearly. Instead the results indicate the differences between Yomiuri’s Jan. 1 and Nikkei’s Aug. 15 editorials.

	Number of tokens		Number of types	
	Jan. 1	Aug. 15	Jan. 1	Aug. 15
Asahi	14, 722	13, 722	5, 364	5, 015
Mainichi	13, 589	14, 313	4, 975	4, 964
Nikkei	11, 879	12, 625	4, 365	5, 296
Sankei	14, 015	11, 265	5, 249	4, 526
Yomiuri	20, 595	13, 120	6, 510	4, 424
mean	14, 960	13, 009	5, 292.6	4, 845

Table 1: Basic results

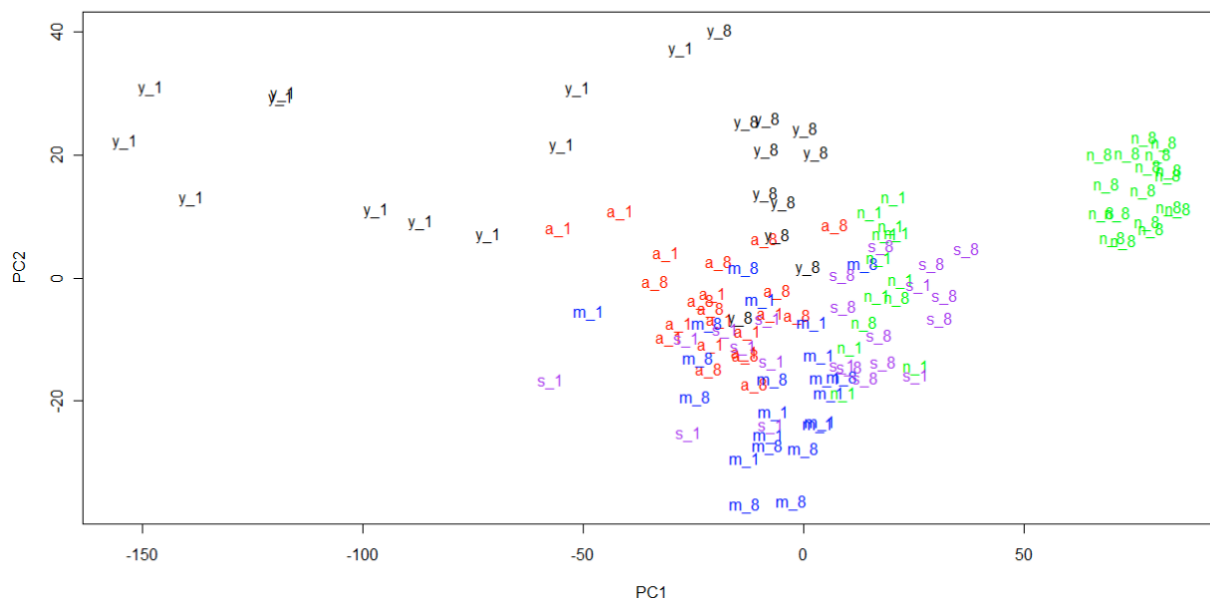


Figure 1: PCA scatter plot (all morphemes)

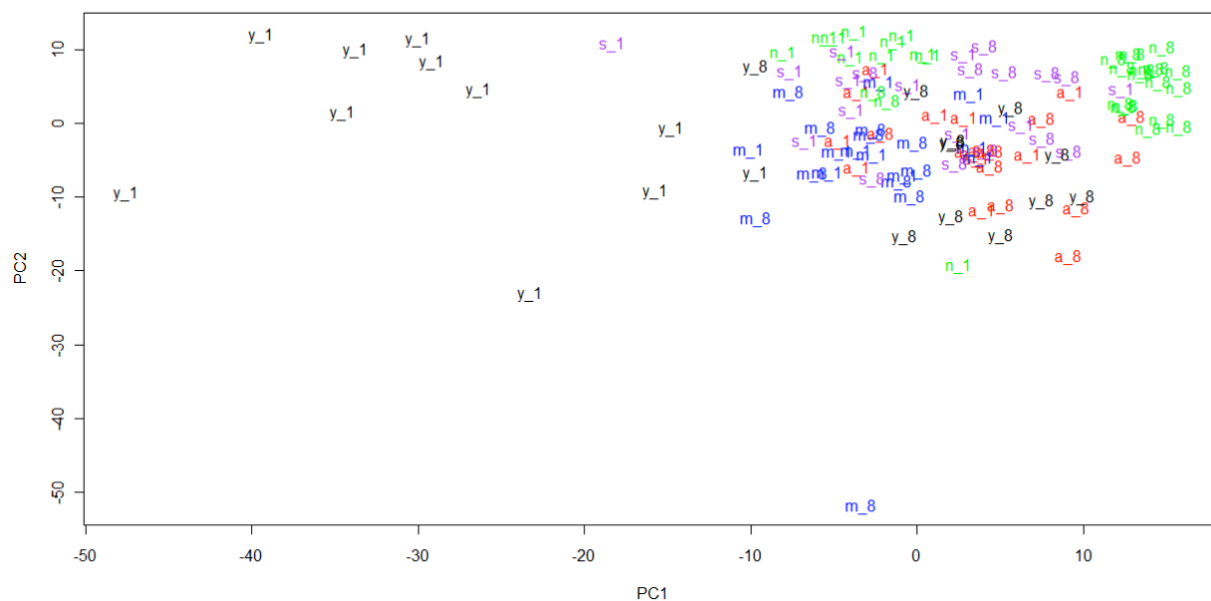


Figure 2: PCA scatter plot (content words)

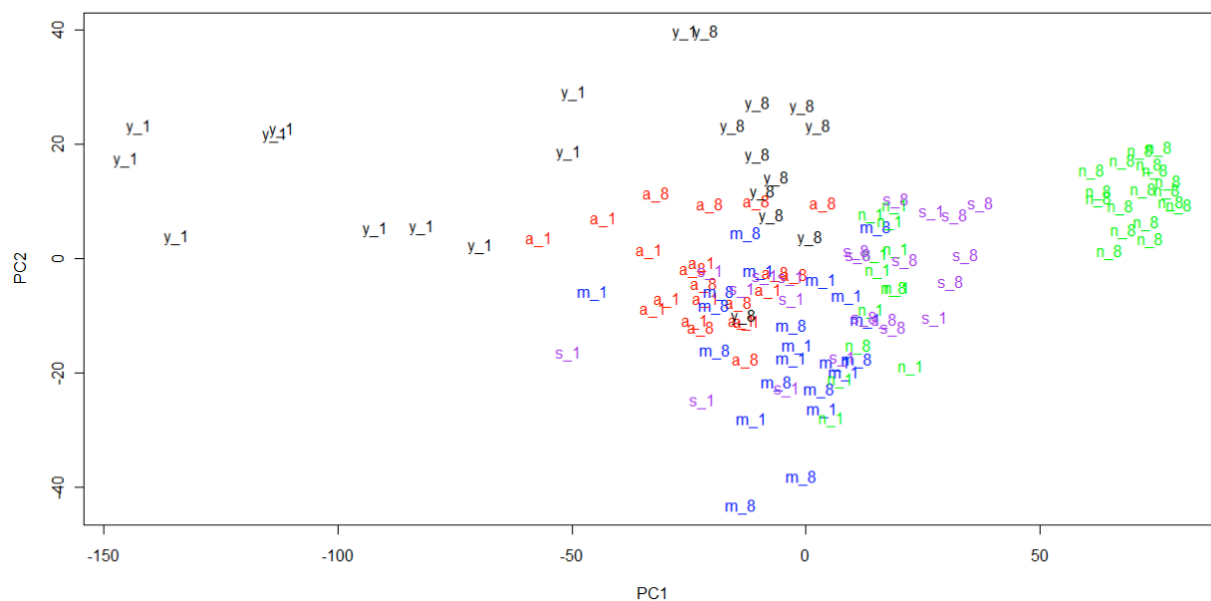


Figure 3: PCA scatter plot (function words)

3.3. Classification experiments by random forests

3.3.1. Experimental results

Table 2 presents the precision, recall rates, and F1 values given by the random forests

classification experiments.⁴ The results show that Exp. 3 (class: date, features: all morphemes) provided the best performance, probably, because the objective of Exp. 3 was to classify the minimum number of classes using the maximum number of features.

⁴ Some values are missing because they can not be calculated.

In the newspaper classification experiments (Exps. 1, 4, and 7), the results show that Exp. 7 (function words) provided better performance than Exp. 4 (content words), which indicates that the main differences among the newspapers are stylistic. This finding is consistent with that given by PCA.

In contrast, in the editorial date classification experiments (Exps. 3, 6, and 9), the results show that Exp. 6 (content words) provided better performance than Exp. 9 (function words), which indicates that content is the main difference between New Year's editorials and the end-of-war anniversary editorials.

The results of the newspaper classification experiments (Exp.1, 4, and 7) show that Nikkei and Yomiuri have special content and styles characteristics because both newspapers obtain rather high classification performances in experiments using three types of features.

The results of the editorial date experiments (Exps. 2, 5, and 8) show that Aug. 15 editorials have higher classification performance than Jan. 1 editorials for all morphemes and content word classifications. However, when using function words as the feature, the former has lower classification performance than the latter, which indicates that Aug. 15 editorials have specific end-of-war content characteristic.

The results of ideology-type classification experiments (Exps. 10, 11, and 12) show that Asahi-Mainichi types have higher classification performance than Yomiuri-Sankei types, which suggest that, with regard to content and styles, Asahi and Mainichi are more similar to each other than they are to Yomiuri and Sankei. The result is also consistent with PCA results.

	precision	recall rates	<i>FI</i>
Exp. 1	88.865	87.377	87.747
Exp. 2	68.015	67.406	66.360
Exp. 3	92.278	92.025	92.118
Exp. 4	74.369	66.029	65.877
Exp. 5	—	46.702	—
Exp. 6	89.763	88.508	88.805
Exp. 7	87.809	87.368	87.424
Exp. 8	67.520	66.963	—
Exp. 9	78.918	79.028	78.919
Exp. 10	91.199	89.377	89.842
Exp. 11	83.295	81.765	81.975
Exp. 12	92.977	91.352	91.901

Table 2: Precision, recall rates, *FI* values

3.3.2. Important classification variables

Table 3 represents the top 20 important variables contributing to classification (Exps. 1-12), with their part of speeches and variable importance values. Among the experiments using all morphemes as features (Exps. 1, 2, and 3), many function words appear in the table for Exps. 1 and 2 (including newspaper classification), while many content words appear in the table for Exp. 3, (editorial date classification). In particular, ‘戦争’ (war) and ‘終戦’ (end of war) appear as the top two variables, representing the special topics that appear in Aug. 15 editorials. These results are consistent with the classification performance given by random forests.

Among the experiments using content words (Exps. 4, 5, and 6), many war-related morphemes appear in Exps. 5 and 6, (including editorial date classification), but not in Exp. 4 (newspaper classification). In Exp. 4, content words that have general meanings or functional roles appear in the lists, which indicates that style affects classification to a greater extent than content even when content words are used as the feature.

Among the ideology-types classification experiments (Exps. 10, 11, and 12), many function words appear in Exp. 10. In Exp. 11, content words that have general meanings or functional roles appear in the list. Therefore, it is evident that the identification of ideological differences is primarily a function of style.

4. Conclusion

This study analyzed the differences among newspaper editorials, focusing on five newspapers, two editorial dates, and ideology types. We applied PCA and constructed scatter plots to observe the overall distribution of these texts and investigated the factors affecting textual characteristics. We also conducted random forests classification experiments using the newspapers, editorial dates, and ideology types as classes to examine the classification performance and identify important features. In these analyses, we used function words and content words as features. These features facilitated the investigation of similarities and differences among the classes and helped determine which of the two characteristics, styles or content, more powerfully affected the classification types.

The PCA results showed that function words affect the textual characteristics more strongly

than content words, Yomiuri's Jan. 1 editorials and Nikkei's Aug. 15 editorials had distinctive characteristics; Asahi's and Mainichi's Jan. 1 and Aug. 15 editorials had similar characteristics. The random forests results showed that function words strongly affect newspaper classification and content words strongly affect editorial date classification. Nikkei and Yomiuri had distinctive style and content characteristics. Asahi-Mainichi types were more similar to each other than Yomiuri-Sankei types.

We clarified the similarities and differences among newspapers, editorial dates and ideology types by textual characteristics. In particular, our results showed that function words had rather important roles for these classifications. This study contributes to text classification studies by deliberately comparing the classification performances determined by different feature sets, function words, and content words. In addition, this study provides empirical evidence that will increase understanding of the characteristics of the five major Japanese newspapers.

In the future, we will investigate the similarities and differences between these five newspapers using a wider variety of editorials and we will compare our results to newspapers in other countries.

Acknowledgements

This study was supported by Grant-in-Aid for Scientific Research 23700288 for Young Scientists (B), from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We would like to express our gratitude for these supports. This research includes revised and expanded content based on graduation thesis presented by Erina Kanou to the Faculty of Sociology, Toyo University. An earlier version of this study was presented at the 19th Annual Meeting of the Association for Natural Language Processing (NLP2013) at Nagoya University. We would like to thank the participants for their useful comments.

References

- Shlomo, Argamon, Casey Whitelaw, Paul Chase, Sobhan Raj Hota, Navendu Garg and Shlomo Levitan 2007. Stylistic text classification using functional lexical features, *Journal of the American Society for Information Science and Technology*, 58(6):802-822.
- Leo Breiman. 1996. Bagging predictors, *Machine Learning*, 24:123-140.
- Leo Breiman. 2001. Random forests, *Machine Learning*, 45:5-32.
- Leo Breiman and Adele Cutler. 2004. Random Forests, www.stat.berkeley.edu/~breiman/RandomForests (7 March 2013 last access).
- Masayuki Goto. 1999. *Mass Media Ron*, Yuhikaku, Tokyo.
- Koichi Higuchi. 2011. Contemporary national newspapers and social consciousness: Efficiency and limitations of newspaper content analysis, *Kodo Keiryogaku (The Japanese Journal of Behaviormetrics)*, 38 (1):1-12.
- Mingzhe Jin and Masakatsu Murakami. 2007. Authorship identification using random forests, *Proceedings of the Institute of Statistical Mathematics*, 55(2):255-268.
- Mingzhe Jin. 2007. *R ni yoru Deta Saiensu*, Morikita Shuppan, Tokyo.
- David J. Newman and Sharon Block. 2006. Probabilistic topic decomposition of an eighteenth-century American newspaper, *Journal of the American Society for Information Science and Technology*, 57(6):753-767.
- Nihon Shinbun Kyokai. 2012. '2011nen Zenkoku Media Sessyoku / Hyoka Cyosa' Hokokusyo, Nihon Shinbun Kyokai, www.pressnet.or.jp/adarc/data/rep/files/2011media.pdf (21 Feb. 2013 last access).
- Shinbun Tsushin Cyosakai. 2009. 2008 Nen Media ni Kansuru Zenkoku Seron Cyosa, Shinbun Tsushin Cyosakai, www.chosakai.gr.jp/notification/pdf/report.pdf (21 Feb. 2013 last access).
- Takafumi Suzuki. 2009. Extracting speaker-specific functional expressions from political speeches using random forests in order to investigate speakers' political styles, *Journal of the American Society for Information Science and Technology*, 60(8):1596-1606.
- Takafumi, Suzuki, Shuntaro Kawamura, Fuyuki Yoshikane, Kyo Kageura, and Akiko Aizawa. 2012. Co-occurrence-based indicators for authorship analysis, *Literary & Linguistic Computing*, 27(2):197-214.
- Tokunaga, Takenobu. 1999. *Joho Kensaku to Gengo Syori*, University of Tokyo Press, Tokyo

ranks	Exp.1			Exp.2			Exp.3		
	variables	pos	VI(acu)	variables	pos	VI(acu)	variables	pos	VI(acu)
1	、	s	0.007347	、	s	0.008540	終戦	n	0.009378
2	。	s	0.007220	て	p	0.008524	戦争	n	0.004674
3	の	p	0.006187	。	s	0.005433	で	p	0.004177
4	て	p	0.006137	の	p	0.005224	地球	n	0.004117
5	は	p	0.006038	と	p	0.004725	日	n	0.004096
6	で	p	0.005795	も	p	0.004528	は	p	0.004028
7	を	p	0.004759	う	av	0.003771	戦没	n	0.004005
8	も	p	0.004098	1	n	0.003551	経済	n	0.003615
9	に	p	0.003836	だ	av	0.003433	。	s	0.003399
10	と	p	0.003699	として	p	0.003427	ある	av	0.003329
11	経済	n	0.003564	を	p	0.00319	世界	n	0.003168
12	が	p	0.003390	は	p	0.003181	が	p	0.003069
13	ない	av	0.003262	た	p	0.003152	な	av	0.002888
14	た	av	0.002878	で	p	0.003145	8月	n	0.002810
15	だ	av	0.002861	いる	v	0.003133	化	n	0.002605
16	終戦	n	0.002749	“	s	0.003120	の	p	0.002493
17	う	av	0.002469	”	s	0.003070	記念	n	0.002491
18	「	s	0.002346	いわゆる	adn	0.002907	財政	n	0.002308
19	」	s	0.002284	が	p	0.002902	追悼	n	0.002306
20	な	av	0.002072	他方	n	0.002748	を	p	0.002205

ranks	Exp.4			Exp.5			Exp.6		
	variables	pos	VI(acu)	variables	pos	VI(acu)	variables	pos	VI(acu)
1	1	n	0.005865	経済	n	0.004497	終戦	n	0.009168
2	いる	v	0.004690	し	v	0.003920	地球	n	0.005378
3	し	v	0.004178	1	n	0.002834	戦争	n	0.004735
4	0	n	0.004095	いる	v	0.002770	経済	n	0.004725
5	られ	v	0.003961	日本	n	0.002714	世界	n	0.004338
6	他方	n	0.003723	必要	n	0.002568	戦没	n	0.003815
7	平成	n	0.003223	終戦	n	0.002556	化	n	0.003486
8	れ	v	0.002953	的	n	0.002298	財政	n	0.003066
9	さ	v	0.002820	世界	n	0.002291	ある	av	0.002929
10	9	n	0.002459	する	v	0.002206	8月	n	0.002735
11	する	v	0.002450	平成	n	0.001967	必要	n	0.002705
12	日本人	n	0.002253	追悼	n	0.001876	先進	n	0.002574
13	なら	v	0.002251	れ	v	0.001829	改革	n	0.002548
14	九	n	0.002181	0	n	0.001805	記念	n	0.002443
15	2	n	0.002077	戦争	n	0.001775	危機	n	0.002408
16	軍事	n	0.001720	られ	v	0.001708	企業	n	0.002292
17	現実	n	0.001692	級	n	0.001640	する	v	0.002096
18	国民	n	0.001548	さ	v	0.001633	日	n	0.001989
19	必要	n	0.001547	地球	n	0.001633	追悼	n	0.001978
20	6	n	0.001512	他方	n	0.001587	成長	n	0.001965

ranks	Exp.7			Exp.8			Exp.9		
	variables	pos	VI(acu)	variables	pos	VI(acu)	variables	pos	VI(acu)
1	、	s	0.027852	、	s	0.022828	日	n	0.023632
2	て	p	0.025036	。	s	0.019897	は	p	0.012206
3	。	s	0.014986	て	p	0.016021	で	p	0.011571
4	の	p	0.012213	の	p	0.015435	。	s	0.010696
5	も	p	0.011241	は	p	0.015083	が	p	0.009095
6	と	p	0.011084	で	p	0.012908	な	av	0.008524
7	”	s	0.009826	を	p	0.010720	あの	adn	0.008178
8	“	s	0.009797	も	p	0.008508	を	p	0.006522
9	だ	av	0.009260	に	p	0.008464	の	p	0.006168
10	いわゆる	adn	0.009048	と	p	0.008185	から	p	0.005435
11	として	p	0.008002	が	p	0.008161	だ	av	0.004824
12	そんな	adn	0.007842	だ	av	0.007798	、	s	0.004327
13	う	av	0.007481	た	p	0.007187	に	p	0.004008
14	が	p	0.007178	日	n	0.006392	ば	p	0.003672
15	は	p	0.006455	な	av	0.005604	なく	av	0.003326
16	た	p	0.006170	A	s	0.005494	ある	av	0.003173
17	を	p	0.005930	いわゆる	adn	0.004757	この	adn	0.002610
18	で	p	0.005830	ない	av	0.004568	ない	av	0.002530
19	「	s	0.005461	「	s	0.004521	A	s	0.002420
20	」	s	0.005459	“	s	0.004468	た	p	0.002257

ranks	Exp.10			Exp.11			Exp.12		
	variables	pos	VI(acu)	variables	pos	VI(acu)	variables	pos	VI(acu)
1	て	p	0.008789	1	n	0.007746	て	p	0.025149
2	。	s	0.006295	0	n	0.005372	。	s	0.017149
3	の	p	0.006063	し	v	0.004070	だ	av	0.014567
4	と	p	0.005060	九	n	0.003963	の	p	0.013545
5	1	n	0.005036	られ	v	0.003553	、	s	0.011640
6	だ	av	0.004911	十	n	0.003462	として	p	0.011371
7	、	s	0.004875	いる	v	0.003348	と	p	0.011303
8	として	p	0.004566	れ	v	0.003225	も	p	0.010391
9	も	p	0.004481	さ	v	0.003062	“	s	0.008768
10	う	av	0.004167	2	n	0.003005	”	s	0.008683
11	を	p	0.003919	五	n	0.002776	が	p	0.008508
12	は	p	0.003807	9	n	0.002385	う	av	0.007858
13	が	p	0.003738	国際	n	0.002329	を	p	0.007856
14	で	p	0.003572	たち	n	0.002291	いわゆる	adn	0.007799
15	九	n	0.003441	なら	v	0.002270	私	n	0.007716
16	0	n	0.003137	6	n	0.002038	は	p	0.007055
17	ない	av	0.003039	日本	n	0.002030	で	p	0.006175
18	”	s	0.003034	三	n	0.001901	こと	n	0.005993
19	いわゆる	adn	0.003024	現実	n	0.001824	ある	av	0.005894
20	こと	n	0.002916	国民	n	0.001760	た	p	0.005048

Table 3: Important variables