Semantic and Word Image Change in Shizuoka Dialect Mamettai −Analyzed by Multinomial Distribution Type Regime Switching Detection

Joy TANIGUCHI [*], Kaisei MINEO[*], Takumi MONDO* and Yuki YAMAGISHI[**]

ABSTRACT

This study examines how the meaning and word image of the Shizuoka dialect mamettai have changed in the past by conducting a large-scale survey on the use of the word and analyzing the results using multinomial distribution type regime-switching detection and multi-category appearance order statistics. The results showed that, in terms of semantic change, the use of mamettai in the sense of "physically hard-working," declined, with this trend was particularly conspicuous among those born in the mid-1980s and later. On the contrary the word's use in the sense of "methodical" has increased. It is suggested that these semantic changes are influenced by standard Japanese. In addition, positive word images were found to be on the decline, while the word images were inclined toward the negative ones. This trend was particularly true for those born in the mid-1970s and later.

INTRODUCTION

Sociolinguistics focus on the correlation between language variation and social factors. Among them, age or generation of the speaker is a major concern. This study examines how the meaning and word image of the word, *mamettai*, have changed using the concept of *Apparent Time Hypothesis* (Chambers and Trudgill, 1988). It is a major analytical method to capture ongoing change and has been used in quantitative sociolinguistics for over 40 years (Bailey, et al., 1991). The concept assumes that differences in language use between generations reflect diachronic changes in language.

The word *mamettai*, used in the Shizuoka dialect, has several meanings. According to Tojo (1951), *mamettai* means (1) to be physically hard-working and (2) to be healthy. In this study, we conducted a pilot survey[2] of native Shizuoka-dialect speakers, which suggested that, in addition to the above two, the dialect is also used in the meanings of (3) restless, mainly used to describe children, and (4) methodical and thorough to the smallest detail. The word, *mamettai*, was selected for survey for the following two reasons. First, this word has multiple meanings, making it suitable for visualizing semantic changes using a new method based on time series data. The second reason is that there is a word with a similar form in Standard Japanese, and we assumed that it would be possible to capture its influence in terms of standardization.

In this study, our objective is to explore ongoing semantic changes in *mamettai* in Shizuoka dialect. We argue two hypotheses based on short sentences written by participants using the target word, *mamettai*.

(1) The semantic change in *mamettai* reflects a diachronic change. Specifically, we suggest that the meaning of *mamettai* is shifting from "physically hard-working" to "methodical."

(2) The word image of *mamettai* has changed from positive to negative, which is an effect of Standard Japanese.

METHODS

Participants

Table 1 describes the survey participants.

Table 1 Number of participants by region and age group

Age	West	Central	Ikawa	East	Total
10s	22	30	4	27	83
20s	111	182	2	117	412
30s	55	140	3	71	269
40s	55	208	1	71	335
50s	40	184	3	44	271
60s	15	93	5	6	119
Over 70s	4	25	23	3	55
Total	302	862	41[2]	339	1,544

The figure shows that participants in the

Received April 1, 2024
[*]Department of Information Design, Faculty of Informatics, [**] Department of Computer System, Faculty of Informatics

prefecture's eastern and western parts were mostly young, whereas those in the prefecture's central part were born between the 1950s and 1970s[1].

Figure 1 shows the age and birth region of the survey participants using multicategory appearance order statistics (see below for details).

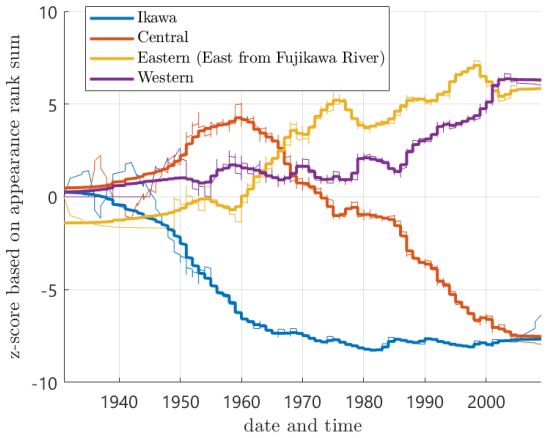

Figure 1　The age and birth region of the participant

Survey

Using a snowball sampling through social networking services, the 1,544 participants responded to a paper- or web-based questionnaire on their gender, year of birth, birthplace, and whether their family used dialects in their households. For birthplace, based on Nakajo (1982), four regions (Eastern, Central, Western, and Ikawa) were selected as options while showing a map. Regarding the use and understanding of *mamettai*, four options were provided: "use it," "do not use it but understand it," "have heard of it but do not understand it," and "have never heard of it". A total of 1,188 participants who answered "have heard of it but do not understand it" or "have never heard of it" in this question will be directed to the end of the questionnaire. On the contrary, 855 participants who answered that they used or understood the target word, *mamettai* were asked to write a short sentence using *mamettai*. We categorized short sentences into the following four categories:

(1) My father works hard in vegetable fields every day, and he is a truly *mamettai* (hard-working).

(2) My grandfather is 90 years old, but he never gets sick and is *mamettai* (=healthy).

(3) She is always moving around and is *mamettai* (restless).

(4) My father is *mamettai* (methodical, does things properly, down to the smallest detail), so he maintains a record of all the money he spends.

Analysis Method

The periodic time-series data used in the present study are complex in terms of their changes, and the amount of information they contain is relatively large, making it difficult to interpret, even if they are visualized. Therefore, this study used the following two analytical methods to simplify the data and estimate the phenomena that occurred in the past and timing of the change.

Multinomial Distribution Type Regime Switching Detection

Multinomial distribution type regime-switching detection is a method for simplifying complex data by generating timelines. In our study, we use multinomial regime-switching detection as a data simplification method to estimate the semantic changes that have occurred in the past and the timing of the change in the Shizuoka dialect *mamettai* as we assume that the data follows a multinomial distribution. The method uses a model selection criterion called the minimum description length, which considers the minimization of code length when data are compressed and transmitted using the model.

While Kleinberg (2002), an existing similar method, specializes in visualizing trend changes of single pieces of information, this method proposed by Yamagishi and Saito (2017) is based on the assumption that multiple pieces of information are handled, and it is shown to be easy to estimate the time when a change occurs.

The data in each regime are assumed to follow a multinomial distribution, and the model parameters and switching time are estimated by maximizing their likelihood. The procedure involves (1) formulating the problem of detecting regime switching, (2) generating a timeline of time-series data based on estimated regime switching, and (3) simplifying and visualizing the data.

Mathematical Expression of Multinomial Regime Switching

Let $\mathcal{D} = \{(s_1, t_1), \cdots, (s_N, t_N)\}$ be a set of answers from the participants, where s and t denotes the answers with J-category and the n-th year of birth, respectively, and $|\mathcal{D}| = N$ indicates the number of answers from the participants. Below, the number n denotes the time step, and let $\mathcal{N} = \{1,2,\cdots,N\}$ be the set of time steps. We express the starting time step of the k-th switching regime as $T_k \in \mathcal{N}$, and let $\mathcal{T}_K = \{T_0, \cdots, T_k, \cdots, T_{K+1}\}$ be the set of switching time steps, where we set $T_0 = 1$ and $T_{K+1} = N + 1$ for convenience. Namely, $T_1, \cdots, T_K$ are the individual

switching time steps to be estimated, where we assume $T_k < T_{k+1}$. Let $\mathcal{N}_k$ be the set of time steps for k-th regime defined by $\mathcal{N}_k = \{n \in \mathcal{N}; T_k \leq n < T_{K+1}\}$ for each $k \in \{0, \cdots, K\}$, where $\mathcal{N} = \mathcal{N}_0 \cup \cdots \cup \mathcal{N}_K$.

We assume that each answer distribution of the regime is modelled by a J-category multinomial distribution. Let $\mathcal{p}_k = \{p_{k,1}, \cdots, p_{k,J}\}$ be the probability vector for the k-th multinomial distribution, and $\mathcal{P}_K = \{\mathcal{p}_0, \cdots, \mathcal{p}_K\}$ be the set of the probability vectors. Then, given $\mathcal{T}_K$, we can define the log-likelihood function as follows:

$$L(\mathcal{D}; \mathcal{P}_K, \mathcal{T}_K) = \sum_{k=0}^{K} \sum_{n \in \mathcal{N}_k} \sum_{j=1}^{J} s_{n,j} \log p_{k,j}, \qquad (1)$$

where $s_{n,j}$ denotes the dummy variables for $s_n \in \{1, \cdots, J\}$ defined by $s_{n,j} = 1$ if $s_n = j$; 0 otherwise. Then, the maximum likelihood estimators of Equation (1) are given by $\hat{p}_{k,j} = \sum_{n \in \mathcal{N}_k} s_{n,j} / |\mathcal{N}_k|$ for $k = 0, \cdots, K$ and $j = 1, \cdots, J$. Therefore, the problem of detecting switching time steps is reduced to the problem of finding the set $\mathcal{T}_K$ that maximizes Equation (1), and we employ a detection method which combines a greedy search and a local search proposed by Yamagishi et al. (2014, 2017, 2018a). In the experiments, we determined an adequate number of switching time steps by introducing the minimum description length (MDL) with degree of freedom $K(J-1)$.

Multi-category Appearance Order Statistics

To explicitly identify time-series changes in data categories and compare them across multiple categories, we transformed the data using statistics with appearance ranks (Yamagishi et al., 2018b). This method is based on Mann–Whitney's U-test, extended to handle multiple groups, and expresses the trend change in the frequency of occurrence of data as a z-score. It can be assumed that long-term changes can be easily captured because the indexes are visualized from a dynamic viewpoint. In addition, because the z-score of each category is based on all other categories, it is easy to compare multiple categories. Burst detection by Kleinberg (2002) is a typical conversion method; however, it is not suitable for continuous trend analysis and comparison among multiple categories. This method can capture long-term trend changes and is suitable for arbitrary multicategory comparisons.

Mathematical Expression of Multi-category Appearance Order Statistics

We resume the problem settings in our regime switching model and consider a J-by-N matrix $Q(q_{j,n} \in \{0,1\})$, where $q_{j,n} = 1$ if $s_n = j$; and 0 otherwise. We can then define the number of appearances of category $j \in J$ in the time steps $\{1, \cdots, n\}$ by $I_{j,n} = \sum_{i=1}^{n} q_{j,i}$. We now extend the U test by Mann and Whitney (1947) and apply the z-score to the set of time steps $\{n \in \mathcal{N}: q_{j,n} = 1\}$ and $\{n \in \mathcal{N}: q_{j,n} = 0\}$ as follows:

$$z_{j,n} = \frac{u_{j,n} - \mu_{j,n}}{\sigma_{j,n}}, \qquad (2)$$

where the U statistics $u_{j,n}$, means of the appearance rank sum $\mu_{j,n}$, and the variances $\sigma_{j,n}^2$ are calculated as follows:

$$u_{j,n} = \sum_{i=1}^{n} q_{j,i} - \frac{I_{j,n}(I_{j,n}+1)}{2},$$

$$\mu = \frac{I_{j,n}(n - I_{j,n})}{2},$$

$$\sigma_{j,n}^2 = \frac{I_{j,n}(n - I_{j,n})(n+1)}{12}.$$

Therefore, we can evaluate Equation (2) with larger (smaller) $z_{j,n}$ being significantly higher (newer) or lower (older) in the appearance rank sum of category $j \in J$ in the time steps $\{1, \cdots, n\}$. For all N answers, we can reduce the computational complexity to $O(NJ)$ with the algorithm proposed by Yamagishi et al. (2018b).

SEMANTIC CHANGE IN SHORT SENTENCES

In this section, we examine the semantic change in *mamettai* based on short sentences written by participants. Of the 1,188 respondents who understood and used the expression *mamettai*, 855 wrote a sentence. Sentences such as "My child who started walking is *mamettai* (=restless) and I cannot take my eyes off him (born in 1971, in the central part of the prefecture)" were classified as meaning (3) "restless". Sentences such as "That person is *mamettai*" were assumed to be "unclassifiable" because of the semantic classification impossibility.

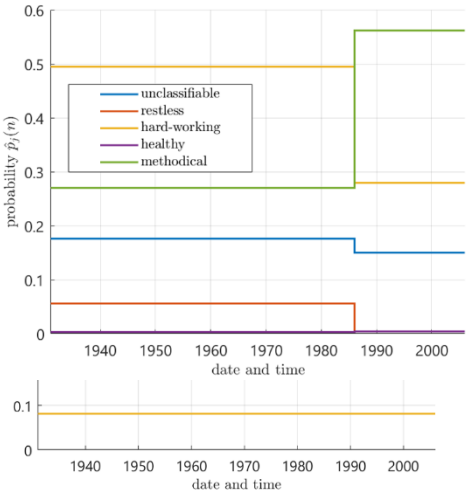

Figure 2 Semantic change by year of birth (time step)

As can be seen in Figure 2, it was observed that the use of the main meaning of *mamettai*, "physically-hard-working" is on a downward trend, while the meaning of "methodical" or "attentive to details" is expanding. The results of the analysis in Figure 3 also shows that this change has become more pronounced since the mid-1970s (the participants' year of birth).

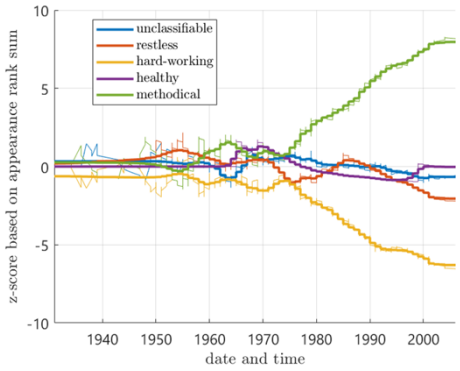

Figure 3 Semantic change by year of birth

Mamettai is a dialect derived from the word *mame*, which in Standard Japanese (hereafter, SJ) means hard-working or diligent. According to the dictionary of the Japanese language, the word *mame* means (1) to be hard-working, (2) to be physically healthy[3]. The meaning of (1) is similar to that of *mamettai* in the Shizuoka dialect; however, based on our pilot study, it also includes the meaning of physical activity. Therefore, the sentence "That person is *mamettai* because he keeps a household accounting book every day" made some participants uncomfortable.

WORD IMAGE CHANGE BASED ON SHORT SENTENCES

In this section, we discuss the results of categorizing the meaning of *mamettai* in the short sentences written by the participants as positive, negative, or neither (unclassifiable). For example, "Your husband is *mamettai* (=hard-working) and takes good care of his children" (born in 1969, in the central part of the prefecture) was classified as "positive," while those such as "Your husband is so *mamettai* (=methodical) that he cares about such details (born in 1981, in the central part of the prefecture)" were classified as "negative". Also, the sentence "Your mother was a very *mamettai* person" (born in 1970, in the western part of the prefecture) was classified as "neither" or "unclassifiable."

Figure 4 shows the time steps related to the word image of *mamettai* based on the participants' year of birth. In the present model selection criteria, it was reasonable to interpret the data as unchanged. This indicates that even if the data were separated at some point, we cannot obtain sufficient amount of information, and it is assumed that there is no difference between generations in terms the image of the word *mamettai*.

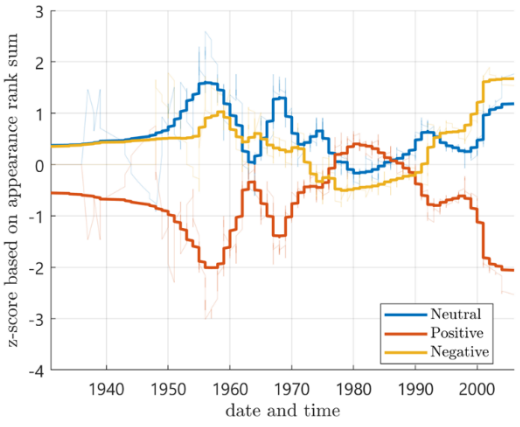

Figure 4 Word image change by year of birth (time step)

Figure 5 shows more detailed data visualization with multi-category appearance order statistics. The analysis revealed that from those born in the mid-1970s to in 1990, *mamettai* was used more frequently in a positive sense, and then it was used again in a negative context. Specifically, it was observed that the younger generation born after 2000 had a negative image of *mamettai*.

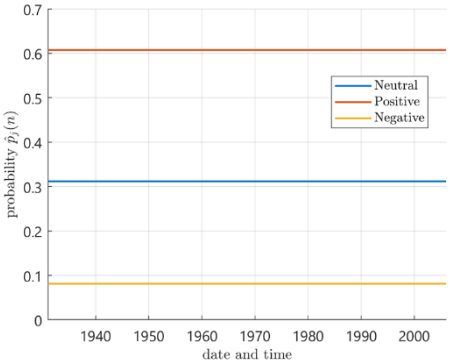

Figure 5 Word image change by year of birth

CONCLUSION

In this study, we attempted to simplify and visualize the data on Shizuoka dialect *mamettai* semantic change using periodic time series data by means of multinomial distribution-type regime switching detection and multi-category appearance order statistics. In the mid-1980s, a significant change in the data structure was observed in the use and understanding of *mamettai* as a result of the estimation of the time when the semantic change of a word occurred. In addition, the observed continuous phenomenon of the z-score, particularly for those born after 1970, and the fact that the *p*-value approximates 0 makes it difficult to conclude that this trend is coincidental. Therefore, it can be inferred that the decrease in the use and understanding of the word in the sense of "physically hard-working" and the increase in the use and understanding of the word in the sense of "being methodical" will continue in the future.

There was no alteration in the data structure for word images, but it was observed that positive word images have been declining and negative images have been increasing since around 1990.

Although the data in this study were difficult to analyze from a general visualization, the methods used made it possible to estimate the timing of the word's semantic change. Further use of these analytical methods in social-linguistic studies is required.

NOTE

1. The majority of participants in Ikawa are older, which is related to the high aging rate of 62.37% in the region.
2. In our pilot survey, we interviewed 70 Shizuoka dialect speakers of various ages to ascertain under what circumstances the target word is used.

3, The word, *mame*, also means (3) to be serious, and (4) to be practical, however the meanings of (3) and (4) have already disappeared in modern SJ.

REFFERENCES

1) Chambers, Jack K, Trudgill, Peter: Dialectology, 2nd ed., Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press. 2019.
2) Bailey, Guy, Tom Wikle, Jan Tillery, and Lori Sand: The Apparent Time Construct. Language Variation and Change 3(3): pp. 241–64. 1991.
3) Mann and Whitney: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other, Ann. Math. Statist., Vol. 18, No. 1, pp. 50–60 (1947)
4) Misao Tojo (Ed.) Japan National Dialect Dictionary, Tokyodo Publishing, Tokyo, 1951.
5) Osamu Nakajo: Research in Shizuoka dialect, Yoshimishoten, Shizuoka, 1982.
6) Jon Kleinberg: Bursty and hierarchical structure in streams. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002), pp. 91–101, 2002.
7) Yuki Yamagishi, Seiya Okubo, Kazumi Saito, Kouzou Ohara, Masahiro Kimura, Hiroshi Motoda: A method to divide stream data of scores over review sites. PRICAI, pp. 913–919, 2014.
8) Yuki Yamagishi, Kazumi Saito: Visualizing switching regimes based on multinomial distribution in buzz marketing sites. In: Kryszkiewicz, M., Appice, A., Ślęzak, D., Rybinski, H., Skowron, A., Raś, Z. (eds) Foundations of intelligent systems. ISMIS 2017, pp. 385–395, 2017.
9) Yuki Yamagishi, Kiyoto Iwasaki, and Kazumi Saito: Simplification of periodic time series data by multinomial distribution-type regime switching detection, Journal of Information Processing, 10, pp. 1-7, 2018a.
10) Yuki Yamagishi, Kiyoto Iwasaki, and Kazumi Saito: Converting stream data based on multi-category appearance order statistics, Transactions on mathematical modeling and its applications, 11(1), pp. 45-52. 2018b.

This research is supported by the Group Research Project of Shizuoka Institute of Science and Technology.
本研究は、静岡理工科大学「グループ研究推進支援費」の助成を受けています。