Comparison of onomatopoeia use among areas using Diet records of Japan

Sachiko Hirata-Mogi† Satoshi Nakamura‡ Takanori Komatsu‡ Kimi Akita*

†School of Medicine, the University of Tokyo, Japan
‡Department of Frontier Media Science, Meiji University, Japan
*Graduate School of Language and Culture, Osaka University, Japan
marshmallow1214@gmail.com

Abstract

Onomatopoeia is believed to be particularly frequent in some regions of Japan. In this study, we examined whether there is really a difference in the use of onomatopoeia among the eleven major regions of Japan. For this purpose, we used conversation data from the most recent 2 decades of Japanese Diet Records, along with the hometowns of the speakers. The results suggested that there are no differences in frequency of use across all types of onomatopoeia in each region. However, one type of onomatopoeia, especially emphatic onomatopoeia, showed regional differences. Further research from a dialectic, micro-scale perspective is needed.

Keywords: Onomatopoeia, Dialect, Region Comparison, Diet Records, Corpus Study

1 Introduction

Japanese onomatopoeic words (also called mimetics, sound-symbolic words, idiophones and iconic words) are more frequent in spoken discourse, especially in informal daily conversations, than in writing. It is a common belief that onomatopoeia is particularly frequent in some areas, such as the Kinki region, which includes Osaka, Kyoto, Nara, Hyogo, Shiga, and Wakayama. To examine the plausibility of this folk dialectology, we investigated the frequency of onomatopoeia in the Minutes of the Diet as a corpus of spoken Japanese.

1.1 Onomatopoeia

Onomatopoeic words imitate sounds (or even soundless situations) or personal feelings. There are several categories of Japanese onomatopoeia; for example, gion-go (e.g., “don-don” as the sound of drum) imitates sounds with voices, and gitai-go (e.g., “doki-doki” as being nervous) imitates the situation or personal feelings. Onomatopoeia forms an important part of the Japanese vocabulary and is widely used in product advertising and in comics. Appropriate contexts and situations for the use of onomatopoeia have been described. For example, previous research has shown that it is not commonly used with older persons or with superiors in a business situation [4]. On the other hand, it is used in conversations with older and younger people in personal relationships, including relatives. In this way, the number of people with whom it is appropriate to use onomatopoeia in daily conversation is relatively limited.

1.2 Onomatopoeia and region

Most Japanese people tend to believe that people who live in the Kinki region have a higher frequency of onomatopoeia use than those in other regions. Descriptions of onomatopoeia can be found in studies of the Osaka/Kansai dialect and Osaka culture. But contrary evidence comes from a dialect investigation, which reported that the amount of onomatopoeia included in conversations in Kinki region was not so great, and the amount was relatively greater in Tohoku and the Kanto region [2].

In a preliminary study, we conducted a web questionnaire to reveal the subjective frequency
of onomatopoeia use in Japan. In this study, 15- to 65-year-old men and women who lived in 47 prefectural and city governmental districts in Japan (total 1100 people) were asked how they usually use onomatopoeia [3]. There were no significant differences in subjective frequencies of onomatopoeia use among the 11 regions of Japan. However, to the question "People in which region do you think use more onomatopoeia?", most respondents reported the people in Kinki region. These results indicated that the folk theory -- "people who live in the Kinki region have a higher frequency of onomatopoeia use than those who live in other regions" -- did not match the facts. We concluded that this bias came from the stereotype of people in the Kinki region provided by the mass media. A limitation of this study is that the subjective data may not have exactly reflected the frequency of individual onomatopoeia use because subjective reports of frequency are not always in accord with real conversation. Moreover, the respondents to a web questionnaire may not have had an adequate understanding of onomatopoeia.

1.3 Purpose of this study

To resolve the subjectivity of frequencies of onomatopoeia use in the previous study [3], we selected the Diet Records of Japan as a massive conversation corpus. Diet Records of Japan is the largest corpus of conversation we could obtain. Moreover, almost all speakers in this record are public people and it is relatively easy to trace their hometown. From these reasons, we selected this record as most appropriate data to demonstrate about the onomatopoeia usage. The purpose of our research was to reveal differences in rates of onomatopoeia usage among regions in Japan based on the Diet Records. For this purpose, we analyzed the onomatopoeia usage rates and the hometowns of the most frequent users of onomatopoeia based on the Diet Records.

2 Methods

2.1 Diet Records of Japan

The Diet Records of Japan include all utterances from all meetings of the Diet held from 1947 to 2011, with accompanying information like speaker names and dates of the meetings. This database is open to the public on the Diet Records Search System website (http://kokkai.ndl.go.jp/). It contains the meeting records of the Houses of Representatives and Councilors, plenary sessions of both Houses, and meetings of the conference committee and the joint examining committee. To compare with our preliminary study in 2011, we limited the data to Houses of Representatives and Councilors plenary sessions from 1992 through 2012 to avoid historical influence of old conversations.

2.2 Onomatopoeia list

We used an onomatopoeia list created from an onomatopoeia dictionary [1]. This list had seven categories of onomatopoeia sorted by morphological characteristics (see Table 1). We collected all the utterances included at least one form of onomatopoeia from the corpus. We calculated the frequency of onomatopoeia use (the number of utterances of each individual which included onomatopoeia divided by his/her total number of utterances). To prevent the extraction of non-onomatopoeia words, we added “-to” to the end of each search word, using the characteristic of onomatopoeia that it frequently used with Japanese suffix “-to”.

2.3 Analysis

Each data record from the Diet Records of Japan consisted of the date and time of the utterance, the name and position of the speaker, and the utterance itself. The data were arranged in chronological order from earliest to latest. First we extracted the date information of from each meeting and converted it to western calendar notation. Then we extracted information about the speakers, their positions, and their utterances. The utterances were broken into sentences, and the database therefore consisted of individual sentences with IDs to identify unique utterances. A sentence was defined as the portion of an utterance between two periods or the point where the speaker changed. The number of sentences stored in the entire database was 48,512,596, and the total number of characters was 3,437,337,516.
Table 1. Categories of Japanese onomatopoeia

<table>
<thead>
<tr>
<th>Category</th>
<th>Example of Onomatopoeia</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive1</td>
<td>ぐいぐい /gui-gui/</td>
<td>strongly</td>
</tr>
<tr>
<td>Repetitive2</td>
<td>ふらふら /hura-hura/</td>
<td>woozily</td>
</tr>
<tr>
<td>Emphatic</td>
<td>しっかり /sikkari/</td>
<td>steadily</td>
</tr>
<tr>
<td>&quot;-n&quot; suffixed</td>
<td>はざり /basari/</td>
<td>readily</td>
</tr>
<tr>
<td>&quot;-Q&quot; suffixed</td>
<td>ぱたっ /potaQ/</td>
<td>trickily</td>
</tr>
<tr>
<td>&quot;-N&quot; suffixed</td>
<td>ぱちん /pachiN/</td>
<td>(with verb &quot;shut&quot;) with the sound &quot;snap&quot;</td>
</tr>
<tr>
<td>Long Vowel</td>
<td>どーん /doN/</td>
<td>reliably</td>
</tr>
</tbody>
</table>

* "Q": double consonant; "N": Japanese syllabic nasal; "n": long vowel

We selected one search word from onomatopoeia list and from the database we acquired all the sentences that included the search word. In addition, we assigned corresponding speaker information to all acquired utterances. Finally, we calculated the total number of utterances of each speaker and the number of utterances which included onomatopoeia. This processing was conducted for all onomatopoeia search words in the database from 1992 to 2012.

2.4 Speakers’ hometowns

We selected speakers who had more than 1000 utterances and investigated their hometowns or birthplaces in internet databases (Wikipedia or their official websites). As a result, the hometowns of 3,606 of 5,803 people (62.12%) who provided more than 1,000 utterances in the Diet minutes were revealed. The speakers whose hometowns were inside Japan were included in the analysis.

3 Results

As in the previous study [3], the speakers were divided into 11 categories according to their hometowns. Table 2 presents the distribution of frequent onomatopoeia users in Japan.

Table 2. Number of speakers in each home region

<table>
<thead>
<tr>
<th>Home regions</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hokkaido</td>
<td>69</td>
</tr>
<tr>
<td>Tohoku</td>
<td>155</td>
</tr>
<tr>
<td>Kanto</td>
<td>492</td>
</tr>
<tr>
<td>Koushin-etsu</td>
<td>99</td>
</tr>
<tr>
<td>Hokuriku</td>
<td>64</td>
</tr>
<tr>
<td>Tohakai</td>
<td>155</td>
</tr>
<tr>
<td>Kinki</td>
<td>262</td>
</tr>
<tr>
<td>Chugoku</td>
<td>120</td>
</tr>
<tr>
<td>Shikoku</td>
<td>73</td>
</tr>
<tr>
<td>Kyusyu</td>
<td>189</td>
</tr>
<tr>
<td>Okinawa</td>
<td>20</td>
</tr>
</tbody>
</table>

3.1 Frequency of all onomatopoeia

Figure 1 is a heat map which expresses the frequencies of all onomatopoeia words in the list in each region using color (red: high frequency; blue: low frequency). To produce this heat map, the number of sentences with onomatopoeia in each prefecture was divided by the total number of sentences in each prefecture, and the number of sentences with onomatopoeia was divided by the total number of sentences by each speaker. Then the average numbers for each prefecture were worked out, and these numerical values were normalized with the maximum to control the bias. Finally, a heat map was rendered based
Because the color distribution of Figure 1 is largely uniform, we concluded that there is no difference in the frequencies of onomatopoeia use among the regions in Japan. For result 2, we conducted an exploratory analysis of the data for certain categories of onomatopoeia.

3.4 Frequency of emphatic onomatopoeia

Figure 2 is the heat map of “emphatic” onomatopoeia. Emphatic onomatopoeia words are in CVCCV-ri form (C: consonant, V: vowel): for example, jikku-ri (“slowly and carefully”), hakki-ri (“clearly”), sikka-ri (“well”). These onomatopoeia are the emphatic form of the basic onomatopoeia like jikuri, hakiri, sikari, so they are called “emphatic onomatopoeia”. For this type of onomatopoeia, the most frequent usage was found in the Kinki (shown in red).

In addition, we analyzed the frequencies of adverbs that are not onomatopoeia (for example, jyu-bun "sufficient" or mata "again") to investigate if differences in frequency among the regions is a phenomenon specific to onomatopoeia. Figure 3 is the heat map of the frequency of such adverbs. The color distribution of this map is also uniform, similar to Fig. 1. From this result, differences in the frequency of adverbs were not observed.

4 Discussion
4.1 Summary

The purpose of this research was to reveal differences in the frequency of use of onomatopoeia among regions in Japan. With the records of the Diet of Japan as a massive conversation data set, we found an absence of frequency differences for all types of onomatopoeia except for the emphatic type, which was most prevalent in the
Kinki region. Moreover, this difference is a specific characteristic of onomatopoeia, because it was not observed with words that play the same role in conversation (i.e., adverbs).

The results of this research are consistent with a study which analyzed talk in local TV programs in the Kinki region [5]. This study pointed out that onomatopoeia such as "ga:Q", "ba:Q" (both also emphatic types of onomatopoeia) were frequently used. With this observation, the folk theory that argues that "onomatopoeia is most frequently used in the Kinki area" may result from observations of a single type of onomatopoeia, rather than observations of all types. To examine this possibility, detailed research focused on the conversation data of Kinki locals is necessary.

4.2 Discussion

This study has several limitations. First, the hometowns of all the selected speakers were not clear. Using various resources we were unable to determine whether the stated hometown was just the birthplace or was the place they spent their childhood. Furthermore, even if the information from the resource was correct, it was not clear if speakers retained the dialect of their hometowns. Older speakers had a typical pattern of moving, for example, living from birth to high school age in their hometowns and then going to Tokyo at the time of university entrance. But it is now very difficult to accurately locate the place where younger people form their language identity, because people nowadays tend to wander from place to place, including to and from foreign countries, from childhood. To prevent these problems, dialectology usually adopts a micro-viewpoint, with detailed descriptions of a small number of speakers. The present study offered what may be called a macro-viewpoint, using large-scale data. Combining these two viewpoints may offer a way to break an impasse in both dialectology and anthropology in the future.

For generalization of this research, we can point out the importance of relationship between dialect and onomatopoeia. The research intended for the distribution of English phonesthemes in England [6] showed clear distributional patterns in certain areas. Even English, which includes less onomatopoeia (=phonesthemes) than Asian Languages, showed the distributional bias inside country. It implies that there is larger impact in Asian languages of the relationship between onomatopoeia and dialects or certain areas. It is getting more important to consider the characteristics of onomatopoeia not only in a language but also within a language to understand and use onomatopoeia efficiently. We hope that this quantitative approach will offer a new viewpoint to the dialectology and also effective use of onomatopoeia in the field of information science.

Acknowledgment

This research was funded by a Grant-in-Aid for Young Scientists (A) to Satoshi Nakamura (23680006).

References