1. Introduction

In this research, I conducted a statistical analysis of the recent synchronic variation in Japanese causatives called \textit{sa}-Insertion. This process adds an extra \textit{sa} to the causative morpheme, as in \textit{yar-as-a-se-ru} as opposed to the standard \textit{yar-ase-ru} ‘let someone do’ and \textit{hair-as-a-se-ru} as opposed to the standard \textit{hair-ase-ru} ‘let someone enter’. \textit{Sa}-Insertion results in an extra phonological sequence \textit{sa} in causative phrases, compared to standard causatives.

As usual with any such variation, \textit{sa}-Insertion has also been referred to as an example of \textit{kotoba-no-midare} “language disturbance” by language purists and school teachers; although Martin (1975: 287) refers to the phe-
nomenon, it was not until Chen (2002) that its grammatical mechanism and sociolinguistic aspects were analyzed in a professional way. *Sa*-Insertion has, thus far, been analyzed based on natural linguistic data and some properties of *sa*-Insertion have been revealed (Chen 2002, Okada 2003, among others). Although previous studies of *sa*-Insertion have made progress, they are insufficient in terms of a formal account of the variation. Also, the possible dialectal differences have been mostly ignored (lack of dialectal control).

The aim of this research, therefore, is to characterize the phenomenon quantitatively and to reevaluate the conclusions of previous studies; specifically, this research focuses on the following points: 1) chronological change of the distribution of *sa*-Insertion (language-external factor); 2) the effect of preceding context on the distribution of *sa*-Insertion; and 3) the effect of following context on the distribution of *sa*-Insertion (language-internal factors). For this purpose, the “On-line full text database of the minutes of the Diet” (henceforth the Diet database), which is characterized by its large-scale, is employed. This database includes all utterances of the members in all sessions and committees from the first Diet held in May, 1947 to the present.1

The *sa*-Insertions and standard causatives collected total 317 and 4,708, respectively; thus, the rate of *sa*-Insertion in all the causative forms amounts to 6.31 percent. The data were then subjected to statistical analysis. This research reveals that the distribution of *sa*-Insertion is significantly affected by language-internal and -external factors. In addition, the factors governing the distribution of *sa*-Insertion were found to interact with each other.

The discussion proceeds as follows: Section 2 identifies the *sa*-Insertion phenomenon, and then introduces the previous studies of *sa*-Insertion and the Diet database, as well as the sampling. In Section 3, I conduct statistical analysis based on the data. Section 4 concludes the discussion.

2. Methodological Preliminaries

In this section, I firstly explore the *sa*-Insertion phenomenon in connection with the causative formation of Japanese. I then introduce the claims of previous studies of *sa*-Insertion as they relate to the present research.

2.1. *sa*-Insertion

*Sa*-Insertion is a variation phenomenon in Japanese causatives. Japanese causatives are formed by attaching causative suffixes to verb stems. The standard causative, which is the traditional variant, comprises the verb stem

---

1 The Diet of Japan is comprised of the plenary sessions and the committees. The committees are further divided into permanent committees of each Ministry and select committees. Each member of the Diet belongs to either the House of Councilors or the House of Representatives (Oyama, 2003).
and the causative suffix *(s)ase;* in contrast, *sa*-Insertion, which is the innovative variant, comprises the verb stem and the causative suffixes *as* and *ase.* I present below some examples of *sa*-Insertion and standard causatives observed in the Diet database.

Standard causatives

(1) hikoo-ki-o mata tukur-ase-ru.
   airplane-ACC again make-CAUS-NONP
   ‘We let (the company) make airplanes again.’
   (Yoshio Namiki, Jun. 11 1952)

(2) iroirona enzetu-o yar-ase-te-itadaita.
   various speech-ACC do-CAUS-TE-AUX.POL.PAST
   ‘I made various speeches.’ (polite)
   (Keigo Oouchi, Jun. 7 1994)

(3) gakkoo-ni ik-ase-nai.
   school-LOC go-CAUS-NEG.NONP
   ‘We allow (students) not to go to school.’
   (Hiroko Mizushima, Feb. 27 2004)

*Sa*-Insertion

(4) tyoosahookokusyo-o yom-as-ase-te-itadakimasita.
   investigation report-ACC read-CAUS-CAUS-TE-AUX.POL.PAST
   ‘I read the investigation report.’ (polite)
   (Seiichi Mizuno, Sep. 27 1995)

(5) situmon-o owar-as-ase-te-itadakimasu.
   question-ACC finish-CAUS-CAUS-TE-AUX.POL.NONP
   ‘Let me finish my question.’
   (Tatsuya Ito, Apr. 11 1997)

(6) kono koosyoo-o torihakob-as-ase-tai.
   this negotiation-ACC advance-CAUS-CAUS-DES.NONP
   ‘I want to let (someone) advance this negotiation.’
   (Sanzo Hosaka, Apr. 10 1998)

2 The Japanese causative suffix shows morphophonemic alternation, according to the types of verb stem: consonant verb takes *are,* as in *yar-are,* while vowel verb takes *ase,* as in *tare-ase* ‘let someone eat’. I will mention the two types of Japanese verb (consonant verb/vowel verb) in Section 2.2.
As exemplified above, each standard causative is formed by attaching *ase* to verb stems, while in *sa*-Insertion *as* and *ase* attach to verb stems, instead of *ase* alone. Thus, *sa*-Insertion contains an extra *sa* in causative phrases, as opposed to the standard causatives.\(^3\)

### 2.2. Previous Studies of *sa*-Insertion

In this section, I introduce previous studies related to the present research. Kikuchi (1997) claims that the grammaticalization process (Meillet 1912, Hopper and Traugott 1993, among others) turned the phrase *-as-ase-te-itadak-* into an independent word, and that *-as-ase-te-itadak-* lost its original meaning and has acquired a new meaning and function.

Chen (2002) analyzed the utterances of all the members of the Diet of Japan (by means of the same corpus as that used in the present research) from 1947 to 1957 and from 1996 to 2001 and concluded that *sa*-Insertion 1) is first observed in 1947 and 2) occurs in almost all the consonant verbs regardless of stem length.

Inoue (2003) points out that *sa*-Insertion is restricted to consonant verbs, and that it does not occur in vowel verbs. Japanese verbs are classified into two types, according to the stem-ending; one type is a consonant verb which ends in a consonant (e.g. *yar-* ‘do’, *hair-* ‘enter’), and the other type is a vowel verb which ends in a vowel (e.g. *mi-* ‘see’, *tabe-* ‘eat’) (Bloch 1946).

<table>
<thead>
<tr>
<th>verb type</th>
<th>stem</th>
<th>standard causative</th>
<th><em>sa</em>-Insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonant verb</td>
<td><em>yar-</em></td>
<td><em>yar-ase-</em></td>
<td><em>yar-as-ase-</em></td>
</tr>
<tr>
<td></td>
<td><em>hair-</em></td>
<td><em>hair-ase-</em></td>
<td><em>hair-as-ase-</em></td>
</tr>
<tr>
<td>vowel verb</td>
<td><em>tabe-</em></td>
<td><em>tabe-sase-</em></td>
<td><em>tabe-sas-ase-</em></td>
</tr>
<tr>
<td></td>
<td><em>mi-</em></td>
<td><em>mi-sase-</em></td>
<td><em>mi-sas-ase-</em></td>
</tr>
</tbody>
</table>

As Table 1 shows, the standard causative can occur with either consonant or vowel verbs, while *sa*-Insertion cannot occur with vowel verbs. Thus, *mi-sas-ase-* or *tabe-sas-ase-*, for instance, are impossible.

Based on this observation, he claims further that the change of *sa*-Insertion would lead to the simplification of the conjugation of verbs in Japanese by analogical leveling (cf. Kiparsky 1978, Bybee 1985, among others).

---

\(^3\) *Sa*-Insertion was given its name based on a phonological aspect of Japanese. Japanese has an open-syllable sound pattern and in principle it does not allow codas. Thus, the sequence *yar-as-ase* is pronounced with CV structure as in *ya.ra.ase*. This results in the auditory perception of an extra *sa* rather than *as*. However, the morphosyntactic investigation revealed that *sa*-Insertion contains an extra causative suffix *as* (Okada 2003, Sano 2006), and the difference between standard causative and *sa*-Insertion cannot simply be attributed to the phonological aspect.
As illustrated in Table 2, the traditional paradigm, which consists of only the standard causative, shows a morphophonemic alternation of the causative suffix between two allomorphs *ase* and *sase* according to the verb types to which it attaches. On the other hand, in the innovative paradigm, which consists of *sa*-Insertion and the standard causative, the causative suffix in vowel verbs has changed from *ase* to *sase*. There is no such alternation in this paradigm, and the causative suffix uniformly takes the form *sase*, regardless of the type of the verb stem. Consequently, the difference in conjugation paradigm disappears.

Okada (2003) points out that 1) the benefactive pattern (*-as-ase-te-itadak* pattern) is especially frequent in *sa*-Insertion (comprising nearly half of the dataset); 2) *sa*-Insertion is a double causative; 3) *sa*-Insertion serves to reinforce politeness, and the newly inserted causative suffix carries this function.

Okada (2004) claims that 1) the intransitive and transitive distinction of verbs is irrelevant to the distribution of *sa*-Insertion; 2) the frequency of *sa*-Insertion in verbs ending in *-s* is low.

### 2.3. Data

In this research, the Diet database, which is characterized by large-scale, is employed: it includes all utterances of the members in all sessions and committees from the first Diet held in May, 1947 to the present. (It continues to be updated.) The large scale of the Diet database enables analysis of the chronological change of the phenomena over sixty years. The present research targets the utterances from May 1947 to April 2005.

In the Diet database, we searched target strings by selecting the Diet number or the date of session/committee in the pull-down menu (this is obligatory) and adding search strings. The target can be specified by adding

---

4 This analysis, however, is problematic because it is primarily based on the *Kana* spelling of Japanese. Although the name *sa*-Insertion is based on the *Kana* spelling, *Kana* is inappropriate for detailed analysis of *sa*-Insertion, and morphophonemic analysis needs to be done.

5 This seems to be associated with the fact that the causative suffix originally carried the intensifier function of the honorific/humble meaning in addition to the causative function in old Japanese (cf. Sato 1977: 317-319, and references cited therein).

further information such as the name of a speaker or the session committee. The target strings are displayed with detailed information per session committee and we directly downloaded the information into our computer.

Based on the list of the members of the Diet by Nambu (2005), I sampled eighty-one members of the Diet who come from Tokyo (dialectal control). The scope of the present research includes all of their utterances from the entire time period. I selected one member for each year of their birth. Thus, the analysis focuses on the utterances of eighty-one members from Tokyo.

In the next section, the data are subjected to a statistical analysis; I conduct a factor by factor analysis by comparing the distributions of sa Insertion and of the standard causative.

3. Analysis of the data

In this section, I conduct analysis of the data, according to each factor. An exhaustive examination of the Diet database brought forth a total of 317 causative forms with sa Insertion, as opposed to a total of 4,708 standard causative forms; thus, the rate of sa Insertion (sa-Insertion/(sa-Insertion + standard causatives) *100) amounts to 6.31 percent, as shown in Table 3 below.

<table>
<thead>
<tr>
<th></th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>sa-Insertion</td>
<td>317</td>
</tr>
<tr>
<td>standard causatives</td>
<td>4,708</td>
</tr>
<tr>
<td>rate (%)</td>
<td>6.31</td>
</tr>
</tbody>
</table>

In Section 3.1., I examine the chronological change of the distribution of sa Insertion. Section 3.2. examines the effect of following context on the distribution of sa Insertion. Finally, Section 3.3. examines the effect of preceding context on the distribution of sa Insertion.

3.1. Chronological change

In this section, I conduct analysis of the chronological change of the distribution of sa-Insertion. Analysis of the chronological change of sa-Insertion, which remains to be explored, is necessary in order to identify the status of

---

7 Sano (2008a, 2008b) describes the criterion concerning the status of sa-Insertion and standard causatives in the data extraction. According to the criterion of Sano (2008a), I excluded the examples which are not qualified as sa-Insertion or standard causatives from the extracted data.
sa-Insertion as an instance of language change. If sa-Insertion is an instance of language change, the distribution of sa-Insertion should show a correlation with the birth-year of the member. To verify this, I first examined the relationship between the distribution of sa-Insertion and the birth-year of the member, where sa-Insertions and standard causatives are sorted according to the birth-year of the members (grouped every ten years). I then considered the relationship between the year-period and the occurrence of sa-Insertion, where sa-Insertions and standard causatives are displayed, grouped for every five years. The distribution of sa-Insertion by birth-year is shown in Figure 1, and the distribution of sa-Insertion by five-year period is shown in Figure 2.

Figure 1. Distribution of sa-Insertion by birth year

![Figure 1](image1.png)

Figure 2. Distribution of sa-Insertion by five-year period

![Figure 2](image2.png)

Figure 1 demonstrates that the more recent the birth-year of the member is, the higher the rate of sa-Insertion. This tendency is remarkable in the birth-years after the 1950s. A significant correlation can be found between the distribution of sa-Insertion and birth-years ($X^2 = 386.96$, d.f.=9, $p<0.001$). This shows that sa-Insertion is an instance of language change. In Figure 2,
the dots which stand for the rate of sa-Insertion for each year of the Diet and the approximated curve are plotted. As in Figure 1, the rate of sa-Insertion is shown to be gradually rising from year to year. In the present data, sa-Insertion is first observed in 1949. This result is consistent with the claim of Chen, and from the first appearance, some sa-Insertion is observed (gradually increasing) until the 1980s. Around 1990, sa-Insertion suddenly exploded. This shows that sa-Insertion is in the beginning of a language change, and although some fluctuations are observed, the shape of the current curve corresponds to the early stage of an S-curve (Bailey 1973, Chambers and Trudgill 1980, among others). Thus, it can be predicted that the change of sa-Insertion will proceed in an S-curve fashion.

Note that sa-Insertion is first observed in 1949 in the present data. This year almost coincides with the birth year of the members who show the most rapid increase (after the 1950s). This fact implies the following scenario: The precursory sa-Insertion which appeared around the 1950s is transmitted to subsequent generations as primary linguistic data and is reproduced by members born after the 1950s. The members born around 1950 are the first generation that could have been exposed to sa-Insertion during the period of acquisition, while most of the members born before the 1950s had already passed the critical period when they were first exposed to sa-Insertion. It follows that the members whose birth year is after 1950 have a grammar with sa-Insertion, and the diffusion is accelerated by these members. This leads to the significant increase of sa-Insertion in the members born after 1950.

3.2. Following context

In this section, I examine the effect of following context on the distribution of sa-Insertion. If Kikuchi’s grammaticalization hypothesis and Okada’s claims are on the right track, sa-Insertion prefers the -as-ase-te-itadak- pattern to other patterns. To verify this hypothesis, I examine the distribution of sa-Insertion and the standard causative by following context. The result is shown below.

<table>
<thead>
<tr>
<th>context</th>
<th>sa-Insertion</th>
<th>standard causative</th>
<th>rate of sa-Insertion(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-te-itadak-</td>
<td>307</td>
<td>1,944</td>
<td>13.64</td>
</tr>
<tr>
<td>-te-</td>
<td>3</td>
<td>573</td>
<td>0.52</td>
</tr>
<tr>
<td>-te-moraw</td>
<td>2</td>
<td>61</td>
<td>3.17</td>
</tr>
<tr>
<td>-ru</td>
<td>2</td>
<td>1,128</td>
<td>0.18</td>
</tr>
<tr>
<td>-ta</td>
<td>1</td>
<td>172</td>
<td>0.58</td>
</tr>
<tr>
<td>-nai</td>
<td>1</td>
<td>257</td>
<td>0.39</td>
</tr>
<tr>
<td>-tai</td>
<td>1</td>
<td>63</td>
<td>1.56</td>
</tr>
</tbody>
</table>
As Table 4 shows, the frequency of sa-Insertion in the -as-ase-te-itadak-pattern is 307, among a total of 317 occurrences of sa-Insertion, and compared to other patterns, the rate of sa-Insertion in the -as-ase-te-itadak-pattern shows an extremely high percentage: 13.64 percent. That is, most of the sa-Insertions occur with the -as-ase-te-itadak-pattern.

As mentioned above, Kikuchi claims that sa-Insertion has become an independent word and Okada claims that the benefactive pattern is especially numerous. Likewise, the results of the present research show that the -as-ase-te-itadak-pattern is predominant. The following context of sa-Insertion, however, is not limited to the -as-ase-te-itadak-pattern. (Other patterns are also observed.) This implies that the grammaticalization process remains to be completed and is currently underway.

To capture the grammaticalization process, I examine the chronological change of the -as-ase-te-itadak-pattern in sa-Insertion.

As Figure 3 shows, few examples of -as-ase-te-itadak-pattern in sa-Insertion are observed until the 1980s. After the 1980s, its occurrence explodes: 95 examples are observed in the 1990s and 203 examples are observed after 2000. This shows that the grammaticalization process follows the same trajectory as the change of sa-Insertion. The grammaticalization is currently underway, accompanied by the change of sa-Insertion and sa-Insertion is contributing to the grammaticalization of the independent word -as-ase-te-itadak-.

Here, I mention the grammaticalization of sa-Insertion. The phrase -as-ase-te-itadak-originally had the literal meaning ‘to get a benefit/permission to do’, but then lost its original meaning and became an independent word with a new grammatical function that adds politeness; the meaning has become semantically unpredictable. That is, -as-ase-te-itadak-has changed
from a causative phrase to a politeness word. In fact, the meaning of -as-ase-te-itadak- is predominantly ‘polite’ and not ‘cause’, at least in the present data. Thus, -as-ase-te-itadak- has arguably undergone grammaticalization.

Incidentally, it is empirically known that sa-Insertion prefers the formal settings. This fact seems to be associated with the grammaticalization process: It is assumed that the -as-ase-te-itadak- pattern, which is an inherently polite expression is used in formal settings. Sa-Insertion, which is accompanied by the -as-ase-te-itadak- pattern, is also used in formal settings.

3.3. Preceding context

In this section, I examine the relationship between the preceding context of sa-Insertion and the distribution of sa-Insertion; Section 3.3.1 discusses the self-controllability restriction; Section 3.3.2, OCP\(^8\) (µ); Section 3.3.3, the effect of verb types; and finally I examine the effect of verb length in Section 3.3.4. In the first three contexts, the distribution of sa-Insertion is categorically rather than quantitatively, restricted.

3.3.1 Self-controllability restriction

Japanese verbs are classified into self-controllable and non-self-controllable verbs (cf. Harada 1973, Tonoike; 1978). Self-controllable verbs describe proactive behavior. In a sentence with a self-controllable verb, a speaker (agent) performs the behavior described by the verb of his/her own volition, as in yar-u ‘do’, hasir-u ‘run’, and ukaga-u ‘ask (honorific)’. In contrast, non-self-controllable verbs describe spontaneous events. In a sentence with a non-self-controllable verb, the speaker (agent) cannot perform the behavior of his/her own volition (e.g. *I am going to suffer.), as in kurusim-u ‘suffer’, komar-u ‘be puzzled’, and kanasim-u ‘grieve’. In the present data, sa-Insertion is restricted to self-controllable verbs and no sa-Insertion occurs with non-self-controllable verbs. Examples such as *kurusim-as-ase-, *komar-as-ase-, *kanasim-as-ase- are not observed.

These results can be explained by the following incompatibility of non-self-controllable verbs with the -te-itadak- component of the -as-ase-te-itadak- pattern: As we saw above, most sa-Insertions occur with the -as-ase-te-itadak- pattern. In a sentence with -te-itadak-, a person who performs a behavior is obligatorily the speaker (agent). Self-controllable verbs, in which the speaker (agent) can perform the behavior voluntarily, are compatible with -te-itadak-. For example, in a sentence with a self-controllable verb such as sore-o yar-as-ase-te-itadakimasu ‘Allow (me) to do it’, the person who performs ‘do’ is the speaker (agent). However, non-self-

\(^8\) Obligatory Contour Principle (OCP), which bars consecutive identical features, was originally proposed by Leben (1973).
controllable verbs, in which the speaker (agent) cannot perform a behavior voluntarily, are incompatible with -te-itadak-. Although the person who performs the ‘suffering’ is the speaker (agent) in a sentence such as *kurusim-as-ase-te-itadakimasu ‘Allow me to suffer’, ‘kurusim’ cannot be performed voluntarily. Thus, the combination of non-self-controllable verbs and -te-itadak- is impossible. The fact that sa-Insertion is restricted to self-controllable verbs is associated with the grammaticalization process.

3.3.2 OCP (µ)
If consonant verbs ending in -s, such as das-u ‘give’, tadas-u ‘correct’, and tobas-u ‘fly’ were to undergo sa-Insertion, the sequence sasa would be created, as in *das-as-ase-, *tadas-as-ase-, and *tobas-as-ase-. Okada (2004) claims that the frequency of sa-Insertion in verbs ending in -s is low.

Although the consonant verbs ending in -s amount to a total of 724 in token frequency, no sa-Insertion with this kind of verbs is observed in the present data. In other words, sa-Insertion which contains the sequence sasa cannot occur. Based on this observation, I introduce the constraint OCP (µ) (Sano 2007).

(7) OCP (µ)
No identical morae are adjacent.

The OCP (µ) bars the adjacent identical morae, such as /*sa.sa/. In this case, the sequence sasa in the matrix of the verb stem and the causative suffix is subject to this constraint. Due to the OCP (µ), sa-Insertion which contains the sequence sasa, as in *das-as-ase-, *tadas-as-ase-, and *tobas-as-ase-, is prohibited.

3.3.3 Effect of verb types
As mentioned above, Japanese verbs are classified into two types: consonant verbs (e.g. yar-, hair-) and vowel verbs (mi-, tabe-). In the present data, sa-Insertion is restricted to consonant verbs and cannot occur with vowel verbs, as Inoue (2003) claims. This derives from the phonological property of sa-Insertion with vowel verbs: any sa-Insertions with vowel verbs obligatorily contain the sequence sasa, as in *tabe-sas-a or *mi-sas-a. Thus, sa-Insertion with vowel verbs is excluded again by the OCP (µ), as in the examples above. The incompatibility of sa-Insertion with vowel verbs has a general phonological ground (it is not a specific property of vowel verbs).

9 OCP (µ) is too powerful, given that there are some nouns containing sasa in Japanese, e.g. sasa ‘bamboo grass’ and Sasaki (proper noun). This constraint has to be refined in terms of the level (range) of application.
The claim of Inoue (2003) is explained by OCP (µ), without appeal to any additional machinery.

3.3.4 Effect of verb-length

In language change (morpho-syntactic change), it is assumed that the change diffuses from shorter verbs to longer verbs (e.g. ra-deletion, Matsuda 1993). If this assumption is also true for the case of sa-Insertion, sa-Insertion prefers the shorter verbs to the longer verbs, with a higher rate in shorter verbs than in longer verbs in the beginning of language change, contrary to Chen’s claim that sa-Insertion occurs in almost all the consonant verbs regardless of stem length. To verify the hypothesis, I examined the distribution of sa-Insertion and the standard causative by verb-length measured by mora. The results are shown below.

<table>
<thead>
<tr>
<th>Token</th>
<th>sa-Insertion</th>
<th>standard causatives</th>
<th>rate of sa-Insertion(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 morae</td>
<td>90</td>
<td>1,411</td>
<td>6.00</td>
</tr>
<tr>
<td>3 morae</td>
<td>157</td>
<td>2,220</td>
<td>6.60</td>
</tr>
<tr>
<td>4 morae</td>
<td>66</td>
<td>878</td>
<td>6.99</td>
</tr>
<tr>
<td>5 morae</td>
<td>2</td>
<td>131</td>
<td>1.50</td>
</tr>
<tr>
<td>6 morae</td>
<td>2</td>
<td>68</td>
<td>2.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>sa-Insertion</th>
<th>standard causatives</th>
<th>rate of sa-Insertion(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 morae</td>
<td>14</td>
<td>55</td>
<td>20.29</td>
</tr>
<tr>
<td>3 morae</td>
<td>17</td>
<td>152</td>
<td>10.06</td>
</tr>
<tr>
<td>4 morae</td>
<td>15</td>
<td>109</td>
<td>12.10</td>
</tr>
<tr>
<td>5 morae</td>
<td>2</td>
<td>55</td>
<td>3.51</td>
</tr>
<tr>
<td>6 morae</td>
<td>2</td>
<td>8</td>
<td>20.00</td>
</tr>
</tbody>
</table>

Token: $X^2=7.952$, d.f.=4, n.s. Type: $X^2=9.788$, d.f.=4, $p<0.05$

As Table 5 shows, the Token frequency of sa-Insertion in 2, 3 and 4 morae verbs is 90, 157 and 66, respectively. On the other hand, the Token frequency of sa-Insertion in 5 and 6 morae verbs is only 2. Likewise, the rate of sa-Insertion in 2, 3 and 4 morae verbs is over six percent, while in 5 and 6 morae verbs, the rate is less than three percent.

These results show that the distribution of sa-Insertion is affected by verb-length, contrary to the claim of Chen. Specifically, sa-Insertion frequently occurs in 2, 3 and 4 morae verbs compared to the standard causative, while in 5 and 6 morae verbs, it seldom occurs. This tendency is also observed in Type frequency. Sa-Insertion prefers the short verb stems. In the beginning of language change, sa-Insertion seldom occurs in verbs of more...
than 5 morae in length and the assumption that the change diffuses from shorter verb to longer verb is true for the case of *sa*-Insertion.\(^{10}\)

4. Conclusion

In this research, I conducted statistical analysis of *sa*-Insertion via the Diet database. As a result, some unexplored properties of *sa*-Insertion were revealed, and claims of previous studies were verified in an objective and empirical manner. Specifically, the results show that *sa*-Insertion: 1) is increasing more and more in recent years; 2) is in the course of grammaticalizing and creating the independent word, -*as-ase-te-itadak*--; 3) is restricted to self-controllable verbs; 4) does not contain the sequence *sasa* (OCP (µ)); 5) is restricted to consonant verbs (OCP (µ)); 6) seldom occurs in verbs more than 5 morae in length. The results partly support Kikuchi’s claim that *sa*-Insertion has contributed to the creation of an independent word and Okada’s claim that the benefactive pattern (-*as-ase-te-itadak*- pattern) is especially frequent in *sa*-Insertion. However, the claim of Chen (2002) that *sa*-Insertion occurs in almost all the consonant verbs regardless of stem length is refuted. Based on the effects of the preceding and following contexts, *sa*-Insertion does not necessarily lead to the simplification of the conjugation of verbs in Japanese, as claimed by Inoue (2003).

I conclude that *sa*-Insertion is in the beginning of a language change; that change will most likely proceed in an S-curve fashion. In addition, the ongoing grammaticalization process and the general tendency of diffusion from particular contexts are observed. The distribution of *sa*-Insertion is greatly affected by language-internal and -external factors. Furthermore, the factors governing the distribution of *sa*-Insertion have been demonstrated to interact with each other (e.g. self-controllability restriction).

The present research captures the beginning of language change. This implies that we can conduct real-time studies of the phenomenon hereafter. Traditionally, the beginning of language change of particular phenomena has already passed when we have noticed the phenomena and initiated the research (e.g. *ra*-deletion, Matsuda 1993; velar nasalization, Hibiya 1995). Therefore, the beginning of language change has been under researched. The continuous real-time study of *sa*-Insertion can reveal various properties of language change that remain to be explored.

---

\(^{10}\) I also observed lexical diffusion (cf. Chen and Wang 1975) of *sa*-Insertion. The change of *sa*-Insertion does not diffuse uniformly among all verbs, but begins with particular verbs and spreads to other verbs. I present the representative verbs with *sa*-Insertion and their frequency: *utsur-u* (65); *ukaga-u* (43); *hair-u*, *owar-u* (29); *yar-u* (22); *yom-u* (21); *kak-u*, *okona-u* (9); *kik-u*, *modor-u*, *tor-u* (7); *i-u* (6), *ik-u*, *kabar-u* (5); *tsuka-u*, *susum-u*, *tob-u* (4).
Appendix: Abbreviations

ACC: Accusative  AUX: Auxiliary  CAUS: Causative
LOC: Locative  NONP: Nonpast  NEG: Negative
DES: Desiderative  PAST: Past  POL: Polite
TE: te-form of the verb

References

On-line full-text database of the Minutes of the Diet <http://kokkai.ndl.go.jp>


