1. Introduction

One of the most prominent case alternations in Japanese is *ga/no* conversion. *Ga/no* conversion (henceforth GNC) is a case alternation where a nominative particle *ga* variably alternates with a genitive particle *no* in certain embedded clauses as (1) but not in main clauses as (2).¹

1.1. GNC in Embedded Clauses

(1)

a. [Ken-*ga/no* kaita] hon
   Ken-NOM/GEN write book
   ‘the book which Ken wrote’

b. [ryoosin-*ga/no* nokosita] isan
   parents-NOM/GEN left assets
   ‘the assets which parents left’

c. [yuki-*ga/no* ooi] kuni
   snow-NOM/GEN heavy country
   ‘the country where there is heavy snow’

d. [inu-*ga/no* neteita] basyo
   dog-NOM/GEN sleep place

* We would like to thank Mark Scott, Philip Spaelti, and 17th ICHL participants, especially Shobhana Chelliah, Reijiro Shibasaki, Sayaka Abe, and Heidi Quinn for their comments and support. This article is an extension of the ideas explored in Nambu (2005a) and Nambu (2005b). The research reported here is partially supported by Grant-in-Aid for Scientific Research No. 17320068 from the Japan Society for the Promotion of Science.

¹ Kyusyu dialect permits *no* as a subject marker in main clauses:

i. Ame-*no* futte kita.
   rain-NOM fall become
   ‘It’s started to rain.’

ii. Doroboo-*no* haitta.
   thief-NOM come in
   ‘The thief came in.’

GNC was first noted by Harada (1971), who derived the genitive construction from the nominative one. In this paper, however, we take GNC as it is and do not take a particular position as to its derivation. Since Harada (1971) a number of studies of GNC emerged in almost every grammatical paradigm proposed so far. What has been missing in the previous approaches, however, is the empirical verification of Harada's (1971) hypothesis that there is an ongoing change, whereby the speakers of Tokyo Japanese increasingly prefer _ga_ to _no_ in the syntactic position. Unfortunately, this intriguing hypothesis has been left untouched for about 30 years since its proposal. In this paper we will attempt to verify Harada's hypothesis quantitatively, and specify the language external/internal factors which affect GNC.

2. Methodology

2.1 Corpus-Based Analysis

To verify whether language change is indeed taking place, we must base our analysis on solid empirical observation. In contrast to the huge number of previous syntactic analyses of GNC, which are mostly based on introspection, there are only a few usage and judgment studies in the history of GNC analyses. Horie and Kang (2000), for example, is a corpus-based study on the variation and Maki et al. (2004) reported results from a questionnaire-based survey. Both of them, however, are purely synchronic in nature, and as such, they do not mention a word about Harada’s hypothesis. Obviously, we need a diachronic corpus of substantial size with a historical depth, covering the past few generations of the same dialect of Japanese, so that the syntactic change, if it exists, can be clearly observed and analyzed.

For this purpose, we adopted the Minutes of the Japanese Diet as our data. The Minutes store records of Diet members’ speeches from every meeting in the Diet in an almost verbatim way, and it is available on the web. Its most prominent feature, for our purpose, is that the Minutes have speech data spanning about 60 years, starting from

---

2 The website is accessible at ‘Full-text Database System for the Minutes of the Diet’ (http://kokkai.ndl.go.jp/).
1947 (Matsuda 2004), and as such it provides us with an ideal dataset. Furthermore, the website features a user-friendly search facility by which the data can be searched by keyword, speaker, date of the session, session name, etc., and the search results can be downloaded to the user’s terminal at once.

2.2 Subjects and Speech Data

Because Harada’s hypothesis was based on Tokyo Japanese, we needed to restrict our data to the speech of members who are also native speakers of the dialect. By checking the hometown of the Diet members, we came up with 182 subjects. Then we sampled one Diet member for each birth year to create a dataset with chronologically equal proportions. This left us with 76 subjects whose birth years range over almost 100 years (1876 to 1970). Their data in the Minutes was downloaded from the website as a text file. For each speaker, we took 100 tokens of the variable, making the final sample size 7,600 tokens of which 948 (or 12.5%) were marked with no.

2.3 Envelope of variation

Before counting the number of ga and no in the data, we needed to delineate the environments where variation is possible (envelope of variation). One such environment is adnominal clauses, which have been discussed as a typical situation where GNC can occur, as in (3).

(3) Ken-wa [[musuko-ga/no yonda] hon]-o katazuketa.
    Ken-TOP son-NOM/GEN read book-ACC clean up
    ‘Ken cleaned up the books which his son read.’

In addition to adnominal clauses, we included made or yori subordinate clauses, following Watanabe (1996) and Kikuta (2002) who claim that these clauses allow GNC as shown below:

---

4 Diet members who experienced special language training (e.g. TV announcers) were excluded from the sample.
5 Data in parentheses or brackets in the Minutes were not extracted, because they are citations from someone else’s speech. Also sections where Diet members clearly read texts were excluded from the sample.
(4)
a. \[[\text{Basu-ga/no kuru] made}\] suwatte iyooka.
   
   bus- NOM/GEN come until sit be
   
   ‘Let's sit until the bus comes.’

b. \[[\text{Kyaku-ga/no kuru] yori}] hayaku nimotu-ga tuita.
   
   customer-NOM/GEN come than earlier package-NOM arrive
   
   ‘The package arrived before the customer came.’

(\text{Kikuta 2002})

Also included in our data set are \textit{toyuu} and \textit{tono} apposition clauses, which have intervening \textit{toyuu} or \textit{tono} between the embedded clause and its head noun. Those clauses are discussed in Inoue (1976) as environments where \textit{no} cannot be used as a subject marker. Contrary to Inoue (1976), however, Ura (1993) showed that \textit{no} can appear in such clauses, if the head noun is a non-derived one such as \textit{uwasa}. This can be seen in (5), which contrasts with (6) where the head noun is a noun derived from the verb \textit{siraseru}.

(5) \[[\text{Handai-ga dansigakusei-ga kinbenda} toyuu] uwasa
   
   Osaka Univ.-NOM male student-NOM diligent-be COMP rumor
   
   ‘the rumor that male students in Osaka University are diligent’

(Ura 1993)

(6) \[[\text{karera-ga buzidatta} toyuu/tono] sirase
   
   they-NOM safety-be COMP news
   
   ‘the news that they were safe’

(Inoue 1976)

3. Results
3.1 Language Change

To verify Harada’s hypothesis, we calculated the rate of \textit{no} and checked its correlation with the Member’s birth year. If his hypothesis is right, the rate should decrease as birth year increases. The result is given in Figure 1, where each dot represents a subject. The gradual decline of the overall trend is pretty clear, so that the Members are gradually
switching to *ga* at the sacrifice of *no* as their birth year approaches the present. The logistic regression line also supports this impression statistically. Thus, Harada’s insight in the early 1970’s was right, and GNC is indeed involved in a change in progress.

![Figure 1](image)

**Figure 1**
Scatterplot of the percentage of *no* and the birth year of the Diet Members

What, then, causes this change? Table 1 shows the historical change in the distribution of *ga* and *no* over the known period of Japanese history (Konoshima 1970, Doi 1982, Matsunaga 1983). At the first stage before the Kamakura and Muromachi periods (~1192), *ga* and *no* show a similar distribution. They then underwent a change during the Kamakura and Muromachi periods (1192-1573), as *ga* emerged in main clauses as a subject marker. In the present time *ga* cannot be used between nominals, and in embedded clauses, as we observed above, a change is now in progress (represented as ‘√→×’ in the table). Notice here that a completion of this change will make the distribution of the two particles perfectly complementary, so that where *ga* can occur, *no* cannot. The picture suggests to us that the current change is the final stage of this millennium-long transition, which was originally motivated by a need for differentiation of two particles with similar syntactic distribution.
Before Kamakura / Muromachi Era (-1192)  
Kamakura / Muromachi Era (1192-1573)  
Present (1573-)

<table>
<thead>
<tr>
<th></th>
<th>Before Kamakura / Muromachi Era (-1192)</th>
<th>Kamakura / Muromachi Era (1192-1573)</th>
<th>Present (1573-)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GA</td>
<td>NO</td>
<td>GA</td>
</tr>
<tr>
<td>Between Nominals</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>EmbeddedClauses</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Main Clauses</td>
<td>×</td>
<td>×</td>
<td>√</td>
</tr>
</tbody>
</table>

Table 1: Historical distribution of ga and no

3.2 Language External/Internal Factors

In this section we will discuss the language external and internal factors which affect GNC. First, we take up language internal factors: adjacency between the subject NP and the verb, Transitivity Restriction, and stativity. We then look into language external factors, Plenary/Committee sessions and House of Representative/Councilors.

3.2.1 Adjacency

Harada (1971) pointed out that the existence of intervening elements between the subject NP and its predicate affects the acceptability of GNC. If there exist some intervening elements, genitive no cannot appear as a subject marker as in (8).

(8) [kodomotati-ga/*no minna de ikioiyoku kakenobotta] kaidan children-NOM/GEN all with swiftly run up stairs ‘the stairs which all children run up swiftly’

(Harada 1971)

Table 2 compares the adjacent case with the non-adjacent one. Here the rate of no in the former case is higher than in the latter one. It shows that adjacency has a crucial effect

6 We also checked other internal factors such as animacy (Silverstein 1976) and negativity (Givón 1979). None of these factors, however, turned out to be significant in our analysis.
on the rate of *no*. But why does the adjacency matter? Here, Shibatani (1975) suggests us a possible explanation. He indicates that the particle *no* has a latent ambiguity between subject marker and possessive marker, and this ambiguity causes a processing problem for the hearer. The problem becomes more serious when the particle stands in the non-adjacent environment, and consequently the interpretation would require more time. All of this would be resolved once *ga* is used instead of *no*.\(^7\)

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>Non-Adjacent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% <em>no</em></td>
<td>17.3%</td>
<td>1.6%</td>
<td>16.2%</td>
</tr>
<tr>
<td>#</td>
<td>907/4,398</td>
<td>36/1,417</td>
<td>943/5,815</td>
</tr>
</tbody>
</table>

Table 2: Probability of *no* by Adjacency

3. 2. 2 Transitivity Restriction

Watanabe (1996) proposes a Transitivity Restriction (TR hereafter) where if a direct object exists as an argument of the predication in the embedded clause, as in (9), genitive *no* cannot appear in the same embedded clause as a subject marker.

(9)

a. [[Ken-*ga* hon-o katta] mise
KEN-NOM book-ACC buy store
‘the store where Ken bought a book’

b. [[hon-o Ken-*ga* katta] mise
book-ACC KEN-NOM buy store
‘the store where Ken bought a book’

If TR indeed affects GNC, the rate of *no* in clauses with a direct object should be lower than in clauses without a direct object. The result is given in Table 3. As it shows, no tokens of *no* have been found with a direct object in the data here. Hence, TR correctly captures an aspect of the distribution of *ga* and *no*.

\(^7\) Interestingly, another particle, *o*, which marks the accusative case and shows variation between *o* and zero in the colloquial speech of Tokyo Japanese, shows a similar effect of adjacency, so that the zero-form is more likely to be used when the verb and the object NP are adjacent to each other (Matsuda 1995).
Horie and Kang (2000) claim, on the basis of their corpus-based analysis, that *ga* is preferred if the predicate is stative. They argue that the acceptability of GNC follows the hierarchy of predicate type: *Verb > Existential predicate, Adjective > Copula* (in the order of ascending stativity). Although their claim is based on *frequency* and not on the *rate* of occurrence, their interesting insight is worth checking with our Minutes data. Here we categorized the predicates into verb, adjective, and nominal adjective. Although their claim is based on frequency and not on the rate of occurrence, their interesting insight is worth checking with our Minutes data. Here we categorized the predicates into verb, adjective, and nominal adjective. Contrary to Horie and Kang (2000), our result, shown in Table 4, indicates that it is the rate of *no* that follows the order of *Adjective > Nominal Adjective, Verb*, which is in the order of descending stativity. Notice that the result is consistent with the result of TR. In fact, TR is a direct antithesis of stativity, and there would be a discrepancy if we took the claim of Horie and Kang (2000) as it is: higher usage of *no* in lower stativity, that is, higher transitivity, does not go with TR.

### Table 3: Probability of *no* by TR

<table>
<thead>
<tr>
<th></th>
<th>With Direct Object</th>
<th>Without Direct Object</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% <em>no</em></td>
<td>0%</td>
<td>13.7%</td>
<td>12.5%</td>
</tr>
<tr>
<td>#</td>
<td>0/656</td>
<td>948/6,944</td>
<td>948/7,600</td>
</tr>
</tbody>
</table>

### Table 4: Probability of *no* by type of predicate

<table>
<thead>
<tr>
<th></th>
<th>Adjective</th>
<th>Nominal Adjective</th>
<th>Verb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% <em>no</em></td>
<td>30%</td>
<td>14%</td>
<td>14%</td>
<td>16.20%</td>
</tr>
<tr>
<td>#</td>
<td>222/718</td>
<td>17/121</td>
<td>704/4,976</td>
<td>943/5,815</td>
</tr>
</tbody>
</table>

### 3.2.5 Speech Style

Nakagawa (1987) states that there is a style difference between *ga* and *no*, and it is often

---

8 Because of space limitations, we do not mention the copula, but we have also investigated the copula. The probability of *no* is very low because of certain reasons, discussed in detail in Nambu (2005a).
said that written language and formal speech promote the use of *no* instead of *ga*. To check the style difference in this research, we considered the kind of session (plenary session versus other committee) and type of House (House of Representatives or *Shugiin* versus Councilors or *Sangiin*). With respect to the former, there is a difference between plenary session and other committees in the degree of formality. A Plenary session provides a more formal environment than others because all Diet members are required to attend it and it is broadcast nationwide. The prediction from this difference is that the probability of *no* in a plenary session would be higher than in others.

Our study, however, failed to find any clue of the style effect in terms of kind of session and type of House. One can think of several possible explanations for this result, but most probably, the style difference may be subtler than is detectable with our broad taxonomy. Indeed, our scheme cannot, in principle, capture the style difference that should exist within the same House or session. In order to pursue the style issue in the Diet Minutes, then, we must come up with some means to precisely locate a given speech on a style gamut, and we leave this for future research.

4. Conclusion

We attempted to verify Harada’s language change hypothesis regarding the GNC variation and demonstrated, using the Diet Minutes data, that the linguist’s insight 30 years ago was correct. The change, we argued, is moving toward the complementarity of the two particles, which originally showed a similar distribution. The GNC was also found to be affected by several internal and external factors. For the internal factors, the adjacency between the subject NP and the predicate, TR, and the stativity of the predicate are all strong factors affecting the rate of *no*. In contrast, only the birth year of the Members had any notable effect on the GNC variation, although the style effect still leaves some room for exploration.

The fact that GNC is involved in an ongoing change suggests several interesting questions. First of all, since GNC itself is a rather famous syntactic phenomenon in Japanese, there rises the question of the reliability of numerous grammatical judgments that have been used as data during the past 30 years. Naturally, younger linguists may well prefer *ga* to *no* in certain syntactic positions than the older ones, but the issue has never been addressed in the field of Japanese syntax. Once we know that GNC itself is in transition, we should be wary of using introspective judgments on GNC without considering the background of the linguist.

Second, GNC provides us with a rare chance for verifying the Constant Rate
Hypothesis (Kroch 1989) with data from a change in progress. Being a hypothesis on syntactic changes, most of the evidence for the hypothesis comes from past changes where only documentary evidence is available. But it is from change in progress that linguists can learn about the way language changes in the most detailed and empirically satisfiable way (Labov 1994). GNC, although it is approaching its final stage (cf. Fig. 1), gives us the most reliable data in this sense. If not for the Minutes data, we could collect natural speech from sociolinguistic interviews, or make a questionnaire-style survey to look into the consciousness of the speakers regarding the change in progress.

Finally, there is the issue of locating the change within an individual. The classical model of language change seeks its source at the time of language acquisition where the child remodels the input data as its own new grammar. Such a model excludes the possibility of language change after acquisition, but there is some evidence that casts doubt on this point (Labov 1982). The fact that GNC is an ongoing change and that the Minutes are a goldmine of transcribed speech data makes it a prime experimental ground to check whether a Member (an adult) shows any significant change in rate of no in the course of his tenure, which could be as long as 30 years. At this stage, we can only give a rough sketch of these problems, but we hope that the current paper managed to mark the beginning of such a research program.

References


