

Corpus-based Study of *Ga/No* Conversion

Satoshi Nambu

1. Introduction

One of the most prominent case alternations in Japanese is *ga/no* conversion (henceforth GNC). As shown in (1) and (2), a genitive particle *no* is variably substituted for a nominative particle *ga* in certain 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
 ‘the place where the dog slept’
- (2) *Ken-ga/*no hon-o kaita.*
 Ken-NOM/GEN book-ACC write
 ‘Ken wrote the book.’

GNC was first noted by Harada (1971), who historically derived the genitive construction from the nominative one. In this paper, however, I 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 have emerged in almost every grammatical paradigm proposed to date. The missing points in previous approaches, however, are empirical verification of Harada's (1971) language change hypothesis, and rigorous specification of which factors affect GNC. I will work on them utilizing quantitative methods.

Harada's (1971) hypothesis says that there is indeed an ongoing change, whereby the speakers of Tokyo Japanese increasingly prefer *ga* to *no* in relevant environments. This intriguing hypothesis has been left untouched for about 30 years. In section 3, I will attempt to verify Harada's hypothesis quantitatively. In a word, I will conduct a real

time verification of change in progress, as many previous studies have done, such as Fowler's restudy of (r) in New York City (Fowler 1986), and the three surveys at 20-year intervals by The National Institute for Japanese Language (The National Institute for Japanese Language 1953, 1974, Yoneda 1997). It is particularly worth noting that the corpus which I have used, as seen in section 2.1, makes it easy to access decades of speech data, compared to the usual studies using real time evidence which face a number of problems, such as the comparability of the data, and the need for time to acquire them (Bailey 2002).

In section 4, I will identify the language external/internal factors which affect GNC by a quantitative method involving the concept of variable rules. There has been some quantitative research such as Horie and Kang (2000) and Maki et al. (2004), but the former discussed GNC based on *frequency* and not on the *rate* of occurrence, and the latter does not consider differences between dialects¹. Furthermore, previous research on GNC discusses several factors which affect GNC, but it does not take into consideration the correlations between them, which mask the independent effects of each factor. Taking care of these problems, I adopt logistic regression to identify the factors from the statistical point of view.

Logistic regression has often been used in the form of VARBRUL in variation theory, and also in other fields such as the research on dative alternation by Bresnan et al. (forthcoming). Logistic regression can specify several factors which affect a certain phenomenon, and compute values allowing a comparison of the degree of each effect. In this paper, I will conduct logistic regression analysis based on the same corpus as that I used in research verifying Harada's (1971) hypothesis, which has the advantage of allowing the selection of speakers by hometown. I consider several factors to untangle the complex correlation between them, taking advantage of accumulated knowledge furnished by recent syntactic analyses of GNC.

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

- i. *Ame-no futte kita.*
rain-GEN fall become
'It's started to rain.'
- ii. *Doroboo-no haïtta.*
thief-GEN come in
'The thief came in.'

(The National Institute for Japanese Language 1989)

2. Data

2.1 Corpus

To verify whether language change is indeed occurring requires solid empirical data. Almost all of the huge number of syntactic treatments of GNC are based on introspection, while only a few rely on usage or surveys of judgments, such as Horie and Kang (2000) and Maki et al. (2004). Both of them, however, are purely synchronic in nature, and as such, they do not mention Harada's hypothesis. Obviously, I need a diachronic corpus of substantial size with historical depth, covering generations of the same dialect, so that any syntactic change can be clearly observed and analyzed.

For this purpose, I draw my data from the Minutes of the Japanese Diet. 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.² Its most prominent feature, for the purpose, is that the Minutes have speech data spanning about 60 years, starting from 1947 (Matsuda 2004), and as such it provides 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, I restrict my data to the speech of members who are also native speakers of the dialect. By checking the hometown of the Diet members,³ I came up with 180 subjects.⁴ Then I sampled one Diet member for each birth year to create a dataset with chronologically equal proportions. This yielded 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, I took 100 tokens of the variable, making the final sample size 7,600 tokens of which 938 (or 12.3%) were marked with *no*.⁵

² The website is accessible at 'Full-text Database System for the Minutes of the Diet' (<http://kokkai.ndl.go.jp/>).

³ The references I used are as follows: *Kizokuin/Sangiin Giin Meikan* (Shugiin/Sangiin 1990a), *Shugiin Giin Meikan* (Shugiin/Sangiin 1990b), *Gendai Seijika Jinmei Jiten* (Nichigai Associates 1999), *Seijika Jinmei Jiten* (Nichigai Associates 2003), *Kokkai Binran* (Nihon Seikei Shimbunsha 1998, 1999, 2000, 2001, 2002, 2003, 2004).

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

2.3 Envelope of variation

Before counting occurrences of *ga* and *no* in the data, I needed to delineate the environments where variation is possible (the envelope of variation). Considering previous syntactic research, the environments which I include in my data as GNC can occur are followings.

(3) Adnominal Clauses

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.’

(4) *Made* or *yoru* subordinate clauses

a. *Taro-ga/no kuru made suwatte iyooka.*
 Taro-NOM/GEN come until sit be
 ‘Let’s sit until Taro comes.’

b. *Hanako-ga/no kuru yori hayaku Taro-ga tuita.*
 Hanako-NOM/GEN come than earlier Taro-NOM arrive
 ‘Taro arrived before Hanako came.’

(5) *To yuu* and *to no* apposition clauses

Taroi-ga buzi-datta toyuu/tono uwasa
 Taro-NOM safe-be COMP rumor
 ‘the rumor that Taro is safe’

3. Ongoing Change in GNC

To test Harada’s hypothesis, I calculated the rate of *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 Diet Member. The gradual decline of the overall trend is fairly 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.

someone else’s speech. Also sections where Diet members clearly read texts were excluded from the sample.

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

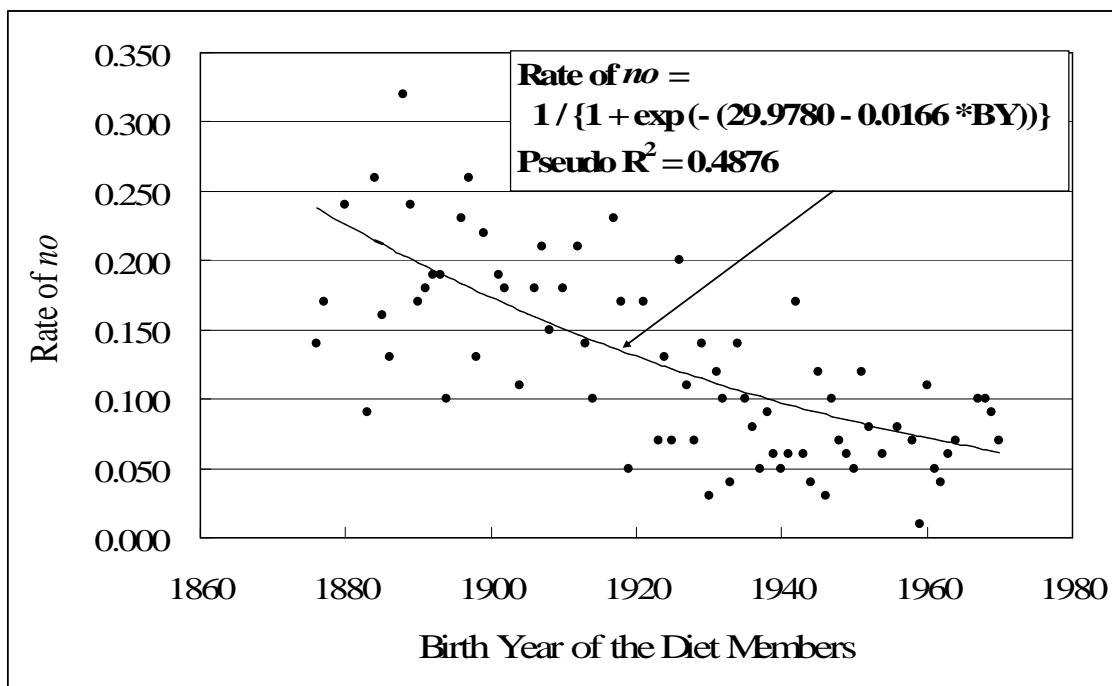


Table 1. Historical distribution of *ga* and *no*

(a black cell represents non-occurrence of the given form in the environment)

	Before Kamakura / Muromachi Era (-1192)		Kamakura / Muromachi Era (1192-1573)		Present	
	<i>GA</i>	<i>NO</i>	<i>GA</i>	<i>NO</i>	<i>GA</i>	<i>NO</i>
Between Nominals						
Embedded Clauses						
Main Clauses						

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).⁶ At the first stage before the Kamakura and Muromachi periods (-1192), *ga* and *no* show a similar distribution, occurring between nominals and in embedded clauses (represented by white cells in the table). They then underwent a change during the Kamakura and Muromachi periods (1192-1573), as *ga* emerged in main clauses as a subject marker. In present-day Japanese *ga* cannot be used between nominals, and in embedded clauses, as I observed above, a change is now in progress (shaded gray).⁷ 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 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.⁸

Despite the observation of ongoing change, I suggest that the change did not begin only in the present time. Here I would like to hypothesize about the time when the change in GNC may have begun by considering its relation to the emergence of *ga* in main clauses. The hypothesis is that the change in GNC began around the same time that *ga* in main clauses emerged. As Ono (1977) and Kikuta (2006) mention, the rise of *ga* as a subject marker has been underway since the Muromachi period. I can thus assume that the change in GNC has progressed following this trend across all environments since that time. Ono (1977) gives evidence to support this hypothesis, through an investigation of two versions of the epic *Heike Monogatari* (*Tale of the Heike*) written in different periods. Comparing two books, *Kakuitibon Heike Monogatari* written in the Kamakura period and *Amakusabon Heike Monogatari* written

⁶ Of course Table 1 is a brief summary of the historical development of *ga* and *no*. When we talk about the historical transition of Japanese, we should keep in mind that there was a migration of the political center from Kyoto to Tokyo (known at the time as Edo). It is necessary to consider to what extent change in this period reflects historical change and how much is due to the shift in dialect.

⁷ Here I show the context in the Edo period (1603-1867), just before the present that I investigate. Yamada (1936), studying the language in “Ukiyoburo” and “Ukiyodoko” by Sanba Shikitei (1776-1822), shows that the proportion of *ga* to *no* in embedded clauses which precede an NP is 12.4% (25/202) *ga* and 87.6% (177/202) *no* in those books. Of course it should be taken into account that Yamada (1936) and my study probably differ in environments where *ga* and *no* are counted as GNC. This suggests that GNC has changed from a situation where *no* was predominant in the relevant clauses to the current state that I have seen.

⁸ At this point, I would like to mention another change in the roles of *ga* and *no*. As pointed out in Kinsui (1984), there was a difference between *ga* and *no* in honorific usage, but the difference disappeared in modern Tokyo dialect. It is reasonable to assume that this disappearance provided an impetus for the change in GNC, since the disappearance brought about an overlap of *ga* and *no* in embedded clauses. However this reasoning might also be reversed, with the change in GNC interpreted as causing the disappearance of the different roles of *ga* and *no*.

in the Muromachi period, Ono observes that the former used *no* as a subject marker in some embedded clauses, but the latter used *ga* in the same clauses. Although I need to examine the relative proportion of the change between those books, the difference implies that the change in GNC had already begun in that period.

4. Logistic Regression Analysis

I conducted a logistic regression analysis in order to identify which factors affect GNC. As in Rietveld and van Hout (1993), logistic regression builds up a regression model which consists of several independent variables and one dependent variable. In the case of GNC, the dependent variable corresponds to the variable which has two values *ga* and *no*, and independent variables correspond to factors which affect GNC. Formulating the model, logistic regression can compute values of each effect of the independent variables, and we can compare their values to specify which factor has a stronger effect than others. In addition, the regression model can evaluate the accuracy rate of predicting the choice of the variables. In Bresnan et al.'s (forthcoming) words, the logistic regression analysis can predict the choice of the variables from several factors.

To conduct the logistic regression analysis, I use the same 7,600 tokens as in the discussion of language change in GNC, and add codes relating factors to the data. Based on the data, logistic regression can statistically determine whether the factors affect GNC or not. Taking advantage of previous research on GNC, I consider several factors to clarify their effects. I briefly describe the language internal/external factors in section 4.1 and 4.2. After glancing over the factors, I explicate the procedure of logistic regression analysis in section 4.3, and see the result in section 4.4.

4.1 Language External Factors

4.1.1 Birth Year

As discussed in section 3, birth year is a measure for the change in GNC. Including it as an independent continuous variable, logistic regression will compute its independent effect and we can confirm the change by looking at the concrete figure.

4.1.2 Session and House Type

Although he does not say clearly what the difference is, Nakagawa (1987) states that there is a style difference between *ga* and *no*, and it is often said that written language and formal speech promote the use of *no* instead of *ga*. To check the style difference in this paper, I consider session type (plenary session versus other committees) and House

type (House of Representatives or *Shugiin* versus Councilors or *Sangiin*). With respect to the former, there is a difference between plenary sessions and other committees in the degree of formality. Plenary session provides a more formal environment than others because all Diet members are required to attend it and it is broadcast nationwide. Although specific effects on GNC by style remains unclear, I consider the above two factors in a pilot study.

4.2 Language Internal Factors

4.2.1 Adjacency

Harada (1971) pointed out that the existence of intervening elements between the subject NP and its predicate affects the acceptability of GNC. He insisted if there exist some intervening elements, genitive *no* cannot appear as a subject marker as in (6). To consider the effect of adjacency, I categorize GNC into two environments, adjacent and non-adjacent.

- (6) *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’

4.2.2 Transitivity Restriction

Watanabe (1996) proposes a transitivity restriction, where if a direct object exists as an argument of the predication in the embedded clause, as in (7), genitive *no* cannot appear in the same embedded clause as a subject marker. Comparing two environments, one with a direct object and the other without, I consider the transitivity restriction as an independent variable.

- (7) 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’

4.2.3 Predicate type

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 a

hierarchy of predicate type: *Verb* > *Existential predicate*, *Adjective* > *Copula* (in 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 the Minutes data. To consider predicate type in the Minutes data, I add adjectival verb as a category in my data to their categorization.

4.2.4 Animacy

Croft (1990) discusses the effect of animacy on grammatical phenomena. Using cross language observation, he gives the animacy hierarchy shown in (8), for number distinction. More highly ranked items on the hierarchy are more likely to show number distinction. It expects NP ranked higher in the hierarchy should be more likely to allow the use of *no*.

(8) Animacy Hierarchy in Croft (1990)

first, second-person pronouns > third-person pronoun > proper names > human
common noun > nonhuman animate common noun > inanimate common noun

4.2.5 Subject NP Type

An effect of the subject NP type should be examined because of the fact that the variable accusative case markers (*o*/zero) in the colloquial speech of Tokyo Japanese is partially determined by the object NP type (Matsuda 1995). The use of the zero-form increases in the order of the hierarchy, *wh*-word > lexical noun > pronoun > nominal clause. It can be inferred that GNC would also have preference for the NP type. I follow the above categorization of NP type, and consider the effect of subject NP type on GNC.

4.2.6 Negativity

Givón (1979) states that an environment which includes negatives is conservative in language change, but only in creative-elaborative change, not in simplificatory change. In the case of GNC, the change would be categorized as simplificatory change as in the discussion in section 3. This suggests that GNC is not related to conservatism in negative environments, but I consider negativity so as to confirm that it does not affect GNC.

4.2.7 *To yuu*/*To no* Clauses

Inoue (1976) observes that GNC cannot occur in *to yuu* and *to no* clauses, as shown in (9). Ura (1993), however, claims that GNC can occur in *to yuu* clauses if the head

noun is not a deverbal noun. To confirm their claim, the existence of *to yuu* and *to no* is one of the factors.

- (9) *karera-ga/*no buzi-dat-ta* *to-yuu/to-no* *sirase*
 they-NOM/GEN safety-COP-PAST COMP news
 ‘the news that they were safe’

4.2.8 Grammatical Noun

The head noun of the embedded clauses where GNC occurs is categorized as either a grammatical noun or other nouns. Masuoka and Takubo (1992) mention that grammatical nouns are bleached and their main role is grammatical. If the grammatical nouns do not have enough characteristics of nouns, the expected result is that the use of *no* with grammatical nouns is lower than that with other nouns, because the adnominal clause is a crucial environment for GNC. The nouns *koto*, *no*, *tokoro*, *wake*, *yoo*, and *mono* are considered as grammatical nouns in this paper.

It should be also considered whether the clause has a head noun or not. Watanabe (1996) and Kikuta (2002) claim that *yori* and *made* clauses which do not have head nouns can exceptionally be environments for GNC.

- (10) a. *Yori* clause
Basu-ga/no kuru made suwatte iyoo-ka.
 bus-NOM/GEN come-PRES until sit be-INTEROG
 ‘Let’s sit until the bus comes.’

- b. *Made* clause
Kyaku-ga/no kuru yori hayaku nimotu-ga tui-ta.
 customer-NOM/GEN come-PRES than earlier package-NOM arrive-PAST
 ‘Let’s sit until the bus comes.’

Kikuta (2002)

4.2.9 No Precedence Environment

“No precedence environment” refers to an instance when *ga/no* follows an NP which has *no* at the end of the word, as in (11). If *ga* alternates with *no*, the concatenation *no ga* changes to *no no*. It can be expected that the sequence of identical sounds *no no* would be avoided in speech.

- (11) *ooki-i mono-ga/no a-ru mise*
 big-PRES thing-NOM/GEN exist-PRES shop
 ‘the shop which has big things’

4.3 Procedure

In the last section, I briefly observed 12 factors which are candidates for independent variables summarized in Table 2. The table shows categories of each variable, except for birth year which corresponds to a continuous variable. Such information is coded in 7,600 tokens.

Before conducting logistic regression, we should pay attention to three things. First, “knockout factors” should be excluded from the analysis. As discussed in Paolillo (2002), a knockout factor means an independent variable which has a category yielding only one fixed value for the dependent variable. Such a variable makes for wrong calculation in the analysis. In Table 2, the transitivity restriction becomes a knockout factor, because I do not have any examples of *no* with a direct object in my data. In addition, there are no examples of *no* with a head noun *wake* or in *yori* clauses. Therefore I do not include the examples of environments with a direct object or *wake* or *yori* clauses to avoid their undesirable effects on the logistic regression analysis.

Second, the logistic regression analysis should not include factors which exist in under 30 tokens in the data, because they also lead to wrong calculation in the output of the analysis (Rietveld and van Hout 1993). Here, examples of a *to no* clause totals 8 and examples of a *made* clause totals 19 in the data, and therefore I exclude the factors “*to no* clause” and “*made* clause” from independent variables.

Third, dependency in independent variables should not be included in the analysis. Dependency indicates that one value of an independent variable is dependent on values of other variables. In logistic regression analysis, the effects are mistakenly calculated. In my data, the dependency is observed in three candidates, subject NP type, animacy, and *no* precedence environment.⁹ Instead of three factors, I consider one complex factor “SAN” multiplied by the three factors. In this factor, the subject NP is categorized as: non-animate common noun, human pronoun, clause, human common noun, non-human common noun with *no* ending, clause with *no* ending, and proper noun.

I also set aside the factor negativity from independent variables. My preliminary study shows that negativity is strongly affected by adjective as a predicate type because

⁹ If the subject NP falls into a personal pronoun in animacy, it always falls into a pronoun in subject NP type, and if the subject NP falls into a pronoun, it always falls into a “without *no* ending” in *No* precedence environment because there is no pronoun with *no* ending.

Table 2. Candidates for Independent Variables and Their Categories

Candidates for Independent Variables	Categories of Independent Variables
Birth Year	(continuous value)
House Type	upper, lower
Session Type	plenary, committees
Transitivity Restriction	with a direct object, without a direct object
Adjacency	adjacent, non-adjacent
Predicate type	adjective, adjectival verb, verb, existential verb, copula
Animacy	first/second-person pronouns, third-person pronoun, proper names, human common noun, nonhuman animate common noun, inanimate common noun
Subject NP Type	wh-word, lexical noun, pronoun, nominal clause
Negativity	Negative, non-negative
<i>To Yuu/To no</i> Clauses	with <i>to yuu/to no</i> , without <i>to yuu/to no</i>
Head Noun	<i>no, koto, tokoro, wake, mono, yoo</i> , others
<i>No</i> Precedence Environment	subject NP with <i>no</i> ending, without <i>no</i> ending

“adjective” includes *nai*, which is also categorized as a negative environment. Eventually, I use 8 factors as independent variables.

As mentioned in Rietveld and van Hout (1993), for research on language variation, the logistic regression analysis has been conducted utilizing VARBRUL in the history of variation theory. In this paper, however, I adopt SPSS instead of VARBRUL, because SPSS can treat continuous independent variables, such as birth year, as independent variables, while VARBRUL can treat only categorical variables. Computing the regression model, the SPSS setting for logistic regression is taken as a stepwise analysis based on likelihood ratio. The stepwise analysis can exclude factors which do not affect GNC from the regression model.

4.4 Result

Table 3 is the output of logistic regression. We realize that three independent variables, House type, session type, and negativity are not included in this table. This is because the stepwise method statistically judged that these factors do not affect GNC, and therefore these factors should be excluded from the regression model. In addition, the output shows 8 factors in order of strength of effect on GNC. Table 3 indicates that

Table 3. The output of logistic regression by SPSS

	B	S.E.	Wald	d.f.	P	Exp(B)	Confidence interval of Exp(B)	
							Lower point	Upper point
Adjacency	2.42	.19	167.27	1	.000	11.19	7.76	16.14
<i>To yuu</i> clause	3.11	.28	122.40	1	.000	22.32	12.88	38.69
Birth Year	-.02	.002	99.81	1	.000	.98	.98	.99
Predicate Type			179.55	4	.000			
(Adjective)	1.025	.33	9.49	1	.002	2.79	1.45	5.35
(Verb)	-.56	.32	3.03	1	.082	.57	.31	1.07
(Existential Verb)	.14	.33	.17	1	.680	1.15	.60	2.20
(Copula)	-2.41	.79	9.23	1	.002	.09	.02	.43
SAN			120.15	6	.000			
a	-1.06	.27	15.73	1	.000	.35	.20	.58
b	.31	.29	1.13	1	.288	1.37	.77	2.43
c	-1.59	.31	27.39	1	.000	.203	.11	.37
d	-.82	.30	7.24	1	.007	.44	.24	.80
e	-2.21	.56	15.57	1	.000	.11	.04	.33
f	-3.24	1.10	8.74	1	.003	.04	.005	.34
Head Noun			401.87	5	.000			
<i>Tokoro</i>	.89	.26	11.57	1	.001	2.42	1.46	4.03
<i>Yoo</i>	-.79	.21	14.22	1	.000	.45	.30	.68
<i>No</i>	-3.42	.25	182.65	1	.000	.03	.02	.05
<i>Koto</i>	-.74	.22	11.03	1	.001	.48	.31	.74
Others	.08	.19	.19	1	.661	1.09	.75	1.57
Constant	27.70	3.29	70.95	1	.000	1.07E+012		

adjacency has the greatest effect. Each category of SAN is a (inanimate common noun), b (human pronoun), c (clause), d (human common noun), e (non-human common noun with *no* ending), or f (clause with *no* ending).

Table 3 gives values of the regression coefficient (B), the standard error for the Wald test (S.E.), Wald test for regression coefficients (Wald), the degree of freedom for the Wald test (d.f.), the statistically significant difference as probability computed by the Wald test (P), odds ratio (Exp(B)), and the confidence intervals of Exp(B). Here, I focus

attention on the values of $\text{Exp}(B)$. The odds ratio is the odds of an event occurring, defined as $p/(1-p)$. Looking at this value, we can figure out the actual effect on GNC by each factor. In the case of a continuous variable birth year, the value 0.98 of $\text{Exp}(B)$ means its rate of change with increase of one year. In the case of the categorical variable, the value of $\text{Exp}(B)$ is calculated by the comparison between the categories of each factor, regarding one of the categories as a reference. For example, adjacency has two categories, adjacent environments and non-adjacent environments, and the non-adjacent one is taken as a reference. The value 11.9 of $\text{Exp}(B)$ indicates that the rate of *no* in the adjacent environments is 11.9 times more than that in the non-adjacent environments. This implies that adjacency has a large effect on GNC. Table 3 shows the other factors as follows. “*To yuu* clause” has its reference in clauses with *to yuu*, and the value of $\text{Exp}(B)$ shows that the existence of *to yuu* has a negative effect on the use of *no* in GNC. “Predicate type” has as reference an adjectival verb, and the value of $\text{Exp}(B)$ shows that the use of *no* is preferable in the order, adjective > existential verb > adjectival verb > copula. The result does not show an effect of stativity as Horie and Kang (2000) claims. SAN has as reference a proper noun. The value of $\text{Exp}(B)$ shows that three factors, subject NP type, animacy, and a *no* precedence environment have independent effects. It shows that the use of *no* is preferable in the order, human pronoun > proper noun > human common noun > inanimate common noun clause > inanimate common noun with *no* ending > clause with *no* ending. “Head noun” has as reference the noun *mono*. The value of $\text{Exp}(B)$ shows that the grammatical nouns as head nouns do not have a common effect on GNC.

The computed regression model is evaluated as a goodness of fit. I conduct the Hosmer-Lemeshow test to calculate the goodness of fit. The value is described as X^2 , and the goodness of fit judges the model as a good model when the test shows a statistically insignificant difference. The result is $X^2 = 7.311$, d.f.= 8, and $p = .503$, and this shows that it is not statistically significant. Therefore, the test shows that the model is good. For other evaluations, the value of Nagelkerke’s R^2 is 0.453 and the probability of prediction of GNC is 87.2%.

5. Conclusion

In the first half of this paper, I verified the existence of ongoing change in GNC, observing that the use of *ga* is increasing. In the latter half, I conducted logistic regression to clarify which factors affect GNC, and considered their correlation. The result ascertained the effects of 8 factors.

In this paper, I used the Minutes of the Japanese Diet for observing language change

because of its recorded speech over such a long period, but the analysis of GNC needs to be done using other types of corpus, so as to compare the results and confirm the factors relevant to language variation. To compare spoken language in the minutes with written language, *Aozora Bunko*, which provides books in computerized formats on its website for free, is useful. As a corpus which has tagged data of spoken language, the Corpus of Spontaneous Japanese can be used to analyze GNC. Furthermore, a questionnaire survey and fieldwork on GNC might contribute to clarifying the difference between language knowledge and its use, and the complex of language change in its real usage.

In addition, the result of this research should be utilized for the theoretical formalization of GNC. In terms of a historical development of case in Japanese, the formalization of GNC will help to determine the cues for the course of language change. The theoretical analyses of GNC have been done by many researchers such as Miyagawa (1993), Hiraiwa (2001), and Ochi (2001). Their generative syntactic analyses are based on introspection, and the statistical knowledge could contribute to confirming the analysis of such non-categorical variation.

References

- Bailey, Guy. 2002. "Real and Apparent Time". *The Handbook of Language Variation and Change* ed. by J.K. Chambers, Peter Trudgill and Natalie Schilling-Estes, 312-332. Oxford: Blackwell.
- Croft, William. 1990. *Typology and Universals*. Cambridge University Press.
- Doi, Yoichi. 1982. "Shukakujoshi [Nominative case marker]". *Koza Kokugoshi 4: Bunposhi* ed. by Hiroshi Tsukishima, 398-402. Tokyo: Taishukan.
- Fowler, Joy. 1986. *The Social Stratification of (r) in New York City Department Stores, 24 years after Labov*. New York University. ms.
- Givón, Talmy. 1979. *On Understanding Grammar*. San Diego: Academic Press.
- Harada, Shin-Ichi. 1971. "Ga-No Conversion and Idiolectal Variations in Japanese". *Gengo Kenkyu* 60, 25-38.
- Hiraiwa, Ken. 2001. "On Nominative-Genitive Conversion". *MIT Working Papers in Linguistics* 39, 66-123.
- Horie, Kaoru & Bongshik Kang. 2000. "Action/State Continuum and Nominative-Genitive Conversion in Japanese and Korean". *Modern Approaches to Transitivity* ed. by Ritsuko Kikusawa & Kan Sasaki, 93-114. Tokyo: Kuroshio Publishers.
- Inoue, Kazuko. 1976. *Henkei Bunpo-to Nihongo [Generative Grammar and Japanese]*.

- Tokyo: Taishukan.
- Kikuta, Chiharu. 2002. "Ga/No Kotaigensho-no Hihaseitekibunseki: Jutsugorentaiki-no Meishisei [Ga-No Conversion in the New Light: A Non-derivational, Mixed Category Analysis]". *Doshisha Studies in English* 74, 93-136.
- _____. 2006. "The Development of the Nominative Case Marker GA: A Competing Constraints Approach". *Doshisha Studies in English* 79, 61-104.
- Kinsui, Satoshi. 1984. "Tenioha-no Keigoho [Honorific usage of Tenioha particles]". *Kenkyu Siryo Nihon Bunpo 9 Keigohohen*, 101-126. Meiji Shoin.
- Konoshima, Masatoshi. 1970. "'Ga'-no Imi/Yoho [Meaning/Usage of 'ga']". *Gekkan Bunpo* 2, 10-17.
- Maki, Hideki, Tamami Morishima & Jessica Dunton. 2004. "A Statistical Analysis of the Nominative/Genitive Alternation in Japanese: A Preliminary Study". *Bulletin of the Faculty of Regional Studies, Gifu University* 14, 87-119.
- Malouf, Robert. 2000. *Mixed Categories in the Hierarchical Lexicon*. Stanford: CSLI Publications.
- Masuoka, Takashi & Takubo Yukinori. 1992. "Kiso Nihongo Bunpo [Basic Japanese Grammar]". Tokyo: Kurosio Publisher.
- Matsuda, Kenjiro. 1995. *Variable Zero-Marking of (o) in Tokyo Japanese*. Ph.D. dissertation, University of Pennsylvania.
- _____. 2004. "Gengo Shiryo-to shitenno Kokkai Kaigiroku Kensaku Shisutemu [The On-line Full-text Database of the Minutes of the Diet: Its Potentials and Limitations]". *Theoretical and Applied Linguistics at Kobe Shoin* 7, 55-82.
- Matsunaga, Setsuko. 1983. *The Development of Case Marking in Japanese*. Master's thesis, Ohio State University.
- Miyagawa Shigeru. 1993. Case-checking and Minimal Link Condition. *MIT Working Papers in Linguistics* 19, 213-254.
- Nakagawa, Yoshio. 1987. "GA-NO Conversion in Japanese Particles". *Bulletin of the Kyoto University of Foreign Studies* 28, 309-318.
- The National Institute for Japanese Language, ed. 1953. *Chiiki Shakai-no Gengo Seikatsu: Tsuruoka-ni okeru Jittai Chosa* [Language Life of a Region: Field Survey of Tsuruoka City]. Tokyo: Shuei Shuppan.
- _____. 1974. *Chiiki Shakai-no Gengo Seikatsu: Tsuruoka-ni okeru 20 Nenmae-tono Hikaku* [Language Life of a Region: A Comparison of the Past and the Present language situation of Tsuruoka City at a 20-year Interval]. Tokyo: Shuei Shuppan.
- _____. 1989. *Grammar Atlas of Japanese Dialects*. Tokyo: Finance Ministry Printing

Bureau.

- Nichigai Associates, ed. 1999. *Gendai Seijika Jinmei Jiten*. Tokyo: Nichigai Associates.
- _____. 2003. *Seijika Jinmei Jiten*. Tokyo: Nichigai Associates.
- Nihon Seikei Shimbunsha, ed. 1998. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 1999. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 2000. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 2001. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 2002. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 2003. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- _____. 2004. *Kokkai Binran*. Tokyo: Nihon Seikei Shimbunsha.
- Ochi, Masao. 2000. "Move F and Ga/No Conversion in Japanese". *Journal of English Asian Linguistics* 10, 247-286.
- Ono, Susumu. 1977. "Shukaku Joshi GA-no Seiritsu [Emergence of GA as a nominative marker]". *Bungaku* 45:7, 102-117. Iwanami Shoten.
- Rietveld, Toni and Roeland van Hout. 1993. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin : Mouton de Gruyter.
- Shugiin/Sangiin, ed. 1990a. *Kizokuin/Sangiin Giin Meikan*. Tokyo: Finance Ministry Printing Bureau.
- _____. 1990b. *Shugiin Giin Meikan*. Tokyo: Finance Ministry Printing Bureau.
- Ura, Hiroyuki. 1993. "L-Relatedness and Its Parametric Variation". *MIT Working Papers in Linguistics* 19 ed. by Colin Phillips, 377-399.
- Watanabe, Akira. 1996. "Nominative-Genitive Conversion and Agreement in Japanese: A Cross-Linguistic Perspective". *Journal of East Asian Linguistics* 5, 373-410.
- Yamada, Masanori. 1936. *Edokotoba-no Kenkyu: Ukiyoburo/Ukiyodoko-no Goho* [the study of Japanese in Edo period: the diction in Ukiyoburo/Ukiyodoko]. Tokyo: Musashino Shobo.
- Yoneda, Masato. 1997. "Survey of Standardisation in Tsuruoka, Japan: Comparison of Results from Three Surveys Conducted at 20-year Intervals". *Japanese Linguistics* 2, 24-39.