The Nominative-to-Accusative Shift in Modern Japanese: A Diachronic Observation

SATOSHI NAMBU
Monash University

SHIN-ICHIRO SANO
Keio University

DAVID Y. OSHIMA
Nagoya University

1 Introduction

Some Japanese predicates, such as potential verbs, allow their object to be marked with the nominative marker *ga*, instead of the accusative marker *o* (which is the default marker of the direct or sole object).[1]

[1] The abbreviations in glosses are: ACC = accusative, ATTR = attributive, COP = copula, NOM = nominative, POT = potential, PRS = present, PST = past, QUOT = quotative particle, TH = thematic *wa* (topic/ground-marker).

Japanese/Korean Linguistics 26.
Edited by Shoichi Iwasaki, Susan Strauss,
Shin Fukuda, Sun-Ah Jun, Sung-Ock Sohn, and Kie Zuraw.

(1) *Mari wa sashimi ga/o tabe-rare-ru.*
M. TH sashimi NOM/ACC eat-POT-PRS
'Mari can eat sashimi.'

The choice between the two particles generally does not lead to a difference in meaning (but see Nambu et al. 2018 and references therein for some semantic and pragmatic effects it may induce).

Shibuya (1993) states that, throughout the known history of Japanese, the nominative marking on an object in the potential construction has been in decline. The reported trend conforms to Eythórsson's (2015) Case Directionality Hypothesis, which states that shift from marked (idiosyncratic) case marking pattern to unmarked (default) case marking pattern is more common than that in the other direction.

This study examines whether the shift from nominative to accusative case ('N-to-A' shift, for short) has progressed in recent years, employing a quantitative approach developed in variationist sociolinguistics (Tagliamonte 2012).

2 Data

Our data were drawn from two corpora developed by the National Institute of Japanese Language and Linguistics: (i) the Balanced Corpus of Contemporary Written Japanese (BCCWJ) (Maekawa et al. 2014) and (ii) the Corpus of Spontaneous Japanese (CSJ) (Maekawa 2004). The BCCWJ consists of approximately one hundred million words collected from texts published between 1975 and 2005 and of various genres such as newspapers, fiction, and online blogs. The CSJ consists of 661 hours of recorded speech, collected between 1999 and 2003 and amounting to approximately 7.5 million words.

We used only the two corpora's 'core' data sections, which come with richer and more reliable annotation. The core section of the BCCWJ consists of approximately one million words, and that of the CSJ consists of approximately half a million. The annotation of both corpora includes birth years of authors/speakers, information essential for identifying and examining ongoing changes based on the apparent-time method (Labov 1963, Cukor-Avila and Bailey 2013).

From these data sources, we extracted direct objects of predicates that have been said to, at least marginally, allow the *ga/o*-alternation ('*ga/o*-predicates' for short), namely: (i) potential verbs with derivational suffix (*r*)*e* or *rare*, (ii) potential verbs of the form: [verbal noun + DEKIRU], (iii) desiderative predicates with the morpheme *ta*, such as TABETAI 'want to

eat', (iv) DEKIRU 'be able to do', (v) WAKARU 'understand', (vi) HOSHII 'want', (vii) SUKI (+ copula) 'fond', and (viii) KIRAI (+ copula) 'not fond' (Tokieda 1950, Kuno 1973, Shibatani 1975, Iori 1995a, 1995b, Ikuta 1996, Sugai and Naruse 2006, Aoki 2008, Fujimura 2009).[2]

From the core data of the BCCWJ, we obtained 1,980 tokens of *ga*- or *o*-marked objects of these predicates. Only 452 among them are associated with a unique birth year; many tokens lack birth year information and some others are associated with multiple birth years because of co-authorship. In some parts of our analysis, only these 452, whose writers' birth years range from the 1890s to the 1970s, were considered. From the core data of the CSJ, we obtained 1,086 tokens, whose speakers' birth years range from the 1930s to the 1970s.

3 Analysis and Results

In our analysis, the following language-external and -internal factors were considered: (i) register (written vs. spoken), (ii) birth year of speaker/writer, (iii) predicate type, (iv) clause type (main vs. subordinate), and (v) adjacency of the predicate and its object.[3]

3.1 The Overall Trends

Figure 1 illustrates how the proportions of *ga*- and *o*-marking change across birth years, in written texts (top; N = 452) and speech (bottom; N = 1,086). The solid lines and shaded areas represent regression lines and 95% intervals, respectively.

These data do not involve a consistent decrease (increase) in the use of *ga* (*o*), and thus do not allow us to conclude that, with respect to the totality of the eight predicate types considered, the N-to-A shift (or a change in the opposite direction) took place in the past 100 years or so.

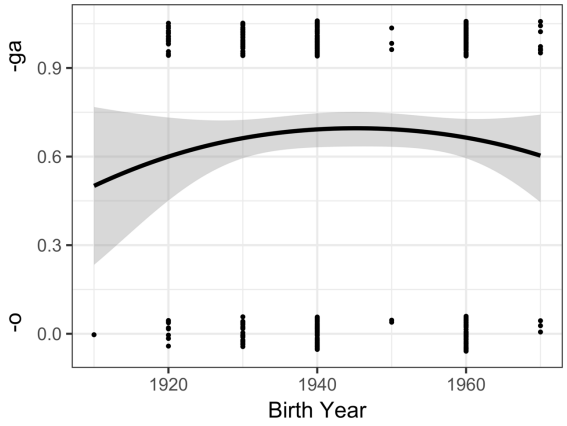
3.2 Potential Verbs

We also investigated how the proportion *ga*- and *o*-marking on objects of potential verbs, which Shibuya (1993) specifically discussed, might have changed. Again, no evidence could be found that the N-to-A shift progressed within the same period. We constructed a logistic regression model using the R environment (R Core Team 2016); the model involved only the data of potential verbs and had the following as factors: (i) birth year, (ii) adjacency, (iii) register, and (iv) clause type. The results did not indicate a sign of change (birth year: $z = 0.24$, $p = 0.81$). (Note that this does

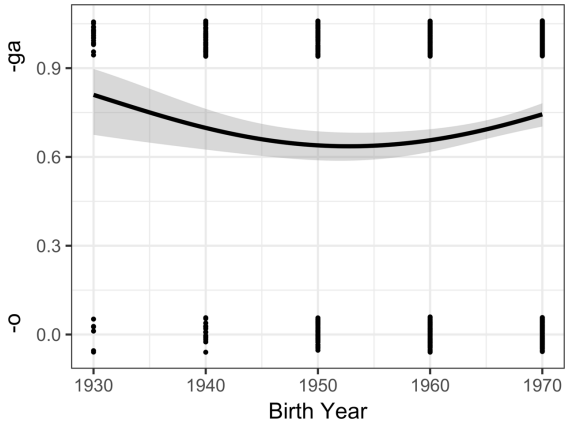
[2] Expressions in small capitals refer to lexemes.

[3] The last factor has been pointed out to sometimes affect the acceptability of a *ga*-marked object (see Nambu et al. 2018 and references therein).

not contradict Shibuya's (1993) remarks, because the periods he considered are different, starting from the Old Japanese period.)



BCCWJ (written)



CSJ (spoken)

FIGURE 1 Use of *ga* and *o* across birth year

3.3 Innovation in Subordinate Clauses

Several noteworthy findings, however, were obtained about the innovative pattern where predicates that are supposed (according to prescriptivism) to select only for a nominative object co-occur with an accusative object. Among the predicate types considered, the five lexical ones—DEKIRU, WAKARU, HOSHII, SUKI, and KIRAI—are such predicates. (2) is an example where SUKI (+ copula), which canonically selects for a *ga*-marked object, occurs with an *o*-marked object.

(2) *omatsuri* ***o*** *suki* *na* *kata*
festival ACC fond COP.ATTR person
'a person who likes festivals' (CSJ: S03F157)

Table 1 presents the breakdown of the frequency of *ga*-marked and *o*-marked objects for each predicate type, where the shaded cells correspond to the innovative pattern.[4]

	BCCWJ (1,980 tokens)		CSJ (1,086 tokens)	
	ga (1229)	*o* (751)	*ga* (757)	*o* (329)
Potentials	285	253	133	48
VN + DEKIRU	88	172	50	47
Desideratives	43	305	7	212
DEKIRU	528	13	271	10
WAKARU	185	4	231	7
HOSHII	31	1	7	1
SUKI	57	3	56	4
KIRAI	12	0	2	0

TABLE 1 *Ga*- and *o*-marked objects of different predicate types

It has been suggested in the literature that grammatical changes tend to take place first in root environments and then are extended to subordinate clauses; for example, (i) changes in word order in Old English and German, and (ii) the emergence of innovative forms of potential verbs such as *mireru* 'be able to see' in Japanese conform to this pattern, which Bybee (2002) puts as 'Main Clauses are Innovative, Subordinate Clauses are Conservative' (see also Matsuda 1998).

Interestingly, the described innovative case-marking pattern in Japanese does not conform to this tendency. Figures 2 and 3 show the numbers of *ga*- and *o*-marked objects of the five lexical predicates; these data show that the new pattern appeared in subordinate clauses earlier than in main clauses. Table 2 presents the difference between main and subordinate clauses in terms of how frequently they involve the new pattern. Figure 2 involves only the tokens with appropriate birth year information, while

[4] For potential predicates (with *(r)e*, *rare*, or DEKIRU) and desiderative predicates, accusative-marking on the direct object is well-established (see (1)), though its frequency (proportion) may have changed over years.

Table 2 involves ones without it as well. We will discuss in Section 4 why subordinate clauses are more innovative with respect to this particular change.[5]

The data presented here also suggest that the change in question has progressed earlier and to a greater extent in speech than in written texts. This arguably can be attributed to the general tendency for written language to be more conservative than spoken language.

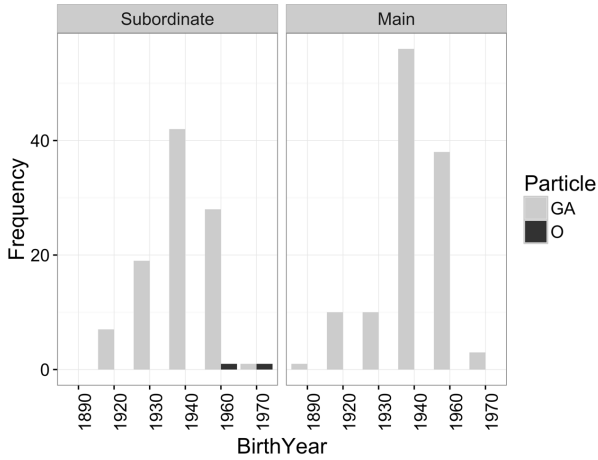


FIGURE 2 Lexical predicates and *ga*/*o*-marking by clause types (BCCWJ)

[5] It was suggested by a JK reviewer that (apparent) alternation between *ga*- and *o*-marking on objects within subordinate clauses might become possible due to the phenomenon known as Exceptional Case Marking (ECM), whereby the semantic subject of the complement clause of a verb such as OMOU 'believe' (an ECM verb) is realized as an *o*-marked nominal in the matrix clause, as in (ib):

(i) a. *Watashi wa [Hiroki ga kashiko-i to] omo-u.*
I TH H. NOM wise-PRS QUOT think-PRS
'I believe that Hiroki is wise.'

b. *Watashi wa Hiroki o [_ kashiko-i to] omo-u.*
I TH H. ACC wise-PRS QUOT think-PRS
'I believe Hiroki to be wise.'

It is a matter of debate whether a (nominative) object too can be the target of ECM. In our collected data, none of the *o*-marked objects of the five lexical predicates was the semantic subject of the complement clause of an ECM verb, and thus the observed innovative pattern cannot be attributed (even partially) to ECM.

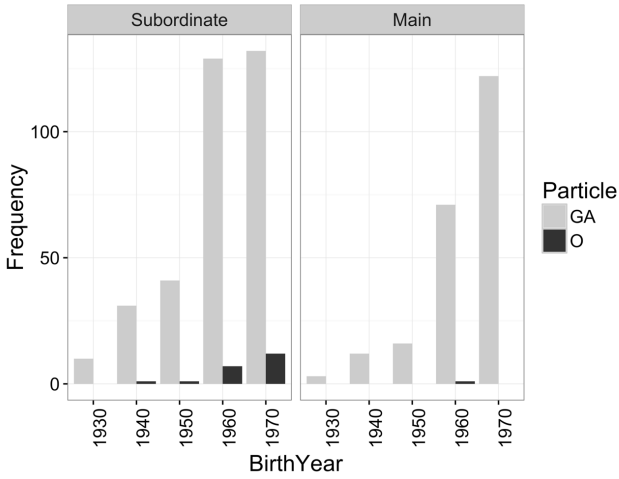


FIGURE 3 Lexical predicates and *ga*/*o*-marking by clause types (CSJ)

	BCCWJ		CSJ	
	Main	Subordinate	Main	Subordinate
ga	423	390	224	343
o	5	16	1	21
	($\chi^2 = 5.44$, $df = 1$, $p < .05$)		($\chi^2 = 9.53$, $df = 1$, $p < .05$)	

TABLE 2 *Ga*- and *o*-marked objects of the five lexical predicates

3.4 The Effects of the Clause Type and the Register

In order to identify factors that affect the choice between *ga*- and *o*-marking on the objects of *ga*/*o*-predicates in general, we constructed a logistic regression model, incorporating the effects of (i) birth year, (ii) clause type (main vs. subordinate), (iii) register (spoken vs. written), (iv) predicate type, and (v) adjacency.

The factor of birth year did not have a significant effect ($z = -1.22$, $p = .22$); this suggests that there was no major shift from nominative to accusative case, or accusative to nominative case, within the time frame considered.

The factor of register had a significant effect ($z = -3.34$, $p < .001$). With respect to *ga*/*o*-predicates as a whole, *ga*-marking was more frequent in speech than in writing; specifically, it accounted for 69.7% in the spo-

ken data (757 tokens out of 1,086) and for only 62.1% in the written data (1,229 tokens out of 1,980) (χ^2= 17.58, *df* = 1, $p < .001$). The factor of clause type too had a significant effect ($z = -5.16$, $p < .001$). We saw in Table 2 that with some lexical predicates (such as WAKARU), *o*-marking on an object is (rare but) relatively frequent in subordinate clauses. A similar pattern holds for *ga*/*o*-predicates in general; the proportion of the choice of *ga* is larger in main clauses than in subordinate clauses. We will discuss briefly why these two factors affect the frequency of *ga*- and *o*-marking in the following section.

4 Further Discussion

We observed above that the innovative case marking pattern, where the object of a predicate such as SUKI (+ copula) 'fond' is *o*-marked, has gained a stronger foothold in subordinate clauses than in main clauses (Section 3.3). It is interesting to inquire why this is the case, especially given that this does not conform to what appears to be the more common pattern such that subordinate clauses are more conservative with respect to grammatical changes.

It appears that this main/subordinate asymmetry in the innovative case marking pattern has to do with the fact that *ga*-marking on an object may induce ambiguity and incur additional processing load. To illustrate, while (3a) with *o* unambiguously mean that some contextually understood individual (possibly the speaker) is able to invite Yumi, (3b) has an additional interpretation where *Yumi* is the subject.

(3) a. *Yumi* *o* *shootai* *deki-ru*.
Y. ACC invite be.able.to.do-PRS
'(He/I/…) can invite Yumi.'

b. *Yumi* *ga* *shootai* *deki-ru*.
Y. NOM invite be.able.to.do-PRS
'(He/I/…) can invite Yumi.' OR 'Yumi can invite (him/me/…).'

This sort of ambiguity is most often resolved instantly with contextual cues, and thus rarely leads to miscommunication; still, it is plausible that the disambiguation requires some processing effort on the part of the hearer.

Arguably, the ambiguity induced by a *ga*-marked nominal tends to be easier to resolve in main clauses, for a reason related to the general strategies and patterns of information-structural coding in Japanese. In a main clause, the subject is very often marked by the particle *wa* in its thematic use; the object too is sometimes marked by thematic *wa*, but this happens much less frequently (Fry 1999, Oshima 2009, and references therein). As

such, upon hearing utterances like (4a,b), the hearer can reasonably infer that 'X *ga*' is likely the object, and 'X *wa*' is likely the subject.

(4) a. X *ga* *suki* *da*.
X NOM fond COP.PRS
'(He) likes *X*.' (more plausible) OR '*X* likes (him).'
b. X *wa* *suki* *da*.
X TH fond COP.PRS
'*X* likes (him).' (more plausible) OR '(He) likes *X*.'

In other words, the presence/absence of *wa* often serves as a clue for resolving the ambiguity induced by a *ga*-marked nominal.

In a subordinate clause, on the other hand, thematic *wa* is used more sparingly, and in some kinds of subordinate clauses including relative clauses, its occurrence is completely blocked. As such, the absence of *wa* in the relative clause in (5) does not help disambiguate the meaning of the sentence.

(5) [X {a. *ga* / b. **wa*} *suki* *na* *hito*] *wa* *Ken*
X NOM TH fond COP.ATTR person TH K.
da.
COP.PRS
'The person {who likes *X*/who *X* likes} is Ken.'

This means that a structure like (5a), where a *ga*-marked nominal and a predicate allowing a *ga*-marked object as well as a *ga*-marked subject co-occur within a subordinate clause, tends to pose a more serious risk of miscommunication than one without embedding. Having the option of marking the object with *o* mitigates the risk, making it possible to single out the 'who *X* likes' interpretation. If the N-to-A shift progresses further and reaches its completion, then (5a) will lose ambiguity and have only the 'who likes *X*' interpretation.

The same reasoning can be applied to the question of why, in general, *o*-marking on the object of a *ga*/*o*-predicate is more frequently observed in subordinate clauses (Section 3.4). When a *ga*/*o*-predicate is embedded, there is a stronger motivation to avoid ambiguity by choosing *o*, and this explains the higher frequency of *o*-marking.

Finally, the observation that *ga*-marked objects are more frequent in speech than in writing too can be connected to the matter of the subject/object ambiguity of a *ga*-marked nominal. It seems fair to assume that written communication requires a higher standard of lucidity than oral communication, because, generally speaking, in the former contextual

cues are relatively scarce and the audience is not able to ask for immediate clarification. Preference for *o*-marking in writing is likely to reflect the stronger need for clarity.

5 Summary

While no evidence was found that the overall frequency of *ga*-marking on the object of *ga*/*o*-predicates decreased within the past 100 years or so, our survey revealed (i) that with some predicates, which used to allow *o*-marking on the object only marginally, the choice of *o* has been increasingly common, and (ii) that the proportions of *ga*- and *o*-marking are affected by the factors of root/subordinate distinction and spoken/written distinction. It was suggested that these patterns were motivated by the functional need to reduce ambiguity, which *ga*-marking on an object may induce.

References

Aoki, H. 2008. Kanoo hyoogen no taishookaku hyooji 'ga' to 'o' no kootai [Alternation between 'ga' and 'o' as object markers for potential expressions]. *Japanese-Language Education around the Globe* 18, 133–146.

Bybee, J. 2002. Main Clauses are Innovative, Subordinate Clauses are Conservative. *Complex Sentences in Grammar and Discourse*, ed. J. L. Bybee and M. Noonan, 1–17. Amsterdam: John Benjamins.

Cukor-Avila, P. and G. Bailey 2013. Real Time and Apparent Time. *The Handbook of Language Variation and Change*, ed. J. K. Chambers and N. Schilling-Estes, 237–262. Oxford: Wiley-Blackwell.

Eythórsson, T. 2015. Syntactic Change. *The Bloomsbury Companion to Syntax*, ed. S. Luraghi and C. Parodi, 365–372. London: Bloomsbury.

Fry, J. 2003. *Ellipsis and Wa-marking in Japanese Conversation*. New York: Routledge.

Fujimura, I. 2009. Tadoosei saikoo: Hidoosashu o hyooji suru 'ga' to 'o' no kootai [The choice of patient markers GA and O in Japanese : A revised definition of transitivity], *Asian and African Studies* XIII(1), 73–102.

Ikuta, Y. 1996. Ganboo hyoogen ni okeru 'o/ga' no kootai ni tsuite [Remarks on the o/ga alternation in stative predicates, with special emphasis on desiderative expressions]. *Nagoya Daigaku Jinbunkagaku Kenkyu* 25, 39–72.

Iori, I. 1995a. 'Ga … shitai' to 'o … shitai': Kaku hyooji no yure ni kansuru ichikoosatsu [GA-SITAI and O-SITAI: A note on case-alternation in Japanese]. *Nihongo Kyōiku* 86, 52–64.

Iori, I. 1995b. 'Ga … shitai' to 'o … shitai': Chokusetsu mokutekigo no kakuhyooji no yure ['Ga … shitai' and 'o … shitai': Variation in case marking on di-

rect objects]. *Nihongo ruigigo hyoogen no bunpoo*, vol.1, ed. T. Miyashima and Y. Nitta, 53–61. Tokyo: Kurosio Publishers.

Kuno, S. 1973. *Nihon bunpoo kenkyuu* [*Studies of the Japanese grammar*]. Tokyo: Taishukan.

Labov, W. 1963. The Social Motivation of a Sound Change. *Word* 19, 273–309.

Maekawa, K. 2004. Design, Compilation, and Some Preliminary Analyses of the Corpus of Spontaneous Japanese. *Spontaneous Speech: Data and Analysis*, ed. K. Maekawa and K. Yoneyama, 87–108. Tokyo: The National Institute of Japanese Language.

Maekawa, K., M. Yamazaki, T. Ogiso, T. Maruyama, H. Ogura, W. Kashino, H. Koiso, M. Yamaguchi, M. Tanaka, and Y. Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation* 48(2), 345–371.

Matsuda, K. 1998. On the Conservatism of Embedded Clauses. *Theoretical and Applied Linguistics at Kobe Shoin* 1, 1–13.

Nambu, S., H. K. Hwang, D. Y. Oshima, and M. Nomura 2018. The Nominative/Accusative Alternation in Japanese and Information Structure. *Journal of East Asian Linguistics* 27(2), 141–171.

Oshima, D. Y. 2009. On the So-Called Thematic Use of *Wa*: Reconsideration and Reconciliation. *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, vol.1, 405–414.

R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Shibatani, M. 1975 Perceptual Strategies and the Phenomena of Particle Conversion in Japanese. *Papers from the Parasession on Functionalism*, ed. R. E. Grossman, L. J. San, and T. J. Vance, 469–480. Chicago: Chicago Linguistic Society.

Shibuya, K. 1993. Nihongo kanoo hyoogen no shosoo to hatten [Aspects and developments of potential expressions in Japanese], *Bulletin of Department of Literature, Osaka University* 33(1), i–262.

Sugai, K. and A. Naruse. 2006. Kiboo hyoogen ni okeru taishoo NP no kaku hyooji ni kansuru oboegaki [Remarks on the case marking of the objective NP in Japanese desiderative expressions]. *Hyogo University of Teacher Education Journal* 29, 49–57.

Tagliamonte, S. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.

Tokieda, M. 1950. *Nihongo bunpoo koogohen* [*Grammar of spoken Japanese*]. Tokyo: Iwanami.