Asia Pacific Journal of Corpus Research
Vol. 4, No. 1, 1-23

http://www.icr.or.kr/apjcr
e-ISSN 2733-8096
https://doi.org/10.22925/apjcr.2023.4.1.1

The Incredible Shrinking Noun Phrase: Ongoing Change in Japanese Word Formation

Kevin Heffernan
(Kwansei Gakuin University)
Yusuke Imanishi
(Kwansei Gakuin University)

Heffernan, K., & Imanishi, Y. (2023). The incredible shrinking Noun phrase: Ongoing change in Japanese word formation. *Asia Pacific Journal of Corpus Research, 3*(2), 1-23.

The Japanese language, as a typical agglutinating language, permits large noun phrases (NP) containing ten or more morphemes. In this paper, we argue that the nature of the NP in Japanese is changing. Our data are drawn from the *Balanced Corpus of Contemporary Written Japanese.* We conduct a series of apparent-time studies of ongoing changes in complex NPs. We first examine the length of compound nouns, followed by the usage of bound suffixes. We then examine ongoing changes in complex NPs that contain genitive case markers. Finally, we examine noun incorporation. All of our studies show a trend towards shorter, less complex NPs. Furthermore, our results suggest that the usage rate of phrases that modify the noun inside the NP (compound nouns, bound nouns, NPs containing genitive case, noun incorporation) appears to be decreasing over time. On the other hand, the usage rate of modifying material outside of the NP (positional phrases, relative clauses) appears to be increasing over time. We conclude by suggesting that our results reflect a diachronic change of decreasing synthetic morphology and increasing analytic morphology. We end by pointing out the implications of this work on our understanding syntheticity and analyticity.

Keywords: Japanese, Noun Phrase, Cyclic Change, Syntheticity, Analyticity

1. Introduction

The Japanese language is known as an agglutinating language. In such languages, complex words such as compound noun phrases (NP) are formed by combining together multiple morphemes, which may be either free or bound. Furthermore, morphemes tend to have a single meaning, and have easily recognizable morphological boundaries (Crowley & Bowern, 2010, p. 220). The Japanese language fits this description (Kageyama, 1999). Importantly, this combining process may be repeated as needed. In this way nouns of notable length can be easily produced, such as the compound noun example given in (1), which consists of twelve components.

Our objective is to report on ongoing changes in the nature of such complex NPs. We argue two points. First, we argue that the repetitive nature of the combining process is weakening in the Japanese language. Second, we argue that the first point arises as a consequence of a gradual shift in the location of the material modifying nouns from inside the NP to outside it.

We base our first argument on the observation that over the last century written Japanese contains fewer and fewer very long NPs such as the compound NP given in (1). Note that our first point makes no claims about compounding in general. As speakers acquire the Japanese language, they learn a wide range of compound nouns. With this knowledge, speakers generalize rules of compounding. Speakers also learn that these compounding rules may be recursively applied to their own output, resulting in compounds of any length. Thus, *kensetu* ‘construction’ may be compounded with *zigyô*

'project' to yield *kensetu-zigyô* 'construction project,' which in turn may be compounded with the compound *hôkoku-syo* 'report' to yield *kensetu-zigyô-hôkoku-syo* 'construction project report,' and so on ad infinitum. We are claiming that the so-called "ad ifinitum" characteristic is what is changing.

(1)

Narita	*sinsen*	*kensetu*	*zigyô*	*chi'nai*
Narita	new.line	construction	project	in.the.ground
maizô	*bunka*	*zai*	*hakkutu*	*chôsa*
buried	culture	treasure	excavation	survey
hôkoku	*syo*			
report	document			

'The report on the survey for excavating cultural relics during the construction of the Narita line' (PB42_00156, 1950)[1]

In order to better understand the nature of this change, compare the Japanese language with the English language. English has a much lower tolerance for long NPs formed through compounding. Rather, English speakers tend to elaborate nouns by adding on adpositional phrases. This point can be clearly seen in the English gloss given in (1). The English gloss uses four prepositions (*on*, *for*, *during*, *of*) whereas the original Japanese expression uses none. Thus, we can characterize English as a language that tends to elaborate nouns externally to the NP (often using prepositional phrases). In contrast, we can characterize Japanese as a language that often elaborates nouns within the NP, as illustrated by the single NP given in (1). Consequently, our second argument is that the Japanese language is slowly shifting from the use of NP-internal elaboration to NP-external elaboration as time passes. In other words, Japanese is making less use of compounding as a means of word formation.

In the next section, we introduce our research methodology. We then present a series of apparent-time studies of ongoing changes in complex NPs. We first examine the length of compound nouns, followed by the usage of bound suffixes. We then examine ongoing changes in complex NPs formed with the genitive case marker. Finally, we examine post-syntactic compounding. We conclude by suggesting that Japanese is becoming less synthetic and more analytic.

2. Methodology

2.1 Data Source

Our data are drawn from several subcorpora of the *Balanced Corpus of Contemporary Written Japanese* (BCCWJ; Maekawa et al., 2014). This corpus is a large-scale corpus of modern Japanese consisting of approximately 100 million words.[2] The corpus is composed of several subcorpora of different genres, such as laws, newspaper articles, and fiction novels. One important characteristic of this corpus is, as the name implies, representativeness. The authors defined the statistical population as all books, magazines, and newspapers published during the four-year period from 2001 to 2005. This population was first divided into genres (newspaper articles, magazine articles, novels, etc.), and

[1] All examples are taken from the *Balanced Corpus of Contemporary Written Japanese*. The information in brackets indicates the text and the birth decade of the author of the specific example. The following abbreviations are used in the glosses: ACC, accusative; DAT, dative; GEN, genitive; INS, instrument; LOC, NOM, nominative; NONEXCL, non-exclusive; PERF, perfective; SUB, subject. If the gloss is missing the morphological information, then all components are nouns.

[2] Technically speaking, the corpus has been parsed in morphemes. However, for the sake of simplicity, we use the term "word" to mean both morpheme and word. Since the Japanese language is agglutinative in nature, a morpheme in Japanese roughly equates to a word in English.

each genre was further divided into subpopulations. For example, the book population was divided into 55 subpopulations, and the magazine population was divided into 30 subpopulations. Texts, defined as of a sequence of characters of a specific length, were then randomly extracted from these subpopulations.

The data have been parsed and tagged with part of speech information. For the purposes of this research, we used the Numtrans version of the data. In this version, numbers originally transcribe in roman numerals have been converted to native Japanese script.

Varying amounts of meta information are provided for each text, such as the genre, author's gender, and author's birth decade. Since we are conducting an apparent-time analysis, the author's birth decade is a crucial datum. We excluded texts missing this information. Furthermore, we only included texts with multiple authors if all of the authors were born in the same decade.

Altogether, we used four subcorpora. We chose these subcorpora since they were both of adequate size, contained adequate meta information, and are representative of modern written Japanese. Of these issues, the lack of meta information was the biggest deciding factor. Some examples of subcorpora that do not include adequate meta information include the laws subcorpus, the newspaper articles subcorpus, and internet blogs subcorpus. The four subcorpora that we used are the library subcorpus, the best sellers subcorpus, the published books subcorpus, and the magazines subcorpus. These subcorpora account for almost 64% of the corpus. Each these are described in turn.

2.1.1 Library Subcorpus

This subcorpus consists of samples taken from books published between 1986 and 2005. The subcorpus was produced by randomly sampling the approximately 350,000 books in circulation in the Tokyo public library system.

2.1.2 Best Sellers Subcorpus

This subcorpus consists of samples taken from books published between 1976 and 2005. The books were selected by randomly sampling the 20 best-selling books in Japan for each year during this period.

2.1.3 Published Books Subcorpus

This subcorpus consists of samples taken from books published in Japan between 2001 and 2005. The subcorpus was produced by randomly sampling the books published during that time that were at least 40 pages in length.

2.1.4 Magazines Subcorpus

This subcorpus consists of samples taken from magazines published in Japan between 2001 and 2005. The subcorpus was produced by randomly sampling from the 1,259 titles in circulation during that time.

Although the BCCWJ contains data produced by authors born between the decades of 1860 and 1980, the number of texts produced by authors born in the decades 1860 through 1890, and in the decade 1980 are extremely few. Therefore, we exclude files produced by authors born during these periods.

The texts have been divided into 17 different genres, such as social science, technology, and reference. We reclassified the texts as either fiction or nonfiction. Text contained in quotation marks, and thus represented speech or thought, was excluded. Table 1 shows the volume of data for each genre by author birth decade, in units of 1,000 words.

Table 1. Word Counts by Genre and Author Birth Decade (Unit: 1,000 Words)

	1900	1910	1920	1930	1940	1950	1960	1970
Fiction	417	388	2,585	3,769	3,442	2,909	2,415	664
Nonfiction	533	887	3,815	6,748	8,620	6,977	4,232	759

Our research takes advantage of the fact that the BCCWJ has been parsed at two levels of granularity, a coarse level and a fine level. In the coarse data, complex nouns, verbs, etc., are treated as monomorphemic units. In the fine data, complex nouns, verbs, etc., are further parsed into their subcomponents and each subcomponent tagged with part of speech information. Thus, we can determine the length of a complex form by simply counting the number of subcomponents that appear in the finely-parsed data.

2.2 Extracting Noun Phrases

We examined all NPs that were tagged as a common noun in the course parsing. Examples are given in (2). Since dates and quantities are tagged as common nouns in the BCCWJ, we included such nouns in our studies. Also, words other than nouns that occurred inside the NP were treated as nouns. Example (2b) illustrates a complex noun that contains the verb *nomu* 'to drink' in the gerund form. Such nominalized forms derived from non-noun words are rare, and account for less than 1% of the data.

The examples given in (2) illustrate the varying length of the NP in Japanese. Example (2a) is a NP composed of one noun, (2b) of two nouns, (2c) of three nouns, and (2d) of six of nouns.

(2) a. *itai*
corpse
'corpse' (OB4X_00182, 1960)

b. *nomi* *ppuri*
drink way
'way of drinking (alcohol)' (PM12_00007, 1930)

c. *kyôiku* *i'in* *kai*
education committee meeting
'education committee meeting' (LBo7_00005, 1920)

d. *tyû* *syô* *kigyô* *taishoku* *kin*
middle small company retirement money
kyôsai
mutual.aid
'retirement fund for small- and medium-sized businesses' (PB53_00521, 1950)

2.3 The Apparent-Time Method

In this study, we use the apparent-time method to observe ongoing language change (Bailey, 2002). This method has over half a century of history (since Labov, 1966/2006), and has been used in countless studies of language variation and change. The apparent-time method is built on the critical age hypothesis. This hypothesis states that once we have obtained the critical age of maturity (around 12 to 15 years of age), then the way that we acquire language fundamentally changes. Young children are capable of effortlessly acquiring subtle linguistic knowledge such as complex syntax and pronunciation. On the other hand, speakers who began acquiring a language after passing the critical age tend to not acquire such details without conscious, ongoing practice, and even then, not the extent of a native speaker.

The critical age hypothesis has consequences for language acquisition. As a consequence of this change in the way that we acquire language, adults tend not to change their grammar and pronunciation after they have passed the critical age, even if the language spoken in the community changes. Variationist linguists capitalize on this tendency for invariance by assuming that an adult's language is representative of the language of their social community when he or she reached the critical age. That is, variationist linguists assume that even if the language spoken in the community changed at some later point in time, that specific adult speaker did not change his or her language. Thus, by looking at the language of an adult, we obtain an approximation of the language of the past. We expand our time depth beyond a single point in time by varying the age of the adults.

As mentioned, our studies use authors born in the decades between the 1900s and the 1970s, allowing us to examine data spanning 70 years. The text samples themselves were published between 1971 and 2005, a period of slightly over three decades. However, the publication date is irrelevant; as long as the author was older than the critical age at the time of writing, then we assume that his or her language usage (at least with regards to subtle grammatical details such as those examined in our studies) remains consistent from one decade to the next.

3. Ongoing Changes in the Length of NPs

3.1 Compound Noun Length

In this section, we examine ongoing changes in the length of compound NPs. We counted the number of nouns in each NP token examined. Table 2 lists the proportional distribution of nouns by the number of nouns in the NP. About 30% of the tokens examined contain more than one noun. The largest NP in our data consists of 34 nouns. However, such large nouns are extremely rare. We found less than five hundred nouns with a length greater than ten, which is less than 0.01% of the data. Such very long nouns tend to be the titles of reports, such as the example given in (1), the names of committees, or the names of positions such as office titles or committee chair titles.

Table 2. Distribution of NPs by Number of Nouns Inside the Phrase

Number of Nouns Inside NP	Count	Proportion
one	6,170,761	69.8%
two	1,895,199	21.5%
three	529,493	6.0%
four	152,056	1.7%
five	63,826	0.7%
six+	24,413	0.3%

We determined the average length of the NP (i.e., the average number of nouns) by author birth decade and genre. Figure 1 shows the result. Note that since our data contains both simple NPs consisting of only one noun and complex NPs consisting of multiple nouns, the overall average is slightly greater than one. From Figure 1 we see that the average length of the NP is gradually decreasing in fiction writing, with the exception of the first decade.[3] In contrast, the size of the NP in nonfiction writing does not begin to decrease until works written by authors born in the 1940s. From

[3] In our opinion, reporting the results of tests of significance, such as regression analysis, is not useful for two reasons. First, the large number of texts makes the tests of significance oversensitive, increasing the risk of a false positive result. Second, many of our reported results show a change in trend, for example from a rising trend to a falling trend. Such results tend to not be significant, but are nevertheless important to our overall argument. Much more important than the significance of any one result is the collective meaning of all of the results together. We argue that all of our reported results reflect the same underlying language change.

that point, the size of the NP gradually decreases.

The results shown in Figure 1 suggest that the overall average length of the NP was gradually increasing in the past, but is now gradually decreasing. Furthermore, this change appears to be more obvious in fiction writing. It maybe the transition from increasing to decreasing has just occurred in the nonfiction data. In other words, fiction writing leads nonfiction writing as the locus of this change.

This result is consistent with two possibilities. The first possibility is that compound NPs in general are gradually being used less frequently. Since we include both nouns of length one and nouns of length greater than one, if the proportion of nouns that are compound nouns decreases, then the overall average length of the NP decreases. However, such a change is not consistent with our claim that the recursive characteristic of compounding is weakening.

The other possibility is that longer compound nouns are being used less frequently, but shorter compound nouns are still being used with more or less the same frequency throughout the time period under observation. If, as we hypothesize, the recursive characteristic of compounding is weakening, then we expect a greater reduction in the usage of longer compounds than in shorter compounds.

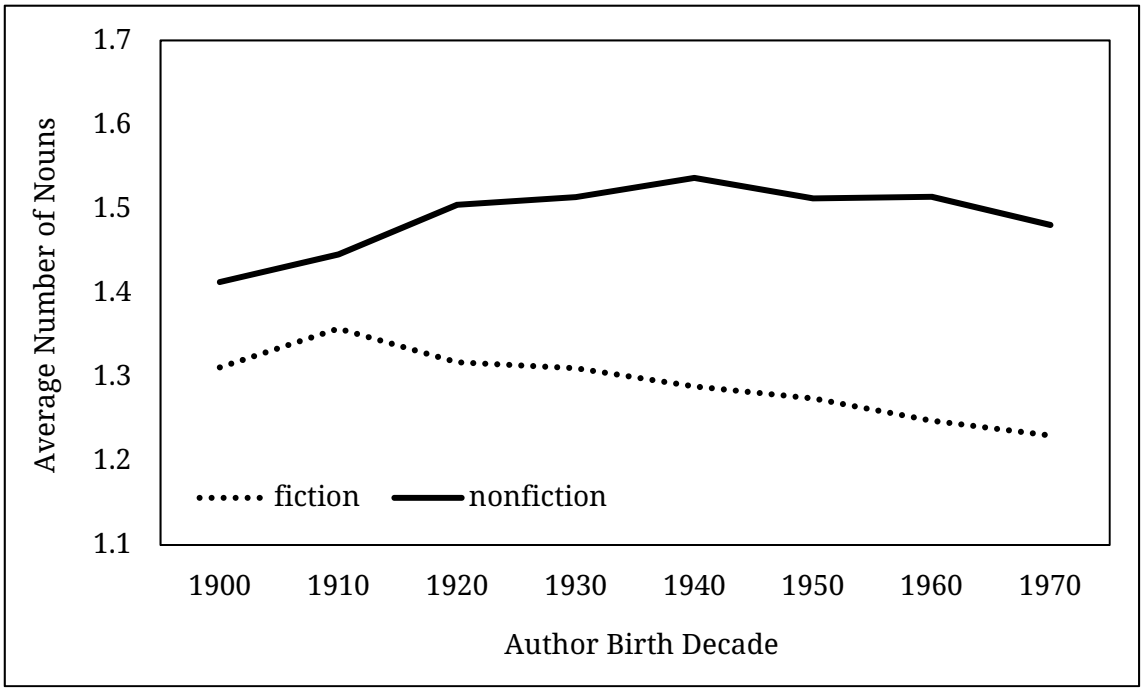

Figure 1. Average Length of Compound NPs by Genre and Author Birth Decade

In order to better understand the nature of the change observed in Figure 1, we examined the relative change in compound noun usage rate by compound noun length. We first determined the rate of usage per one million nouns for compound NPs of a given length by authors born in the 1900s. This value became our baseline. We then measured the relative change compared to the baseline. As a hypothetical example, consider the case in which the rate of compound nouns of length two is 150,000 occurrences per one million nouns in the texts produced by authors born in the 1900s. If the rate of compound nouns of length two increases to 165,000 occurrences per one million nouns in the texts produced by authors born in the 1910s, then the relative change between 1900 and 1910 is (165,000-150,000)/150,000 x 100% = +10%. We calculated the relative change compared to the baseline (the value of the 1900s) for each decade, and by genre. The results are shown in Figure 2 for the fiction texts and in Figure 3 for the nonfiction texts.

The data from the fiction genre (Figure 2) shows the expected results. From 1910 onward, we see a relative decrease in usage rate for compound nouns of all sizes, but the larger the NP, the greater the rate of decrease. Compared to the results for the data from the fiction genre, the results for the data

from the nonfiction genre are not as clear. The relative rate of usage of compound nouns in nonfiction texts does not begin to decrease until we reach texts written by authors born in the 1940s. Once we pass this point, then the plot lines for the larger noun sizes show a greater decrease in usage than the plot line for the nouns of size two. Once again, our results suggest that the changes are taking place in the fiction data before they are taking place in the nonfiction data. Thus, these results are consistent with our claim that the iterative characteristic of compounding is weakening, with the caveat that we mostly see the change in the fiction writing.

3.2 Changes in the Usage Rate of Bound Nouns

One question that naturally arises from our results is as follows. If the compound NP is gradually becoming shorter, then how are authors expressing concepts that are normally expressed through lengthy compounds? Presumably authors are expressing the same concepts with word sequences that span multiple syntactic phrases. Recall the example given in (1). The Japanese expression consists of a single NP. In contrast, the English gloss consists of a mix of NPs and prepositional phrases. Simply put, we hypothesize that Japanese is gradually becoming more like English.

We investigate further by examining the trend in the usage rate of bound nouns. The Japanese language contains nouns that occur only as bound suffixes, such as *-go* 'the language of' and *-hen* 'a chunk of'. Such words are coded as bound morphemes in their part of speech data. Overall, we found 690 different bound noun types in our data.

Figure 4 shows the proportion of nouns inside compound NPs that are bound nouns. Our results suggest that usage rate of bound nouns is gradually decreasing as time passes. We do not observe a clear difference between fiction writing and nonfiction writing.

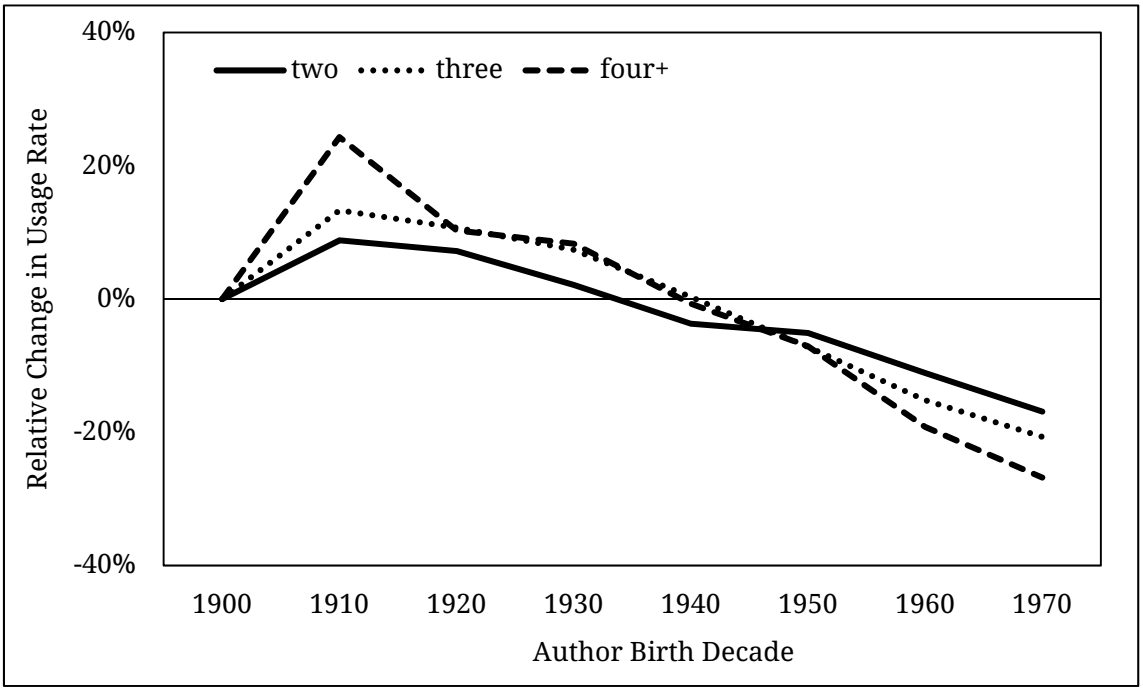

Figure 2. Relative Change in Usage Rate of Compound NPs by Number of Words and Author Birth Decade (Fiction Texts)

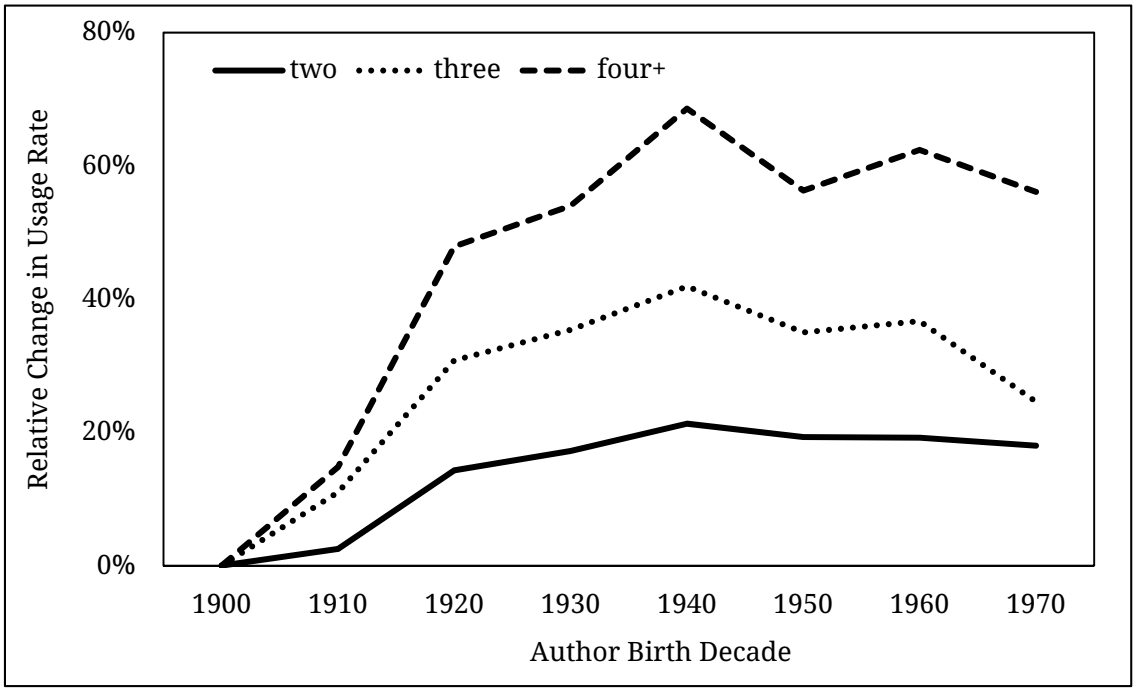

Figure 3. Relative Change in Usage Rate of Compound NPs by Number of Words and Author Birth Decade (Nonfiction Texts)

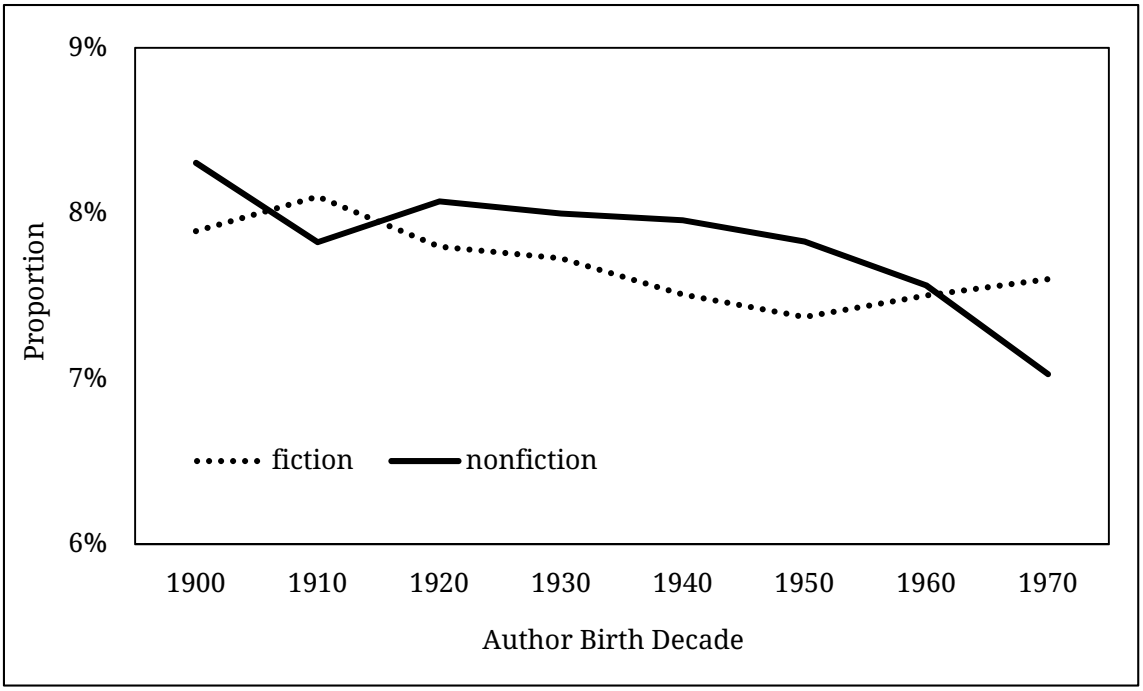

Figure 4. Proportion of Nouns That Are Bound, by Author Birth Decade and Genre

In order to better understand this ongoing change, let us consider two such bound nouns in more detail: *-rai* 'since' and *-irai* 'since'. Both of these words follow a phrase expressing a specific point in time. They indicate a continuous band of time starting at that specific point and continuing to the present. Example NPs containing *-rai* are given in (3a), and containing *-irai* in (3b). We choose these words because the Japanese language also contains one more way to express the same meaning: the bound postposition *-kara* 'since'. Note that all three word are bound morphemes. However, whereas *-rai* and *-irai* are nouns, *-kara* is a postposition. This syntactic difference is important. The use of either *-rai* or *-irai* results in compounding nouns together to make a complex NP consisting of several nouns.

In contrast, the use of *-kara* results in a NP followed by a prepositional phrase. In other words, the NP is shorter in the case of *-kara* than in the cases of *-rai* and *-irai*.

(3) a. *sen* *kyû* *hyaku* *nana* *zyû* *kyû* *nen* *rai*
thousand nine hundred seven ten nine year since
'since 1979' (LBi2_00076, 1910)
ni *zyû* *nen* *-rai*
two ten year since
'for the past twenty years' (PB37_00047, 1930)
sakunen *rai*
last year since
'since last year' (PM35_00010, 1960)

b. *sen* *hap* *pyaku* *roku* *zyû* *go* *nen* *irai*
thousand eight hundred six ten five year since
'since 1865' (LBb2_00031, 1930)
roku *sai* *no* *toki* *irai*
six years old of time since
'since six years old' (LBn2_00013, 1940)
kekkon *irai*
marriage since
'since getting married' (LBo3_00079, 1930)

In order to investigate the possibility that *-rai* and *-irai* are gradually being replaced by *-kara*, we extracted all occurrences of these words after the bound noun *-nen* 'year' (see the examples given). Furthermore, an important difference between *-kara* and the other two words is *-kara* that can follow expressions denoting future or unrealized events (irrealis), whereas *-rai* and *-irai* cannot. In order to account for this difference, we further limited that data to only items denoting dates before the year of publication of the text. In this way, we extracted a total of 2,010 tokens from the data. We did not investigate the two genres separately due to the small token counts and because the results presented in Figure 4 suggest that there is little difference between fiction and nonfiction with regards to bound noun usage rates.

Figure 5 shows the proportional usage of each of the words *-rai*, *-irai*, and *-kara* by author birth decade. As expected, the usage rate of *-kara* appears to be increasing over time whereas the usage rates of *-rai* and *-irai* appear to be decreasing over time, in this investigated context. This result is consistent with our observation that bound nouns in general seem to be decreasing over time.

4. Ongoing Changes in NPs with Genitive Case

In this section, we examine ongoing changes in complex NPs containing the genitive case marker. Other than direct compounding, NPs may also be elaborated with genitive case through inflection. Furthermore, such NPs also show the recursive nature seen in compounding. The example given in (4) illustrates this point. This single complex NP is composed of 16 words, including six genitive case markers.

Elaboration via genitive case may also occur ad infinitum. Consequently, we hypothesize that this recursive quality of genitive case elaboration is also weakening. We began confirming our hypothesis by counting the number of nouns modified by genitive case. Approximately 16.7% of the NPs are modified by genitive case. We next determined the proportion of NPs modified by genitive case by

genre and author birth decade. Figure 6 shows the results. In general, genitive case is used more in nonfiction writing than in fiction writing. Importantly, younger authors use the genitive case marker less than older authors, suggesting an ongoing change in the usage rate.

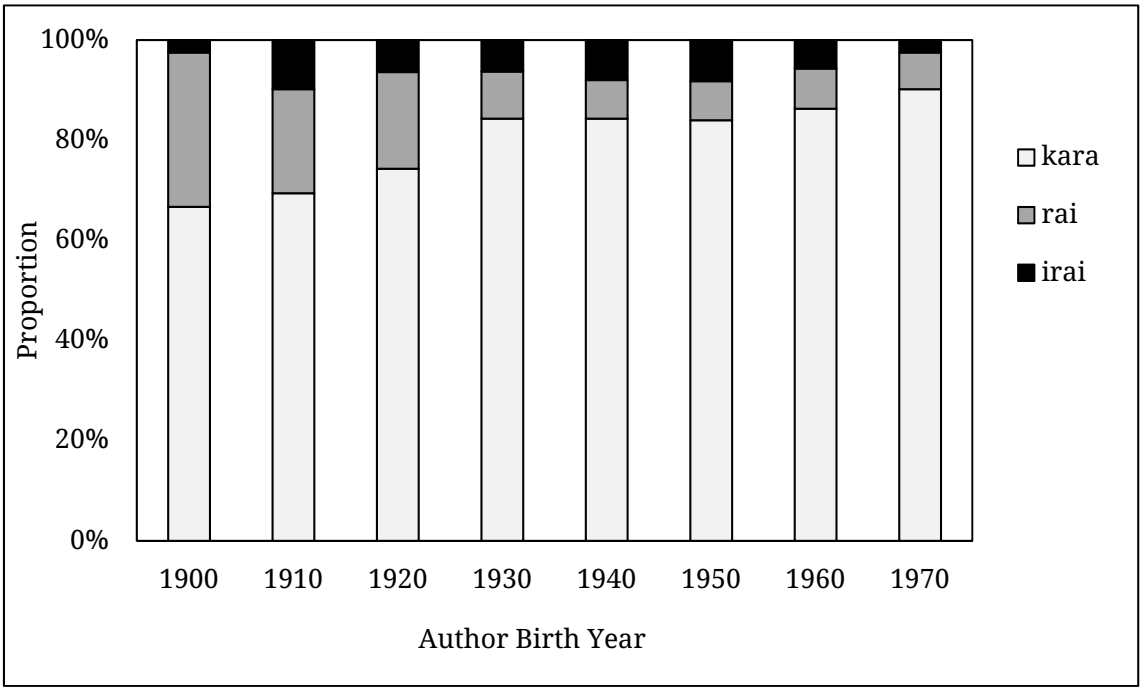

Figure 5. Proportional Usage of *-rai*, *-irai*, and *-kara* after the Word *-nen* 'year'

(4)	*tokutê*	*no*	*zigyô*	*yô*	*sisan*	*no*
	specific	of	business	use	assets	of
	kaikae	*no*	*ba'ai*	*no*	*jôto*	*syotoku*
	purchase	of	case	of	transfer	income
	no	*kazê*	*no*	*tokurê*		
	of	tax	of	special.case		

'the special case of the transferred income tax on the purchasing of specific assets for business use' (PB53_00175, 1940)

As was the case for the compound nouns, we are again faced with two possibilities. The first possibility is that there is an overall reduction in the use of genitive case, regardless of the size of the complex NP. The second possibility is that the reduction in genitive case usage seen in Figure 6 is a consequence of the reduced usage of long complex NPs such as the example give in (5). In order to determine which pattern better describes the ongoing change, we followed a similar procedure as the previous section. We first established the size of the complex NP in terms of the number of genitive case markers it contains. The example given in (4) contains six genitive case markers and therefore has a size of six. Table 3 lists the proportional distribution for complex NPs with one, two, and three or more genitive case markers. From Table 3, we see that the vast majority of complex NPs contain only one genitive case marker. Nevertheless, very long phrases are possible, and the longest phrase in our data contains twenty genitive case markers. The very long complex NPs tend to be found in fiction, and are creative or humorous in nature.

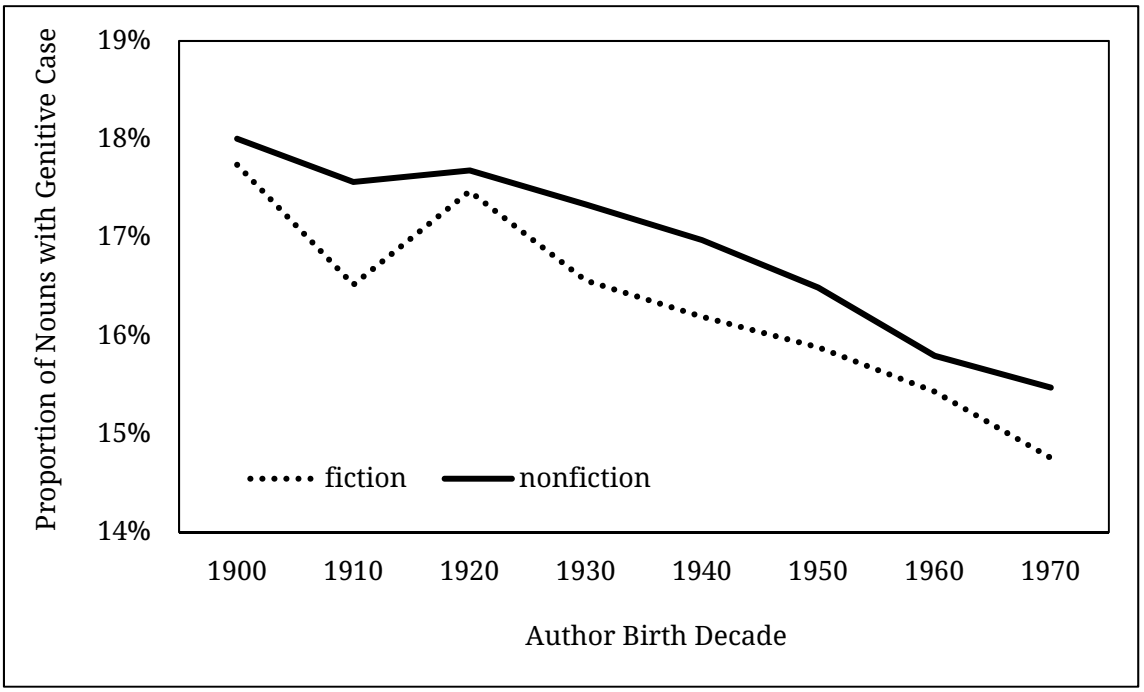

Figure 6. Proportion of Nouns with Genitive Case, by Author Birth Decade and Genre

Table 3. Distribution by Number of Genitive Case Markers

Number of Markers	Count	Proportion
one	1,038,814	92.4%
two	79,953	7.1%
three+	5,346	0.5%

We next determined the baseline value as the rate of usage per one million nouns for complex nouns of a given length by authors born in the 1900s. We then measured the relative change compared to the baseline from one decade to the next. The results are shown in Figure 7 for the fiction texts and in Figure 8 for the nonfiction texts. In both Figure 7 and Figure 8, we see the relative rate of usage of a complex NP containing a single genitive case marker is stable. In contrast, the relative rate of usage of a complex NP containing more than one genitive case marker decreases as the age of the author decreases. Furthermore, the rate of decrease differs. The rate of usage of complex NPs containing three or more genitive case markers decreases faster than the rate of usage of complex NPs containing two genitive case markers. Thus, we see that the reduced usage of the genitive case marker is not from its overall reduced usage, but rather specifically from the reduced usage in long complex NPs.

In order to better understand the ongoing changes in the usage of the genitive case marker, let us consider in detail changes in the usage of the noun *ba'ai* 'case, situation'. This word occurs 31,869 times in our data. Of those occurrences, *ba'ai* is modified by a noun with a genitive case marker 10,423 times (32.7%), and by a verb 16,120 times (50.6%). Examples of *ba'ai* modified by a noun with a genitive case marker are given in (5), and by a verb in (6). We have intentionally chosen example expressions containing verbal nouns to emphasize the flexibility in word choice that an author has. In the case of a verbal noun such as *sibô* 'die/death,' a writer can choose to use the word as either a noun or a verb. Consequently, the author can choose to use genitive case (5) or a relative clause (6).

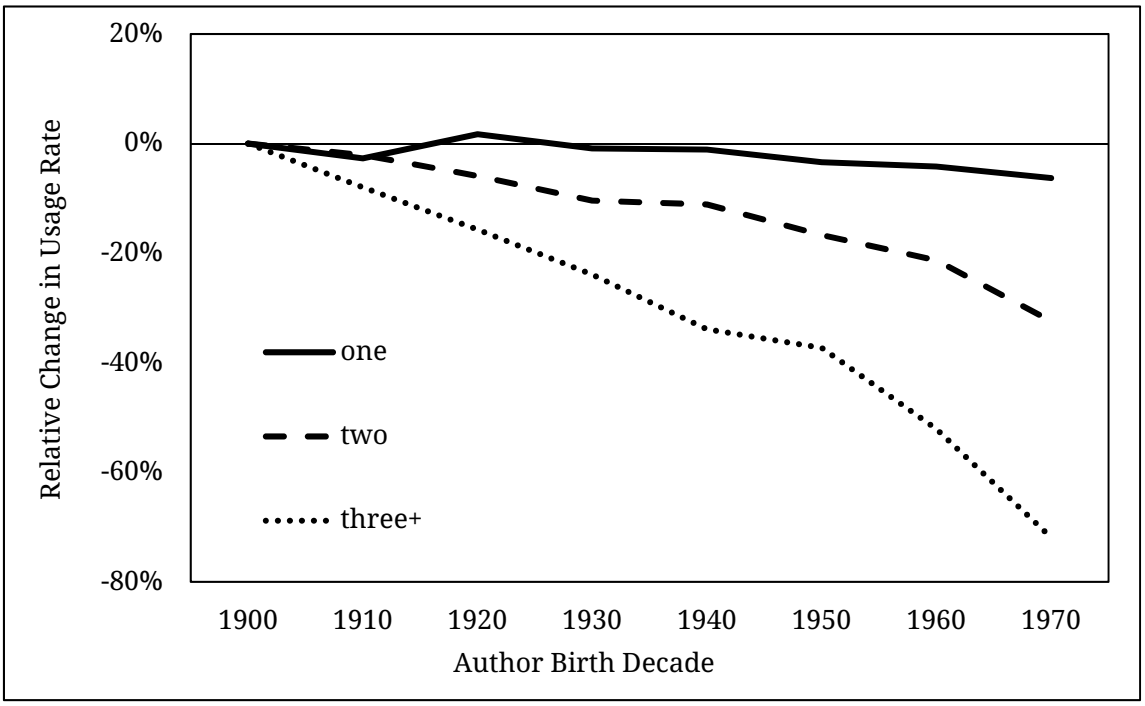

Figure 7. Relative Change in Usage Rate of Complex NPs Containing One, Two, and Three+ Genitive Case Markers by Author Birth Decade, for the Nonfiction Texts.

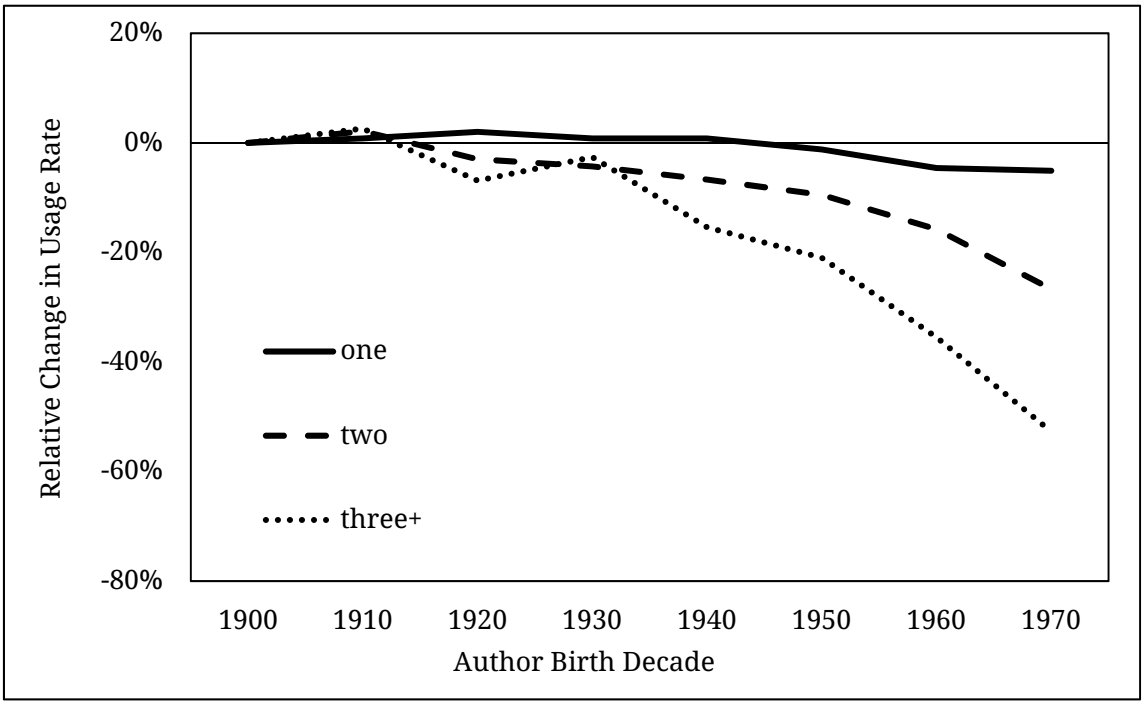

Figure 8. Relative Change in Usage Rate of Complex NPs Containing One, Two, and Three+ Genitive Case Markers by Author Birth Decade, for the Fiction Texts.

(5) *syôgai* *ya* *sibô* *no* *ba'ai*
injury and.NONEXCL death GEN case
'in cases such as death or injury' (PB43_00747, 1950)
tôyo *ga* *hituyô* *no* *ba'ai*
administration NOM necessary GEN case
'in the case that administration (of medicine) is necessary' (LBm4_00050, 1950)

(6) *kuruma* *ziko* *de* *sibô* *si-ta* *ba'ai*
car accident INS death do-PERF case
'in the case of death in a car accident' (PB43_00471, 1950)
settai *no* *hituyô* *ga* *aru* *ba'ai*
reception GEN necessary NOM exist case
'in the case that reception (of a guest) is necessary' (LBa3_00037, 1930)

The expressions *sibô-no-ba'ai* and *sibô-si-ta-ba'ai* both have the meaning of 'in case of death'. In the first expression, the noun *ba'ai* is modified by another noun, resulting in a single complex NP. In second expression, the noun *ba'ai* is modified by the empty verb *suru* 'do', and the resulting expression consists of both NPs and verb phrases. We hypothesize that the rate of usage of complex NPs is decreasing over time, and that the location of expressions modifying a noun is gradually shifting from inside the NP to outside the NP. If that is the case, then we predict a decrease in the usage rate of noun + *no-ba'ai*, and increase in the usage rate of noun + verb + *ba'ai*. In order to investigate our hypothesis, we counted the occurrences of verbal noun + *no-ba'ai* and verbal noun + suru 'do' + *ba'ai*. We also included tokens inflected for perfective aspect: verbal noun + *si-ta* 'do' + *ba'ai*.

Figure 9 shows the proportional usage of these two cases relative to all tokens of *ba'ai*, by author birth decade. Once again, due to the small number of tokens and the lack of a notable difference, we present the two genres together as one. As can be seen from Figure 9, the usage rate of noun + *no ba'ai* decreases as the age of the writer decreases. Again, we observe a gradual shift from NP-internal elaboration to NP-external elaboration.

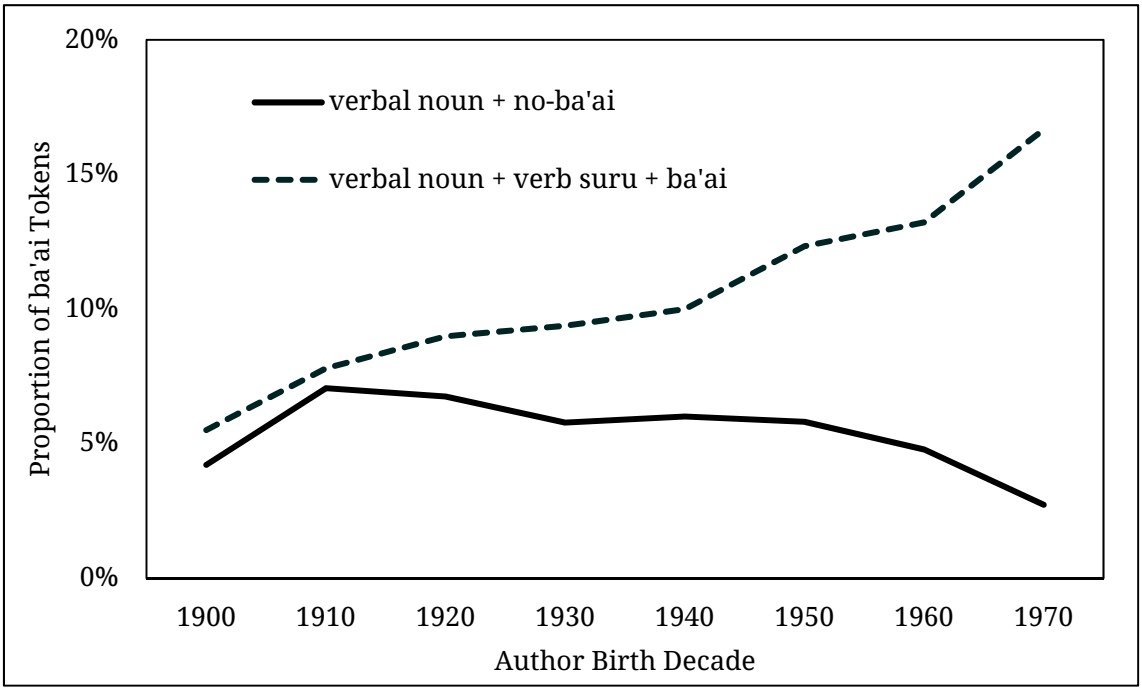

Figure 9. The Proportion of *ba'ai* Tokens Modified by a Verbal Noun and Either the Dummy Verb *suru* or Genitive Case, by Author Birth Decade

5. Ongoing Changes in Noun Incorporation

Our results so far have suggested that complex NPs are undergoing change. We have reported on compound nouns, bound nouns, and NPs containing genitive case. In all cases we saw that the complex NP is shortening. We pointed out that the words modifying the noun tend to appear more frequently

outside the NP as time passes. Another source of complex NPs in Japanese is noun incorporation (henceforth NI; Kageyama, 1993, 2009, 2016; Shibatani & Kageyama, 1988). NI phrases are formed by compounding either a verb that is converted to a noun (7) or a verbal noun (8) with its internal argument. We limited the data to the cases of a nominal noun argument of an unaccusative verb / verbal noun (7a, 8a), a dative noun argument of a motion verb / verbal noun (7b, 8b), or an accusative noun argument of a transitive verb / verbal noun (7c, 8c). The examples given in (7) and (8) show both the elaborated form in which syntactic relationships are indicated by case markers, and the semantically-equivalent NI form, indicated by an arrow. Following Kageyama (2016), the square brackets indicate NI compounds, and the colon indicates a phonological boundary (see examples). In the following discussion, we investigate NI phrases formed with verbs separately from those formed with verbal nouns.

(7) a. *ko* *ga* *aru*
child SUB exist
→ [*ko ari*] (LBp9_00201, 1940)
'(a mother) with a child'

b. *Ise* *ni* *mairu*
Ise Shrine DAT visit (humble)
→ [*Ise mairi*] (PB19_00552, 1930)
'visiting Ise Shrine'

c. *iki* *o* *koraeru*
breath ACC hold.in
→ [*iki* : *korae*] (PB25_00059, 1950)
'holding one's breath'

(8) a. *kêzai* *ga* *sêtyô* *suru*
economy NOM grow do
→ [*kêzai sêtyô*] (PB12_00079, 1920)
'economic growth'

b. *Sizuoka.eki* *ni* *tôtyaku* *suru*
Shizuoka.Station DAT arrive do
→ [*Sizuoka.eki* : *tôtyaku*] (PB59_00539, 1960)
'arrival at Shizuoka Station'

c. *isi* *o* *hyôzi* *suru*
intention ACC indicate do
→ [*isi hyôzi*] (PB23_00025, 1940)
'indication of intention'

The results of our investigation into complex NPs containing bound nouns and containing genitive case suggest a gradual reduction in the usage of phrase-internal elaboration. Kageyama (2016) presents evidence that NI phrases are single compound NPs. In that case, the NI process results in elaboration of a verbal noun or a verb via the incorporation of its internal argument.[4] In other words, NI is phrase-internal elaboration. In this way, NI compounds are similar to our studies of the bound noun suffixes *-rai* 'since' and *-irai* 'since', and our study of *suru* verb + *ba'ai* / *no-ba'ai* 'the case of' in that NI compound phrases have two forms: a form with phrase-internal elaboration and form with phrase-external elaboration. Consequently, we hypothesize that the usage rate of NI phrases is also

[4] See Kageyama (2016) and the references therein for evidence that the NI found in Japanese bears resemblance to NI found in polysynthetic languages such as Mohawk as discussed in Baker (1988).

decreasing as time passes.

Kageyama and Shibatani (Kageyama 1993, 2009; Shibatani & Kageyama, 1988) argue that there are two types of NI compounds in Japanese, those that are formed in the lexicon and those that are formed in the (post-)syntax: the latter is what Shibatani and Kageyama term post-syntactic compounds. We will use post-syntactic and syntactic interchangeably for compounds as nothing hinges on the choice of the terms in our discussion. They point out that lexicon-based compounds have word accent (e.g., *kibun-tenkan* 'change of pace'), whereas post-syntactic compounds have phrasal accent with a slight phonological break indicated as ":" between the two components of a compound (e.g., *hisaiti:sisatu* 'inspecting the stricken area'). In post-syntactic NI compounds, for example, each of the components bears a word accent, thereby displaying two accentual peaks: hiSAITI and siSATU, where capitalization indicates where a word accent falls (Kageyama, 2016). By contrast, lexical NI compounds are pronounced with one accentual peak even though each of the components may have its own accent (and peak) in when spoken in isolation. For example, when yaMA 'mountain' and noBORI 'climbing' are compounded, the resulting compound is pronounced as yaMA-NObori 'mountain-climbing' with only one accentual peak (Shibatani & Kageyama, 1988).

Another argument they give is co-occurrence with certain bound morphemes. Only syntax-based compounds can occur with morphemes such as *-go* 'after' (e.g., *zikken-shûryô-go* 'after conclusion of the experiment').

NI compounds in general are rare. Our data contain over 8 million noun tokens, of which 82,746 tokens (0.94%) were classified as NI compounds. Of these NI compounds, 997 tokens (1.2%) are NI compounds containing a verb. The remainder are NI compounds containing a verbal noun. Due to the sparsity of data for NI phrases containing a verb, we first report on the more-frequently-occurring case of NI phrases containing a verbal noun. We then report on NI phrases containing a verb.

5.1 NI Phrases Containing a Verbal Noun

We investigated the usage rate of NI phrases containing a verbal noun as follows. We first created a list of NI compound candidates by extracting all noun + case marker + verbal noun + empty verb sequences (e.g., *chika-ga kôtô-suru* 'land prices soar') that occur in the data. We defined a NI compound as either of the two sequences given in (9).

(9) i. *noun + verbal noun + bound noun*
ii. *noun + verbal noun + case marker*

Note that the third component of the sequence, either a bound noun or a case marker, may only follow a NP, guaranteeing that the previous two words have been produced as a NP by the writer. Such a requirement is necessary as it is also possible to produce a noun-verbal noun sequence in which the verbal noun is syntactically a verb, and the case marker has been omitted. (Case marker's omission is very rare in written language, but common in spoken language.) Finally, we also required the noun + verbal noun pair to occur in the NI compound candidates list, ensuring that that every NI compound has occurred in its elaborated form at least once in our data. Table 4 lists the top ten most-frequently-occurring verbal noun NI phrases.

We determined NI compound type (either lexical or syntactic) as follows. We identified the 65 most-commonly-occurring clitics that only occur with syntactic NI compounds. Based on this list, we classified each token as either lexical or syntactic. If the NI compound occurs in the corpus with one or more of the bound nouns on our list created for this purpose, then the compound was coded as syntactic. Otherwise, the compound was coded as lexical. This procedure yielded 11,734 tokens of syntactic NI compounds and 63,149 tokens of lexical NI compounds.

Table 4. The Top Ten Most-Frequently-Occurring Verbal Noun NI Compounds and Their Type

Rank	Nonfiction	Type
1	*ningen* : *kankê* 'human : relations'	lexical
2	*kêzai* : *sêtyô* 'economy : growth'	lexical
3	*isi* : *kettê* 'intention : decision'	lexical
4	*songai* : *baisyô* 'damages : compensation'	lexical
5	*anzen* : *hosyô* 'public.safety : preservation'	lexical
6	*sôgo* : *sayô* 'mutual : effect'	lexical
7	*gakkô* : *kyôiku* 'school : education'	lexical
8	*pêzi* : *sansyô* 'page : reference'	lexical
9	*tosi* : *kêkaku* 'city : planning'	syntactic
10	*kêzai* : *hatten* 'economy : development'	lexical

Similar to the previous studies, we conducted an apparent-time study of the usage rate for each NI compound type. We plotted usage rate per 1,000 words against author birth decade for the fiction texts (Figure 10) and the nonfiction texts (Figure 11). Our results show that NI compounds are almost exclusively limited to nonfictional writing. Within the nonfiction texts, lexical NI compounds initially increase with as the author birth decade become more recent, but peaks and then transitions to a decreasing trend. In contrast, syntactic NI compounds do not show a clear decreasing trend. Thus, in conclusion, the usage rate of NI compounds in writing is decreasing, but this decrease is limited to lexical NI compounds, and tends to occur in the texts written by younger authors.

5.2 NI Phrases Containing a Verb

We investigated the usage rate of NI phrases containing a verbal noun as follows. We first automatically extracted all noun + verb gerund form + particle sequences. Note that the third component, a particle, ensures that the preceding noun + verb gerund form sequence was produced as a NP by the writer. The output was then checked by hand and tokens that were incorrectly included due to mistakes in the original part of speech data were excluded. The second author also impressionistically determined if the token was a syntactic compound or a lexical compound. This procedure yielded 191 syntactic verb NI tokens and 997 lexical verb NI tokens.

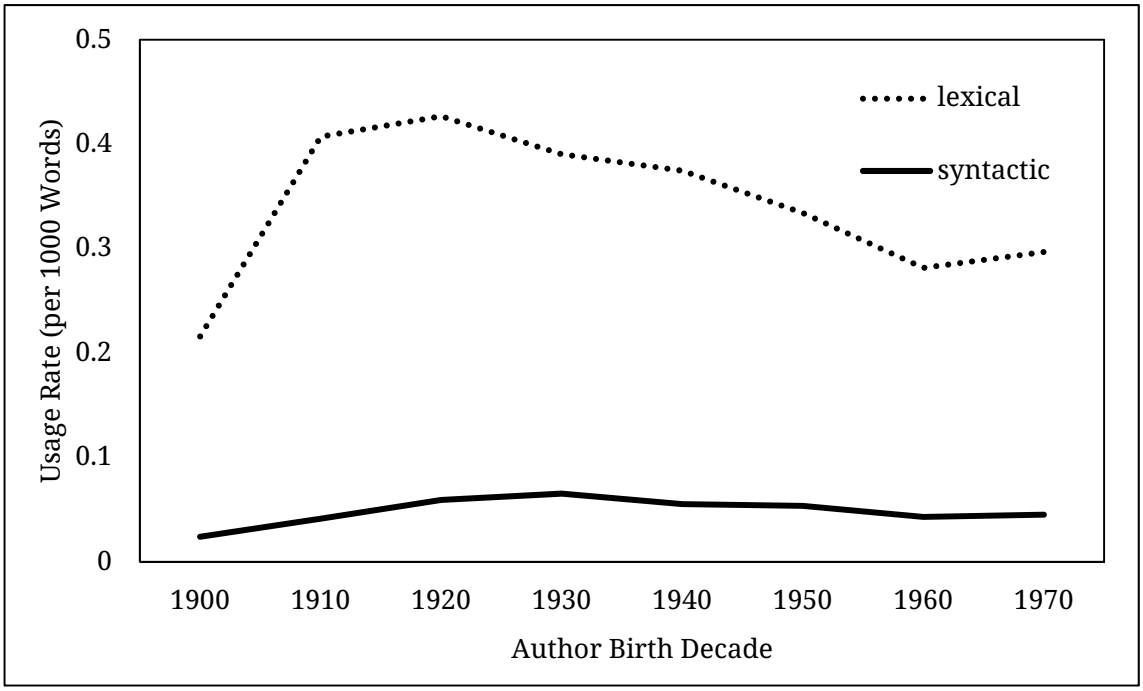

Figure 10. Usage Rate of Verbal Noun NI Phrases Against Author Birth Decade (Fiction Texts)

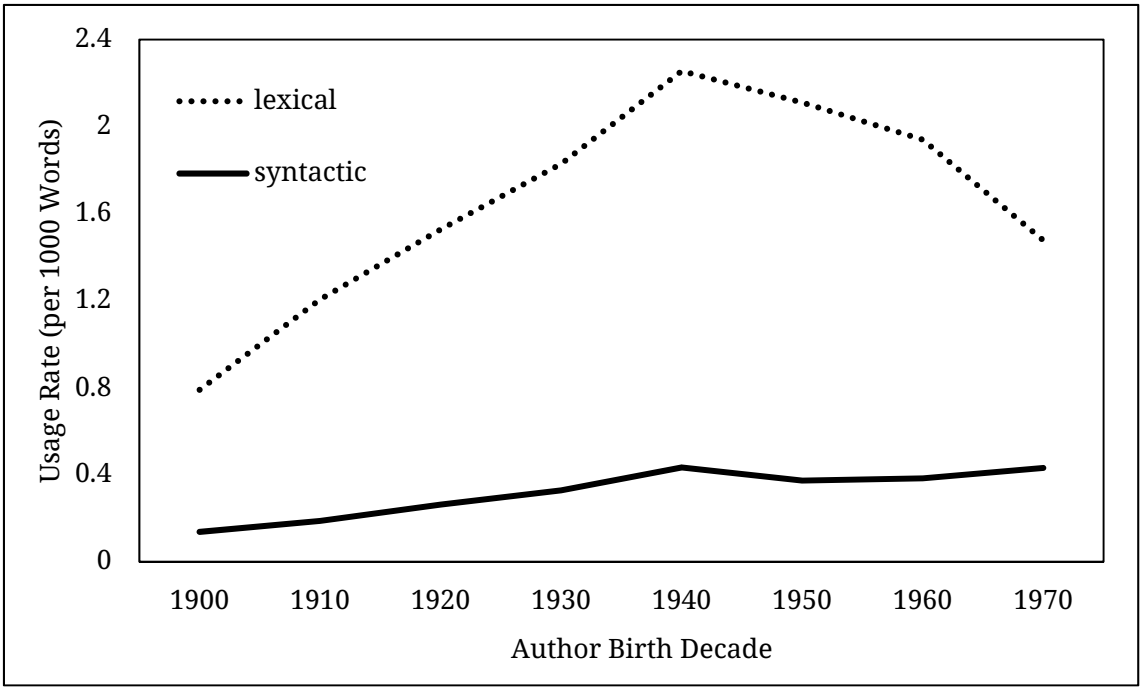

Figure 11. Usage Rate of Verbal Noun NI Phrases Against Author Birth Decade (Nonfiction Texts)

Table 5. The Top Ten Most-Frequently-Occurring Verb NI Compounds

Rank	Compound
1	*wasi* : *dukami* 'eagle : grasping'
2	*ido* : *hori* 'well : digging'
3	*aisô* : *dukasi* 'fondness : exhaustion = being fed up'
4	*idensi* : *kumikae* 'genetics : modifying'
5	*furosiki* : *tutumi* 'bath cloth : wrapping'
6	*serii* : *tuki* 'chord : 'série : striking = playing a series of musical notes'
7	*hagai* : *jime* 'wings : pinning = chock hold'
8	*kûki* : *susuri* 'air : sipping'
9	*zibun* : *sagasi* 'self : searching'
10	*se* : *nobasi* 'back : stretching'

The extremely small number of syntactic verb NI tokens limits the value of further analysis of them. However, our results of verbal noun NI compounds suggests that only lexical NI compounds are undergoing change. Consequently, we limit our investigation to lexical verb NI compounds. Table 5 lists the top ten most-frequently-occurring verb noun NI compounds.

We investigated the usage rate of lexical verb NI compound phrases for the fiction texts and the nonfiction texts (Figure 12). In general, the usage rate of NI phrases containing a verb is reducing as the birth year of the author increases. Unlike the NI compounds containing a verbal noun, Figure 12 does not show a differentiation between the fiction texts and the nonfiction texts. NI compounds containing a verb seem to be equally rare in both genres, and in both genres the overall trend in the usage rate is similar.

Together, the results for NI compounds are consistent with our hypothesis that the usage rate of NI compounds is decreasing with time. These results are also consistent with our claim that complex NPs are undergoing a gradual shift away from phrase-internal elaboration.

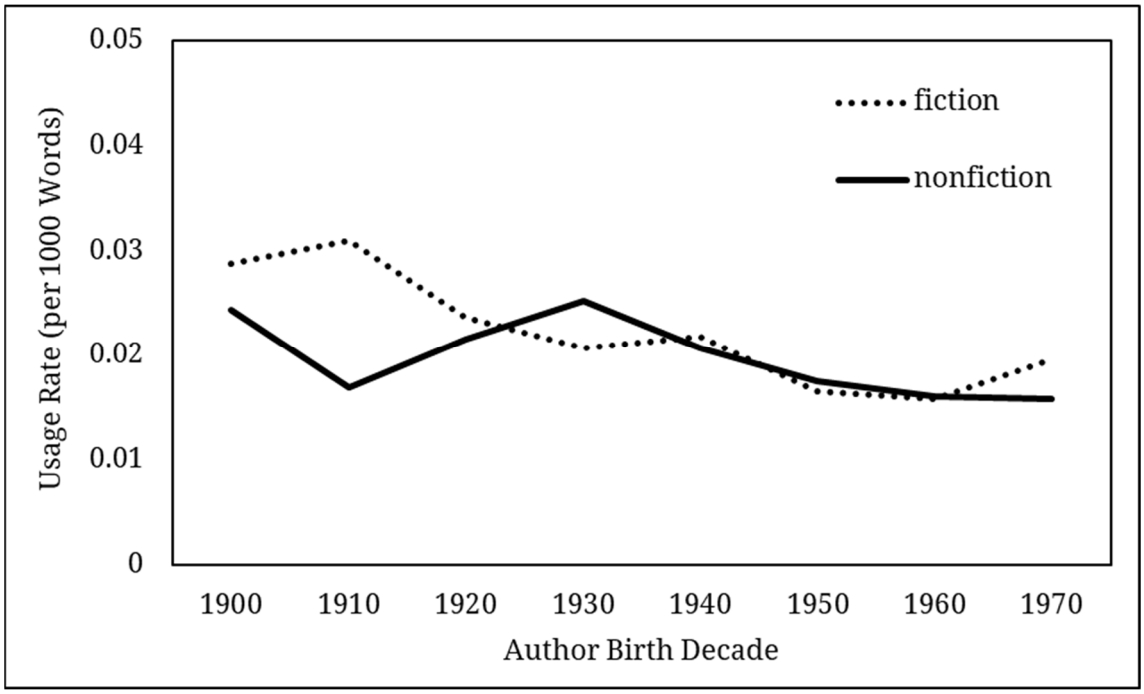

Figure 12. The Usage Rate of Lexical Verb NI Compound Phrases by Author Birth Decade and Genre

6. Discussion and Conclusion

In this paper we have presented several apparent-time studies of ongoing changes in NPs in written Japanese. We examined the usage rate of complex NPs in a corpus of written Japanese by author birth year. Following the Apparent Time Hypothesis, we interpret generational differences in the usage rates of the phrases that we investigated as reflecting ongoing diachronic change.

We demonstrated the following points:

- In general, the number of words used in compound NPs is gradually decreasing.
- Specifically, the relative usage rate of longer compound NPs is decreasing faster than that of shorter compound NPs.
- The proportion of nouns that are bound suffixes is decreasing.
- In general, the proportion of nouns modified with a genitive case marker is decreasing.
- Specifically, the relative usage rate of longer NPs containing genitive case makers is decreasing faster than that of shorter NPs containing genitive case makers.
- The usage rate of NI phrases shows a complex pattern dependent on NI compound type and genre. Syntactic NI phrases seem to not be undergoing diachronic change in usage rate. On the other hand, lexical NI phrases are clearly undergoing change. Our results show an initial increase in usage rate for verbal noun NI phrases, followed by a later decrease in usage rate. The usage rate of verb NI phrases also seem to be decreasing.
- In general, the fiction genre appears to lead the changes, and the nonfiction texts lag behind. However, the difference between the two is small.

Based on these results, we suggest that collectively these observed trends in usage rates reflect an ongoing change in the way nouns are elaborated in Japanese. Specifically, there seems to be a shift from the use of noun-internal elaboration to noun-external elaboration. We illustrated this point with studies of specific words. Our first such study looked at the suffixes *rai* ‘since’ and *irai* ‘since’. The two suffixes are bound morphemes, and occur within the NP. Our investigation showed that they are gradually being replaced by the postposition *kara* ‘since’, which occurs outside of the NP.

Our second such study looked at the word *ba'ai* 'case, situation'. This word often follows a noun marked with genitive case, as in the examples given in (5), or as the head of a relative clause, as in the examples given in (6). Our investigation revealed that its usage with a noun marked with genitive case is gradually decreasing, whereas its usage with a relative clause is increasing. The unifying point of these two studies is that they both show a gradual shift away from the use of modifying words inside the NP towards the use of modifying words outside the NP.

We also demonstrated ongoing changes in usage rates of NI phrases. A NI phrase is a single compound NP. Therefore, NI is another example of phrase-internal elaboration. The observed decreases in usage rates of NI is consistent with the claim that Japanese is undergoing a gradual shift away from phrase-internal elaboration.

Our claim is general enough that we predict to see related ongoing changes elsewhere in the Japanese language. Following are two examples of such predictions. The first example concerns cases of noun incorporation phrases that use adjectival nouns. Adjectival nouns such as *koyû* 'inherent' can form a compound by incorporating its internal argument, as seen in (10) (Kageyama, 1993, 2016 among others). Example (10a) is the baseline example for the adjectival noun *koyû*, with the relationship between the adjectival noun and its argument indicated by the dative case marker. Example (10b) is the semantically-equivalent noun incorporation phrase with noun *koyû* as its head.

(10) a. *Nihongo* *ni* *koyû* *no* *tokutyô*
Japanese DAT inherent GEN feature

b. [*Nihongo* : *koyû*] no *tokutyô*
Japanese inherent GEN feature
'features inherent in Japanese'

As shown by Kageyama's work and the references cited therein, NI phrases with adjectival nouns such as (10b) display the same range of properties as NI phrases with verbs or verbal nouns. Given this, we predict that the usage rates of the noun incorporation phrases with adjectival nouns as the head noun are also decreasing.

Our second example concerns compound verbs. Given our observation that the length of the compound noun appears to be shortening with time, we predict that the usage rate of compound verbs to also be reducing. In general, the verb phrase in Japanese is not as tolerant to compounding. Nevertheless, compound nouns are commonly used, and there are verbs that can be productively compounding with many other verbs, such as *tukusu* 'exhaust, complete' seen in the compound verbs *uri-tukusu* 'sell off completely' and *tabe-tukusu* 'completely eat up'. Another example of this pattern is VERB-*makuru* 'to VERB repeatedly', as to *asobi-makuru* 'to have a lot of fun'. If verbs are indeed showing a similar pattern of reduced usage of compounding, then we predict that the usage rate of such expressions to be gradually reducing. In that case, we also expect such expressions to be replaced by adverbs or serial verb sequences. Historically, the Japanese language has already undergone such a change in honorifics. Many honorific forms that were expressed as compound verb phrases in Old Japanese are now expressed as serial verb sequences. For example, the Old Japanese compound verb pattern VERB-*safurafu*, as in *kaeri-safurafu* 'to return (honorific)' is now expressed as the serial verb sequence VERB-*ni naru*, as in *kaeri-ni naru* 'to return (honorific)'. Note that the modern form also contains two verbs (*kaeru* and *naru*) but the verbs are separated by the dative case marker *ni*, indicating two separate verb phrases (morphological similar to the NP-*no*-NP pattern). Thus, we predict that patterns such as VERB-*tukusu* 'to VERB completely' and VERB-*makuru* 'to VERB repeatedly'. We leave the investigation of these predictions to future work.

6.1 A Shift Away from Synthetic Features Towards Analytic Features

This gradual shift away from phrase-internal elaboration most likely reflects a more general language-wide change, that is, a gradual reduction in synthetic language features and a gradual increase in analytic language features. Over half a century ago, Hodge (1970) argued that languages are continuously undergoing cyclic change, which he called the Linguistic Cycle Hypothesis. This hypothesis claims that the extent to which a language encodes grammatical relationships between constituents via free morphemes with grammatical and semantic meaning (i.e., analytic) or bound morphemes that only have a syntactic role (i.e., synthetic) slowly oscillates like a pendulum swinging back and forth. Imagine that at a given point in history, a certain language leans towards the analytic. At earlier point in time, that that language was relatively more synthetic in nature. Likewise, at a later point in time, that language will become relatively more synthetic in nature.

Szmrecsanyi (2012) gives empirical evidence supporting the linguistic cycle hypothesis with a study of English. Szmrecsanyi measures grammatical analyticity by the rate of occurrence of words from closed word classes such as complementizers, conjunctions, and modals, and grammatical syntheticity by the rate of occurrence of bound grammatical markers. By plotting these two measures for English for the past nine centuries against time, he showed that the English has essentially gone full circle. Grammatical syntheticity was the highest in the 12th century and then decreased until it reached its lowest score in the 15th century. The score for the 20th century English is close to that of 12th century. Grammatical analyticity was at its lowest in the 13th century, after which it abruptly shot up to its highest value in the 14th century. From there, the value has gradually decreased so that the score for 20th century English is very close to that of 12th century English. That is, if you only consider the grammatical syntheticity and analyticity scores, then 20th century English resembles 12th century English more than any other period.

The time scale of the data used in this study is a fraction of that reported in Szmrecsanyi. Thus, it is not surprising that we for the most part report a unidirectional change. However, a more nuanced view of our results suggests a more complicated pattern. While some of our results show unidirectional change, other results show a transition. Regardless, our results are unified by their consistency in the direction of the change.

One notable divergence in methodology from Szmrecsanyi (2012) and our study is the treatment of lexical words. Szmrecsanyi's measures are limited to grammatical words and lexical words with inflection. He specifically notes that purely lexical word tokens are not relevant. In contrast, we have included purely lexical words in our study, and argue that the ongoing observed changes apply regardless of the lexical status of the word. Our example given in (1) does not contain any grammatical markers per se. Yet our results reported in Figure 1 suggest that words such as the example given in (1) are becoming rarer. This divergence in methodology particularly matters when we consider the expressions *sibô no ba'ai* 'case of death', given in (5), and *sibô si-ta ba'ai* 'case of death', given in (6). We claim that the former phrase is obsolescing, and we suggest that this change reflects increased analyticity. The former phrase contains the function word *no*, which indicates genitive case, and the later phrase contains the verbal inflectional suffix *-ta*, which indicates perfective aspect. Szmrecsanyi classifies functional markers as analytic and inflectional affixes as synthetic. Consequently, if Japanese is becoming more analytic, then we expect the opposite result, viz., the former expression should be increasing and the latter expression decreasing.

We reconcile this discrepancy by noting the common ground between traditional definitions of syntheticity-analyticity and our study: boundedness. Studies such as Szmrecsanyi uses the rate of occurrence of bound grammatical markers to measure syntheticity, and although not defined as such, the rate of occurrence of unbound grammatical words to measure grammatical analyticity. Thus, one notable difference between indicators of syntheticity and indicators of analyticity is the status of the boundedness of the constituent, be it a word, a morpheme, or an ablaut vowel change. Specifically,

indicators of syntheticity are more bound than indicators of analyticity. We argue that bound nouns are decreasing and unbound nouns are increasing in Japanese. Thus, in so far as boundedness is concerned, the change in Japanese also equates to reduced syntheticity and increased analyticity.

6.2 Implications for Our Understanding of Syntheticity-Analyticity

The boundedness of the morpheme has long been a major theme of syntheticity-analyticity research. Schwegler (1990) argues that the concept is paramount, and that syntheticity-analyticity research must always consider the morphemic interdependency of speech units. Schwegler examined a number of diachronic changes in the Romance languages, such as the development of the modern-day French first person singular pronoun *je* from Latin *ego*. Schwegler argues that to if we are to fully account for this diachronic change, then it is not enough to simply list the phonological and morphosyntactic changes that took place—we must also note that *je* has a tighter morphological relationship with the verb than *ego*, and then we must "seek to relate the pertinent individual diachronic changes which *together* create this new morphological bond" (Schwegler, 1990, p. 116, original emphasis).

In contrast, recently researchers such as Ledgeway (2017) and Haspelmath and Michaelis (2017) argue that syntheticity-analyticity research must abandon the concept of boundedness because it is based on the concept of the *word*, which is "not well-defined except in trivial orthographic sense" (Haspelmath & Michaelis, 2017, p. 4). Certainly, many languages, of which Japanese is one, do not indicate words even in an orthographic manner. Another such example is Arabic. Alsager and (2021) contrasted measures of analyticity and syntheticity for a corpus of 10th century Arabic with those for a corpus of modern Arabic, and concluded that modern Arabic is more analytic. They closely follow the methodology of Szmrecsanyi with one notable exception: their treatment of the *word*. They focused only on the syntactic role of morphemes, and ignored whether the morpheme was free or bound.

Haspelmath and Michaelis go further, redefining syntheticity-analyticity in terms of diachronic change involving grammaticalization, and do not refer to the *word* or boundedness in at all. Specifically, they claim that the term *analytic* should be understood as meaning having recently undergone (re)grammaticalization (Haspelmath & Michaelis, 2017, p. 7). Note that their definition results in notably different results than previous work. For example, the English past tense marker *-ed*, traditionally treated as an indicator of syntheticity, is by their definition an indicator of analyticity. They point out the advantages of not relying on the ambiguously-defined notion of the word.

Such an approach works well in some cases, such as Hyman's (2017) demonstration of increasing analyticity in Bantoid languages of the Nigeria-Cameroon borderland area. Generally speaking, Bantoid language such as the Bantu varieties are highly synthetic, containing multiple verb extensions used to indicate manner, valency, etc. Hyman examined the variation in the degree to which varieties had lost the applicative verb extensions (i.e., a synthetic feature) and gained an analytic equivalent. Hyman gives examples that include the grammaticalization of nouns and pronouns into prepositions that replace the roles of the lost verb extensions. Thus, Hyman's work strongly supports the desire of Haspelmath and Michaelis to redefine analyticity and syntheticity in terms of diachronic changes.

However, the definition proposed by Haspelmath and Michaelis does not fit our observations about ongoing changes in the noun phrase in Japanese. For example, their proposal does not help us understand our observation that the length of the morphological phrase is gradually shortening. Nor does their proposal explain the increasing preference of nominal verbs to be used with the empty verb *suru* instead used the genitive case marker *-no*. In contrast, returning to Schwegler (1990), considering our results in terms of "morphological 'tightness'" (p. 192), allows us to not only present a unified explanation of our results, but also link our results to other research on syntheticity-analyticity.

6.3 Conclusion

We conclude by agreeing with Schwegler's argument that that the concept of morphological tightness is central to syntheticity-analyticity research. We further argue that such a concept is more important than syntactic versus morphological—the approach taken by Aldawood, Hodge, Szmrecsanyi, and many others. Such a shift would expand syntheticity-analyticity research and consequently the Linguistic Cycle to include lexical word classes as well as grammatical classes. Continued work on non-Western languages should clarify if this is indeed the right direction to go.

Finally, returning back to the observation of the shrinking noun phrase, we ask the question following: Is the Japanese language going to eventually become like English in this regard? To answer this question, we again turn to Szmrecsanyi (2012). Szmrecsanyi compared the cloud of data points resulting from plotting each of the 604 texts individually with samples of modern-day varieties of Italian, German, Bulgarian, and Russian. The historical variation of English as a whole does not overlap at all with the other European languages in the synthetic dimension, and only somewhat in the analytic dimension. Thus, although English shows clear diachronic change in its degree of analyticity and syntheticity, the extent of this intra-language variation is much smaller than the extent of inter-language variation. Simply put, languages do not leave their syntheticity-analyticity comfort zone, and that comfort zone tends to not overlap with the comfort zone of a taxonomically unrelated language. We assume that Szmrecsanyi's result extends to morphological phrase size. We conjecture that the Japanese language has a morphological phrase size comfort zone, and that even though the tolerated size of the phrase is shrinking, that size will always be larger than that of a language such as English. Further research on typological and diachronic variation in phrase size is needed.

Acknowledgments

We would like the express our gratitude to the participants at the following conferences who gave us feedback: *International Corpus Linguistics 2019*, *LACUS 2021*, *NWAV-AP 7*, and *Linguistic Society of Japan 165*. The second author's research has been funded by the JSPS Grant-in-Aid for Young Scientists (18K12388).

References

Alsager, H., & Aldawood, A. (2021). Syntheticity and analyticity in Arabic: A corpus analysis. *The International Journal of Communication and Linguistic Studies*, *19*(2), 49-60.

Baker, Mark C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Bailey, G. (2002). Real and apparent time. In Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (Eds.), *The Handbook of Language Variation and Change* (pp. 312-332). Oxford: Blackwell.

Crowley, T., & Bowern, C. (2010). *An Introduction to Historical Linguistics*. 4th ed. Oxford: Oxford University Press.

Haspelmath, M., & Michaelis, S. M. (2017). Analytic and synthetic: Typological change in varieties of European languages. In Buchstaller, I., & Siebenhaar, B. (Eds.) *Language Variation - European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe* (pp. 3-22). Amsterdam: Benjamins.

Hodge, C. T. (1970). The linguistic cycle. *Language Sciences*, *13*(7), 1-7.

Hyman, L. M. (2017). Multiple argument marking in Bantoid: From syntheticity to analyticity. In Bisang, W. & Malchukov, A. (Eds.), *Unity and Diversity in Grammaticalization Scenarios* (pp. 67-95). Berlin: Language Science Press.

Kageyama, T. (1993). *Bumpoo to Gokeisei* [Grammar and Word Formation]. Tokyo: Hitsuji.

Kageyama, T. (1999). Word formation. In N. Tsujimura (Ed.), *The Handbook of Japanese Linguistics*. Malden, Mass. and Oxford: Blackwell Publishers.

Kageyama, T. (2009). Isolate: Japanese. In Lieber, R., & Stekauer, P. (Eds.) *The Oxford Handbook of Compounding* (pp. 512-526). Oxford: Oxford University Press.

Kageyama, T. (2016). Noun compounding and noun-incorporation. In Kageyama, T., & Kishimoto, H. (Eds.), *Handbook of Japanese Lexicon and Word Formation* (pp. 237-272). Berlin: De Gruyter.

Labov, W. (1966/2006). *The Social Stratification of English in New York City*. 2nd ed. Washington, DC: Center for Applied Linguistics.

Ledgeway, A. (2017). Syntheticity and Analyticity. In Dufter, A., & Stark, E. (Eds.), *Manual of Romance Morphosyntax and Syntax* (pp. 839-886). Berlin: De Gruyter.

Maekawa, K., Yamazaki, M., Ogiso, T., et al. (2014). Balanced corpus of contemporary written Japanese. *Language Resources & Evaluation*, *48*, 345-371.

Schwegler, A. (1990). *Analyticity and Syntheticity. A Diachronic Perspective with Special Reference to Romance Languages*. Berlin: De Gruyter.

Shibatani, M., & Kageyama, T. (1988). Word formation in a modular theory of grammar: Postsyntactic compounds in Japanese. *Language*, *64*(3), 451-484.

Szmrecsanyi, B. (2012). Analyticity and syntheticity in the history of English. In Nevalainen, T., & Traugott, E. C. (Eds.) *The Oxford Handbook of the History of English* (pp. 654-665). Oxford: Oxford University Press.

THE AUTHORS

Kevin Heffernan is a professor at Kwansei Gakuin University. He is the creator of the *Corpus of Kansai Vernacular Japanese*, and he has written several articles and books about the Japanese language, including the corpus-based grammar, *The Grammar of Kansai Vernacular Japanese*.

Yusuke Imanishi is a professor at Kwansei Gakuin University. His research interests center around case, agreement, ergativity, nominalization, and *wh* constructions among others. He has conducted fieldwork research on Kaqchikel (Mayan) and Kikai (Ryukyuan). His works have appeared in international journals such as *Natural Language and Linguistic Theory* (Springer) and *Studia Linguistica* (Wiley).

THE AUTHORS' ADDRESSES

First and Corresponding Author
Kevin Heffernan
Professor
Kwansei Gakuin University
1 Gakuen-Uegahara, Sanda, JAPAN
E-mail: kevin@kwansei.ac.jp

Co-author
Yusuke Imanishi
Professor
Kwansei Gakuin University
1 Gakuen-Uegahara, Sanda, JAPAN
E-mail: yimanishi@kwansei.ac.jp

Received: 22 May 2023
Received in Revised Form: 8 August 2023
Accepted: 14 August 2023