P-2 Ongoing change in noun elaboration via bound suffixes in written Japanese

Kevin Heffernan (Kwansei Gakuin University)

Yusuke Imanishi (Kwansei Gakuin University)

Abstract

We argue that the use of bound nouns in written Japanese is gradually decreasing. Our data source is the *Balanced Corpus of Contemporary Written Japanese*. We determined the proportion of bound and free nouns for 15,303 text samples and found that 7.7% of nouns are bound. Each text was coded for author birth decade, which ranged from 1900 to 1970. A simple regression analysis revealed that the usage rate of bound nouns decreases as author birth decade increases. We hypothesize that our result reflects a general shift in noun elaboration from inside the NP to outside the NP. We provide a case study of the collocations *-nen-rai*, *-nen-irai*, and *-nen-kara* 'since the year', supporting this claim.

1. Introduction

The Japanese language is a well-known example of an agglutinating language. In such languages, complex words such as compounds are readily formed by annexing together multiple morphemes, which may be either free or bound. Example (1) illustrates a complex noun phrase consisting of four free nouns. Example (2) illustrates a complex noun consisting of two free nouns and the bound noun *-tû* 'while'.

(1) *Narita*   *sinsen*   *kensetu*   *zigyô*
     Narita   new.line   construction   project
     'The construction project for the Narita new line' (PB42_00156, 1950)[1]

(2) *mêru*   *sakusê*   *tyû*
     email   creation   middle
     'while composing an email' (PM41_00115, 1960)

Our objective is to report on ongoing changes in the nature of such complex noun phrases. We argue two points. First, we argue that the usage rate of bound nouns is gradually decreasing as time passes. Second, we argue this decrease in bound noun usage rate reflects a more general diachronic change in the Japanese language. Specifically, we suggest that the noun elaboration is shifting from inside the noun phrase to outside the noun phrase.

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 noun phrases formed through compounding. Rather, English speakers tend to elaborate nouns with adpositions (*on*, *for*, *during*, *of*, etc.). This point can be seen in the English glosses given in (1) and (2). In order to sound natural, the

---

[1] All examples are taken from the *Balanced Corpus of Contemporary Written Japanese*. The information in brackets indicates the text file and the author birth decade.

English glosses use adpositions whereas the original Japanese expressions do not. Thus, we can characterize English as a language that tends to elaborate nouns externally to the noun phrase. In contrast, we can characterize Japanese as a language that often elaborates nouns within the noun phrase. The examples given in (1) and (2) are single noun phrases. Consequently, our second argument is that Japanese language is slowly shifting from the use of noun phrase-internal elaboration to noun phrase-external elaboration as time passes.

In the next section, we introduce our research methodology. We then present an apparent-time study of ongoing change in bound nouns. In order to support our second argument, we present a case study of the collocations *-nen-rai*, *-nen-irai*, and *-nen-kara*, which all mean 'since the year'. We conclude with some implications of the second argument for the Japanese language.

2. Methodology

2.1 Data source

Our data is 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. The data have been parsed and tagged with part of speech information. Each of the subcorpora are composed of texts, and varying amounts of meta information is 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 all texts missing this information. Furthermore, we only retained 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 and contained adequate meta information. The four subcorpora that we used are the library subcorpus, the best sellers subcorpus, the published books subcorpus, and the magazines subcorpus.

Although the BCCWJ contains data produced by authors born between the decades of 1860 and 1980, the number of texts produced by authors 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 classified into 17 different genres, such as social science, technology, and reference. We reclassified the texts as either fiction or nonfiction. Finally, text contained in quotation marks, and thus represented speech or thought, was excluded.

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

---

[2] Technically speaking, the corpus has been parsed in morphemes. However, for the sake of simplicity, 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.

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 corresponding finely-parsed data.

2.2 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. Consequently, once speakers have grown older than 15 years old, they no longer change their language. Therefore, we can assume an adult speaker's language usage reflects the way language was used in the community when that speaker was a child, regardless of how old that speaker is.

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 (which we assume is true), then his or her language usage (at least with regards to subtle grammatical details such as those examined in our studies) reflects the language of the community when he or she was a child.

3. Ongoing changes in the use of bound nouns

3.1 Overall trend

In this section, we investigate trends in the usage rate of bound nouns. The Japanese language contains many nouns that occur only as bound suffixes, such as *-go* 'the language of' and *-hen* 'a chunk of'. The part of speech data for each morpheme indicates if it is a bound morpheme or a free morpheme. Overall, we found 690 different bound nouns (i.e., types) in our data. On average, 7.8% of noun morphemes are bound.

We conducted a simple regression analysis on the proportion of bound nouns, with birth decade and genre as independent variables. Our results show a significant relationship ($F_{(2,15299)} = 51.4$, $p < .001$). The proportion of bound nouns is 0.22% more in nonfiction than in fiction, and decreases by 0.15% per author's birth decade. Our results, shown in Figure 1, suggest that the use of bound nouns is slowly decreasing as time passes.
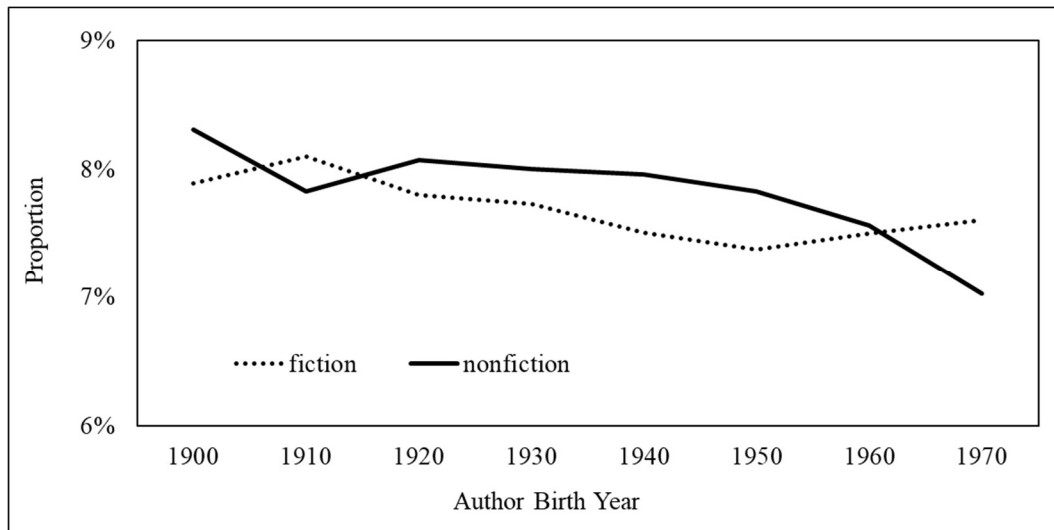
Figure 1: Proportion of nouns that are bound, by genre and author birth year.

3.2 A case study

In order to better understand the overall trend, let us consider one bound word in more detail: *-rai* 'since'. The bound noun *-rai* follows a noun expressing a specific point in time and indicates a continuous band of time starting at the specific point and ending at the present. Example noun phrases containing *-rai* are given in (3a).

The same meaning may alternatively be expressed in two other ways. The first way is by using the free noun *irai* 'since'. Similar to *-rai*, *irai* also follows a noun expressing a specific point in time, as shown by the examples in (3b). Finally, the same meaning may also be expressed by the adposition *kara* 'from'. When preceded by a noun that expresses a specific point in time, such as in the examples in (3c), *kara* has a meaning similar to that 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   happyaku   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)

c.   *nigatu*   *yokka*   *kara*

      February   fourth   from

      'from February 4th' (PM31_00764, 1940)

      sen   *kyû*   *hyaku*   *hati*   *jû*   *nana*   *nen*   *kara*

      thousand   nine   hundred   eight   ten   seven   year   from

      'since 1987' (PB15_00055, 1960)

In order to ensure that the morphemes *-rai*, *irai*, and *kara* are indeed interchangeable, we examined the usage rates of each of these following the bound noun *-nen*, which indicates a specific year within a date, such as the year 1987 seen in the second example given in (3c). Thus, all three collocations, that is, *-nen-rai*, *-nen-irai* and *-nen-kara* have the same meaning of 'since the year'. Importantly, note that the third collocation uses the adposition *-kara*, which occurs outside the NP. In contrast, the other two collocations use nouns, which occur inside the NP.

We calculated the proportional usage of each collocation by author birth decade. Our result is shown in Figure 2. Our result shows that *-nen-kara* is gradually replacing the other two collocations. This result agrees with the claim that noun elaboration is gradually shifting from the inside the noun phrase to outside the noun phrase.
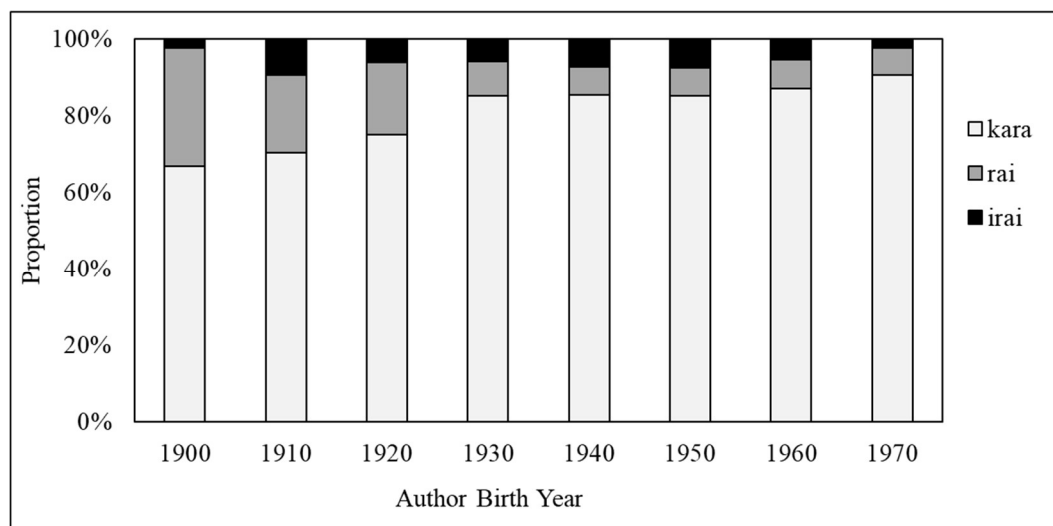

Figure 2: Proportional usage of -rai, *-irai*, and *-kara*, following *-nen*, by author birth year.

4. Conclusion

Using a corpus of written Japanese, we have shown that the usage of bound nouns to elaborate noun phrases is gradually decreasing. Furthermore, we also showed that the two collocations *-nen-rai* and *-nen-irai*, which consist of elaboration internal to the noun phrase, are gradually being replaced by *-nen-kara*, which consists of elaboration external to the noun phrase. Thus, our results seem to suggest that internal noun phrase elaboration is gradually decreasing and external noun phrase elaboration is gradually increasing.

We make two predictions. First, we predict a gradual decrease in the usage rate of genitive case. This prediction is motivated by utterances such as those seen in (4). The expression *sibô-no-ba'ai* 'in case of death' seen in (4a) contains the genitive case marker *-no*. An expression with the equivalent meaning but without the genitive case marker is seen in (4b). In the case of (4a), the noun *sibô* is modified by the noun *ba'ai* within a single noun phrase. In the case of (4b), the noun *sibô* is again modified by the noun *ba'ai*, however in this case *ba'ai* is in a relative clause, and therefore outside the NP containing *sibô*. We predict that cases such as (4a) are decreasing and cases such as (4b) are increasing.

(4) a. *syôgai   ya   sibô   no   ba'ai*
      injury   and.NONEXCL   death GEN case
      'in cases such as death or injury' (PB43_00747, 1950)

   b. *kuruma   ziko   de   sibô   si-ta   ba'ai*
      car   accident INS   death do-PST   case
      'in the case of death in a car accident' (PB43_00471, 1950)

Another grammatical phenomenon in which nouns are elaborated within a single noun phrase is noun incorporation. (henceforth NI; Kageyama, 1993, 2009, 2016; Shibatani & Kageyama, 1988). NI phrases are formed by compounding either a verb (5a) or a verbal noun (5b) with its internal argument. The examples given in (5) 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.

(5) a. *ko   ga   aru*
       child   SUB   exist
   → [*ko : ari*] (LBp9_00201, 1940)
       '(a mother) with a child'
       Ise   ni   mairu
       Ise Shrine   DAT   visit (humble)
   → [*Ise : mairi*] (PB19_00552, 1930)
       'visiting Ise Shrine'

    b.   *kêzai*   *ga*   *sêtyô*   *suru*

       economy   NOM   grow   do

→ [*kêzai* : *sêtyô*] (PB12_00079, 1920)

     'economic growth'

     Sizuoka.eki   *ni*   *tôtyaku*   *suru*

     Shizuoka.Station   DAT   arrive   do

→ [*Sizuoka.eki* : *tôtyaku*] (PB59_00539, 1960)

     'arrival at Shizuoka Station'

In the case of the NI form, the second noun elaborates the first noun within a single noun phrase. If, as we argue, noun phrase-internal elaboration is decreasing, then we predict that the usage rate of NI in Japanese should also be decreasing as time passes.

We conclude by speculating about the future of the Japanese language. The 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 analytical 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 syntactic information (i.e., analytic) or morphological information (i.e., synthetic) slowly oscillates like a pendulum swinging back and forth. Elaboration of a noun external to the noun phrase requires an increased use syntactic information to encode the relationship. Thus, in conclusion, we speculate that Japanese is gradually becoming more analytic in nature.

References

Bailey, G. (2002). Real and apparent time. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 312–332). Oxford: Blackwell.

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

Kageyama, T. (1993). *Bumpoo to gokeisei* [Grammar and word formation]. Tokyo: Hitsuji.

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

Kageyama, T. (2016). Noun compounding and noun-incorporation. In T. Kageyama & H. Kishimoto (Eds.), *Handbook of Japanese lexicon and word formation* (pp. 237–272). Berlin: De Gruyter Mouton.

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

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

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