Dynamic Shift of Word Frequency Effect in the Course of Linguistic Change*

Shin-ichiro Sano
International Christian University

1. Background

1.1. Neogrammarian regular sound change, lexical diffusion and word frequency

In historical linguistics, Neogrammarians proposed the regularity principle, based on the empirical findings (see Campbell 2004 and references cited therein). The principle consists of two components. The first component is the claim that the sound change is regular or exceptionless; that is, all relevant sounds in every lexical item are subject to the change. The second component is the claim that sound change spreads throughout the lexicon in an abrupt manner; that is, the change affects the subjects all at once, and accordingly there is no gap among lexical items with respect to the degree of progress. A vast amount of examples can be observed such as Grimm's Law and Great Vowel Shift that follow the regularity principle. However, there are also exceptions, such as palatalization (> ∢) found in the change from Latin to Standard French. In this example, some lexical items were not affected by the change, namely the change was not regular. Thus, these sorts of exceptional changes were summarized as lexical diffusion, in which the progress of some changes differs according to lexical items, creating the irregularity (Wang 1969, 1977; Wang & Cheng 1977).

Several theories of lexical diffusion have thus far been proposed that attempt to find root causes that create the exceptions. I will consider two. The first is Labov's (1981, 1994) account of the irregularity in terms of the roles of social factors. He explains the exceptions as follows. If the change is driven solely by internal factors, the change would be regular; however, if social factors come into play in addition to internal factors, then the change would be irregular, showing lexical diffusion. The former case is an instance of the *change from below* where no one is conscious about the change, and social factors do not come into play. But once the change is recognized by speakers, it becomes the target of social factors, and irregularity emerges.

The other proposal that I consider is based on the word frequency. Namely, changes diffuse from words with higher frequency to those with lower frequency.[1] This suggests that words with higher frequency are susceptible to changes; on the other hand, words with lower frequency are resistant to changes. In other words, words with higher frequency are more

likely to be the locus of changes or the target of rule application. For example, Bybee (2002) presents English t/d-deletion as one of the manifestations of word frequency effect. She demonstrates that the progress of the change differs from one lexical item to another due to the influence of word frequency.

However, there are also some aspects of word frequency effect that remain unexplored. For example, the dynamic aspects of the influence of word frequency, namely the change of the effect itself has been understudied. To date, previous research has mainly focused on synchronic aspects, and has paid little attention to diachronic aspects. Therefore, the goals of this research are to 1) demonstrate the dynamic aspects of word frequency effect in ongoing linguistic change, and 2) demonstrate the mechanism of lexical diffusion.

1.2. *Ra*-deletion: ongoing morphophonological change in Japanese

This section presents some background information on the change that is investigated. Japanese verbs are classified into two types. The first type is called consonant verbs, in which stem-final segments are consonants, as in *hasir-* 'run,' and *yar-* 'do.' The second type is called vowel verbs, in which stem-final segments are vowels, as in *mi-* 'see,' and *tabe-* 'eat.' The potential suffix in Japanese shows allomorphy; namely, the potential suffix undergoes morphophonological alternation according to the type of verb stem to which they attach: consonant verbs take *-e-*, as in *hasir-e-* 'can run,' and *yar-e-* 'can do'; on the other hand, vowel verbs take *-rare-*, whose initial segment is consonant, as in *mi-rare-* 'can see,' or *tabe-rare-* 'can eat.' This alternation is partly driven by the strong preference for CV syllable structure in Japanese. Importantly, *ra*-Deletion occurs only in vowel verbs; and accordingly, consonant verbs are outside the scope of this research.

A variant of the potential suffix is *-re-*, in which the *ra* syllable is not produced (Matsuda 1993; Ito & Mester 2004). Examples of this innovative form are shown in (1) and (2).

(1) oisii mono-ga tabe-re-ru. (cf. traditional: tabe-rare-)
delicious stuff-NOM eat-POT-NONP
'(We) can eat delicious foods.' (S00M0002)[2]

(2) onazi keekoo-ga mi-re-masu. (cf. traditional: mi-rare-)
same tendency-NOM see-POT-POLITE.NONP
'(We) can observe the same tendency.' (A01M0565)

2. Data

This section explains how data was gathered. Data was retrieved from the *Corpus of Spontaneous Japanese* (abbreviated as CSJ, NIJL 2008). The CSJ is a large-scale spontaneous speech corpus of common Japanese with rich annotation (Maekawa 2004; Sano & Hibiya 2012). I extracted all potential forms of vowel verbs from the corpus. Potential verb forms were collected from 1,286 speakers, whose birth-years ranged from the 1910s to the 1980s. In total, 8,518 tokens of vowel verbs with the potential suffix were extracted. Of these, 543 tokens exhibited *ra*-Deletions (6.66%), as shown in Table 1.

Table1: Summary of the retrieved data

traditional potential	7,615
ra-Deletion	543
probability of *ra*-Deletion (%)	6.66

In Figure 1, I illustrate the chronological transition of potential forms, according to the birth-year of speakers. As Figure 1 shows, the probability of *ra*-Deletion gradually rises as the speakers' birth-year becomes more recent. The result shows that in the present data the first *ra*-Deletion is observed in speakers born in 1920s. The result is consistent with the claims of previous research (Matsuda 1993, among others). Furthermore, we can argue that the change of *ra*-Deletion is a change in progress, and it is currently in the intermediate stage.

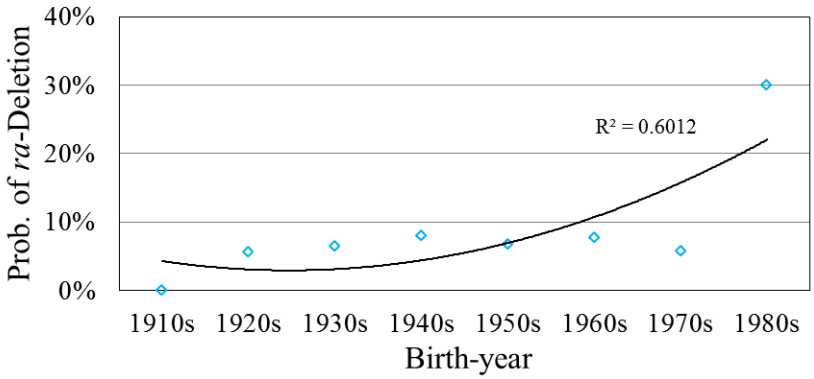

Figure 1: Chronological change of the probability of *ra*-Deletion

3. Examination of word frequency effect

3.1. Method

The procedure is summarized as follows; 1) for each verb stem, I calculated the probability of *ra*-Deletion, 2) for each verb stem, I assigned the word frequency with reference to NIJL (2005),[3] 3) I classified each token into four birth-year periods (grouped

every 20 years): from the 1910s to the 1920s, from the 1930s to the 1940s, from the 1950s to the 1960s, and from the 1970s to the 1980s, namely I conducted the analysis assuming the apparent-time methodology, in which the difference in speakers' birth-year corresponds to the apparent flow of time, 4) I calculated the chronological transition of the correlation between the probability of *ra*-Deletion and the frequency of verb stem.[4] As mentioned above, I focused on every potential form of vowel verbs, I mean, the forms that can be *ra*-Deleted. However, I excluded the verb *i-ru* from the data set as an outlier, since the frequency of *i-ru* in NIJL is exceptionally high (8,642) compared with the mean frequency (365) in the whole dataset, which skewed the results.

In Table 2, I summarized the probability of *ra*-Deletion in CSJ and the frequency in NIJL for some representative verbs. For each time-period, I examined the correlation between the probability of *ra*-Deletion and the frequency in the whole data set. The significance of correlation was tested with Kendall's rank correlation tau.

Table2: Probability of *ra*-Deletion and frequency of verb stems (entire time-period)

verb	probability of *ra*-Deletion in CSJ (%)	frequency in NIJL
ku-ru 'come'	76.09	1,845
ne-ru 'sleep'	60.00	53
de-ru 'go out, participate in'	45.56	613

3.2. Results

In Figure 2, I plotted the obtained correlation between the probability of *ra*-Deletion and the frequency of verb stem for each time-period. The vertical axis represents the probability of *ra*-Deletion, and the horizontal axis the frequency of verb stem. By comparing each of the correlations in four time-periods, we can estimate the chronological transition of correlation. Having a closer look at Figure 2, we can observe two kinds of correlations. Firstly, in every time-period we can observe the positive correlation between the probability of *ra*-Deletion and the frequency of verb stem, namely, as the frequency of verb stem increases, the probability of *ra*-Deletion also increases. Secondly, the degree of correlation is represented by the slope of each line; namely, the steeper the slope, the stronger the correlation. If we compare the lines by their slopes, we can observe the tendency that the slope gets progressively steeper, as the time-period becomes more recent. Here again, a positive correlation is attested.

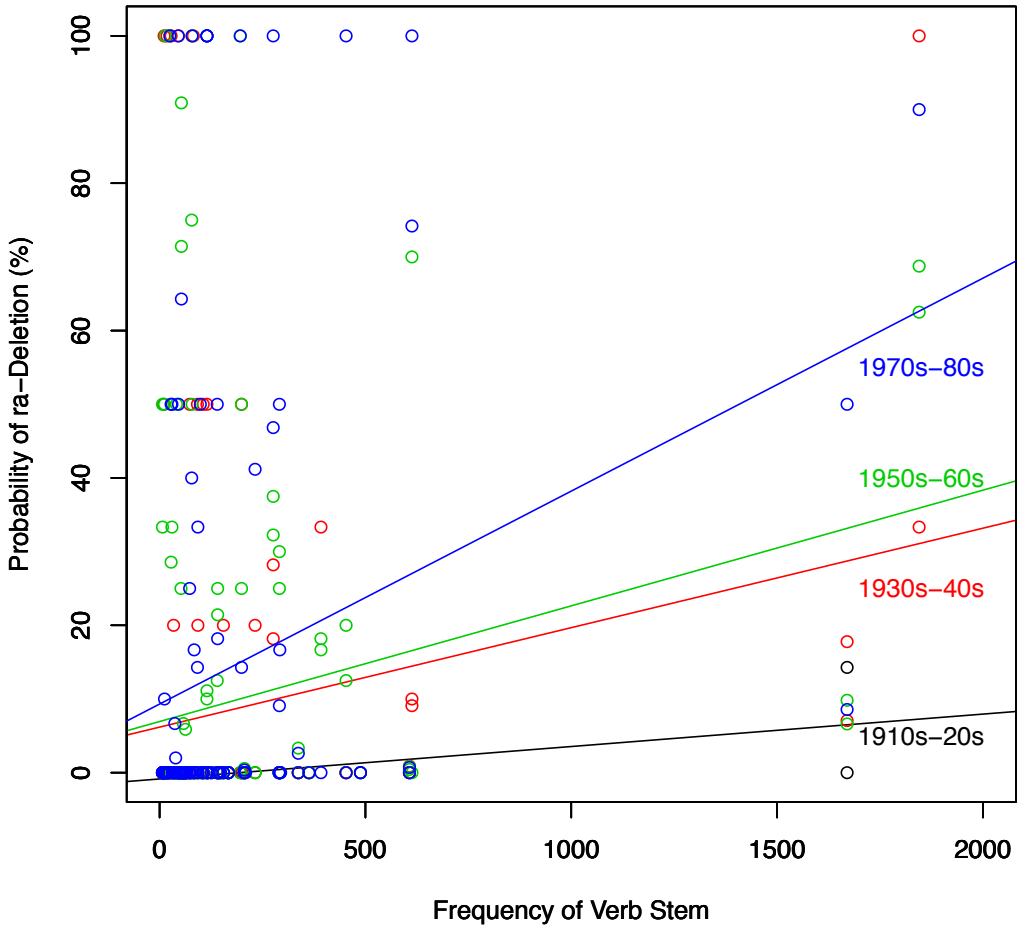

Figure 2: Correlation between probability of *ra*-Deletion and frequency of verb stem

In Table 3, each of the slopes is represented in the form of formula. From the oldest time-period to the latest time-period, the coefficient that represents the slope becomes increasingly higher. Furthermore, the results of the test for significance of correlation show that the correlation is more likely to be significant as the time-period becomes more recent. In the oldest time-period, the correlation was not significant; however after the 1930s to the 1940s the P-value gets progressively lower. This indirectly suggests that the correlation between the probability of *ra*-Deletion and the frequency of verb stem is getting stronger.

Table3: Transition of the degree and of the significance of correlation

time-period	degree of correlation	significance of correlation
1910s-1920s	y = 0.0044x -0.854	(n.s.)
1930s-1940s	y = 0.0135x -6.1776	(p<0.05)
1950s-1960s	y = 0.0157x -6.9389	(p<0.002)
1970s-1980s	y = 0.0289x -9.3121	(p<0.0005)

3.3. Discussion

Based on the results, I consider the relationship between the word frequency effect and the lexical diffusion. The following two points are consistent with the previous research (Bybee 2002; Phillips 2006). Firstly, the progress of the change and the word frequency show a positive correlation; Secondly, words with higher frequency are the preferred context for the change; namely, innovative forms are more likely to occur in words with higher frequency than in words with lower frequency. However, the tendency does not always hold true. In the case of *ra*-Deletion, at the outset (up to 1910s-1920s) *ra*-Deletions are less likely to occur in verbs with higher frequency as well as those with lower frequency; in other words, there is little gap between the verbs with higher frequency and those with lower frequency with respect to the progress of the change of *ra*-Deletion. However, along with the progress of the change (from 1930s-1940s on) it becomes more likely to occur in verbs with higher frequency; namely, there is a widening gap between these verbs. This suggests that as the change progresses, *ra*-Deletion diffuses into verbs with higher frequency, but not into verbs with lower frequency. In other words, the change starts at the same level across the lexicon; subsequently, it develops in a direction toward contexts with higher frequency.

These points can be generalized as follows. Words with higher frequency are not always the preferred context; namely, at the outset words with higher frequency are not preferred context or resistant to the change; in later stages, however, words with higher frequency change to the preferred context or are susceptible to the change. On the other hand, words with lower frequency are consistently not the preferred context; that is, consistently resist the change. Furthermore, based on the generalization above, we can refine the model of lexical diffusion: At the outset, both lower-frequency items and higher-frequency items are resistant to the change, and no lexical gap arises. In contrast, at later stages, lower-frequency items are still resistant to the change, however higher-frequency items are susceptible to the change, and here the lexical gap occurs. In summary, the word frequency effect is not stable, but dynamically changes. The effect is strengthened as the change progresses.

Finally, taking a broader view point, I will consider two kinds of factor effects. The Constant Rate Effect (Kroch 1989) predicts that internal factors affect the change in such a way that the rate of progress is stable or same in all contexts, because the change is driven by a single underlying rule; on the other hand, given that the word frequency effect is an external factor, we can argue that external factors affect the change in such a way that the rate of progress is different in each context, because the change is driven by a variable or dynamic factor. Here, we can parameterize the factor effects: The effects of internal factors are stable,

rendering the constant progress of the change; while, external factors are dynamic, and the progress becomes uneven.

4. Conclusion

In this paper, I demonstrated the dynamic or variable effects of word frequency. I examined the effects by means of the ongoing morphophonological change in Japanese potential forms, called *ra*-Deletion. The correlation between the probability of *ra*-Deletion and the frequency of verb stem was attested, and the correlation showed the chronological transition. The results show the positive correlation that words with higher frequency are preferred context for the change; word frequency effect itself is variable or dynamic; the effect is strengthened along with the progress of the change. Specifically, words with lower frequency are consistently resistant to the change, while those with higher frequency change from resistant to susceptible to the change. Consequently, as the change progresses, innovative forms are more likely to occur in words with higher frequency. This leads to the gap between lower-frequency items and higher-frequency items, being reflected in the lexical diffusion. Based on the Constant Rate Effect and the dynamic effect, two kinds of factors render the different outcomes. Internal factors with the Constant Rate Effect render regular changes; on the other hand, external factors with the dynamic effect render irregular changes; that is, lexical diffusion.

References

Bybee, Joan. (2002). Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change. Language Variation and Change 14: 261–290.

Bybee, Joan. (2007). Frequency of Use and the Organization of Language. New York: Oxford University Press.

Bybee, Joan & Paul Hopper (eds.). (2001). Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Campbell, Lyle. (2004). Historical Linguistics: an Introduction, 2nd edition. Edinburgh: Edinburgh University Press, and Cambridge, MA: MIT Press.

Hooper, Joan. (1976). Word Frequency in Lexical Diffusion and the Source of Morphophonological Change. In Christie, W. (eds.) Current Progress in Historical Linguistics, 96–105. Amsterdam: North Holland.

Ito, Junko & Armin Mester. (2004). Morphological Contrast and Merger: Ranuki in Japanese. Journal of Japanese Linguistics 20: 1–18.

Kroch, Anthony. (1989). Reflexes of Grammar in Patterns of Language Change. Language Variation and Change 6: 199–244.

Labov, William. (1981). Resolving the Neogramarian Controversy. Language 57: 267–308.

_____. (1994). Principles of Linguistic Change: Internal Factors. Oxford: Basil Blackwell.

Maekawa, Kikuo. (2004). Nihongo hanashikotoba koopasu-no gaiyoo [An Overview of the Corpus of Spontaneous Japanese]. Nihongo Kagaku 15: 111–133.

Matsuda, Kenjiro. (1993). Dissecting Analogical Leveling Quantitatively: The Case of the Innovative Potential Suffix in Tokyo Japanese. Language Variation and Change 5: 1–34.

National Institute for Japanese Language (NIJL). (2005). A Survey of Vocabulary in Contemporary Magazines (1994). Tokyo: National Institute for Japanese Language.

_____. (2008). The Corpus of Spontaneous Japanese, Ver. 2. Database.

Phillips, Betty. (2006). Word Frequency and Lexical Diffusion. New York: Palgrave Macmillan.

Sano, Shin-ichiro, & Junko Hibiya. (2012). Nihongo hanasikotoba koopasu-o tukau [Usage of the Corpus of Spontaneous Japanese]. In Junko Hibiya (ed.), Hajimete manabu syakaigengogaku [Introduction to sociolinguistics], 80–99. Tokyo: Minerva Shobo.

Wang, William S-Y. (1969). Competing Changes as a Cause of Residue. Language 45: 9–25.

_____. (1977). The Lexicon in Phonological Change. The Hague: Mouton.

Appendix: list of abbreviations

NOM: Nominative POT: Potential POL: Polite NONP: Nonpast

* I would like to thank Kevin Heffernan, Miriam Meyerhoff, Satoshi Nambu, Lawrence Reid, Naomi Nagy, Natsuko Tsujimura, Fumio Inoue, and other participants of NWAV-AP2 for their invaluable comments. I am also grateful to the Department of Language Change and Variation at NINJAL for their help organizing the conference. Any remaining faults are, of course, mine.

[1] The direction of the effect can be reversed depending on the size of the domain: unlike the present case that is categorized as phonological or morphological change, and smaller in its domain, if the domain is large such as syntax, then words with lower frequency are more likely to be the locus of changes or the target of rule application (Phillips 2006).

[2] The alphanumeric character annotated to the end of each example (e.g. S00M0002) is the 'speech ID' that is used as the index of each speech. See Maekawa (2004) and Sano & Hibiya (2012) for more detail.

[3] NIJL lists 48,000 lexical items sampled from the magazines published in 1994 with properties such as the frequency and the grammatical information.

[4] In this analysis, the frequency of verb stem is constant across every time-period.