Homomorphemic Diffusion in Japanese Nonce Lexemes

Mark Irwin

1. Introduction

Modern Japanese (hereafter ModJ) exhibits apparently irregular allomorphic behavior within one subset of its S[ino]-J[apanese] morphemes. The morphemes in question are those which are bimoraic and end in *-ki*, the irregular allomorphy occurring when they appear as the initial morpheme in a SJ bimorphemic compound whose second morpheme begins in *k*-.[1] This is illustrated in (1), where the six bimorphemic SJ lexemes have an initial morpheme ending in *-ki* and all are listed in the four major dictionaries Nihon Daijiten Kankōkai 1972–76, Tōdō et al. 2001, Shinmura 2003, and Watanabe, Skrzypczak, and Snowden 2003.

(1)	a.	石灰	*se**Q.k**ai*	(**se**ki.k**ai*)	‘quicklime’
	b.	積極	*se**Q.k**yoku*	(**se**ki.k**yoku*)	‘positive’
	c.	僻見	*he**ki.k**eN*	(**he**Q.k**eN*)	‘bias’
	d.	色界	*si**ki.k**ai*	(**si**Q.k**ai*)	‘realm of form (Budd.)’
	e.	適格	*te**Q.k**aku* ~ *te**ki.k**aku*		‘competence’
	f.	激高/激昂	*ge**Q.k**oo* ~ *ge**ki.k**oo*		‘exasperation’

All other bimoraic SJ morphemes derived from Early Middle Chinese codas in **-k* or **-t* (for a reconstruction see, *inter alia*, Pulleyblank 1991) —those ending in *-ku* (2a), *-tu* (2b) or *-ti* (2c)—show a clearly rule-based allomorphy before morphemes beginning with *k*-, whereby apocope of the final high vowel is accompanied by assimilation of the preceding consonant to the initial *k*- of the following morpheme. This assimilation can be analysed phonemically as the generation of the mora obstruent *Q*.[2] Examples of this regular “*Q*-generation” allomorphy are:

Japanese Language and Literature **42** (2008) 45–61

(2) a. *ku.k > Q.k*

*sya**ku***	+	***k**iN*	>	*sya**Q.k**iN***
借		金		借金
'borrow(ing)'		'money'		'debt'

b. *tu.k > Q.k*

*bu**tu***	+	***k**yoo*	>	*bu**Q.k**yoo***
仏		教		仏教
'Buddha'		'teach(ing)'		'Buddhism'

c. *ti.k > Q.k*

*ni**ti***	+	***k**i*	>	*ni**Q.k**i***
日		記		日記
'day'		'record'		'diary'

As is clear from (1) above, however, no such clear rule-based allomorphy can be posited for SJ morphemes ending in *-ki* (henceforth SJ {-ki} morphemes), where *Q*-generation appears to be irregular: SJ compounds with *Q*-generation (1a,b), without *Q*-generation (1c,d), and with variants in both (1e,f). These three types of SJ compounds I will henceforth refer to as *Q*-ful, *Q*-less and *Q*-variant, respectively. The apparent irregularity does not stop here, however. Even within individual sinographs—in (3) below within that of 石 *seki* 'stone'—we witness *Q*-ful (3a), *Q*-less (3b) and *Q*-variant (3c) SJ compounds (all lexemes listed in Shinmura 2003):

(3) a. 石鹸 *se**Q.k**eN* (**se**ki.k**eN*) 'soap'
b. 石国 *se**ki.k**oku* (**se**Q.k**oku*) toponym
c. 石果 *se**Q.k**a* ~ *se**ki.k**a* 'drupe, stone fruit'

In this paper I employ the notion of the homomorph, which I define as a group of morphemes all of which are pronounced identically but which are semantically (and in Japanese also orthographically) distinct. Thus, for example, {SEKI} refers to many different semantically and orthographically distinct {seki} morphemes (e.g., {seki} 石 'stone', {seki} 赤 'red', {seki} 席 'seat', etc.), which together form the {SEKI} homomorph.[3] My definition of the homomorph parallels that of the traditional homophone, but at the level of the morpheme rather than the orthographic word, while in addition being used to define a group rather than an individual entity: under my definition a homomorph is a group of homophonous, semantically distinct morphemes.[4]

After reviewing the relatively scant amount of previous research on the allomorphy outlined in (1)–(3), Irwin (2006:118–152) conducted a synchronic analysis of SJ {-ki} morphemes utilizing the four major dic-

tionaries already mentioned above in connection with (1). The results of this large corpus were necessarily complex but, to summarize, statistically significant trends which emerged among the ten SJ {-KI} homomorphs attested in the synchronic corpus ({EKI, GEKI, SEKI, TEKI, HEKI, BEKI, REKI, SIKI, ZIKI, RIKI}) included:[5]

(i) {SEKI} has an overwhelming tendency towards *Q*-generation.

(ii) {TEKI} has a strong tendency towards *Q*-variation.

(iii) If {BEKI}, {RIKI} and {ZIKI} are disregarded as statistically valueless, of the remaining five homomorphs, only {GEKI} exhibits a noticeable (although still low) level of *Q*-generation.[6]

(iv) The remaining four homomorphs, {EKI}, {HEKI}, {REKI} and {SIKI}, show an overwhelming tendency to eschew *Q*-generation.

Furthermore, a diachronic corpus (Irwin 2006:153–199), which utilized 20 written primary sources dating from the 12th to the late 19th centuries, revealed there to be only one isolated case of a lexeme of the type in (1) or (3) exhibiting *Q*-generation prior to the Meiji Restoration of 1867. This lexeme contained a {SEKI} homomorph and appeared in the *Unpo Irohashū* of 1548.

For those lexemes in both the synchronic and diachronic corpora where *Q*-generation does occur, the following sound change can be readily posited (where all morphemes belong to the SJ vocabulary stratum):

(4) {(C)Vki} → {(C)VQ} / ____ {k(y)V(C)(V)}

This can be viewed as an extension of the morphophonological rules within SJ lexemes that apply to {-ku} morphemes (2a) and thus to the morphophonological rules that apply to all SJ {(C)VCV} structures (2). Irwin (2006:153–162, 200–204) explains the varying degrees to which different homomorphs appear to have undergone and be undergoing the sound change in (4) as an example of the hitherto unrecognized phenomenon of homomorphemic diffusion, under which Labov's (1981:296) criteria for lexical diffusion must be viewed in a homomorphemic, rather than a lexical or morphemic, light in order to yield the correct output.[7] Entire groups of homophonous morphemes are evincing the spread of an identical sound change to a differing degree from other groups of homophonous morphemes: thus, homomorphs appear to be behaving in lexical diffusionist terms just as individual lexemes or morphemes might be

expected to. Based on a number of factors, Irwin (2006:200) goes on to posit a date for the onset of this diffusion of approximately the early part of the 19th century.

Lexical diffusion—and thus, by extension, homomorphemic diffusion—may be represented using an S-curve model which predicts that change will be most rapid between the 20% and 80% stages (the sweet spot) of a given process. A synthesis of the synchronic and diachronic data indicates that the isolated 1548 lexeme exhibiting *Q*-generation would mark the first plottable point in the {SEKI} S-curve, the diachronic Meiji data evidence of the slower initial 20% change (the starting friction), and the synchronic corpus data the sweet spot. The remaining 20% of the curve (the diminishing returns) has still to run its course, as evinced by the small number of {seki} morphemes which still remain *Q*-less or exhibit *Q*-variation in the synchronic corpus. By contrast, {TEKI} and {GEKI} would appear to show their first plottable S-curve points in the diachronic Meiji data, the large amount of *Q*-variation (though still only very few wholly *Q*-ful lexemes) in the synchronic corpus evidence of starting friction. The remaining homomorphs are still at a very early stage in the diffusional process—some (e.g. {EKI}) slightly more advanced, and others (e.g. {SIKI}) hardly registering at all. This can be somewhat crudely illustrated in Figure 1, where both the dates and the percentage of lexemes affected should be taken as gross approximations.

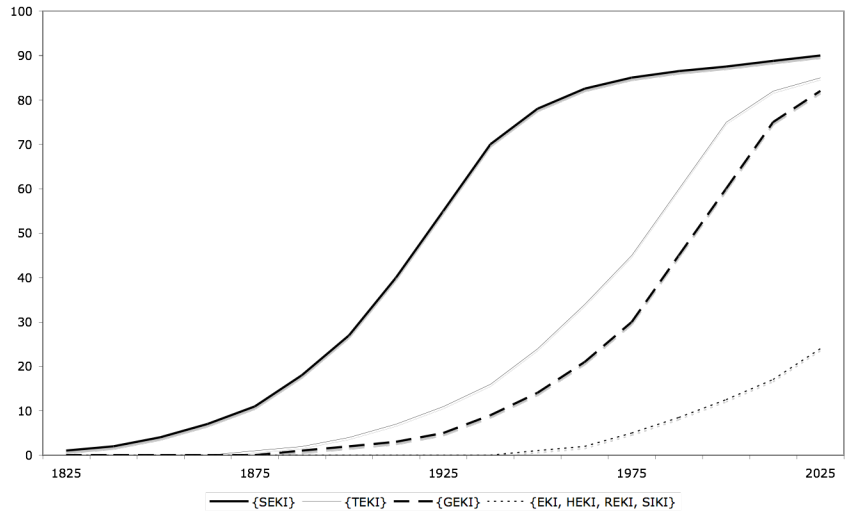

Figure 1. Homomorphemic Diffusion Over Time

Irwin (2006:205–209) provided two possible explanations for the advanced diffusion of {SEKI} homomorphs. These may be found in work carried out by Bybee (2000, 2001, 2002) and others (see Phillips 2006 for a thorough review) on the links between reductive sound change and lexeme frequency, as well as research by Ohno (2000) on "native speaker reference" or "native speaker memorization." What Irwin left for future study, however, is what I will examine in this paper, namely, to ascertain the correctness of the homomorphemic diffusion over time hypothesis, shown in Figure 1, through seeking confirmation outside dictionary entries, to which the corpora in Irwin 2006 restricted themselves. This I will do by means of a survey designed to elicit native speaker readings of nonce SJ bimorphemic compounds containing {-ki} morphemes.

The remainder of this paper proceeds as follows. First, in §2, I will discuss the theoretical rationale behind the survey. In §3 and §4, respectively, I will present its content and results. Finally, in §5, I will consider the sociolinguistic implications of these results and offer some conclusions on the homomorphemic diffusion hypothesis.

2. Theoretical Rationale

Were the sound change in (4) one of lexical diffusion or even one of analogy, we would expect to witness a great deal of variation in *Q*-generation right across the gamut of extant SJ bimorphemic compounds, this variation having no correlation with the compound's initial {-ki} morpheme.[8] Previous research (Martin 1952:30; Wenck 1957:117; McCawley 1968:116; Aoki 1981:89–93; Vance 1987:160) had indeed implied just this, and taken the matter no further. Irwin's claim as to the existence of both the homomorph and of homomorphemic diffusion, however, predicts that we find an underlying homomorphemic order to the variation in *Q*-generation witnessed in both corpora, an order which receives gross synchronic and diachronic simplification in Figure 1 above. If, in the mind of a native speaker, SJ bimorphemic compounds with an initial {SEKI} homomorph really are considerably more likely to undergo *Q*-generation than those with an initial {GEKI}, and these latter compounds are, in turn, considerably more likely to undergo *Q*-generation than those with an initial {EKI}, which are in fact unlikely to show any *Q*-generation at all, then we should expect similar levels of *Q*-generation to obtain in nonce SJ bimorphemic compounds also.

Creating such nonce lexemes (which for the remainder of this paper I

refer to as "Type A lexemes") is, however, not a task as simple as it initially may appear. Firstly, the lexemes must refer to concepts or objects which are realistically possible: total incomprehension on the part of respondents increases the likelihood of invalid responses. Additionally, for the same reason, the sinographs employed to write both morphemes of these semantically transparent nonce lexemes must themselves be semantically transparent, as well as occurring with a degree of frequency in the written language great enough to enjoy instant recognition on the part of the native reader. Sinographs which have two or more different, yet still semantically transparent, meanings (e.g., 劇 *geki* 'severe', as in 劇痛 *gekituu* 'severe pain', or 'drama, theater', as in 劇作家 *gekisaQka* 'playwright') must also be avoided in order to maximize the transparency of the compound as a whole. Moreover, although it was recognized from the outset that a survey administered to such a large number of respondents would necessarily elicit a wide range of often illogical and even impossible readings, in order to minimize such responses only sinographs with a single *on'yomi* (SJ reading) and as few *kun'yomi* (N[ative] J[apanese] readings) as possible (ideally, no *kun'yomi*) were to be utilized. Bearing in mind the semantic constraints above, however, this proved impossible to achieve in practice, although it was adhered to as closely as possible.

Although respondents were to be told before taking the survey that they were going to be asked to provide readings to nonce lexemes, it was felt that the overall statistics might become biased should they become aware too soon after beginning the questionnaire that what they were being surveyed for was the presence or absence of *Q*-generation. For this reason, the survey would also contain a second type of lexeme (hereafter "Type B lexemes"), namely SJ bimorphemic nonce lexemes whose initial morpheme was not of the {-ki} type, but which conformed to both the semantic and reading constraints imposed on the Type A lexemes. The fact that these Type B lexemes too were clearly and unambiguously SJ binoms would, it was hoped, make respondents more likely to believe that they were simply being surveyed for SJ binom readings in general, thus reducing the number of mixed *on-kun* or pure *kun* readings elicited for the vital Type A lexemes.

3. Survey Content and Administration

The survey consisted of 20 lexemes in total, 12 Type A and eight Type B.

Type B lexemes occurred more frequently in the first half of the survey (5/10) as opposed to the second half (3/10) in order to reduce potential statistical bias resulting from any realization of *Q*-generation elicitation on the part of the respondents. The survey was administered in written form and in Japanese. After being informed that the survey was part of a research project into Japanese phonology, the respondents were requested to indicate their age, gender, Japanese native speaker status, and whether or not they had already completed the survey. Japanese non-native speaker surveys were ignored, as were those which indicated the respondent had already completed a survey. Respondents were then requested to write in *katakana* how they would read 20 nonce lexemes, being told only that these SJ lexemes were nonce.[9] Table 1 below shows the 20 nonce lexemes in the order they appeared in the survey, along with an indication of whether they are Type A or B, an English nonce gloss, the expected *Q*-less SJ reading and, for Type A lexemes only, the expected *Q*-ful SJ reading.

Table 1. Nonce Lexemes Used in Survey

Nonce Lexeme	Type	English Nonce Gloss	Standard SJ Readings	
			Q-less	*Q*-ful
軍区	B	military district	*guN.ku*	
抹没	B	eradicate (rub out + rid)	*matu.botu*	
撃旗	A	attack flag	*geki.ki*	*geQ.ki*
敵血	A	enemy blood	*teki.ketu*	*teQ.ketu*
労圧	B	pressure of work	*roo.atu*	
駅犬	A	station dog	*eki.keN*	*eQ.keN*
適刑	A	appropriate punishment	*teki.kei*	*teQ.kei*
累税	B	extra tax	*rui.zei*	
石球	A	stone ball	*seki.kyuu*	*seQ.kyuu*
害愛	B	harmful love	*gai.ai*	
益婚	A	marriage for profit	*eki.koN*	*eQ.koN*
孤婚	B	lonely marriage	*ko.koN*	
赤菌	A	red fungus	*seki.kiN*	*seQ.kiN*
式館	A	ceremonial hall	*siki.kaN*	*siQ.kaN*
猟帽	B	hunting cap	*ryoo.boo*	
劇客	A	theater customer	*geki.k(y)aku*	*geQ.k(y)aku*[10]
暦期	A	calendar period	*reki.ki*	*reQ.ki*
慶娯	B	joy (delight + pleasure)	*kei.go*	
壁刻	A	wall carving	*heki.koku*	*heQ.koku*
潟江	A	lagoon inlet	*seki.koo*	*seQ.koo*

As already mentioned above, NJ readings or mixed SJ-NJ or NJ-SJ readings were also expected and these will be discussed in §4 below. Of the 12 Type A lexemes, three had an initial {SEKI} homomorph, two had {TEKI}, two {GEKI}, two {EKI}, one {SIKI}, one {REKI} and one {HEKI}, a ratio approximating to the frequency of their distribution in Irwin's (2006) synchronic corpus. The survey was administered to a total of 631 students at Yamagata University who attended English classes during April and May 2004. As the majority of English classes at the university are provided for first-year students, this age group formed the bulk of the respondents. Limiting the target age group to approximately 18–21 has advantages and disadvantages: although we fail to gain any information on levels of *Q*-generation in other age groups, we do obtain a reasonably unadulterated snapshot, not only of a single age range but of an age range whose members all have a broadly similar educational ability.

4. Survey Results

Table 2 shows the age and gender of the respondents. Overall, 89% (564/631) of respondents came from the 18–19 age group, with the bulk of the remainder coming from the 20–21 age group. Male students were in the majority at 56% (356/631) overall, as well as being the dominant gender in both major age groups.

Table 2. Age and Gender of Respondents

	18–19	20–21	22–23	24+	Total
♂	318	33	4	1	356
♀	246	25	3	1	275
Total	**564**	**58**	**7**	**2**	**631**

It is apparent from the raw data in Table 3 below that levels of *Q*-generation vary greatly across the 12 Type A nonce lexemes, a fact which will be examined in greater detail shortly. Both the number of non-standard readings and invalid responses also exhibited large levels of variation.[11] The data for 潟江 'lagoon inlet', which accrued only 5 (< 1%) standard SJ readings, is severely aberrant. The mostly likely reasons for this may be an ignorance on the part of the respondents as to the SJ reading for 潟 'lagoon' and the relative infrequency of this sinograph in

general, which is mainly confined to the toponym 新潟 *nii.gata*, where it appears in *rendaku* NJ form.[12] In addition, its level of invalid responses was also, at 14% (86/631), much higher than the overall average. Although tiredness or boredom was doubtless an issue here (invalid response rates increase towards the end of a survey for obvious reasons), one suspects that the factors of ignorance and infrequency already cited were also at play. The data for 潟江 were so aberrant that it was clear the decision to include the lexeme in the survey had been grossly infelicitous, that statistically it was valueless, and that to include it in averages and totals would seriously skew the overall data. For this reason it was decided to expunge data for 潟江, and it does not appear in any of the further tables below.

Table 3. Type A Nonce Lexeme Response Data

Nonce Lexeme	Standard SJ Readings			Non-Standard Readings	Invalid Response	Total
	Q-ful	*Q*-less	Subtotal			
撃旗	151	361	**512**	96	23	**631**
敵血	188	187	**375**	236	20	**631**
駅犬	27	302	**329**	266	36	**631**
適刑	66	515	**581**	31	19	**631**
石球	391	132	**523**	63	45	**631**
益婚	42	496	**538**	72	21	**631**
赤菌	228	281	**509**	113	9	**631**
式館	6	540	**546**	62	23	**631**
劇客	47	501	**548**	52	31	**631**
暦期	127	337	**464**	129	38	**631**
壁刻	15	512	**527**	62	42	**631**
潟江	0	5	**5**	540	86	**631**
Total	1,288	4,169	**5,457**	1,722	393	**7,572**

Turning now to non-standard responses in general, readings such as *eki.inu* and *hati.keN*[13] helped push the non-standard reading rate for 駅犬 'station dog' up to as high as 42% (266/631), while at the other end of the spectrum 適刑 'appropriate punishment' evinced a non-standard response rate of a mere 5% (31/631). Clearly, as expected and noted in §3, the availability of an alternative SJ or NJ reading was a major factor in a higher non-standard reading rate, with nonce lexemes such as 益婚 'marriage for profit' and 式館 'ceremonial hall', where neither sinograph has any alternative readings (whether NJ or SJ) at all, showing low rates, and those such as 敵血 'enemy blood' and 赤菌 'red fungus', where one or

both sinographs have alternative readings, showing high rates. With the 潟江 'lagoon inlet' data expunged, the average non-standard reading rate was 17% (107/631) or, more importantly when viewed from a different angle, the average standard SJ reading rate was 79% (496/631), with only 駅犬 'station dog' showing a rate substantially lower than this.

Table 4 below shows *Q*-generation rates amongst standard SJ readings for the remaining 11 type A lexemes. Here, the only two lexemes to exhibit a *Q*-generation rate of over 50% were 石球 'stone ball' and 敵血 'enemy blood', while a total of five (式館 'ceremonial hall', 壁刻 'wall carving', 益婚 'marriage for profit', 駅犬 'station dog' and 劇客 'theater customer') showed rates of under 10%.

Table 4. Type A Nonce Lexeme *Q*-Generation Rates

Nonce Lexeme	*Q*-Generation
石球 'stone ball'	74.8%
敵血 'enemy blood'	50.1%
赤菌 'red fungus'	44.8%
撃旗 'attack flag'	29.5%
暦期 'calendar period'	27.4%
適刑 'appropriate punishment'	11.4%
劇客 'theater customer'	8.6%
駅犬 'station dog'	8.2%
益婚 'marriage for profit'	7.8%
壁刻 'wall carving'	2.8%
式館 'ceremonial hall'	1.1%

Grouping the data in Table 4 by homomorph, however, allows us to make the fundamental comparison between native speaker attitudes towards *Q*-generation in nonce lexemes and the proposed chronology of homomorphemic diffusion outlined in Figure 1. From Table 5 below, it is clear that the only homomorph to exhibit a *Q*-generation of over 50% is {SEKI}, while {EKI}, {HEKI} and {SIKI} all show rates lower than 10%. More important, however, is Table 6, which compares the nonce lexeme *Q*-generation rates in Table 5 with synchronic corpus dictionary entry *Q*-generation rates in Irwin 2006:149, from which the proposed chronology in Figure 1 is extrapolated.[14] While the higher overall rates evident in the synchronic corpus are in all probability an outcome of the different media employed and need not concern us here, what is striking is the extremely close correlation in terms of homomorph rank between

the nonce lexeme survey and the synchronic corpus in Irwin 2006. With a Pearson coefficient (*r*) of 0.91, it is only {REKI}, with a *Q*-generation rate in the nonce lexeme survey higher than that in the synchronic corpus, and {SIKI}, which shows the reverse phenomenon, that exhibit an unexpected result.

Table 5. Type A Nonce Lexeme *Q*-Generation Rates by Homomorph

Homomorph	Nonce Lexemes	*Q*-ful Readings	Standard SJ Readings	*Q*-Generation
{SEKI}	石球 赤菌	619	1,032	60.0%
{REKI}	暦期	127	464	27.4%
{TEKI}	敵血 適刑	254	956	26.6%
{GEKI}	撃旗 劇客	198	1,060	18.7%
{EKI}	駅犬 益婚	69	867	8.0%
{HEKI}	壁刻	15	527	2.8%
{SIKI}	式館	6	546	1.1%
All		**1,288**	**5,452**	**23.6%**

Table 6. Nonce Lexeme and Synchronic Corpus *Q*-Generation Rates by Homomorph

Homomorph	Nonce Lexemes	Synchronic Corpus Dictionary Entries
{SEKI}	60.0%	85.3%
{TEKI}	26.6%	44.7%
{GEKI}	18.7%	30.2%
{REKI}	27.4%	9.5%
{SIKI}	1.1%	8.6%
{EKI}	8.0%	8.0%
{HEKI}	2.8%	2.2%

5. Conclusions and Further Implications

The implications of the correlation apparent in Table 6 are clear. The levels of *Q*-generation across homomorphs for a range of nonce lexemes mirror closely the proposed chronology of homomorphemic diffusion shown in Figure 1. It would appear that, to a very great extent, the phenomenon of homomorphemic diffusion is a real one; that {SEKI} does indeed hold an extremely advanced position vis-à-vis other homomorphs; that *Q*-generation rates for {TEKI} and {GEKI} are lagging behind although advancing fast; and that rates for {SIKI}, {HEKI} and {EKI} are

still evincing starting friction. Why {REKI} should have recorded an unexpectedly high level of *Q*-generation in the nonce lexeme survey reported here is a difficult question. Clearly, a further survey containing more than one nonce SJ lexeme with an initial {REKI} homomorph is necessary to validate the result obtained here and to ascertain whether it should hold a more advanced position on the S-curve than proposed.

An interesting corollary of the data obtained in this survey is the difference between the levels of *Q*-generation obtained from male and female respondents. Table 7 gives the *Q*-generation rates by homomorph shown in Table 5 split according to gender, with homomorphs ranked by overall *Q*-generation rate. In all four advanced homomorphs ({SEKI}, {TEKI}, {GEKI} and {REKI}) female speakers exhibit higher levels of *Q*-generation, while the opposite is generally true (with the exception of {SIKI}) for the remaining three.

Table 7. Type A Nonce Lexeme Q-Generation Rates by Gender

Homomorph	*Q*-Generation Male	*Q*-Generation Female	*Q*-Generation All
{SEKI}	57.3%	63.1%	60.0%
{REKI}	24.2%	31.3%	27.4%
{TEKI}	24.3%	29.4%	26.6%
{GEKI}	27.1%	32.6%	18.7%
{EKI}	9.2%	6.5%	8.0%
{HEKI}	3.7%	1.7%	2.8%
{SIKI}	0.6%	1.7%	1.1%

A common statistical pattern revealed by sociolinguistic research is that women's data conform more closely to a prestige or standard variant than do men's (see, *inter alia*, Wolfram 1969, Trudgill 1974, Milroy and Milroy 1978, Coates 1993).[15] This patterning frequently receives its explanation in terms of status (i.e., women's subordinate role and consequent powerlessness in most societies leads to higher linguistic consciousness and susceptibility to 'pressure from above'; see Labov 1972 and Trudgill 1974), although Milroy and Milroy (1993) suggest that female forms are themselves the source of prestige in a society because of an "ideology of the feminine." Whatever the reasons for female data mirroring more closely prestige or standard forms may be, the data in Figure 7 suggest strongly that the sound change in (4) is yet another example of this sociolinguistic patterning, with the four advanced homomorphs all showing higher female scores. This should not be seen as surprising:

since *Q*-generation is complete in SJ morphemes ending in *-ku*, *-tu*, and *-ti* in identical environments (see (2) above), what we are witnessing is simply SJ morphemes ending in *-ki* beginning to conform to the same "standard" allomorphy with female speakers in the vanguard. Whether this patterning also extends to other environments involving *Q*-generation, where, for example, morphemes are not restricted to the SJ stratum (e.g. *buti.komu~buQ.komu*) or where the *Q*-generation is morpheme-internal and of a non-emphatic nature (e.g. *ati~aQti*) is a question only further research can resolve.

NOTES

1. Prefixed SJ bimorphemic compounds (副大臣 *huku-daiziN* 'vice minister'), compound SJ bimorphemic compounds (総理大臣 *soori-daiziN* 'prime minister'), prefixed compound SJ bimorphemic compounds (前総理大臣 *zeN-soori-daiziN* 'ex prime minister'), and further combinations thereof, are all possible. There is a major constituent break between a prefix and a SJ bimorphemic compound, or between two SJ bimorphemic compounds, and in such cases my definition refers to the second morpheme after such a break. *Q*-generation never occurs before a major constituent break, e.g., the prefixed SJ bimorphemic compound 核開発 *kaku+kai.hatu* (not **kaQkaihatu*) 'nuclear development' or the compound SJ bimorphemic compound 教育改革 *kyoo.iku+kai.kaku* (not **kyooiQkaikaku*) 'education reform'.
2. Phonetically, a *QC* phoneme cluster is a long or geminate consonant, exhibiting "suspension of articulatory movement" (Komatsu 1980:566) or "prolongation of consonantal articulation" (Shibatani 1990:167–168). See Vance 1987:39–47 for an overview.
3. I use braced majuscule for the transcription of homomorphs.
4. The standard definition of homophone implies not just semantic but also orthographic difference and thus in some circles a notion of "homomorph," different from my own, has been created to describe homophones with an identical orthography. By this definition a "homomorph" is a homophonous, semantically distinct, orthographically identical word. In Japanese, where sinography is still an indispensable component of a complex orthographical system, my homomorphs happen to be also orthographically distinct, as they would be in Chinese. In the other Sinoxenic languages, however, where sinography has been replaced either completely (Vietnamese, Korean as prescribed in North Korea) or largely (Korean as prescribed in South Korea) by an alphabet, my homomorphs would not be orthographically distinct, bringing them closer to the other definition of homomorph just described. I shelve the issue of orthographical distinction for future debate.

5. There are actually 16 permissible SJ ModJ {-KI} homomorphs, of which {KEKI, DEKI, IKI, TIKI, HIKI, BIKI} fail to occur in the synchronic corpus, i.e., while there are SJ morphemes of these six shapes, none occurs bound as the initial morpheme in a SJ bimorphemic compound whose second morpheme begins in *k*-.
6. All appear in only five lexemes or fewer (<2% of the synchronic corpus) and in only 10 tokens or fewer (again, <2% of the corpus).
7. Although first noted by Sturtevant (1917:82), the appellation "lexical diffusion" is usually credited to Wang (1969). While discussing the relative chronology of competing sound changes in, primarily, the Amoy and Lóngxī dialects of Chinese, Wang argued that sound change is phonetically abrupt and lexically gradual, a position diametrically opposite to the Neogrammarian position. The ensuing debate was lengthy, complex, and is still ongoing. A full treatment is beyond the scope of this paper. Labov (1994: 542–543), however, sought to resolve the Neogrammarian-diffusionist paradox by predicting that "the realms of regular sound change and lexical diffusion [will] display complementary distribution." This is illustrated in the table below, which is adapted from Labov 1981:296, with an additional column on the far right indicating into which category the trends visible from the historical corpus in Irwin 2006 fall. The order of the criteria has been altered from those cited by Labov for ease of exposition.

Labov's Criteria	Neogrammarian Change	Lexical Diffusion	Irwin (2006) Historical Corpus
discrete	no	yes	yes
dictionary entries	1	2	2
phonetic conditioning	fine	rough	rough
grammatical conditioning	no	yes	n/a
lexical exceptions	no	yes	yes
social effect	yes	no	no
predictable	yes	no	no
learnable	yes	no	no
categorized	no	yes	yes
lexical diffusion	no	yes	yes

8. See Irwin (2006:153–162) for a detailed discussion on analogy with regard to (4).
9. Eliciting readings in *katakana* rather than *hiragana* reinforces further the SJ rather than the NJ element, since traditionally *on'yomi* are written using the former and *kun'yomi* using the latter. An anonymous reviewer has pointed out that, while it is probably true that *katakana* has a strong association with *on'yomi,* eliciting readings in *katakana* may also have had the unintended effect of making respondents feel they were writing foreign words (*gairaigo*). While his/her point is valid, it would seem—fortunately—that this was only an unintended side-effect and that the scales tipped overwhelmingly in favour of *on'yomi*. Only two respondents showed any noticeable

"*gairaigo* effect" and then not even in the majority of their responses. These included *sutoresu* 'stress' for 労圧 'pressure of work', *monyumeNto* 'monument' for 壁刻 'wall carving', and *korera* 'cholera' for 赤菌 'red fungus'. However, the same two respondents also showed a large number of unique (and rather amusing) NJ responses: *makeinu* 'sad loser' for 孤婚 'lonely marriage', *huriN* 'adultery' for 害愛 'harmful love', and *siawase* 'happiness' for 益婚 'marriage for profit'. Total *gairaigo* responses were in the low teens (<0.2% of the total) and, apart from the two respondents just cited, overwhelmingly confined to readings for 害愛 'harmful love': *eizu* 'AIDS', *gei* 'gay', *esuemu* 'S&M', *sutookaa* 'stalker', etc.

10. 客 has two standard SJ readings, *kyaku* and *kaku*.
11. "Non-standard readings" included NJ readings, mixed SJ-NJ readings, and mixed NJ-SJ readings, while "invalid responses" included answers written in sinographs or the Roman alphabet, incomplete responses, and no responses.
12. The likelihood of this second possibility receives strong support form the fact that the most popular response (249/631) for 潟江 was *gata.e*, where both readings are NJ but the former has mysteriously undergone *rendaku* (which is, of course, a Japanese allomorphic phenomenon impossible in word-initial position).
13. The seemingly inexplicable reading of *hati* for 駅 probably derives from associating the idea of 'station dog' with Hachikō, a famous dog who waited years for his deceased master at Ueno Station in Tokyo—a tale all Japanese children learn in school.
14. Note that Irwin 2006 does not confine itself to the seven homomorphs shown here. The paucity of data for {BEKI}, {RIKI} and {ZIKI} homomorphs renders them statistically worthless, however, and thus they were included neither in Figure 1 nor in the nonce lexeme survey.
15. Researchers working in the field of gender and language have noted that the links between gender and prestige forms are complex and that they vary in different cultural and discourse contexts. For discussion of this issue, see James 1996.

REFERENCES

Aoki, Paul Ken. 1981. Gemination in Japanese. Ph.D. diss., University of Washington.

Bybee, Joan. 2000. The phonology of the lexicon: Evidence from lexical diffusion. *Usage-based models of language*, ed. by Michael Barlow and Suzanne Kemmer, 65–85. Stanford: CSLI.

———. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

———. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261–290.

Coates, Jennifer. 1993. *Women, men and language*. Harlow: Longman.

Irwin, Mark. 2006. Mora obstruent allomorphy in Sino-Japanese morphemes in final -/ki/: A case of homomorphic diffusion in Modern Japanese. Ph.D. diss., University of Sheffield.

James, Deborah. 1996. Women, men and prestige speech forms: A critical review. *Rethinking language and gender research: Theory and practice*, ed. by Victoria L. Bergvall, Janet M. Bing, and Alice F. Freed, 98–125. Harlow: Addison Wesley Longman.

Komatsu, Hideo. 1980. Sokuon. *Kokugogaku daijiten*, ed. by Kokugo Gakkai, 566–567. Tokyo: Tōkyōdō.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

——— 1981. Resolving the Neogrammarian controversy. *Language* 57:267–308.

——— 1994. *Principles of linguistic change: Internal factors.* Oxford: Blackwell.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Martin, Samuel E. 1952. Morphophonemics of standard colloquial Japanese. *Language* 28(3), Part 2, Supplement.

——— 1987. *The Japanese language through time*. New Haven and London: Yale University Press.

Milroy, James, and Lesley Milroy. 1978. Belfast: Change and variation in an urban vernacular. *Sociolinguistic patterns in British English*, ed. by Peter Trudgill, 19–36. London: Edward Arnold.

——— 1993. Mechanisms of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics* 3:57–78.

Nihon Daijiten Kankōkai, ed. 1972–76. *Nihon kokugo daijiten*. Tokyo: Shōgakukan.

Ohno, Kazutoshi. 2000. The lexical nature of *rendaku* in Japanese. *Japanese/Korean linguistics 9*, ed. by Mineharu Nakayama and Charles J. Quinn, Jr., 151–164. Stanford: CSLI.

Phillips, Betty. 2006. *Word frequency and lexical diffusion*. Basingstoke: Palgrave Macmillan.

Pulleyblank, Edwin. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: University of British Columbia Press.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Shinmura, Izuru, ed. 2003. *Kōjien*. Tokyo: Iwanami Shoten. [CD-ROM version on *Sūpā tōgō jiten 2003*, Fujitsū Corporation]

Sturtevant, Edgar. 1917. *Linguistic change: An introduction to the historical study of language*. Chicago: University of Chicago Press.

Tōdō, Akiyasu, Akira Matsumoto, Akira Takeda, and Yoshimitsu Kanō, eds. 2001. *Kanwa daijiten*. Tokyo: Gakushūkenkyūsha. [CD-ROM version]

Trudgill, Peter. 1974. *The social differentiation on English in Norwich.* Cambridge: Cambridge University Press.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: State University of New York Press.

Wang, William S-Y. 1969. Competing changes as a cause of residue. *Language* 45:9–25.

Watanabe, Toshirō, Edmund Skrzypczak, and Paul Snowden, eds. 2003. *Kenkyūsha's new Japanese-English dictionary*, 5th ed. Tokyo: Kenkyūsha.

Wenck, Günther. 1957. *Japanische Phonetik Vol. 3: Die Phonetik des Sinojapanischen*. Wiesbaden: Otto Harrassowitz.

Wolfram, Walter. 1969. *A sociolinguistic description of Detroit negro speech*. Washington: Center For Applied Linguistics.