ASIAN EFL JOURNAL
The EFL Professional's Written Forum

Volume 7. Issue 2
Article 2

Frequency Effects on Japanese EFL Learners' Perception of Morphologically Complex Words

Takahiro, Ioroi, Kochi Women's University and Masayoshi, Otagaki, Naruto University of Education

1. Introduction

1.1 Studies on Word Structure in Formal Linguistic Theory

Looking back over the history of the theory of generative linguistics, we notice that the structure of the lexicon or the lexical component of the grammar, which was once considered mere storage filled with unpredictable information, has drawn a number of researchers' attention since early 1970s. From the late 1970s through the mid-1980s, a theory called Lexical Morphology (Siegel, 1974; Kiparsky 1982; Mohanan 1986, among others) took shape and played an undeniably important role in explicating a number of intricate phenomena related to the interface between phonology and morphology. The theory seemed successful in explaining the following observations. First, there are two types of suffixes in English with respect to their phonological properties: Class I suffixes and Class II suffixes. Class I suffixes may affect word stress assignment while Class II suffixes show no such effect. Second, suffixation is subject to a certain ordering relation. Put simply, a Class I suffix is always attached to the stem before any Class II suffixation takes place. To handle these observations, the theory was so constructed as to have a multi-layered structure which allows Class I suffixation to take place at the first layer (Level 1) and Class II suffixation at the second layer (Level 2), thereby ruling out such forms as **neighborlity*. This ill-formed word offers a contrast to its well-formed counterpart *neighborliness* in that *-ly*, a Class II suffix, can precede another Class II suffix *-ness* for the attested form *neighborliness*, while for the unattested form **neighborlity*, *-ity* is no longer available at Level 2 due to its Class I status. The internal structures of *neighborliness* and **neighborlity* are represented in (1) and their licit and illicit derivational processes are illustrated in (2).

(1) $[[[\text{neighbor}]_N\text{ly}]_A\text{ness}]_N$
$*[[[\text{neighbor}]_N\text{ly}]_A\text{ity}]_N$

(2)

Underlying Representation		/neighbor/	/neighbor/
Level 1:	*-ity*	NA	NA
Level 2: 1st cycle	*-ly*	[neighbor-ly]	[neighbor-ly]
2nd cycle	*-ness*	[[neighbor-ly]-ness]	NA

Output Representation [neighborliness] --------

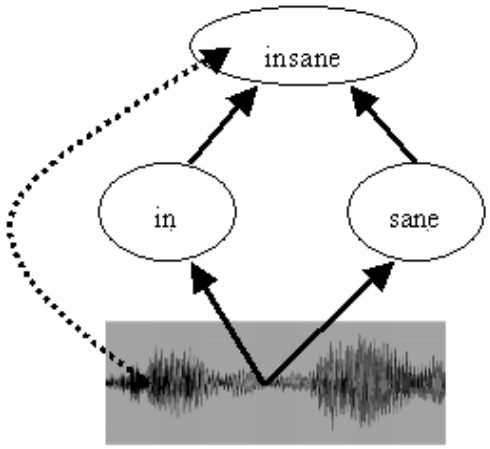
It should be pointed out, however, that the validity of building up a theory of mental lexicon purely on formal grounds does not seem well-justified, given that we consider the nature of word formation and the lexicon in conjunction with issues on the psychological reality of word accessibility/decomposability. For example, in such a framework, we do not seem capable of explaining the following questions: (i) why native speakers of English almost always answer that *perceptiveness* sounds better and *perceptivity* sounds awkward or strange although both of them are possible (Aronoff, 1976: 35); (ii) why they judge the word *pipeful*, for example, to be more complex (i.e., easier to decompose) than the word *bowful* despite the fact that they share the same suffix *-ful*. (Hay, 2000: 280).

1.2 Frequency-Based Theory of Word Structure and Decomposability

In contrast to formal approaches to word structure in generative theory, we see alternative theories in psycholinguistics seeking to capture the way in which speech streams are perceived, recognized and organized into word forms, by means of lexical frequencies. Regarding how a given word is accessed through one's mental lexicon, there are some important proposals worth reviewing. Cutler (1981) asserts that the acceptability of new words relies crucially on the degree of their phonological transparency. Bybee (2001:28) argues that "high-frequency items grow strong and therefore are easier to access" and "little-used items will tend to fade in representational strength and grow more difficult to access". Among others, Hay's (2000) study on word access in derivational morphology seems particularly important for the present study because of her argument for the assumption that the difference in the frequencies of a given suffixed word and its base form has an influence on the distribution, phonological alternations, and productivity of the word. This type of frequency relation is termed "relative frequency", to which we will turn later.

First, it is helpful to provide a brief description of the theory of lexical decomposability and then summarize the conclusion that Hay (2000) drew from her experiment with L1 subjects. In the diagram in Figure 1, Hay (2000: 9) presents a schema of the dual-route model of morphological access.

Figure 1: Schema of the dual-route access model



The solid lines above the fast phonological preprocessor indicate the decomposed route, and the

broken line indicates the direct access route. Frauenfelder and Schreuder (1991) assume that the parsing route is facilitated by both phonological and semantic transparency. For example, despite the fact that *insane* and *sanity* both contain the base form *sane*, *insane* exhibits a greater degree of transparency both phonologically and semantically than *sanity* does. Hay (2000), taking advantage of information on lexical frequency obtained from the CELEX database,[1] demonstrates that a choice between the two routes for a given complex word relies heavily on the information on the frequency of the word; she also emphasizes that what is crucial for such a choice is not the surface frequency of the complex word but the relative frequency of the word against that of the base form it contains.

Hay's analysis is based on the responses that 16 Northwestern University undergraduate students made in completing a lexical decomposability task which asked them to indicate which member of the pair they considered more "complex", i.e. easier to break down into parts, for a set of stimuli composed of 64 word pairs (30 fillers included). She states that "[f]orms which were more frequent than the bases they contained were consistently rated less complex than their counterparts, which were less frequent than the bases they contained" (p. 126). Her conclusion is drawn from the fact that for prefixed forms, 65% of the responses judged forms that were less frequent than their bases to be more complex (easier to decompose), and for suffixed forms 66% of the responses made the same judgment.

The present paper is organized as follows. Section 3 presents a brief description of the material we employed, the participants in the experiment, and the procedure of the experiment. Section 4 examines the results of the experiment, starting with an overview of the entire data, followed by discussions on the relevance of vocabulary size to the learners' perception of morphological complexity. Section 5 concludes this exploration and comments on its practical aspect and a remaining issue for future studies.

2. Method

Material: Forty-nine pairs of morphologically complex words were selected from the Longman Defining Vocabulary (LDV), which we considered would provide a range of words familiar and approachable to EFL learners.[2] The selection of these words was intended to make sure that the members of each pair exhibit surface frequencies close to each other but differ in that one member is more frequent than its own base form while the other member shows the opposite type of frequency profile.[3] Table 1 presents a description of the frequency profiles of two out of the 49 pairs: *insurance* and *formal* and *infection* and *determination*.

Table 1: Samples of stimuli

ID #	Derivative	Frequency	Base	Frequency	Type
#10	*Insurance*	5704	*Insure*	342	D > B
	Formal	6111	*Form*	33246	D < B
#28	*Infection*	2636	Infect	106	D > B
	Determination	2667	*Determine*	3852	D < B

The task sheet was completed with a random transposition and sequencing of these 49 word pairs.

Subjects: 96 first- to third-year students at a national university and a prefectural university in Japan.

Procedure: A statement describing instructions for the subjects was prepared. They were asked to read a description of the notion "complex word" first and say which one of the two members of each pair they considered more decomposable.

3. Results and Discussion

3.1 An Overview

In Table 2 below, the cells in the column headed by "D>B", on the one hand, show the numbers of responses chosen as derivatives (D) which are of higher frequency than their base words (B) and more decomposable. The cells in the column headed by "D<B", on the other hand, show the responses judging the derivatives as less frequent than their base words as more decomposable.

Table 2: Results (all subjects, N=96)

Pair	D > B	D < B	Pair	D > B	D < B	Pair	D > B	D < B
#1	71	25	#21	64	32	#41	76	20
#2	18	78	#22	71	25	#42	31	65
#3	93	3	#23	56	40	#43	26	70
#4	72	24	#24	22	74	#44	76	20
#5	59	37	#25	61	35	#45	41	55
#6	38	58	#26	52	44	#46	23	73
#7	57	39	#27	61	35	#47	75	21
#8	35	61	#28	76	20	#48	26	70
#9	70	26	#29	44	52	#49	56	40
#10	75	21	#30	75	21	N	49	49
#11	28	68	#31	54	42	Total	2667	2037
#12	68	28	#32	66	30	%	56.70	43.30
#13	28	68	#33	60	36	M	54.43	41.57
#14	89	7	#34	12	84	SD	21.46	21.46
#15	70	26	#35	42	54	*t*-value	2.097*	
#16	75	21	#36	78	18	*$p < .05$		
#17	70	26	#37	74	22			
#18	40	56	#38	50	46			
#19	19	77	#39	41	55			
#20	22	74	#40	81	15			

The total number of responses judging "D>B" derivatives as more decomposable is 2667 (56.70%) and the number of the responses calling for the opposite type is 2037 (56.70%). A T-test conducted against this result indicates that the differences between these two categories are statistically significant at the .05 level. This seems to support the hypothesis that the relative frequency between the surface frequency of a suffixed word and that of its base form exerts a certain influence on the learners' acquisition of complex words.

3.2 Relevance of Vocabulary Size

Next, we examine the possibility of some correlation between the vocabulary sizes of learners and the results of their decomposability judgments. Making use of Mochizuki's (1998) vocabulary-size measuring tests at the 1000 through 5000 word-levels, we obtained the results reported in Table 3.

Table 3: The make up of the upper and lower groups

Voc Size	N	M	SD	*t*-value
Upper	54	4147.59	258.79	11.17**
Lower	42	3388.07	377.13	**p<.01

The participants revealed approximately 3800 words as their mean vocabulary size (N=96, M= 815.302, SD=492.144). This point divides all the subjects into two groups, the upper and the lower. Table 3 shows the make up of these two groups, for which Welch's T-test indicates a difference statistically significant at the .01 level. Then, the data in Table 2 was split into two sets, as shown in Table 4 and Table 5, according to the distinction in vocabulary size based on the two groups. Table 4 represents the results from the 54 subjects in the upper group, and Table 5 the 42 subjects in the lower group.

Table 4: Results (upper group) (N=54)

Pair	D > B	D < B	Pair	D > B	D < B	Pair	D > B	D < B
#1	41	13	#21	42	12	#41	41	13
#2	8	46	#22	40	14	#42	16	38
#3	51	3	#23	38	16	#43	16	38
#4	38	16	#24	11	43	#44	46	8
#5	33	21	#25	33	21	#45	23	31
#6	23	31	#26	27	27	#46	13	41
#7	33	21	#27	35	19	#47	42	12
#8	19	35	#28	46	8	#48	13	41
#9	38	16	#29	25	29	#49	33	21
#10	42	12	#30	43	11	Total	1492	1154
#11	10	44	#31	28	26	%	56.39	43.61
#12	35	19	#32	37	17	M	30.45	23.55
#13	17	37	#33	32	22	SD	12.67	12.67
#14	52	2	#34	5	49	*t*-value	1.91	n.s.
#15	41	13	#35	27	27			
#16	39	15	#36	41	13			
#17	43	11	#37	40	14			
#18	22	32	#38	26	28			
#19	9	45	#39	22	32			
#20	12	42	#40	45	9			

Table 5: Results (lower group) (N=42)

Pair	D > B	D < B	Pair	D > B	D < B	Pair	D > B	D < B
#1	30	12	#21	22	20	#41	35	7
#2	10	32	#22	31	11	#42	15	27
#3	42	0	#23	18	24	#43	10	32
#4	34	8	#24	11	31	#44	30	12
#5	26	16	#25	28	14	#45	18	24
#6	15	27	#26	25	17	#46	10	32
#7	24	18	#27	26	16	#47	33	9
#8	16	26	#28	30	12	#48	13	29
#9	32	10	#29	19	23	#49	23	19
#10	33	9	#30	32	10	Total	1175	883
#11	18	24	#31	26	16	%	57.09	42.91
#12	33	9	#32	29	13	M	23.98	18.02
#13	11	31	#33	28	14	SD	9.28	9.28
#14	37	5	#34	7	35	*t*-value	2.247*	
#15	29	13	#35	15	27	*$p < .05$		
#16	36	6	#36	37	5			
#17	27	15	#37	34	8			
#18	18	24	#38	24	18			
#19	10	32	#39	19	23			
#20	10	32	#40	36	6			

The numbers of responses favoring the "D>B" type and the "D<B" type are each accounted for by 56.39% and 43.61% of the total of 2546 responses for the upper group and by 57.09% and 42.91% of the total of 2058 responses for the lower group. The "D>B" type outnumbers the "D<B" type for both sets of subjects, but the T-tests we performed against these results did not show a statistically significant difference between the two relative frequency types although it did indicate a statistically significant difference at the p< .5 level.

Finally, we consider the data in Table 6, which summarizes the scores presented in Table 4 and Table 5 for the "D>B" type.

Table 6: Rates in favor of the decomposability of the "D>B"- type words

Pair	Upper	Lower	Pair	Upper	Lower	Pair	Upper	Lower
#1	0.759	0.714	#21	0.778	0.524	#41	0.759	0.833
#2	0.148	0.238	#22	0.741	0.738	#42	0.296	0.357
#3	0.944	1.000	#23	0.704	0.429	#43	0.296	0.238
#4	0.704	0.810	#24	0.204	0.262	#44	0.852	0.714
#5	0.611	0.619	#25	0.611	0.667	#45	0.426	0.429
#6	0.426	0.357	#26	0.500	0.595	#46	0.241	0.238
#7	0.611	0.571	#27	0.648	0.619	#47	0.778	0.786
#8	0.352	0.381	#28	0.852	0.714	#48	0.241	0.310

#9	0.704	0.762	#29	0.463	0.452	#49	0.611	0.548
#10	0.778	0.786	#30	0.796	0.762	N	49	49
#11	0.185	0.429	#31	0.519	0.619	Total	27.63	28.00
#12	0.648	0.786	#32	0.685	0.690	M	0.56	0.57
#13	0.315	0.262	#33	0.593	0.667	SD	0.23	0.22
#14	0.963	0.881	#34	0.093	0.167	*t*-value	-0.550	n.s.
#15	0.759	0.690	#35	0.500	0.357			
#16	0.722	0.857	#36	0.759	0.881			
#17	0.796	0.643	#37	0.741	0.810			
#18	0.407	0.429	#38	0.481	0.571			
#19	0.167	0.238	#39	0.407	0.452			
#20	0.222	0.238	#40	0.833	0.857			

This table presents the rates of the relevant responses divided by their total score for each of the two frequency types. The two groups each exhibited totals of 27.63 and 28.00 and mean scores of 0.56 (SD=0.23) and 0.57 (SD=0.22). We conducted a T-test on these two groups and gained no statistically significant level of difference; therefore, the relevance of vocabulary size to decomposability perception was not confirmed.

4. Conclusion

In this paper we have argued that lexical frequency information is likely to affect the way in which Japanese EFL learners organize morphologically complex words in their mental lexicon. This result has much in common with Hay's (2000) precursory experiment on native speakers of English, leading us to conclude that it is likely that EFL learners perceive morphologically complex words that are less frequent than their bases as single sequences and those of the opposite type as sequences composed of separate substrings.

Next, a practical aspect of the findings of the present study should be pointed out. Selection and presentation of vocabulary items to learners, especially of morphologically complex words, should take into consideration the frequencies of the words in relation to their base words. By this, it is meant that the learning and storage of suffixed words may not always develop in an analytic way based on formal properties. Teachers can benefit from being able to make distinctions between items that can or should be handled as single units and those which are dissoluble into constituent structures, in that they can avoid overloading the learner with an unnecessary range of morphological information.

Finally, we would like to touch on a residual issue yet to be further investigated. It should be recalled that not all the "D>B" words in Table 6 attracted positive responses regarding their decomposability judgment. In fact, we find some which show the reverse tendency in both upper and lower groups: *pressure* (0.148, 0.238), *existence* (0.352, 0.381), *division* (0.315, 0.262), *dependent* (0.167, 0.238), *approval* (0.222, 0.238), *arrangement* (0.204, 0.262), *invitation* (0.093, 0.167), *cheerful* (0.96, 0.357), *admiration* (0.296, 0.238), *amusing* (0.241, 0.238), *infectious* (0.241, 0.31). This state of affairs seems to indicate that we need re-examine these rather unwanted results thoroughly in order to confirm whether they suggest the existence of some hierarchical relations among the suffixes that may override frequency effects or they can simply be attributed to the source of frequency information that could be tested against Japanese EFL learners.

Notes

1. Abundant information on the CELEX database is available at http://www.ru.nl/celex/index.html#Info.
2. The Longman Defining Vocabulary is made up of about 2250 words.
3. The frequencies of the derived words and their bases were obtained from a frequency list extracted from the British National Corpus.

Appendix

The list of the complex words from which the stimuli presented to subjects were made

	Type "D>B"	Surface frequency	Base frequency	Type "D<B"	Surface frequency	Base frequency
1	production	14858	10535	personal	15967	24496
2	pressure	11579	10436	recently	11512	14735
3	manager	10976	3925	successful	10464	12823
4	operation	9236	3977	industrial	9852	17168
5	employment	9103	1697	effective	9646	22894
6	failure	7257	3295	powerful	6975	29977
7	protection	6450	4962	possibility	6968	33430
8	existence	6449	5355	possibly	6558	33430
9	famous	6204	1044	difficulty	6169	21465
10	insurance	5704	342	formal	6111	33246
11	relative	5673	2486	reasonable	6096	17699
12	opposite	5462	851	dangerous	5462	5664
13	division	5211	1566	appearance	5262	10738
14	attractive	4978	2504	historical	5066	16535
15	introduction	4959	3394	careful	4788	22729
16	intention	4647	2021	explanation	4585	7436
17	association	3981	1071	criminal	4040	6413
18	comfortable	3790	2948	northern	3932	9672
19	dependent	3713	3441	specialist	3931	19244
20	approval	3640	1030	improvement	3916	6095
21	arrival	3278	2800	friendly	3645	11329
22	electricity	3216	2742	southern	3631	9373
23	exciting	3196	205	establishment	3581	5139
24	arrangement	3178	2082	representative	3458	4519
25	permission	3097	1594	healthy	3382	17585
26	comparison	3076	2141	judgment	3139	5654
27	preparation	3018	2740	helpful	3084	35710

28	infection	2636	106	determination	2669	3852
29	destruction	2240	1914	childhood	2620	23128
30	instruction	2135	335	desirable	2068	5218
31	punishment	2021	456	addition	2010	7205
32	entertainment	1799	641	refusal	1831	2367
33	complaint	1772	1313	influential	1802	10252
34	invitation	1758	1177	attendance	1624	3522
35	explosion	1603	355	servant	1614	4929
36	declaration	1572	914	peaceful	1557	7725
37	loyalty	1550	1245	kingdom	1549	5853
38	probability	1507	1171	tourist	1527	5432
39	disappointment	1433	181	favorable	1450	3256
40	dismissal	1408	802	encouragement	1416	4946
41	advertisement	1074	491	fashionable	1101	4304
42	cheerful	1064	682	foolish	1060	1714
43	admiration	913	745	faithful	925	4553
44	decoration	887	321	harmful	804	2861
45	explosive	770	355	similarity	787	17365
46	amusing	764	219	sticky	785	3999
47	enclosure	556	423	musician	577	13186
48	infectious	452	106	careless	512	22729
49	annoyance	442	227	slippery	423	2252

References

Bod, R., Hay, J. & Jannedy, S. (2003). *Probabilistic Linguistics*. Cambridge: The MIT Press.

Bybee, J. (2001). *Phonology and LanguageUuse*. Cambridge: Cambridge University Press.

Cutler, A. (1981). Degrees of transparency in word formation. *Papers from the 26th regional meeting, Chicago Linguistic Society*, 73-77.

Frauenfelder, U. & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. E. Booij and J. van Marle (Eds.), *Yearbook of Morphology 1993* (pp.165-83). Dordrecht: Kluwer.

Hay, J. (2000). *Causes and Consequences of Word Structure*. Ph.D dissertation, Northwestern University, Chicago, Illinois.

Kiparsky, P. (1982). Lexical morphology and phonology. In I. S. Yang (Ed.) *Linguistics in the morning calm*. (pp. 3-91). Seoul: Hanshin Publishing Co.

Mohanan, K. P. (1986) *The Theory of Lexical Phonology*. Dordrecht: Reidel.

Plag, I. (1999) *Morphological Productivity: Structural Constraints in English Derivation*. Berlin and New York: Mouton de Gruyter.

Plag, I. (2003) *Word-formation in English*. Cambridge: Cambridge University Press.

Siegel, D. (1974) *Topics in English morphology*. Ph.D dissertation, MIT, Cambridge, Mass.