Lexical factors affecting modality differences in word translation by Japanese university EFL learners

天野修一
Shuichi Amano

1. Introduction

Listening in a second language (L2) is not a mere passive skill that requires little effort from learners in the process of comprehension; rather, it is an active skill that requires learners to continuously derive the meaning of an input by extensively using two types of information: bottom-up information—phonological, lexical, syntactic, semantic, and pragmatic knowledge—and top-down information—knowledge about the topics, contents of the inputs, or general knowledge of the world (see e.g., Rost, 2002). Stæhr (2009) argues that L2 listening would be challenging for listeners who fail to efficiently integrate both types of cues. Nevertheless, Bonk (2002) states that bottom-up cues are more fundamental because utilizating top-down cues depends at least partly on information obtained through bottom-up cues. In other words, when learners try to activate prior knowledge, they need to recognize a certain number of words contained in connected speech. Therefore, it is reasonable to argue that auditory word recognition is an essential construct for L2 listening.

However, despite such importance of auditory word recognition in L2 listening, there is insufficient knowledge about its detailed process. This study argues that knowing the meanings of a word does not ensure learners' successful word recognition. This is because, as is often the case with Japanese learners of English as a foreign language (EFL), learners cannot always recognize words in connected speech even if they can recognize the same words in written texts. Therefore, the present study focuses primarily on auditory word recognition and explores the gap between visual and auditory word recognition in learning EFL.

2. Literature Review

Word recognition can be regarded as a process of matching presented inputs with lexical representation in the mental lexicon. The mental lexicon stores various information about vocabulary; Levelt's (1993) model, which is divided into four categories—semantic, syntactic, morphological, and phonological representations—is well known. If a stimulus is presented visually, the input will be matched with morphological representation, and if a stimulus is presented auditorily, the input will be matched with phonological representation. Subsequently, when a semantic (or syntactic) representation is activated through either representation, the stimulus word will be recognized. Therefore, the type of input modality, visual or auditory, appears to be of great importance. It is highly unlikely that there is no gap between visual and auditory word recognition, because first, these two types of input modalities are different, and second, different types of representations are activated even when the word itself is the same. The processes of visual and auditory word recognition might have common and contrasting aspects, because several studies have clearly shown that participants activate not only orthographic but also phonological representation of the stimulus in visual word recognition and vice versa in auditory word recognition (see e.g., Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008).

The outcomes of studies that have examined modality differences in L1 word recognition vary depending on the research objectives and methods (Connine, Mullennix, Shernoff, & Yelen, 1990; Taft, 1986; Woutersen, de Bot, & Weltens, 1995), and to my knowledge, modality differences have not been fully investigated. Despite insufficient support from L1 research, some studies have investigated modality differences in L2 word recognition. This is probably because language instructors have observed that L2 learners face many difficulties in recognizing spoken words, for example, during dictation exercises, especially in foreign language contexts. Although such studies were conducted within the paradigm of vocabulary acquisition rather than word recognition, these research procedures do not significantly differ; thus, reviewing previous studies can provide us with useful information. Ishizaki and Iimura's (2007) study targeted Japanese high school learners of English. They presented visual and auditory stimuli to participants and asked them to respond by translating the stimuli into L1 equivalents. The result showed that the average score for visual stimuli was higher than that for auditory stimuli. Ishikawa (2009) also investigated modality differences in L2 word recognition. In her translation experiment, participants responded to the stimuli by choosing one appropriate answer given four choices. The result contradicted that of Ishizaki and Iimura's (2007) study, that is, there was little difference between the average scores for the visual and auditory stimuli.

If there is a difference in the difficulty in L2 word recognition between modalities, language instructors should determine the factors that promote better understanding among learners. However, as described above, the results of previous studies are not consistent. This inconsistency may arise at least partly because they only considered the mean scores. Adequately controlling variables may be difficult because the types of lexical factors that affect modality differences is not yet clear; consequently, there may be uncontrolled factors that affect modality differences in their stimulus lists. Furthermore, because of the complex array of potential variables, a simple comparison of the mean scores is not adequate for examining modality differences in L2 word recognition. It is necessary to investigate not only the mean scores for all stimuli but also the respective scores for each stimulus to clarify the detailed difference in the process of matching both types of inputs with mental representation. Since few studies examine the effects of lexical factors on the gap between visual and auditory word recognition, I conduct a study to identify the factors that affect this gap from among those that have been shown to affect word recognition (e.g., frequency, familiarity, and neighborhood density). This preliminary exploratory study primarily identifies factors that should be controlled and discovers new hypotheses worth pursuing in future studies. If some types of words are easier or more difficult to recognize compared with others in either modality, a deeper understanding of the phenomenon will benefit both research and teaching.

3. Experiment

3.1 Participants

The participants were mixed-ability EFL undergraduates (N=64) enrolled in the compulsory language course at a private Japanese university in Aichi Prefecture. They were all native speakers of Japanese and non-English major students. None of them reported a history of speech or hearing problems, or experience of living in English-speaking countries for over a month. Although all of them were exposed to formal EFL instruction in their junior and senior high schools for six years, they did not take any English-language tests during the university admission procedure. According to the results of a placement test for the course, their proficiency ranged roughly between EIKEN grades three and five.

3.2 Stimuli

The thirty stimulus words were prepared as follows. Words with grater than 4.50 familiarity ratings were randomly extracted from the L2 visual word familiarity database developed by Yokokawa (2006). Then, variables such as L2 auditory familiarity (Yokokawa, 2009), written frequency, spoken frequency (Baayen, Piepenbrock, & van Rijn, 1993), and number of phonemes and phonological neighbors (Davis, 2005) were assigned to the stimuli. The stimuli were recorded by a female American English speaker, and the resulting sound files were normalized and presented in 44.1 KHz, 16-bit format.

3.3 Method

Previous studies adopted several tasks such as dictation tasks (Yoneyama & Munson, 2010), multiple-choice translation tasks (Ishikawa, 2009), and word translation tasks (Ishizaki & Iimura, 2007) implemented simultaneously for a large number of participants in order to assess word recognition. These are considered lightly loaded tasks appropriate for less proficient learners. Among these, the multiple-choice translation and dictation tasks are unsuitable for the present study's aim of verifying modality differences in L2 word recognition. In the multiple-choice translation task, decision branches might compensate for listeners' insufficient ability of L2 phonological perception. In the dictation task, on the other hand, when a participant responds incorrectly to a stimulus, it is difficult to assess whether the incorrect response was because the participant made a spelling error, was unable to recognize the stimulus, or did not know the stimulus. Furthermore, correct responses to a dictation task do not always mean that participants recognize the word; they might be able to form the responses by simplely dictating the word as they listened to it, even without actually knowing it. A primary problem with a word translation task is that it does not consider participants' confidence in their response. As Ishizaki and Iimura (2007) noted, not distinguishing between confident and non-confident responses appears to be problematic. Therefore, I use a word translation task that includes confidence ratings.[1]

The criterion for correct answers in the present study is whether the answers are included as Japanese equivalents either in ***Sanseido's Wisdom English-Japanese Dictionary*** (Inoue & Akano, 2007) or in ***Taishukan's Genius English-Japanese Dictionary*** (Konishi & Minamide, 2006). The scoring rubric is presented in Table 1. Correct answers with confidence were given 2 points, correct answers with no opinion on confidence were given 1.5 points, and correct answers with no confidence were given 1 point. Incorrect or no answers were not given any points regardless of their confidence ratings.

3.4 Procedure

Table 1

Scoring Rubric for Word Translation Task

Types of response	Confidence rating	Points
correct answer	confident	2
correct answer	no opinion	1.5
correct answer	not confident	1
incorrect answer/no answer	confident	0
incorrect answer/no answer	no opinion	0
incorrect answer/no answer	not confident	0

All participants in the same class were concurrently tested in a quiet classroom. The auditory items were presented through an audio speaker. The experimenter explained the procedure by reading the instructions aloud. The participants were instructed to write an L1 equivalent and indicate their confidence by marking a three-alternative forced choice (confident, no opinion, or not confident) on the answer sheet.

The stimuli were presented by the following steps. Two seconds after the start signal, the first auditory stimulus was presented. Each stimulus was presented twice with a two-second interval. The participants were then given a twelve-second interval to answer. Following this, the next stimulus was automatically presented, preceded by a signal. This sequence was repeated until all thirty stimuli in the list were presented. The entire auditory session required approximately seven minutes and twenty seconds, and it was followed by the visual session. For this session, the same stimuli from the auditory session were listed on the sheet, but in a different order. The participants were asked to translate the words by looking at their spellings. As in the auditory session, they were given seven minutes and twenty seconds to answer. This procedure was designed largely by referring to Ishikawa (2009). Three practice trials were conducted before the thirty main trials. Finally, the participants answered a participant questionnaire to clarify their English learning experience.

4. Results and Discussion

Table 2 summarizes the descriptive statistics of the scores for visual and auditory word recognition (hereafter the VWR and AWR scores, respectively). The rightmost column indicates the difference between the VWR and AWR scores (calculated by adding the average values of the VWR scores minus the AWR scores of each stimulus, and then dividing them by the total number of stimuli; the results are hereafter referred to as the V−A scores). There is a higher average for the VWR than the AWR score. Furthermore, Wilcoxon W test shows that the difference was statistically significant ($z=-3.49$, $p<.01$), and the effect size was large ($r=.64$), that is, even a simple comparison of the average scores in the present study shows modality differences in word translation. This finding is consistent with that of Ishizaki and Iimura (2007), who employed a similar word translation task, but inconsistent with that of Ishikawa (2009), who employed a multiple-choice translation task.

Table 2

Descriptive Statistics of Scores for Visual and Auditory Stimuli

	visual	auditory	visual-auditory
Mean	0.98	0.50	0.48
SD	0.52	0.37	0.40
Maximum	1.81	1.30	1.36
Minimum	0.08	0.02	−0.16

Examining the scores for each stimulus in further detail, I discover the advantage of visual word recognition, as shown in Figure 1 and 2. The VWR scores for all the stimuli were higher than the AWR scores, except for *danger* and *safety* (see Figure 2 and Appendix). This is represented by almost all the dots in Figure 1 being concentrated in the top left half of the plot. The Spearman correlation between the VWR and AWR scores was strong (ρ=.71, p<.01). It should also be noted that there are considerable differences in the stimuli's V−A scores despite the consistent advantage of visual word recognition. Some stimuli showed larger differences than expected between the VWR and AWR scores (e.g., *king, river, egg, key,* and *wave*).

To identify possible factors for the difference between the VWR and AWR scores, the Spearman correlation between the three types of scores and six lexical factors was computed. Table 3 summarizes the results. The items that show significant differences are enumerated below.

1) The higher L2 visual familiarity is, the better the VWR (ρ=.78), AWR (ρ=.66), and V−A (ρ=.53) scores are.
2) The smaller the number of phonemes is, the larger the VWR (ρ=−.49) and V−A (ρ=−.42) scores are.
3) The larger the phonological neighborhood density is, the larger the VWR (ρ=.46) and V−A (ρ=.47) scores are.

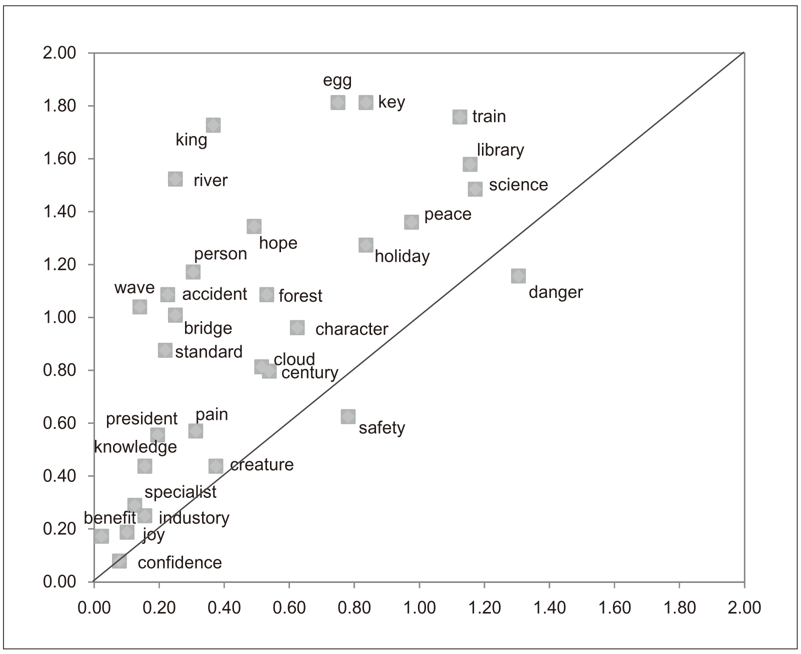

Figure 1. Relationship between VWR and AWR scores.

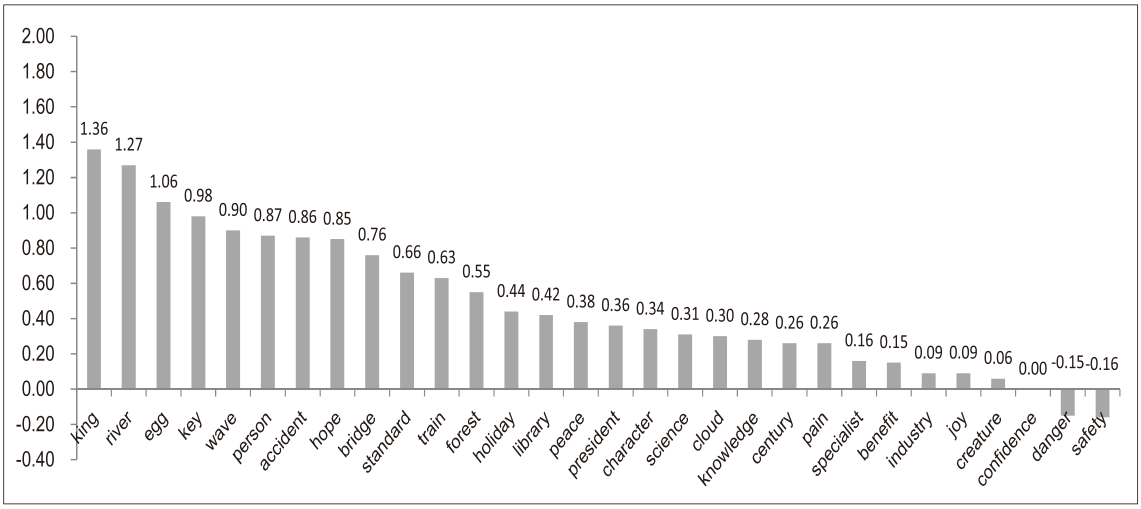

Figure 2. V−A scores for each stimulus, ranked in descending order.

Table 3

Correlation between the Three Types of Scores and Lexical Factors

Lexical Items (ρ)	visual	auditory	visual-auditory
CELEX Written Frequency	.06	.04	.07
CELEX Spoken Frequency	.10	.11	.09
L2 Visual Familiarity	.78***	.66***	.53**
L2 Auditory Familiarity	.08	31	−.16
Number of Phonemes	−.49**	−.25	−.42*
Number of Phonological Neighbors	.46*	.18	.47**

Note: *p<.05, **p<.01, ***p<.001

Owing to the small sample size in this study, I discuss the relationship between the scores and the factors on the basis of these correlations. First, neither written nor spoken frequencies were correlated with any of the VWR, AWR, and V−A scores. This may indicate that written and spoken frequencies, which reflect the actual use of English by native speakers, do not significantly influence the present participants, EFL learners with no experience of residing in English-speaking countries. In addition, it may indicate that the English they are exposed to in classrooms has a different word frequency distribution compared with that of authentic English.

There was a fairly strong positive correlation between L2 visual familiarity and the VWR score (ρ=.78, $p<.01$). L2 visual familiarity also strongly correlated with the AWR score (ρ=.66, $p<.01$) and the V−A score (ρ=.53, $p<.01$). Meanwhile, L2 auditory familiarity was correlated with any of the VWR, AWR, and V−A scores. These results also indicate that the participants with less daily exposure to spoken English did not have good auditory word recognition ability compared to visual word recognition ability, which can be readily gained even in EFL settings. In other words, they had sufficient ability to translate the stimuli and be strongly influenced by familiarity in response to visually presented stimuli, but they had less ability to identify the L1 equivalents of auditory stimuli even when the stimuli were auditorily familiar words.

There was a moderate positive correlation between the V−A score and the number of phonemes (ρ=−.42, $p<.05$). Since a smaller number of phonemes implies a smaller amount of available acoustic-phonetic information for listeners, I assumed that it could help L2 listeners recognize spoken words, because processing a smaller amount of acoustic-phonetic information may lead to a lesser likelihood of misperception or misprocessing. However, an opposite correlation is found. There are two possible reasons for this unexpected result. One is that the smaller number of phonemes prevents L2 listeners, who generally have insufficient ability of phonological perception, from compensating for the imperceptible parts of the stimulus by perceiving the other parts.

Another reason is the close relationship between the number of phonemes and phonological neighborhood density, which had a positive correlation with the VWR (ρ=.46, p<.05) and V−A(ρ=.47, p<.01) scores. The Spearman correlation coefficient shows a very strong negative correlation between the number of phonemes and phonological neighborhood density among the stimuli used in this study(ρ=−.83, p<.01). This indicates that stimuli with a smaller number of phonemes also have higher phonological neighborhood density; thus, their influence on word recognition is in the same direction. In other words, higher phonological density causes listeners to confuse the true target word with its neighbors during auditory word recognition. This in turn makes auditory word recognition more difficult and the V−A scores larger, clearly suggesting the importance of phonological perception in L2 auditory word recognition.

However, there is another aspect that needs to be considered in the correlation. The view supporting the impact of phonological neighborhood density is problematic. In some cases, considering participants' L2 proficiency, it is unreasonable to assume that they knew the phonological neighbors of the stimuli. For example, it is difficult to assume drain [dreɪn] and trace [treɪs], phonological neighbors of the stimulus train [treɪn], as known words for the majority of the present participants. In such a case, confusion between the stimuli and their neighbors cannot occur, and the influence of phonological neighborhood density may not be the result of simple confusion between them. This issue needs to be examined in future studies.

The present study is exploratory. Because of the small sample size, the findings are limited. Future confirmatory studies should use more sophisticated experimental methods with individual experiments. In addition, it is also possible that the experimental procedure in this study contributed to the advantage of visual word recognition, because the participants were exposed to the auditory stimuli before the visual stimuli. However, despite such limitations, the present study yielded several findings that could be considered when designing future studies toward a deeper understanding of the gap between visual and auditory word recognition in learning EFL. Although frequency is an initial factor that is usually carefully controlled in word recognition research, it might be a low-priority factor depending on the learners' L2 proficiency. In contrast, L2 familiarity, particularly L2 visual familiarity, is an important factor that should be controlled during stimulus selection. The number of phonemes and phonological neighborhood density are two other factors that should be controlled. In addition, I believe that phonetic factors should be considered in future work. Even though major lexical factors such as frequency, familiarity, and neighborhood density were considered in this study, various phonetic factors such as vowel quality, consonant cluster, syllable structure, and position of stressed syllables were not. These factors should be included in future studies.

*An earlier version of this study was presented at the 76th Conference of the Japan Association for Language Education and Technology, Chubu Chapter, at the Hashima Campus of Gifu Shotoku Gakuen University, November 28, 2010. This study has been thoroughly revised and expanded on the basis of the discussion at the presentation session.

Notes

1. Note that when a word translation task is adopted for word recognition research, analysis of the results should carefully consider whether word translation can be regarded as equivalent to word recognition. Although word recognition seems to be a precondition of word translation and as having a common mechanism, at least partly, they should strictly be considered different because listeners may be able to recognize stimulus words even when thay are not able to translate them into other languages. This task might require listeners to perform unnecessary translations in order to respond.

Reference

Baayen, R. H., Piepenbrock, R., & van Rijn, H.(1993). ***The CELEX Lexical Database [CD-ROM]***. Linguistic Data

Consortium, University of Pennsylvania.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14,* 14-31.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084-1096.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and either psycholinguistic statistics. *Behavior Research Methods, 37,* 65-72.

Inoue, N., & Akano, I. (eds.) (2007). *Sanseido's wisdom English-Japanese dictionary. Second Edition.* Tokyo: Saiseido.

Ishikawa, Y. (2009). Shinin goi to chokai goi: Goi taipu goireberu to 2 shurui no juyo goiryoku no kankei [The effect of word type and word level on the phonological vocabulary and orthographical vocabulary]. *Journal of the Chubu English Language Education Society, 38,* 47-54.

Ishizaki, T., & Iimura, H. (2007). Gakushusha no goichishiki no 'fukasa' ga risuningu tesuto no chokaido ni oyobosu eikyo [On depth of vocabulary knowledge as a determinant of L2 listening]. *Language Education & Technology, 44,* 107-115.

Konishi, T., & Minamide, K. (eds.) (2006). *Taishukan's genius English-Japanese dictionary. Fourth Edition.* Tokyo: Taishukan Shoten.

Levelt, W. J. M. (1993). *Speaking: From intention to articulation.* Cambridge: MIT Press.

Rost, M. (2002). *Teaching and researching listening.* London: Longman.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31,* 577-607.

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes, 1,* 297-308.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58,* 366-379.

Woutersen, M., de Bot, K., & Weltens, B. (1995). The bilingual lexicon: Modality effects in processing. *Journal of Psycholinguistic Research, 24,* 289-298.

Appendix

List of Stimulus Words Used in the Experiment

SN	words	VWR	AWR	V−A	SN	words	VWR	AWR	V−A
1	accident	1.09	0.23	0.86	16	key	1.81	0.84	0.98
2	benefit	0.17	0.02	0.15	17	king	1.73	0.37	1.36
3	bridge	1.01	0.25	0.76	18	knowledge	0.44	0.16	0.28
4	century	0.80	0.54	0.26	19	library	1.58	1.16	0.42
5	character	0.96	0.63	0.34	20	pain	0.57	0.31	0.26
6	cloud	0.81	0.52	0.30	21	peace	1.36	0.98	0.38
7	confidence	0.08	0.08	0.00	22	person	1.17	0.30	0.87
8	creature	0.44	0.38	0.06	23	president	0.55	0.20	0.36
9	danger	1.16	1.30	−0.15	24	river	1.52	0.25	1.27
10	egg	1.81	0.75	1.06	25	safety	0.63	0.78	−0.16
11	forest	1.09	0.53	0.55	26	science	1.48	1.17	0.31
12	holiday	1.27	0.84	0.44	27	specialist	0.29	0.13	0.16
13	hope	1.34	0.49	0.85	28	standard	0.88	0.22	0.66
14	industry	0.25	0.16	0.09	29	train	1.76	1.13	0.63
15	joy	0.19	0.10	0.09	30	wave	1.04	0.14	0.90

Yokokawa, H.(ed.)(2006). *Nihonjin eigo gakushusha no eitango shinmitsudo mojihen: Kyoiku kenkyu no tameno dainigengo detabesu [Familiarity for English words of Japanese EFL learners, Visual version: L2 Database for Research and Education].* Tokyo: Kuroshio Shuppan.

Yokokawa, H.(ed.)(2009). *Nihonjin eigo gakushusha no eitango shinmitsudo onseihen: Kyoiku kenkyu no tameno dainigengo detabesu [Familiarity for English words of Japanese EFL learners, Auditory version: L2 Database for Research and Education].* Tokyo: Kuroshio Shuppan.

Yoneyama, K., & Munson, B.(2010). The influence of lexical factors on word recognition by native English speakers and Japanese speakers acquiring English: A first report. *Journal of the Phonetic Society of Japan, 14 (1)*, 35-47.