Behav Res
DOI 10.3758/s13428-013-0389-z

Cross-linguistic similarity norms for Japanese–English translation equivalents

David Allen · Kathy Conklin

Abstract Formal and semantic overlap across languages plays an important role in bilingual language processing systems. In the present study, Japanese (first language; L1)–English (second language; L2) bilinguals rated 193 Japanese–English word pairs, including cognates and noncognates, in terms of phonological and semantic similarity. We show that the degree of cross-linguistic overlap varies, such that words can be more or less "cognate," in terms of their phonological and semantic overlap. Bilinguals also translated these words in both directions (L1–L2 and L2–L1), providing a measure of translation equivalency. Notably, we reveal for the first time that Japanese–English cognates are "special," in the sense that they are usually translated using one English term (e.g., コール /kooru/ is always translated as "call"), but the English word is translated into a greater variety of Japanese words. This difference in translation equivalency likely extends to other nonetymologically related, different-script languages in which cognates are all loanwords (e.g., Korean–English). Norming data were also collected for L1 age of acquisition, L1 concreteness, and L2 familiarity, because such information had been unavailable for the item set. Additional information on L1/L2 word frequency, L1/L2 number of senses, and L1/L2 word length and number of syllables is also provided. Finally, correlations and characteristics of the cognate and noncognate items are detailed, so as to provide a complete overview of the lexical and semantic characteristics of the stimuli. This creates a comprehensive bilingual data set for these different-script languages and should be of use in bilingual word recognition and spoken language research.

Keywords Bilingual · Cross-linguistic similarity · Cognates · Japanese

Words within a language can have formal (phonological [P] and orthographic [0]) and/or semantic (S) overlap (e.g., *bat/bat* [+P, +0, –S], *tear/tear* [–P, +0, –S], *break/brake* [+P, –0, –S], *couch/sofa* [–P, –0, +S]). Importantly, research has shown that such overlap can increase activation and speed processing, or create competition and slow processing (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Jared, McRae, & Seidenberg, 1990; McClelland & Rumelhart, 1981). Words in different languages can also have such overlap. For example, the English word *ball* and the Japanese word ボール /booru/ overlap in terms of S, although not all of the senses of English *ball* are associated with the Japanese word /booru/. This issue of degree of S overlap is a crucial aspect of the present research and is discussed in more depth later. The two words overlap a great deal in P, although there is no O overlap, as the two languages are written in different scripts. Such overlap, or what we will refer to in the present research as *cross-linguistic similarity*, plays an important role in bilingual language processing (Allen & Conklin, 2013; Dijkstra, Miwa, Brummelhuis, Sappeli, & Baayen, 2010). Moreover, in the present research we show that this overlap is perceived by bilinguals to be continuous in nature.

In the literature on bilingual word processing, words that share both form and meaning are usually referred to as *cognates* (Dijkstra, 2007). This is because, until recently, most of the research has investigated the processing of European

Electronic supplementary material The online version of this article (doi:10.3758/s13428-013-0389-z) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

D. Allen (✉) · K. Conklin
School of English, University of Nottingham, Nottingham, UK
e-mail: david.allen@nottingham.ac.uk

Published online: 18 September 2013

languages (e.g., Catalan, Dutch, English, French, German, Italian, and Spanish). Thus, when words had form and meaning overlap (e.g., English *night*, French *nuit*, German *nacht*), this was in fact due to the modern words having a common historical root (e.g., Latin *nocte*); therefore, they were cognates. In the case of the Japanese word ボール /*booru*/ "ball," it is more appropriately called a *loanword/borrowing*. However, in this case not the historical origin of such words in a language, but instead their cross-linguistic similarity, influences their processing. Thus, for the purposes of this article, we will refer to any cross-linguistic word pairs that share form and meaning as *cognates*. Although Japanese–English cognates can easily be identified by simply considering the overlap of form and meaning across languages, a much more precise definition of "cognateness" can be determined by the use of bilingual measures of perceived similarity. This is discussed further in the following sections.

Cognates have been central to psycholinguistic research into bilingual language processing. Traditionally, bilinguals have been referred to by such classic distinctions as "compound" and "additive" by scholars such as Weinreich (1953). However, in the psycholinguistic literature, bilinguals are characterized by their proficiency in both languages. Thus, if a native speaker of Japanese also speaks English as a second language to some degree of proficiency, the speaker can be referred to as bilingual. An important question about bilingual language processing has been whether bilinguals could selectively activate a single language or whether both of their languages are activated nonselectively. In other words, when processing one language, is it possible to turn the other language "off"? Cognates have provided an ideal way to investigate this question, as they had a great deal of formal and S overlap. When bilinguals perform a task such as lexical decision, in which all words are presented in one language, cross-linguistic overlap should only influence processing if language activation is nonselective. That is, if both languages are activated during single-language processing, cognates should facilitate processing. Alternatively, if only a single language is activated during language processing, shared cross-linguistic S–O–P features of cognates should not influence processing relative to words that have no S–O–P overlap.

A considerable amount of research has shown that bilingual word recognition is fundamentally nonselective in nature (e.g., Dijkstra & Van Heuven, 2002; Van Heuven, Dijkstra, & Grainger, 1998; see Dijkstra, 2007, for a review). When bilinguals use a second language, cognates have been shown to speed their responses, relative to matched noncognate controls, in a wide variety of tasks, such as word naming (Schwartz, Kroll, & Diaz, 2007), word translation (Christoffels, de Groot, & Kroll, 2006), and lexical decision (Dijkstra, Grainger, & Van Heuven, 1999). Moreover, similar findings have been found for languages that differ in script (e.g., lexical decision with Hebrew–English, Gollan, Forster, & Frost, 1997; picture naming with Japanese–English, Hoshino & Kroll, 2008; masked priming lexical decision with Japanese–English, Nakayama, Sears, Hino, & Lupker, 2012; or masked priming lexical decision and word naming with Korean–English, Kim & Davis, 2003). Typically, in such studies cognates and noncognates are matched on important characteristics such as frequency, length, phonological onset, and phonological neighborhood size (for naming) or orthographic neighborhood size (for lexical decision).

Although cognates typically speed responses in L2 tasks such as lexical decision, picture naming, and translation, Dijkstra et al. (1999; Dijkstra et al., 2010) showed that in language decision tasks, in which bilinguals had to decide whether the targets were either Dutch or English, cognates were inhibited relative to noncognates.Thus, for tasks in which cross-linguistic similarity is disadvantageous, as in language decision, cognates can actually slow processing.

Even in sentence processing tasks, in which semantic and syntactic constraints may be more likely to induce language-selective processing, cognate facilitation has been observed relative to noncognates (e.g., Schwartz & Kroll, 2006; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; Van Hell & de Groot, 2008). Although cognate effects are typically more prominent in the L2 than in the L1 for unbalanced bilinguals (i.e., bilinguals who are not equally proficient in both languages, and typically are more proficient in the L1), due to the boosted activation of the more dominant L1, L2 cognate effects have also been observed in the L1 (Duñabeitia, Perea, & Carreiras, 2010; Van Assche et al., 2009). Thus, even when bilinguals are more dominant in an L1, it is still possible to observe cross-linguistic similarity effects in both L1 and L2 processing.

Although much of the previous research into bilingual processing has defined cognates and noncognates as being dichotomous, a growing number of studies have reported that bilinguals are sensitive to the degree of similarity, above and beyond a simple binary distinction (e.g., Allen & Conklin, 2013; Dijkstra et al., 2010; Van Assche et al., 2011; Van Assche et al., 2009). Using mixed-effects modeling with multiple independent variables, these studies have revealed that continuous measures of cross-linguistic similarity are indeed predictive of bilinguals' responses in L2 tasks. Most relevant for the present study, Allen and Conklin found that Japanese–English bilinguals responded to English words faster in lexical decision, depending on the degree of P similarity between the English and Japanese words. For example, whereas both *bus* /basu/ and *radio* ラジオ /rajio/ are cognates and were responded to more quickly than noncognate matched controls, *bus* was rated as being more phonologically similar to バス /basu/ than *radio* was to ラジオ /rajio/, and *bus* was responded to significantly more quickly than *radio*. This study used mixed-effects modeling with multiple predictors,

including word length and word frequency. Because such predictors are correlated with each other and also with P similarity, residualization was used to orthogonalize the predictors prior to model fitting. Collinearity was removed between all correlated variables, and then the residuals of these predictors were used to predict response times (RTs). The orthogonalized predictors showed that length, word frequency, and P similarity accounted for significant, but independent, portions of the variance in RTs. These facilitatory effects of P similarity were observed in L2 English lexical decision and picture naming with Japanese–English bilinguals. In addition, S similarity was shown to be an important predictor of responses to cognates in picture naming, with more semantically similar cognates being responded to more quickly than less semantically similar cognates. In an English lexical-decision experiment, S similarity had a reverse effect, relative to picture naming: Less semantically similar word pairs were responded to faster, due to such items having more senses, which apparently boosted activation of the lexical representation, leading to speeded responses relative to more semantically similar words (which tend to have fewer senses). These results highlighted the importance of task effects in language processing, but they also underscore the importance of continuous measures of cross-linguistic similarity as crucial indicators of bilinguals' processing performance.

Despite the importance of cross-linguistic measures of word similarity in bilingual language processing research, to our knowledge only one previous study has collected bilingual measures and made them available to researchers. Tokowicz, Kroll, de Groot, and van Hell (2002) conducted a large-scale study on 1,003 word pairs with Dutch–English bilinguals who rated translation equivalents for cross-linguistic O, P, and S similarity. They also elicited translations in order to determine translation equivalency and to assess the number of translations that a word has. Bilinguals translated words in both directions (i.e., from L1 to L2 and from L2 to L1), and this was used to determine whether a word has one or more translations in each language, as the number of translations that a word has has been shown to influence bilingual processing. The number of translations also can provide a metric of the amount of S overlap between words in two languages: If a word that has a number of senses in one language is translated into a single word in the other language, then both words are likely to be used in similar contexts in the two languages. However, if a word in one language has multiple senses that lead to different translations in the other language, that word is likely to be translated into more than one word. Thus, the words will be used in a variety of contexts, and likely will have less complete S overlap. An important implication of this is that when bilinguals translate a word presented in isolation, for many words multiple translations are likely to be given. Therefore, determining a "best" translation for most words is problematic. Providing a number-of-translations measure, on the other hand, provides information about the likelihood that bilinguals will be considering single or multiple translations for any particular word.

For researchers interested in L2 and bilingual language processing, it is critical to have norms for cross-linguistic S similarity in order to control for the influence of the "other" language during language-processing tasks. Moreover, bilingual ratings may be more suitable measures of S overlap than are dictionary measures of the number of meanings/senses in each language, because dictionaries vary greatly in their methods of quantifying meanings/senses, and they also reflect the total number of senses that *exist* in the language, as opposed to those known by the average bilingual (see Gernsbacher, 1984, for a similar argument).

To our knowledge, currently no measures of cross-linguistic similarity are available for languages other than Dutch–English. The present study thus provides cross-linguistic norming data for Japanese–English translations. Research into Japanese–English bilingual processing is particularly important, not only because relatively little bilingual research has focused on languages that differ in script (in comparison to research on same-script languages), but also because of the importance of English in Japanese society. Compulsory education and tertiary institutions place a strong emphasis on language education, and English is the most widely learned second language in Japan. The Japanese language has many thousands of loanwords borrowed from English that have entered the language since Japan's ending of its self-imposed isolation in the mid-nineteenth century. Many of these loanwords are in regular and general use; however, the majority are reserved for technical and academic uses. The proportion of loanwords in the 5th edition of the *Koujien* (1998), a comprehensive Japanese dictionary, is 10.2 %, which equals around 23,000 word entries (Kawaguchi & Tsunoda, 2005; cited in Igarashi, 2008). Moreover, around 90 % of loanwords are borrowed from English (Shinnouchi, 2000). Therefore, a better understanding of how Japanese–English bilinguals process these loanwords/cognates is an important area for research.[1]

The primary goal of the present study was thus to provide a range of cross-linguistic similarity measures of Japanese–English translation equivalents. To this end, ratings were

[1] Some research has also been conducted with English–Japanese bilinguals regarding the learning and use of cognates in Japanese (e.g., Prem, 1991; Tomita, 1991; see Kess & Miyamoto, 2000, for an overview). However, little solid empirical research using English–Japanese bilinguals overlaps with the research discussed in this article, and therefore English–Japanese bilinguals are not discussed further. Moreover, whereas one may assume that perceived cross-linguistic similarity may be comparable for both Japanese–English and English–Japanese bilinguals, no research has put this idea to the test (with any bilinguals), and thus the present data set should be considered as being applicable only to Japanese–English bilinguals.

collected to assess P and S similarity. Also, participants were asked to translate words in order to provide estimates of the number of translations and meanings that were known by the bilinguals. A second aim was to collect additional norming data that will be critical for designing experiments to investigate bilingual processing, yet that are not publicly available for all of the Japanese words in this study. Because most studies have focused on high-frequency words (e.g., Yokokawa, 2009, investigated the top 3,000 words in the British National Corpus), there are few measures for many of the cognates that are ubiquitous in the Japanese language, but that tend to be of lower frequency. Thus, information about the perceived age of acquisition (AoA) and concreteness of L1 Japanese words was collected. Concreteness is particularly useful for researchers, as it is typically highly correlated with grammatical class, such that verbs tend to refer to abstract events or actions, whereas nouns often refer to concrete objects as well as to abstract entities. Although grammatical class is problematic as a norming measure, because many words can be read as either verbs or nouns (e.g., *call*, *run*, *telephone*), concreteness can be used as a measure of the intrinsic S properties of the item, which may include both the verbal and nominal uses of items. In addition, we collected bilingual ratings of L2 (English) word familiarity. To create a more complete set of information about Japanese–English cognates that could be useful to researchers, we also provide the following information in the present database: word length, number of English senses (WordNet), number of Japanese senses (Meikyo Japanese Dictionary, 2008 edition), English word frequency (Balota et al., 2007), and Japanese word frequency (Amano & Kondo, 2000). Finally, a set of descriptive statistics is presented, as well as a correlation analysis of the ratings and the collected measures.

Method

Participants

One hundred and sixty-six first- and second-year undergraduate university students participated in the present research. The participants were recruited from two Japanese universities: the University of Tokyo and Waseda University. All participants were enrolled in English language courses in one of the two institutions. All recruitment and participation procedures for the studies reported in this article were approved by the ethics committee at the School of English, University of Nottingham. All participants received course credit for taking part, and no participant took part in more than one study. All of the participants were native Japanese speakers who had studied English prior to their university education. In order to qualify for the rating studies, all participants confirmed that they considered themselves native Japanese speakers, who had lived in Japan for the majority of their lives and had received their education in Japan. Thus, L1 proficiency data were not collected, as all participants were at the native-speaker level. Details about the participants, as well as the number of participants in each study, are shown in Table 1. Participants were asked to rate their own perceived English language proficiency in reading, writing, speaking, and listening on a scale of 0–10, with 0 being *no ability at all* and 10 being *native-speaker-level ability*. The scores from

Table 1 Number and mean age of participants; age at which they began learning English; time learning English; their proficiencies in reading, writing, speaking, and listening; and overall proficiency scores

	P and S Rating (Concrete Items)	P and S Rating (Abstract Items)	Number of Translations Task	English (L2) Word Familiarity Rating	Concreteness Rating	Age-of-Acquisition Rating
Number of participants	33	36	38	19	18	22
Age	20.4 (4.5)	20.2 (4.4)	18.8 (0.8)	18.4 (0.5)	20.6 (4.4)	19.1 (0.9)
Age began learning L2[a]	11–15 years	11–15 years	11–15 years	11–15 years	–	–
Time learning L2[b]	5–9 years	5–9 years	3–7 years	3–7 years	–	–
L2 reading proficiency	6.8 (1.1)	6.2 (1.6)	6.2 (1.2)	6.7 (1.5)	–	–
L2 writing proficiency	5.3 (1.3)	4.8 (1.6)	5.1 (1.5)	5.0 (1.6)	–	–
L2 speaking proficiency	4.2 (2.0)	3.8 (1.7)	4.1 (1.7)	4.4 (1.8)	–	–
L2 listening proficiency	5.4 (2.2)	4.7 (1.9)	4.6 (1.9)	5.4 (1.9)	–	–
Overall L2 proficiency[c]	5.4 (1.3)	4.9 (1.4)	5.0 (0.9)	5.4 (1.0)	–	–

Standard deviations appear in parentheses. [a] "Age began learning" is derived from self–selected categories (0, 1–5, 6–10, 11–16, 17–21, and 21 years or above), and the data provided above are the mode responses for the participants. [b] "Time learning L2" is simply "Age" minus "Age began learning." [c] "Overall proficiency" is the mean of reading, writing, speaking, and listening proficiency measures.

each component for each participant were averaged in order to calculate an overall proficiency score.

Stimuli and apparatus

A total of 198 words were selected for the study. Our aim was to collect ratings for both concrete and abstract words in order to create a more representative stimulus set that can be used in a variety of tasks, such as picture naming (which typically uses concrete nouns) and comprehension tasks (which may include both concrete and abstract words). Moreover, because bilingual studies often make use of cognates due to their unique characteristics of having both formal and S similarity, approximately half of the words in the database were cognates. Loanwords in Japanese are all written in a separate script, katakana, making it relatively easy to determine "cognate" status. The cognates were all loanwords in Japanese that the authors determined shared obvious P and S similarity with their English translations. It was not necessary to do more than this, as the ratings themselves will show how similar the words are across languages. Five items were removed from analyses because they were not the same grammatical class (e.g., *expect* (verb) and 期待 /kitai/ (noun) "expectation"). This reduced the total number of items in the final study to 193. The concrete cognate and noncognate items (n = 94; cognate = 48; noncognate = 46) were selected from Nishimoto, Miyawaki, Ueda, Une, and Takahashi's (2005) picture naming norming study. By selecting items from Nishimoto et al.'s study, which were taken from Snodgrass and Vanderwart's (1980) picture naming norms in English (also see Szekely et al., 2004), the stimuli are suitable for research in both English and Japanese languages.[2] Because English loanwords in Japanese are often low-frequency and to ensure that participants would know the items (i.e., that they are lexicalized in Japanese), all of the abstract cognate and noncognate words (n = 99; cognate = 50; noncognate = 49) were selected from a high-frequency word list derived from a 330-million-word Japanese Web corpus (Kilgariff, Rychly, Smrz, & Tugwell, 2004). For the cognates, two professional Japanese–English translators confirmed that the Japanese and English words were translation equivalents, although the translation was not always the most likely translation (e.g., the English word *call* has many possible translations, with the cognate コール /kooru/ being one of them).

For the similarity rating task, the items were randomized and compiled into lists for P and S ratings. An additional 20 nontranslation filler pairs were added to the S rating task to encourage use of the full scale for similarity ratings. More specifically, because all item pairs are translation equivalents, they would be rated as similar to some degree across languages; in order to get participants to use the *completely different* end of the scale, non-translation-equivalents (e.g., *door*–雨 /ame/ "rain" in Japanese) were also included in the S rating task. Fillers were not necessary in the P similarity part of the study, as the use of both cognate and noncognate pairs ensures that the full scale will be utilized. The set of materials and ratings are provided in Appendix A (and in the supplemental materials). The filler items were removed from the analysis and were not used in any of the other tasks reported here. Two groups of participants completed the rating studies, one for concrete items and another for abstract items.

[2] Response latency data for picture naming with Japanese and English monolinguals are available from Nishimoto et al. (2005) and Szekely et al. (2004), respectively.

Procedure

All participants completed informed consent forms prior to beginning the experimental procedure. All surveys were administered using the online survey tool (www.surveymonkey.com). Fifteen participants were removed from the tasks due to due to incomplete responses or misunderstanding of the task. The total number of participants included in the tasks is shown in Table 1.

P and S similarity rating Each item was rated on a 5-point scale ranging from (1 = *completely different* to 5 = *identical*). A 5-point scale was used instead of the typical 7-point scale because in a pilot study using a 7-point scale the participants found it difficult to discriminate between some of the levels (i.e., the difference between 5 and 6, or that for 2 and 3). Instructions were provided in Japanese to ensure understanding of the task. A brief explanation and examples were provided at the beginning of each survey. Participants were asked to decide how similar the word pairs sounded on the basis of their intuition and were encouraged to say the words aloud if necessary to help them decide. The examples provided for the P similarity task included *band*–バンド (/bando/), *stress*–ストレス (/sutoresu/), *bird*–鳥 (/tori/), which were rated as similar/very similar (4–5), somewhat similar/similar (3–4) and very different/different (1–2), respectively. For the S similarity-rating task, participants were asked to decide how similar in meaning the words in each pair were. The instructions asked participants to consider differences in senses shared and not shared between the languages, and also differences in use between the two languages. They were told not to use a dictionary, but to complete the task on the basis of their intuition (i.e., their knowledge of the words). The examples provided were *triangle*–三角 (/sankaku/), *fan*–扇子 (/sensu/), and *clock* − 壁 (/kabe/ "wall" in Japanese), which were rated as very similar (5), somewhat similar (3), and very different (1), respectively. Additional explanatory text was included to make clear the basis for the ratings of the examples: The words *triangle*–三角 have one meaning that is almost identical in both languages, thus having considerable S similarity;

fan has a range of meanings in English, whereas 扇子 in Japanese has only one meaning that is similar to that of a (hand-held) *fan*, therefore they have some S similarity, but also differ in some senses; finally, *clock* and 壁 do not share word meanings, and therefore these words have no S similarity.

To ensure that all parts of the scale were used, 20 nontranslation equivalents were included in the stimulus list. (All nontranslation equivalents were rated as 1, or *completely different*, in terms of S similarity). All Chinese characters that may have been unknown to the participants were transcribed in the hiragana phonetic script. Because of the large number of items that required ratings for both P and S similarity, and the likelihood of "survey fatigue," each participant rated half of the words for each type of similarity, but no participant rated a pair of words for both types of similarity. Each individual item was rated for both P and S similarity by between 16 and 18 different participants.

Number of translations task Because bidirectional translation data are desirable for bilingual research, two lists were created with half of the items being translated from the L2 to the L1, and the other half being translated from L1 to L2. These lists were counter-balanced across participants and items were presented in random order; each item was only translated once (i.e., either from L2 to L1, or from L1 to L2) by an individual participant. Participants were asked to think of the first translation that comes to mind for each item and to enter that word in the space provided. Instructions were in Japanese and examples were provided in both forward and backward translation tasks; these examples included both cognates and noncognates, and were reversed for each language direction: for instance, L1 to L2, 鳥 (/tori/)–*bird*, ストレス (/sutoresu/)–*stress*; and L2 to L1, *bird*–鳥, *stress*–ストレス.

Age-of-acquisition rating Participants were asked to rate Japanese words on a scale of 1–7 indicating the age at which they had learned the words in Japanese: The seven response categories included (1) *0–2 years*, (2) *3–4 years*, (3) *5–6 years*, (4) *7–8 years*, (5) *9–10 years*, (6) *11–12 years*, and (7) *13 years or later*. Participants were asked to focus on when they acquired knowledge of the word itself rather than the written form, as this may vary depending on the script (i.e., kana or kanji). Instructions were in Japanese and an example provided for respondents was お母さん (/okaasan/ "mother") whose meaning would be learned between the ages of 0–2 years, whereas its written form would typically be acquired between 3–6 years, with the kana form preceding the kanji form.

Concreteness rating Participants were asked to rate Japanese word items on a scale of 1–7: response categories ranged from *very abstract* (1) to *very concrete* (7). Participants were asked to consider whether an item was easily pictured in their mind, making it concrete, or whether it was difficult to picture, in which case it was more abstract. No examples were provided with this task.

L2 familiarity rating Participants were asked to rate English word items on a scale of 1–7: response categories ranged from *very unfamiliar* (1) to *very familiar* (7). Participants were asked to consider how often they use the words in speaking and writing and also in reading and listening. Instructions were in English and examples were provided (*signature* and *abolish* are not used every day, whereas *book* may well be). A clarification was made to consider the words only in English, not loanwords in Japanese (e.g., サッカー /sakkaa/, "soccer"). Because participants were asked to focus on their use of the words, this familiarity survey is similar to a subjective frequency survey (e.g., Gernsbacher, 1984).

Results and discussion

In this section we first describe the cross-linguistic measures (P and S similarity, number of translations), followed by the norming data (AoA, concreteness, L2 familiarity) and finally the additional data that we are including in the data set (L1/L2 frequency, L1/L2 number of senses, L1/L2 word length (number of characters/ number of syllables). The descriptive statistics of all cross-linguistic, norming, and additional data are presented in Table 2. In what follows we make a distinction between cognate and noncognate items (on the basis of both the script used (i.e., katakana for cognates and hiragana/kanji for noncognates) and the obvious P and S similarity between the words) for the purposes of illustrating the characteristics of the stimuli. However, the cross-linguistic similarity ratings provided in this research will allow for more precise measurements of "cognateness" in future empirical Japanese–English bilingual studies. We provide the Bayes factors (using the Jeffrey–Zellner–Siow prior and Cauchy distribution on effect size) for comparisons between means of cognate and noncognate words (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The advantage of using a Bayes factor over a traditional *t*-test is that the factor represents both the presence and strength of an effect (see Jeffreys, 1961).

P and S similarity Respondents used all parts of the scale in both the P and S similarity rating tasks (Fig. 1). Cognate items were clearly distinguishable from noncognates on the basis of P ratings, with a Bayes factor that suggests a decisive rejection of the null hypothesis ($BF < 0.001$). S similarity ratings were skewed to the right side of the scale indicating that items were mainly rated as being highly semantically similar across languages (Fig. 2; note that nontranslation fillers were removed

Table 2 Descriptive statistics (range, mean, standard deviation) of all ratings and additional standardization measures for all items; means (and standard deviation) for cognate and noncognate items; and Bayes factor for the comparison of cognate and noncognate means

	Range	Mean (*SD*)	Cognate Mean (*SD*)	Noncognate Mean (*SD*)	Bayes Factor (BF)
Mean P similarity ratings	1.0–4.3	2.3 (1.2)	3.5 (0.4)	1.1 (0.3)	<0.001**
Mean S similarity ratings	2.3–5.0	4.2 (0.5)	4.3 (0.4)	4.2 (0.5)	5.41
Number of translations					
L1–L2 translation	1.0–6.0	1.6 (1.0)	1.1 (0.3)	2.0 (1.3)	<0.001**
L2–L1 translation	1.0–12.0	2.8 (2.1)	3.0 (2.1)	2.5 (2.0)	2.86
Number of meanings					
L1–L2 translation	1.0–3.0	1.2 (0.4)	1.1 (0.3)	1.2 (0.5)	0.53
L2–L1 translation	1.0–4.0	1.4 (0.6)	1.5 (0.6)	1.3 (0.6)	2.12
Mean AoA ratings	2.8–5.4	3.8 (0.6)	4.0 (0.5)	3.9 (0.6)	5.85
Mean concreteness ratings	1.9–6.1	4.5 (1.2)	4.4 (1.2)	4.5 (1.3)	7.22
Mean L2 familiarity ratings	2.0–5.4	3.8 (0.7)	4.1 (0.5)	3.6 (0.6)	<0.001**
Japanese word frequency (raw)	0.0[a]–156,283	7,012.6 (19,013.6)	6,706.7 (18,768.0)	10,421.2 (25,027.3)	0.18
Log-transformed Japanese word frequency	0.0–12.0	7.3 (1.7)	7.3 (1.6)	7.7 (1.7)	0.03*
English word frequency (per million)	1.1–861.4	70.3 (139.6)	80.0 (151.2)	60.6 (126.4)	5.66
Log-transformed English word frequency	0.1–6.8	3.2 (1.5)	3.2 (1.5)	3.1 (1.3)	6.57
Japanese number of senses	1.0–15.0	2.1 (1.9)	1.9 (1.3)	2.4 (2.3)	2.09
English number of senses	1.0–45.0	8.0 (7.4)	8.6 (8.3)	7.2 (6.4)	2.86
Japanese word length (morae)	2.0–6.0	3.3 (0.9)	3.5 (0.9)	3.1 (0.8)	0.02*
English word length (letters)	3.0–10	5.1 (1.3)	5.3 (1.4)	5.0 (1.3)	4.12
English word length (syllables)	1.0–4.0	1.5 (0.7)	1.6 (0.8)	1.5 (0.6)	4.81

* Very strong evidence against H_0. ** Decisive evidence against H_0 (see Jeffreys, 1961; Kass & Raftery, 1995). [a] One item (シェアー /sheaa/ *share*) was not found in the Amano and Kondo (2000) corpus; hence, a single zero frequency is included in the data set.

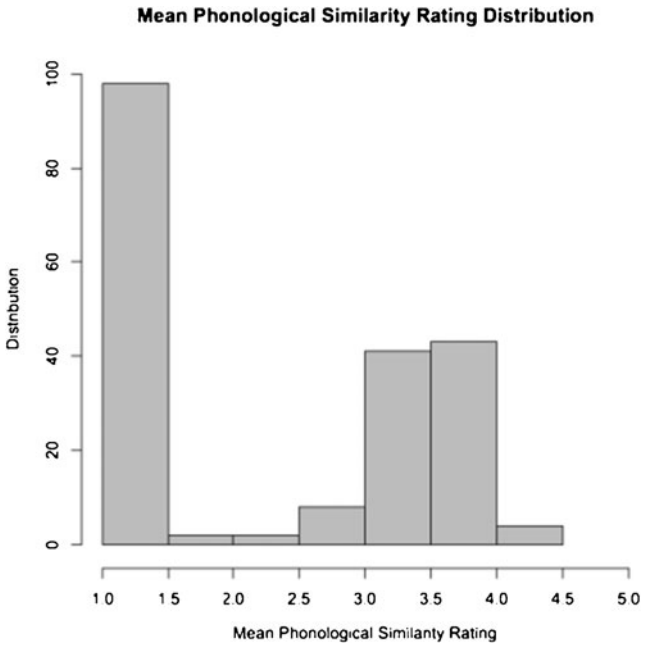

Fig. 1 Distribution of mean P similarity ratings for all items. The *x*-axis shows the mean ratings on a 5-point scale, with 1 being *completely different* and 5 being *identical*. The *y*-axis shows the number of translation pairs that fall into each mean rating band

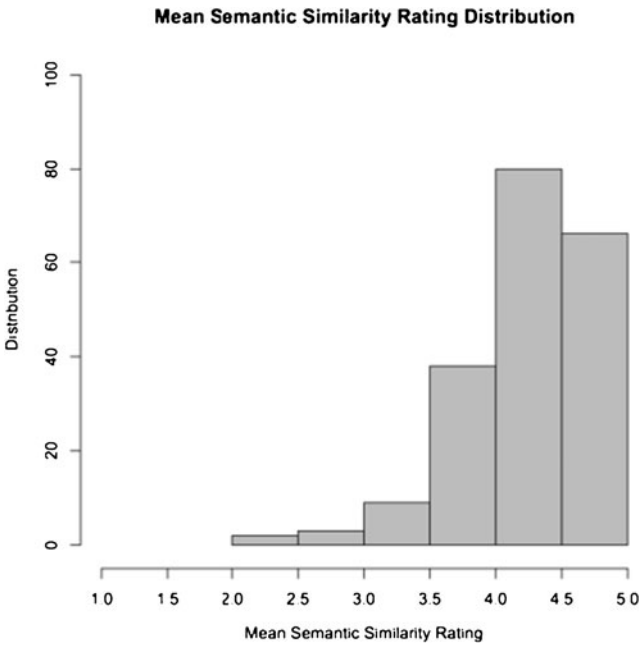

Fig. 2 Distribution of mean S similarity ratings for all items (nontranslation filler items removed)

from the analysis). The S ratings showed no difference between cognates and noncognates. This was expected as the primary distinction between cognates and other translation equivalents is that cognates share both form and meaning, whereas noncognate translation equivalents share only meaning. This finding supports those of Tokowicz et al. (2002), who found a similar result for Dutch–English translations, and thus refutes the assumption made by Van Hell and de Groot (1998) that cognates are more likely to share meaning because they share formal features. The present study shows that for languages that differ in script, formal (P) similarity does not make it more likely that words will share a greater amount of S similarity across languages.[3]

Number of translations Two professional Japanese–English translators determined the accuracy of translations in both directions (L1–L2, L2–L1). Correct translations were then coded for whether they were the expected translation (i.e., that provided by Nishimoto et al., 2005, for concrete items [the picture naming stimuli], or the translation assigned in the initial item selection stage—e.g., *ball*–ボール—or an alternative translation). The number of distinct meanings provided as translations was also determined and added to the database. (We did not count verb uses of nouns, adjectival uses of nouns, and so on, as different meanings. Also, where meanings were not easily distinguishable, such as in the case of *find* and *locate* for 見つける /mitsukeru/ in Japanese, they were treated as the same meaning; thus, our number of meanings measure is somewhat conservative as only distinct meanings were coded as being different). Additional data for the translation task are included in a separate sheet in the database (see the supplemental materials).

Table 2 shows the descriptive statistics for the translation tasks in both directions. As expected, when translating from the L1 into the L2, there were more errors than when translating from the L2 into the L1 (11.5 % vs. 8.4 %). More information on error rates for translations of items in each direction are provided in the supplemental materials for this article. Also, the mean number of translations and the mean number of meanings provided was smaller when translating into the L2 relative to translating into the L1 (mean translations, 1.6 vs. 2.8; mean meanings: 1.2 vs. 1.4). Interestingly, when comparing the number of translations of cognates versus noncognates across the two tasks, one difference emerged: When cognates are translated from Japanese to English, there is usually only one translation (M = 1.1, SD = 0.3), which is the English cognate (e.g., クラス /kurasu/ is translated as *class*); however, when translating the same cognates from English to Japanese, there is a greater range of translations (M = 3.0, SD = 2.1), which may or may not include the Japanese cognate translation. Furthermore, for concrete items such as *television*, which have only one translation in Japanese, these are translated using the Japanese cognate form (テレビ /terebi/); however, more abstract words, such as *class* and other verbs, which can have multiple translations in Japanese, are translated using multiple Japanese words (e.g., クラス /kurasu/, 学級 /gakkyuu/, or 等級 /toukyuu/). The Bayes factor (BF < 0.001) indicates decisive evidence for a difference between the mean numbers of translations for cognates and noncognates in the L1-to-L2 direction, indicating that when bilinguals translate cognates into English, they use significantly fewer translations than when translating noncognates into English. Noncognates had more than one translation on average, regardless of direction (L1 – L2, M = 2.0, SD = 1.3; L2 – L1, M = 2.5, SD = 1.8).

Age of acquisition (AoA) All parts of the scale were used, although few participants rated learning words in the earliest category (0–2 years). The mean AoA was 3.8, which is between the third and fourth categories (5–6 and 7–8 years; SD = 0.6; Table 2). There was no difference in AoA ratings between cognate and noncognate items. To test the reliability of the ratings, they were compared with Nishimoto et al.'s (2005) Japanese AoA ratings for picture stimuli and with Kuperman, Stadthagen-Gonzalez, and Brysbaert's (2012) AoA ratings for English words. Only the items that existed in both the present and previous data sets could be subject to this analysis. Because Nishimoto et al.'s ratings focused on picture stimuli, only the concrete items occurred in both data sets. Kuperman et al.'s data set, however, was much larger and covers most of the concrete and abstract words in the present data set. Correlations for the Japanese picture stimulus items were reasonable (r = 0.26, CI = 0.04, 0.46) but stronger for the English word AoA ratings (r = 0.47, CI = 0.35, 0.58). The weaker correlation between our AoA ratings and Nishimoto et al.'s ratings reflects the difference in task requirements. In Nishimoto et al., participants rated the AoA for the concepts depicted in the picture stimuli, whereas our measure reflects the acquisition of word knowledge, which may be acquired later than conceptual knowledge. In sum, the AoA ratings appear most comparable to those collected from English native speakers by Kuperman et al. AoA for words thus appears to have some overlap across languages.

Concreteness All parts of the scale were used showing that the stimuli included a variety of concrete and abstract words (M = 4.5, SD = 1.2). There was no difference in concreteness ratings between cognate and noncognate items. Correlations with concreteness and imageability ratings for those items that

[3] A reviewer suggested that perhaps no difference appears in the S similarity ratings for cognates and noncognates because of the inclusion of abstract cognates, such as *work*–/waaku/. However, a comparison of Bayes factors revealed no differences between abstract cognates and noncognates (BF = 6.28) or concrete cognates and noncognates (BF = 5.41), demonstrating that S similarity ratings were not different for items, regardless of their cognate status or concreteness.

could be cross-referenced (n = 76) taken from the MRC database (Coltheart, 1981) revealed strong correlations (r = 0.91, CI = 0.86, 0.94, and r = 0.84, CI = 0.76, 0.90, respectively), indicating that the present concreteness ratings collected with Japanese speakers are highly comparable to those collected with English speakers.

English (L2) familiarity All parts of the scale were used. The Bayes factor ($BF < 0.001$) for familiarity ratings for cognate and noncognate words provides clear evidence of a difference between the means with the cognate familiarity (M = 4.1, SD = 0.5) being considerably higher than the noncognate familiarity (M = 3.6, SD = 0.6). This shows that English words that are cognate with Japanese were rated as significantly more familiar than those that are noncognate (see Yokokawa, 2009, for a similar finding). To test the reliability of these ratings they were compared with Yokokawa's L2 familiarity ratings for visually presented English words collected from Japanese learners of English. The correlation was high (r = 0.77, CI = 0.68, 0.84) suggesting that the present ratings are a comparable and reliable resource.

Typically, norming data are collected from monolingual groups for use in monolingual studies. Such data can also be used as measures of one of a bilingual's languages. However, a bilingual's language processing system is not simply a combination of two monolingual systems (Grosjean, 1989). Research shows that a bilingual does not process language by accessing one lexicon exclusively depending on the language being used (Dijkstra, 2007). In contrast, nonselective access in language processing by bilinguals suggests that cross-linguistic activation influences performance a great deal in a wide variety of language tasks (Dijkstra, 2007). Here we show that in an L2 rating task, a bilingual's first language (the nontarget language) can modulate responses, demonstrating cross-linguistic influences in tasks that are not response-speed-dependent (i.e., for which RT is not the primary dependent variable). Thus, when researchers collect L2 norming data, such as familiarity, from bilinguals, they must consider the impact of cross-linguistic influences on such ratings.

Thus, the present L2 word familiarity measure incorporates bilingual participants' familiarity with both of their languages. Although it is primarily a measure of L2 familiarity, this is clearly influenced by the L1 (as evidenced by the significantly higher familiarity ratings for cognate translations, which share form and meaning with the L1, than for noncognates, which only share meaning). Therefore, it is likely that this measure will be particularly predictive of bilinguals' responses in word recognition tasks in the L2, at least for the particular sample population (i.e., mid-proficiency Japanese–English bilinguals). Because cross-linguistic influences tend to be more prominent in the weaker language (L2) than the dominant language (L1; Dijkstra & Van Heuven, 2002), the measure may be most predictive in L2 word recognition and or production tasks. Moreover, this bilingual measure of L2 familiarity should be more predictive of word recognition responses for Japanese–English bilinguals than a monolingual measure of English word familiarity.

Additional data for items

Japanese word frequency Word frequency in Japanese was taken from the Amano and Kondo (2000) database, which consists of word frequencies from all issues of the *Asahi* Japanese newspaper between 1985–1998 (see Appendix B and the supplemental materials for all additional data for items). The corpus has a total type frequency of 341,771 morphemic units and a total token frequency of 287,792,797 morphemic units (cf. Tamaoka & Makioka, 2009). When Japanese words were used in more than one script (e.g., *camel*-ラクダ/駱駝 /rakuda/, the frequencies of the word in each script were totaled. When words had more than one reading (e.g., *head*–頭, in which the Japanese as a stand-alone noun is read /atama/ and when used in a compound it is pronounced /gashira/ or /tou/), frequency of the stand alone noun only was used). Descriptive statistics for raw frequencies are provided in Table 2. We could not provide occurrences per million, as we only had the token count for morphemic units, which overestimates the actual number of "words" (which often have two or more morphemes) in the corpus. Log-transformed frequencies, which increase normality and reduce random variance, are also provided.

The Bayes factor (BF = 0.03) provides strong evidence for a difference between cognate and noncognate log-transformed word frequencies, though there was less evidence for such a difference using the raw frequencies (BF = 0.18). Thus, although our cognates were selected from a high frequency wordlist of katakana loanwords in Japanese, they are still lower in frequency than the noncognates in the present sample. This may partially be due to the fact that cognates tend to have one borrowed meaning (i.e., few senses). This is especially true for borrowed verbs, adjectives, and adverbs, as native words often exist, and the borrowed words fill narrow lexical gaps. The implication of this is that it is difficult for researchers to match cognate and noncognate items in languages in which the cognates are all borrowed words. Therefore, mixed-effects modeling, which can account for multiple continuous variables such as word frequency and number of senses as well as P and S similarity, might be most suitable for analyses with Japanese–English cognates.

English word frequency Word frequency per million words in English was taken from the SUBTLEX corpus of film and television subtitles (Brysbaert & New, 2009) available from the Elexicon Project (Balota et al., 2007). Log-transformed frequencies, which increase normality and reduce random

variance (Baayen, 2008), are also provided (logSUBTLEX). There was no difference in English word frequency or log-transformed frequency for cognate and noncognate items.

In addition to the frequencies from the subtitles corpus (SUBTLEX) for English and the newspaper corpus (Amano & Kondo, 2000) for Japanese, we provide an additional set of corpus frequencies taken from large web-corpora for each language. These two corpora were obtained from the Sketch Engine website (www.sketchengine.co.uk; Kilgariff et al., 2004); the English corpus (UkWaC) contains 1,318,612,719 words, and the Japanese corpus (JpWac) contains 333,246,192 words. The advantage of using these corpora is that they are comparable in terms of their derivation: Both are derived from the Web—specifically, from shopping and commercial websites, blogs, and discussion forums. The log-transformed frequencies are included in Appendix B (and the supplemental materials). The UkWac corpus log frequencies significantly correlate with the SUBTLEX corpus log frequencies (r = 0.78, CI = 0.72, 0.83) and the JpWac log frequencies correlate strongly with the log frequencies from the Japanese newspaper corpus (r = 0.71, CI = 0.63, 0.77). The two Web corpora also correlated (r = 0.70, CI = 0.60, 0.75) to a much higher degree than the English subtitles and Japanese newspaper corpora (r = 0.35, CI = 0.22, 0.47). Thus, whereas within-language corpora correlations are strong for both languages, the Web corpora appear to better correlate across languages, indicating that they are utilizing similar text resources as the basis for the frequencies. Thus, these may also prove to be valuable resources for studies of Japanese–English bilingual language processing. All log-transformed frequencies are provided in Appendix B (and the supplemental materials).

Number of English senses The total number of senses regardless of class (verb, noun, etc.) was taken from the online version of WordNet (Princeton University, 2010). There was no difference in the numbers of English word senses between cognate and noncognate items.

Number of Japanese senses The total number of senses for Japanese words was taken from *MeikyoKokugoJiten* (*Meikyo Japanese Dictionary*, 2008 edition). In four cases, the Japanese loanword was not listed as a single entry (i.e., only as a compound entry) in the selected dictionary; therefore, the number of senses for these items was taken from a second dictionary—*Koujien*, 6th edition (2008)—in which the items were listed as single entries. Similar to the number of English senses, there was no difference between cognates and noncognates in terms of the number of Japanese senses.

Japanese word length Japanese word length was calculated as the total number of morae in each word. A mora is the basic phonemic unit in Japanese, roughly corresponding to a syllable. For example, 魚 /sakana/ "fish" is written in *kanji* (Sino-Japanese characters) and contains three morae, which can be visualized by transcribing the word using the phonetic script, hiragana: さかな, /sa/, /ka/, /na/. On the other hand, カンガルー /kangaruu/ "kangaroo," is written in katakana, which is used for writing loanwords, and contains five morae in Japanese (/ka/, /n/, /ga/, /ru/, and /u/) , even though the English word contains only three syllables. This exemplifies how the Japanese phonemic system determines the resulting phonetic constitution of the borrowed word, while also briefly illustrating the use of the three scripts of the Japanese language. The Bayes factor (BF = 0.02) provides very strong evidence for a difference in the numbers of morae in cognate and noncognate words, such that the former were longer on average. This is not surprising given that loanwords, which are rephonalized into Japanese from English, tend to be longer than native Japanese words, which typically contain 2–4 morae.

English word length The number of letters and syllables in each English translation were used as two separate measure of English word length. As expected, the word lengths did not differ for the English translations of cognates and noncognates, whether we looked at the number of letters, or syllables, in each word.

Table 3 Intercorrelations among factors

Factor	1	2	3	4	5	6	7	8	9
1. Semantic similarity	–	–0.29	–0.41	–0.20	–0.14	0.40	–0.10	–0.31	–0.09
2. Number of translations from L1 to L2		–	0.13	0.01	0.42	–0.31	0.10	0.05	–0.43
3. Number of translations from L2 to L1			–	0.22	0.12	–0.51	0.11	0.37	0.13
4. Number of meanings of items translated into L2				–	–0.09	–0.13	0.10	0.22	0.12
5. Number of meanings of items translated into L1					–	–0.11	0.18	0.01	–0.11
6. Concreteness						–	–0.20	–0.37	–0.04
7. Number of senses in Japanese							–	0.37	–0.09
8. Number of senses in English								–	0.12
9. Phonological similarity									–

Correlations between ratings and collected measures

S similarity It is likely that the S similarity is related to other lexical–semantic characteristics, such as concreteness or number of translations. In order to assess whether this is the case, a number of predictors were selected for a correlation analysis with the S similarity measure derived in this study: number of translations (in both directions), number of meanings translated (in both directions), concreteness, number of senses in Japanese and English, and P similarity (Table 3). First, S similarity was strongly negatively correlated with the number of translations measures in the L1 to L2 direction ($r = -0.29$, $CI = -0.41, -0.16$) and in the L2 to L1 direction ($r = -0.41$, CI= $-0.52, -0.29$). This shows that as the number of translations increases, S similarity decreases, which is similar to Tokowicz et al.'s (2002) finding for Dutch–English translations. Second, S similarity was negatively correlated with the number of meanings translated in the L1 to L2 direction ($r = -0.20$, $CI = -0.33, -0.06$) and less so in the L2 to L1 direction ($r = -0.14$, $CI = -0.28, 0.00$). Again, the negative correlation shows that words translated with more meanings were rated as less semantically similar across languages. The number of translations measures (L1–L2, L2–L1) were not strongly correlated ($r = 0.13$, $CI = -0.01, 0.27$); this was also the case for the number of meanings ($r = -0.09$, $CI = -0.23, 0.05$). This reflects the fact that the degree of S knowledge varies across languages, with participants having a greater knowledge of S characteristics of words in the L1 relative to the L2. Third, concreteness was highly correlated with S similarity ($r = 0.40$, $CI = 0.27, 0.51$), such that the more concrete the words were rated, the more semantically similar across languages they are (this is similar to Tokowicz et al., 2002). Fourth, S similarity was highly negatively correlated with the number of English senses ($r = -0.31$, $CI = -0.43, -0.18$) and but much less so with the number of Japanese senses ($r = -0.10$, $CI = -0.24$, 0.04). The discrepancy may well be due to the different degrees of sense disambiguation in the English and Japanese sources (WordNet vs. *Meikyo Japanese Dictionary*), the former tending to provide many senses, and the latter tending to be more conservative. Nevertheless, the two measures of number of senses were strongly correlated ($r = 0.37$, $CI = 0.24, 0.49$). Taken together, the numbers of translations, meanings, and senses, along with concreteness, appear to be important S characteristics that determine cross-linguistic S similarity.

In addition, the role of P similarity was explored in order to determine whether there was any relationship between it and S similarity; however, the two similarity measures were not strongly correlated ($r = -0.09$, $CI = -0.23, 0.05$). This supports the finding of Tokowicz et al. (2002) who reported a similar finding for Dutch–English translations. Interestingly, P similarity was highly correlated with number of translations in the L1–L2 direction ($r = -0.43$, $CI = -0.54, -0.31$) but much less so with the L2–L1 direction ($r = 0.13$, $CI = -0.01$, 0.27). This shows that more phonologically similar items (i.e., cognates) had fewer translations in the L2 than did phonologically dissimilar items (i.e., noncognates); for example, コール /kooru/ "call" is usually translated into English using the cognate translation only (i.e., *call*). Finally, P similarity was not correlated with the numbers of meanings in the L2 ($r = 0.12$, $CI = -0.02, 0.26$) or in the L1 ($r = -0.11$, $CI = -0.25$, 0.03), which demonstrates that although fewer different translations were provided for cognates than for noncognates in the L1–L2 direction, the numbers of meanings provided did not differ depending on cognateness or direction of translation.

The present study is the first to report this interesting difference in the numbers of translations for language pairs that do not share etymological origins but are instead loanwords. This characteristic of borrowed words is also likely to be observable in languages pairs such as Korean–English. Thus, when bilinguals translate Korean loanwords into English, they are likely to use a single translation, but this will not be the case when translating from English into Korean. To illustrate, the English word *style* can be translated into various Korean words: 스타일 /sutail/, 모양 /moyang/, 품격 /pumkyek/, or 문체 /munche/. However, when translating the Korean loanword 스타일 /sutail/, Korean–English bilinguals will use only the English word *style*. Because the number of translations influences bilingual processing, this feature of loanwords in such languages is thus important for understanding bilingual processing mechanisms.

Table 4 Intercorrelations among factors

Factor	1	2	3	4	5	6	7
1. Phonological similarity	–	–0.96	–0.24	0.06	0.24	–0.05	–0.02
2. Cognate status		–	0.25	–0.06	–0.25	–0.09	–0.08
3. Log-transformed Japanese word frequency			–	0.35	–0.24	–0.30	–0.11
4. Log-transformed English word frequency				–	–0.30	–0.42	–0.42
5. Japanese word length (mora)					–	0.45	0.39
6. English word length (letters)						–	0.78
7. English word length (syllables)							–

P similarity and cognates In most research to date, words have been dichotomized as cognate or noncognate on the basis of the degree of formal and S overlap. However, as we have shown here, words that are typically classed as cognate can vary in terms of their cross-linguistic P overlap. Because formal overlap across languages has been shown to influence processing bilingual tasks, both as a dichotomous "cognate status" variable (Hoshino & Kroll, 2008; Taft, 2002) and as continuous measures of P and/or O overlap (Allen & Conklin, 2013; Dijkstra et al., 2010), it is crucial to investigate the role of overlap in bilingual processing.

As can be seen in Table 4, P similarity was highly correlated with cognate status (r = −0.96, CI = −0.97, –0.95), showing that the two measures are predicting much of the same characteristic. The almost complete correlation between P similarity and cognate status demonstrates just how well P similarity can categorize items as either cognate or noncognate. Importantly, because bilinguals have been shown in this research to be sensitive to the degree of P similarity between translations across languages, as opposed to simply knowing that words are either cognate or noncognate, P similarity is a superior measure of bilinguals' actual word knowledge and thus should prove to be a more valid measure of bilingual performance in tasks that investigate cross-linguistic processes.

Also, whereas Japanese log word frequency was highly correlated with P similarity (r = −0.24, CI = −0.37, –0.10), it was not correlated with English log word frequency (r = 0.06, CI = −0.08, 0.20). The same pattern is apparent for cognate status and the two log word frequency measures. This highlights the fact that in Japanese, cognates are typically of lower frequency than noncognates, even though we specifically selected items from a high-frequency word list in Japanese.

Finally, whereas both the number of English letters and English syllables were not correlated with P similarity, the number of mora in Japanese was (r = −0.25, CI = −0.38, –0.11). This highlights the fact that Japanese cognates, which are loanwords from English, tend to have a greater number of mora than native Japanese words (i.e., noncognates).

Conclusions

The goal of this study was to provide cross-linguistic norming data for Japanese–English translation equivalents, which will be a useful resource for researchers of bilingual processing of Japanese and English languages. This is the first study to provide such rich resources for languages that differ in script. The data may be used for norming items for use in production tasks such as picture naming (see also Nishimoto et al., 2005; Szekely et al., 2004), word naming and translation, and also comprehension tasks, such as lexical decision, sentence-context reading studies, and studies using progressive demasking techniques or the masked priming paradigm (e.g., Nakayama et al., 2012).

In addition, we highlight a number of important features of cross-linguistic similarity for Japanese–English translations. First, we showed that P similarity ratings are varied for translation equivalents and distinguish between cognates and noncognates as well as within the cognates category. Thus, P similarity is more likely to reflect the processing mechanisms of bilinguals than a dichotomous all or nothing categorization of similarity, even though cognate status and P similarity are very highly correlated predictors.

Second, we showed that although S similarity ratings do not differ significantly for cognate and noncognate items (contra the assumptions of Van Hell & de Groot, 1998), the number of translations varies by direction. Specifically, when Japanese loanwords are translated into English, one translation is unanimously preferred. However, when English words that have loanword equivalents in Japanese are translated, bilinguals use not only the Japanese loanwords but other words as well. This interesting feature may well be present in other languages that borrow from English but do not share its etymological origins, such as Korean–English and Chinese–English. Such knowledge is crucial for selecting stimuli for experiments that test theories of bilingual processing and representation.

We also provided measures of standardization that are not freely available for all of the Japanese items in the present study (age of acquisition and concreteness) and bilingual norming data for English word familiarity. In the L2 familiarity study we observed language transfer effects that resulted in English cognates receiving higher familiarity ratings than noncognates, which is likely due to the effect of cross-linguistic similarity. This further stresses the important role of cross-linguistic similarity in offline, as well as online, tasks. Finally, additional information (frequency, number of senses, and word length) was provided. Cognates tend to be lower in L1 frequency and longer in the number of Japanese characters (or morae), whereas these factors are no different for cognates and noncognates in the L2 (English).

To deal with these inherent differences between Japanese cognates and noncognates, bilingual research that uses cognates might benefit from the use of mixed-effects modeling, as this method can account for multiple continuous variables, such as frequency, length, and number of senses, as well as the researchers' particular variables of interest. All in all, the present data set provides the richest cross-linguistic lexical resource currently available for bilingual studies with different-script languages.

Author note This work was supported in part by a Global Studies grant, Japanese Ministry of Education, Culture, Sports, Science and Technology, to the first author.

Appendix A

Table 5 Cross-linguistic and norming measures data (P and S similarity, Age-of-acquisition, L2 familiarity, concreteness)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
access	アクセス	akusesu	アクセス	No	C	4.5	4.8	3.4	1	5	1	1	3.0	3.9
acid	酸	san	サン	No	NC	4.8	3.5	4.0	1	3	1	1	1.2	4.5
aid	援助	enjo	エンジョ	No	NC	4.4	3.3	2.9	5	7	1	2	1.4	4.1
arm	腕	ude	ウデ	Yes	NC	4.0	3.6	5.4	1	2	1	2	1.0	4.2
arrow	矢印	yajirushi	ヤジルシ	Yes	NC	3.4	3.4	4.4	4	1	2	1	1.0	3.3
ashtray	灰皿	haizara	ハイザラ	Yes	NC	4.2	2.0	5.4	1	2	2	1	1.1	4.1
balloon	風船	fuusen	フウセン	Yes	NC	3.6	3.3	5.6	1	2	1	2	1.0	4.7
banana	バナナ	banana	バナナ	Yes	C	3.3	4.0	5.7	1	1	1	1	3.4	4.9
bed	ベッド	beddo	ベッド	Yes	C	3.9	4.8	5.7	1	1	1	1	3.7	4.7
bench	ベンチ	benchi	ベンチ	Yes	C	3.7	4.1	5.6	1	2	1	1	4.2	4.1
bicycle	自転車	jitensha	ジテンシャ	Yes	NC	4.0	4.1	5.7	2	1	1	1	1.0	4.7
blank	空白	kuuhaku	クウハク	No	NC	4.6	4.1	4.3	2	3	1	1	1.1	3.7
bone	骨	hone	ホネ	Yes	NC	4.1	3.2	4.9	1	1	1	1	1.0	4.6
bricks	レンガ	renga	レンガ	Yes	NC	3.4	2.9	5.7	2	3	2	1	1.1	3.7
broom	ほうき	houki	ホウキ	Yes	NC	3.4	2.7	5.5	3	1	1	1	1.0	2.3
brush	ブラシ	burashi	ブラシ	Yes	C	3.5	4.0	5.4	1	2	1	2	3.1	4.2
bus	バス	basu	バス	Yes	C	3.8	3.1	5.8	1	1	1	1	4.1	4.7
bust	胸	mune	ムネ	No	NC	4.3	3.2	4.9	2	3	1	1	1.2	4.2
button	ボタン	botan	ボタン	Yes	C	3.8	4.1	5.6	1	1	1	1	3.0	4.2
cake	ケーキ	keeki	ケーキ	Yes	C	3.5	4.5	5.4	2	1	1	1	2.6	4.2
call	コール	kooru	コール	No	C	4.6	4.7	2.7	1	4	1	2	3.0	3.6
camel	ラクダ	rakuda	ラクダ	Yes	NC	3.2	3.5	5.7	1	2	1	1	1.1	4.1
care	ケア	kea	ケア	No	C	4.6	4.1	2.8	1	8	1	2	3.9	4.1
career	キャリ	kyaria	キャリ	No	C	4.4	3.7	3.1	1	5	1	2	2.9	3.6
carrot	人参	ninjin	ニンジン	Yes	NC	3.6	3.1	6.1	1	1	1	1	1.0	4.8
case	ケース	keesu	ケース	No	C	4.3	4.5	3.7	1	6	1	2	3.3	3.8
caution	注意	chuui	チュウ	No	NC	4.3	3.6	2.5	4	3	2	2	1.1	4.1
cherry	さくらんぼ	sakuranbo	サクランボ	Yes	NC	2.9	4.0	5.9	1	2	1	1	1.0	4.1
chimney	エントツ	entotsu	エントツ	Yes	NC	3.0	2.1	5.3	1	1	1	1	1.0	3.7
class	クラス	kurasu	クラス	No	C	4.0	4.8	4.3	2	2	1	1	3.5	3.7
classic	クラシック	kurashikku	クラシック	No	C	4.4	4.3	4.2	1	2	1	2	3.7	3.7
clear	クリアー	kuriaa	クリアー	No	C	4.4	4.2	2.6	1	6	1	3	3.7	3.6
clue	手がかり	tegakari	テガカリ	No	NC	4.3	3.6	2.8	3	2	1	1	1.1	3.9

Table 5 (continued)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
cool	クール	kuuru	クール	No	C	4.3	4.6	2.7	1	4	1	2	3.9	3.8
coral	サンゴ	sango	サンゴ	No	NC	4.2	2.8	5.3	1	1	1	1	1.2	3.7
core	コア	koa	コア	No	C	4.9	3.5	2.6	1	4	1	1	3.7	4.1
course	コース	koosu	コース	No	C	4.3	4.5	3.4	1	3	1	2	3.9	4.5
cow	牛	ushi	ウシ	Yes	NC	3.8	3.3	5.7	3	2	2	1	1.0	4.6
crime	犯罪	hanzai	ハンザイ	No	NC	4.3	3.5	3.4	3	3	2	1	1.1	4.6
cross	クロス	kurosu	クロス	No	C	4.3	3.7	3.2	1	4	1	2	3.5	4.1
cure	治る	naoru	ナオル	No	NC	4.2	4.0	3.3	5	4	1	2	1.1	3.7
curtain	カーテ	kaaten	カーテ	Yes	C	3.3	3.9	5.5	1	1	1	1	3.4	4.5
cycle	サイクル	saikuru	サイクル	No	C	4.6	4.0	2.5	1	5	1	2	3.3	4.0
deer	シカ	shika	シカ	Yes	NC	3.5	3.1	5.8	2	1	1	1	1.0	4.7
demand	要求	youkyuu	ヨウキュウ	No	NC	4.6	4.0	2.9	6	3	2	1	1.1	4.3
desk	机	tsukue	ツクエ	No	NC	4.4	4.6	5.7	1	1	1	1	1.0	4.4
dolphin	イルカ	iruka	イルカ	Yes	NC	3.0	3.4	5.7	1	1	1	1	1.0	4.9
door	ドア	doa	ドア	Yes	C	3.6	5.4	5.4	1	2	1	1	3.6	4.6
dress	ドレス	doresu	ドレス	Yes	C	3.8	4.3	5.4	1	4	1	3	3.2	4.1
dresser	たんす	tansu	タンス	Yes	NC	3.0	2.8	5.6	5	3	1	3	1.0	3.1
eagle	ワシ	washi	ワシ	Yes	NC	3.5	3.0	5.7	3	2	1	1	1.2	4.5
elephant	象	zou	ゾウ	Yes	NC	3.4	3.2	5.8	1	1	1	1	1.0	4.9
exit	出口	deguchi	デグチ	No	NC	4.3	4.2	3.9	1	2	1	2	1.1	4.1
fail	失敗	shippai	シッパ	No	NC	4.7	4.0	2.5	4	2	1	1	1.1	4.1
find	見つける	mitsukeru	ミツケル	No	NC	4.0	4.6	3.1	3	2	1	2	1.1	4.3
firm	会社	kaisha	カイシヤ	No	NC	4.0	3.6	3.8	3	5	1	3	1.2	2.9
fish	魚	sakana	サカナ	Yes	NC	4.1	4.6	4.3	1	1	1	1	1.0	4.6
flag	旗	hata	ハタ	Yes	NC	4.3	3.8	5.2	1	1	1	1	1.0	4.3
flute	フルート	furuuto	フルート	Yes	C	3.9	4.0	5.7	1	1	1	1	3.8	4.1
fool	バカ	baka	バカ	No	NC	3.4	3.7	2.7	3	2	1	1	1.0	4.1
foot	足	ashi	アシ	Yes	NC	3.5	4.3	5.6	2	1	2	1	1.0	3.8
fork	フォーク	fooku	フォーク	Yes	C	3.1	3.8	5.6	2	1	2	1	3.9	4.1
fox	キツネ	kitsune	キツネ	Yes	NC	3.1	3.5	5.6	1	1	1	1	1.0	4.6
frog	カエル	kaeru	カエル	Yes	NC	3.1	3.1	5.8	1	1	1	1	1.0	4.6
front	前	mae	マエ	No	NC	4.0	4.4	3.5	4	3	3	2	1.1	4.3
fuel	燃料	nenryou	ネンリョウ	No	NC	4.7	3.6	3.7	3	2	2	2	1.1	4.5
fund	資金	shikin	シキン	No	NC	4.5	3.5	3.6	5	5	1	1	1.1	4.1
future	将来	shourai	ショウライ	No	NC	4.4	4.8	2.4	1	2	1	1	1.1	4.4

Table 5 (continued)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
genre	ジャンル	janru	ジャンル	No	C	4.4	3.2	2.6	3	2	1	1	2.5	4.1
giraffe	キリン	kirin	キリン	Yes	NC	3.0	3.0	6.1	1	1	1	1	1.0	4.3
glass	グラス	gurasu	グラス	No	C	3.7	4.4	5.3	1	3	1	2	3.7	4.1
goal	ゴール	gooru	ゴール	No	C	3.7	4.6	3.9	1	4	1	2	3.6	4.6
goat	ヤギ	yagi	ヤギ	Yes	NC	2.8	2.6	5.8	1	3	1	1	1.0	4.5
gorilla	ゴリラ	gorira	ゴリラ	Yes	C	2.8	3.6	5.9	1	1	1	1	3.3	4.7
grapes	ブドウ	budou	ブドウ	Yes	NC	3.0	3.4	5.8	1	1	1	1	1.0	4.6
guitar	ギター	gitaa	ギター	Yes	C	3.7	4.3	5.6	1	1	1	1	3.1	4.9
hammock	ハンモック	hanmokku	ハンモック	Yes	C	4.0	2.6	5.3	1	1	1	1	3.3	5.0
hanger	ハンガー	hangaa	ハンガー	Yes	C	3.1	4.4	5.7	2	1	2	1	3.6	4.1
hate	憎む	nikumu	ニクム	No	NC	4.9	4.0	2.4	1	3	1	1	1.1	4.1
head	頭	atama	アタマ	No	NC	3.8	4.7	5.2	1	1	1	1	1.1	4.1
heart	ハート	haato	ハート	Yes	C	4.4	4.5	3.3	1	3	1	2	3.2	3.7
helicopter	ヘリコプター	herikoputaa	ヘリコプター	Yes	C	3.6	4.0	5.8	1	1	1	1	3.4	4.9
helmet	ヘルメット	herumetto	ヘルメット	Yes	C	3.4	3.7	5.4	1	1	1	1	3.6	4.9
hope	希望	kibou	キボウ	No	NC	4.9	4.9	1.9	1	4	1	2	1.3	4.1
ideal	理想	risou	リソウ	No	NC	4.6	3.5	2.2	2	1	1	1	1.2	4.1
iron	アイロン	airon	アイロン	Yes	C	3.6	3.6	5.7	1	2	1	2	3.2	2.8
jar	つぼ	tsubo	ツボ	No	NC	4.4	2.6	5.2	1	3	3	1	3.6	4.6
joint	関節	kansetsu	カンセツ	No	NC	4.4	3.0	4.5	1	5	1	2	1.1	3.1
joke	ジョーク	jooku	ジョーク	No	C	4.5	4.5	3.4	1	2	1	1	3.3	4.6
jury	陪審	baishin	バイシン	No	NC	5.4	2.6	4.1	1	2	1	1	1.1	4.0
kangaroo	カンガルー	kangaruu	カンガルー	Yes	C	3.3	3.6	5.9	1	1	1	1	3.5	4.9
kick	キック	kikku	キック	No	C	3.4	4.5	4.4	1	2	1	1	3.9	4.6
kiss	キス	kisu	キス	No	C	4.0	4.3	4.4	1	2	1	1	3.6	4.7
ladder	はしご	hashigo	ハシゴ	Yes	NC	3.2	2.8	5.2	1	3	1	1	1.0	4.1
learn	習う	narau	ナラウ	No	NC	3.9	4.5	3.1	1	2	1	1	1.1	4.4
left	左	hidari	ヒダリ	No	NC	3.6	4.3	4.3	1	2	1	2	1.1	4.4
lemon	レモン	remon	レモン	Yes	C	3.1	4.2	5.8	1	1	1	2	3.4	4.9
lesson	レッスン	ressun	レッス	No	C	4.2	5.1	3.8	1	5	1	1	3.5	4.3
lion	ライオン	raion	ライオン	Yes	C	3.0	3.8	5.6	1	2	1	2	3.7	4.9
lips	唇	kuchibiru	クチビル	Yes	NC	3.6	3.7	5.5	1	1	1	1	1.0	4.7
loan	ローン	roon	ローン	No	C	4.9	3.7	3.7	1	6	2	1	4.0	3.9
lobster	ザリガニ	zarigani	ザリガニ	Yes	NC	3.0	3.1	5.8	2	5	1	1	1.0	3.5
local	ローカル	rookaru	ローカ	No	C	4.3	4.8	3.4	1	3	1	1	3.4	3.9

Table 5 (continued)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
loose	緩い	yurui	ユルイ	No	NC	4.4	3.8	2.8	3	6	1	1	1.2	3.9
lucky	ラッキー	rakkii	ラッキー	No	C	4.0	5.1	2.4	1	3	1	2	3.8	4.3
matter	物事	monogoto	モノゴト	No	C	4.6	4.1	3.0	4	3	2	1	1.1	3.6
maze	迷路	meiro	メイロ	No	NC	3.9	2.2	4.6	2	2	1	2	1.1	2.9
morale	モラル	moraru	モラル	No	C	4.6	3.2	2.7	1	5	1	1	2.9	4.1
naked	裸	hadaka	ハダカ	No	NC	3.4	3.1	4.9	2	3	1	1	1.0	4.3
necklace	ネクレス	nekuresu	ネクレス	Yes	C	3.9	4.5	5.5	1	3	1	1	3.6	4.8
normal	普通	futsuu	フツウ	No	NC	4.1	4.4	2.1	4	2	1	1	1.1	4.4
nose	鼻	hana	ハナ	Yes	NC	3.9	4.1	5.5	1	1	1	1	1.1	4.9
past	過去	kako	カコ	No	NC	4.6	4.1	2.2	2	2	1	1	1.0	4.3
peanut	ピーナツ	piinatsu	ピーナツ	Yes	C	3.3	4.0	5.8	1	2	1	1	3.3	4.9
pelican	ペリカン	perikan	ペリカン	Yes	C	3.3	2.7	5.8	1	1	1	1	3.4	4.8
pencil	鉛筆	enpitsu	エンピツ	Yes	NC	3.6	3.9	5.5	1	1	1	1	1.1	4.8
penguin	ペンギン	pengin	ペンギン	Yes	C	3.1	3.7	5.8	1	1	1	1	3.3	5.0
pig	豚	buta	ブタ	Yes	NC	3.6	3.5	5.7	1	1	1	1	1.1	4.6
pipe	パイプ	paipu	パイプ	Yes	C	4.0	3.1	5.0	1	4	1	1	3.7	4.4
place	場所	basho	バショ	No	NC	4.2	4.6	3.3	5	1	1	2	1.0	4.3
plain	明白	meihaku	メイハク	No	NC	4.9	3.6	2.4	3	12	2	1	1.0	3.3
pool	プール	puuru	プール	Yes	C	3.8	3.6	5.5	1	3	1	3	3.6	4.3
prison	刑務所	keimusho	ケイムショ	No	NC	4.4	3.1	5.4	2	3	1	2	1.2	4.5
profit	利益	rieki	リエキ	No	NC	4.5	3.3	2.9	4	2	2	1	1.0	4.5
pyramid	ピラミッド	piramiddo	ピラミッド	Yes	C	3.7	3.2	5.3	1	3	1	1	3.4	4.6
rabbit	ウサギ	usagi	ウサギ	Yes	NC	3.0	3.7	6.1	1	2	1	1	1.0	5.0
rabbit	ウサギ	usag	ウサギ	Yes	NC	3.0	3.7	6.1	1	2	1	1	1.0	5.0
race	レース	reesu	レース	No	C	4.5	4.2	4.4	1	3	1	1	3.2	3.8
radio	ラジオ	rajio	ラジオ	Yes	C	3.9	4.2	5.4	1	1	1	2	2.5	4.6
rain	雨	ame	アメ	No	NC	3.6	4.5	5.1	1	2	1	1	1.3	4.1
rank	ランク	ranku	ランク	No	C	4.1	4.4	3.2	1	7	1	1	3.7	4.1
real	リアル	riaru	リアル	No	C	4.5	4.3	2.5	1	4	1	2	3.3	4.2
regular	レギュラー	regyuraa	レギュラー	No	C	4.2	4.6	3.2	1	5	1	2	3.3	3.7
release	リリース	ririisu	リリース	No	C	4.6	3.7	3.1	1	7	1	1	3.7	3.9
rental	レンタル	rentaru	レンタル	No	C	4.2	4.5	3.9	1	8	1	2	3.3	4.4
return	リターン	ritaan	リターン	No	C	4.3	4.2	3.2	1	8	1	1	3.4	4.3
ring	指輪	yubiwa	ユビワ	Yes	NC	3.9	4.0	5.3	1	5	1	2	1.1	4.0
rocket	ロケット	roketto	ロケット	Yes	C	3.8	3.5	5.2	1	1	1	3	3.6	4.5

Table 5 (continued)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
roll	ロール	rooru	ロール	No	C	3.9	3.6	3.1	2	6	2	1	3.2	4.0
rule	ルール	ruuru	ルール	No	C	4.1	4.5	3.1	1	7	1	1	3.5	4.3
sailor	水兵	suihei	スイヘイ	No	NC	5.3	2.7	5.2	2	9	1	2	1.1	4.1
scale	スケール	sukeeru	スケール	No	C	4.5	3.6	2.8	1	5	1	1	3.3	4.0
scissors	はさみ	hasami	ハサミ	Yes	NC	3.3	2.7	5.7	1	1	1	4	1.1	4.7
score	スコア	sukoa	スコア	No	C	4.5	4.2	3.4	1	4	1	1	3.7	4.3
screen	スクリーン	sukuriin	スクリーン	No	C	4.3	4.5	4.9	1	2	1	2	3.8	4.4
screw	ネジ	neji	ネジ	Yes	NC	3.2	3.1	3.0	1	4	1	1	1.0	3.5
sense	センス	sensu	センス	No	C	4.3	4.8	2.7	1	4	1	2	3.8	3.8
share	シェア	shea	シェア	No	C	4.4	4.7	5.4	1	4	1	2	3.9	4.0
shark	サメ	same	サメ	Yes	NC	3.1	3.1	5.2	2	2	1	1	1.1	4.9
shirt	ワイシャツ	waishatsu	ワイシャツ	Yes	C	3.8	4.4	2.7	2	1	1	2	1.7	3.4
shock	ショック	shokku	ショック	No	C	4.4	4.4	5.4	1	5	1	1	3.7	4.3
show	ショー	shoo	ショー	No	C	4.2	4.4	3.6	1	5	1	1	3.4	4.0
shower	シャワー	shawaa	シャワー	No	C	3.6	4.5	5.3	1	2	1	2	3.6	4.1
sign	サイン	sain	サイン	No	C	4.4	4.6	4.0	2	8	2	2	4.0	3.3
single	シングル	shinguru	シングル	No	C	4.3	4.3	3.1	1	6	1	2	3.2	4.1
size	サイズ	saizu	サイズ	No	C	4.0	4.7	3.3	1	2	1	2	3.7	4.3
ski	スキー	sukii	スキー	Yes	C	4.0	4.1	5.6	1	1	1	1	4.1	4.6
skill	スキル	sukiru	スキル	No	C	4.6	4.5	3.2	1	4	1	1	3.5	4.3
skirt	スカート	sukaato	スカート	Yes	C	3.5	4.0	5.6	1	1	1	1	3.2	4.8
slipper	スリッパ	surippa	スリッパ	Yes	C	3.3	3.5	5.6	1	2	1	2	3.4	4.2
slow	スロー	suroo	スロー	No	C	4.2	4.3	2.7	1	2	1	1	3.6	4.1
smell	香り	kaori	カオリ	No	NC	4.2	3.7	2.9	4	2	2	1	1.0	4.0
snake	ヘビ	hebi	ヘビ	Yes	NC	3.2	3.4	5.6	1	1	1	2	1.1	4.9
snowman	雪だるま	yukidaruma	ユキダルマ	Yes	NC	2.8	3.0	5.2	1	3	1	1	1.0	4.6
sock	靴下	kutsushita	クツシタ	Yes	NC	3.5	3.0	5.8	1	2	1	2	1.0	4.6
solid	固体	kotai	コタイ	No	NC	4.7	3.0	3.2	3	4	1	1	1.2	4.1
spoon	スプーン	supuun	スプーン	Yes	C	2.8	4.4	5.9	1	3	1	2	4.3	4.8
stroke	なでる	naderu	ナデル	No	NC	3.5	3.2	4.2	2	7	1	1	1.0	2.3
style	スタイル	sutairu	スタイル	No	C	4.5	4.5	2.3	1	8	1	4	3.5	3.9
swan	白鳥	hakuchou	ハクチョウ	Yes	NC	3.7	3.1	5.6	1	1	1	1	1.0	4.6
tank	戦車	sensha	センシャ	Yes	NC	4.1	3.6	5.5	3	4	2	1	1.0	3.4
task	タスク	tasku	タスク	No	C	4.6	4.1	2.9	1	4	1	2	3.9	3.9
telephone	電話	denwa	デンワ	Yes	NC	4.0	4.3	5.1	4	1	2	2	1.0	4.9

Table 5 (continued)

English name	Japanese name	Transcription	Katakana reading	Picture naming	Cognate status	Japanese (L1) age-ofacquisition	English (L2) familiarity	Concreteness (L1)	No. Trans L1-L2	No. Trans L2-L1	NoM L1	NoM L2	Phonological similarity	Semantic similarity
television	テレビ	terebi	テレビ	Yes	C	3.6	4.4	5.7	1	1	1	1	2.4	4.9
tent	テント	tento	テント	Yes	C	3.7	3.4	5.2	1	1	1	1	3.9	4.8
tiger	トラ	tora	トラ	Yes	NC	3.2	3.4	5.7	1	1	1	1	1.7	4.8
toaster	トースター	toosutaa	トースター	Yes	C	3.5	3.6	5.8	1	1	1	1	3.5	4.7
tomato	トマト	tomato	トマト	Yes	C	3.8	4.2	5.9	1	1	1	1	2.8	4.9
tractor	トラクター	torakutaa	トラクター	Yes	C	3.7	2.7	5.2	1	1	1	1	3.6	4.8
trap	ワナ	wana	ワナ	No	NC	4.0	3.5	3.9	1	1	1	1	1.0	4.3
truck	トラック	torakku	トラック	Yes	C	4.1	3.6	5.6	1	1	2	1	3.5	3.9
trumpet	トランペット	toranpetto	トランペット	Yes	C	3.6	3.5	5.8	1	2	1	1	3.7	5.0
turtle	カメ	kame	カメ	Yes	NC	3.4	3.3	5.7	1	1	1	1	1.0	4.8
umbrella	傘	kasa	カサ	Yes	NC	3.3	3.7	5.6	1	1	1	1	1.0	4.9
vest	ベスト	besuto	ベスト	Yes	C	3.8	3.4	3.6	1	2	2	1	3.4	4.2
view	眺め	nagame	ナガメ	No	NC	4.6	4.2	3.2	3	11	2	1	1.0	4.1
violin	バイオリン	bairorin	バイオリン	Yes	C	3.7	3.7	5.7	1	1	1	2	3.3	4.8
wake	覚める	sameru	サメル	No	NC	4.2	3.6	2.8	2	3	1	1	1.0	4.1
warm	暖かい	atatakai	アタタカイ	No	NC	4.1	4.0	3.3	2	3	1	1	1.0	4.6
waste	浪費	rouhi	ロウヒ	No	NC	5.0	3.6	3.2	2	5	2	2	1.0	4.0
wolf	オオカミ	ookami	オオカミ	Yes	NC	3.8	3.0	5.4	1	1	1	2	1.1	4.9
work	ワーク	waaku	ワーク	No	C	4.7	4.8	3.1	1	2	1	1	3.6	4.0
youth	ユース	yuusu	ユース	No	C	4.9	3.5	3.4	2	3	2	1	3.4	3.9
zebra	シマウマ	shimauma	シマウマ	Yes	NC	2.9	3.1	5.7	1	1	1	2	1.0	4.7

English name, Japanese name, the item pair in English and Japanese; *Transcription*, the roman letter transcription of the Japanese word; *Katakana reading*, the transcription of the Japanese word into katakana for confirming the phonetic reading; *Picture naming*, whether or not the item was selected from Nishimoto et al's (2005) picture naming study, if not, the item was selected from a high-frequency wordlist derived from a large web-corpus (Kilgariff et al., 2004); *Cognate status*, whether the item can be termed a cognate (C) or noncognate (NC), based on its obvious phonological and semantic similarity in English and Japanese; *Japanese (L1) ageof-aquisition*, the mean rating for age-of-acquisition gained from 22 bilinguals using the following scale: 1) 0-2 years, 2) 3-4 years, 3) 5-6 years, 4) 7-8 years, 5) 9-10 years, 6) 11-12 years and 7) 13 years or later; *English (L2) familiarity*, the mean rating gained from 19 bilinguals using a scale of 1-7: response categories ranged from very unfamiliar (1) to very familiar (7); *Concreteness (L1)*, the mean concreteness rating for Japanese words gained from 18 bilinguals using a scale of 1-7: response categories ranged from very abstract (1) to very concrete (7); *No. Trans L1-L2*, the total number of different accurate English translations given for the Japanese word; *No.Trans L2-L1*, the total number of different accurate Japanese translations given for the English word; *NoML1*, the total number of different meanings translated for the L2 (English) word into the L1 (Japanese), i.e., only one meaning of the English word *class* was translated into Japanese; *NoML2*, the total number of different meanings translated for the L1 (Japanese) word into the L2 (English), i.e., only one meaning of the Japanese word クラス /kurasu/ 'class' was translated into English; *Phonological similarity*, the mean phonological similarity rating for item pairs using a scale from 1 (competely different) to 5 (identical); *Semantic similarity*, the mean semantic similarity rating for item pairs using a scale from 1 (competely different) to 5 (identical).

Appendix B

Table 6 Additional measures data (Word frequencies, word length, number of senses)

English name	Japanese name	SUBTLWF	log SBTLWF	AK	logAK	UkWac	logUK	JpWac	logJP	Length	Syll	Mora	ENoS	JNoS
access	アクセス	31.7	3.46	1901	7.55	424272	12.96	29653	10.30	6	2	4	8	2
acid	酸	10.0	2.30	431	6.07	32830	10.40	4218	8.35	4	2	2	5	4
aid	援助	13.9	2.63	35098	10.47	87524	11.38	18824	9.84	3	1	3	6	1
arm	腕	65.4	4.18	5656	8.64	41717	10.64	11039	9.31	3	1	2	8	5
arrow	矢印	7.8	2.06	213	5.36	9330	9.14	1239	7.12	5	2	4	2	1
ashtray	灰皿	3.3	1.18	364	5.90	564	6.34	560	6.33	7	2	4	1	1
balloon	風船	8.7	2.16	804	6.69	8837	9.09	427	6.06	7	2	4	4	2
banana	バナナ	10.7	2.37	1264	7.14	5774	8.66	2770	7.93	6	3	3	2	1
bed	ベッド	187.1	5.23	4571	8.43	91169	11.42	9427	9.15	3	1	3	13	2
bench	ベンチ	9.7	2.27	2940	7.99	15246	9.63	2817	7.94	5	1	3	9	1
bicycle	自転車	6.6	1.89	7809	8.96	9351	9.14	14077	9.55	7	3	4	2	1
blank	空白	9.7	2.27	2627	7.87	18551	9.83	1894	7.55	5	1	4	8	2
bone	骨	26.1	3.26	6312	8.75	32268	10.38	8365	9.03	4	1	2	6	6
bricks	レンガ	3.9	1.37	1488	7.31	9020	9.11	1095	7.00	6	1	3	2	1
broom	ほうき	4.8	1.56	394	5.98	1691	7.43	524	6.26	5	1	3	5	1
brush	ブラシ	14.2	2.65	385	5.95	12780	9.46	988	6.90	5	1	3	14	1
bus	バス	74.2	4.31	11801	9.38	94182	11.45	29002	10.28	3	1	2	7	1
bust	胸	27.6	3.32	113755	11.64	5510	8.61	15323	9.64	4	1	2	10	7
button	ボタン	28.3	3.34	2558	7.85	72516	11.19	11724	9.37	6	2	3	9	2
cake	ケーキ	45.1	3.81	942	6.85	18535	9.83	7745	8.95	4	1	3	4	1
call	コール	861.4	6.76	3021	8.01	269805	12.51	4161	8.33	4	1	3	41	4
camel	ラクダ	5.0	1.61	542	6.30	2837	7.95	362	5.89	5	2	3	1	2
care	ケア	485.3	6.18	1730	7.46	365146	12.81	8244	9.02	4	1	2	11	2
career	キャリ	45.2	3.81	1525	7.33	135195	11.81	8783	9.08	6	2	4	3	3
carrot	人参	3.8	1.34	1812	7.50	3902	8.27	936	6.84	6	2	4	4	2
case	ケース	282.4	5.64	22587	10.03	495699	13.11	27323	10.22	4	1	3	22	2
caution	注意	5.2	1.64	12308	9.42	13690	9.52	34338	10.44	7	2	3	5	3
cherry	さくらんぼ	13.6	2.61	708	6.56	4301	8.37	15	2.71	6	2	5	5	1
chimney	エントツ	4.2	1.43	635	6.45	7924	8.98	687	6.53	7	2	4	2	1
class	クラス	117.4	4.77	9927	9.20	234399	12.36	24518	10.11	5	1	3	9	2
classic	クラシック	16.2	2.78	1729	7.46	60857	11.02	5237	8.56	7	2	5	5	2
clear	クリアー	171.8	5.15	954	6.86	295974	12.60	814	6.70	5	1	4	45	4
clue	手がかり	17.6	2.87	1316	7.18	10913	9.30	1960	7.58	4	1	4	3	2
cool	クール	195.9	5.28	193	5.26	53108	10.88	4169	8.34	4	1	3	11	2
coral	サンゴ	2.4	0.86	1078	6.98	6962	8.85	378	5.93	5	2	3	5	2
core	コア	9.8	2.28	195	5.27	82620	11.32	5098	8.54	4	1	2	11	5
course	コース	487.2	6.19	10728	9.28	620418	13.34	23421	10.06	6	1	3	13	4
cow	牛	25.5	3.24	4585	8.43	11145	9.32	11445	9.35	3	1	2	4	1
crime	犯罪	71.2	4.27	12694	9.45	100196	11.51	26146	10.17	5	1	4	2	1
cross	クロス	55.0	4.01	258	5.55	84840	11.35	4570	8.43	5	1	3	16	3
cure	治る	20.8	3.04	2035	7.62	17772	9.79	1574	7.36	4	1	3	5	8
curtain	カーテ	10.3	2.33	962	6.87	7198	8.88	2513	7.83	7	2	4	3	2
cycle	サイクル	5.9	1.77	1396	7.24	69463	11.15	4319	8.37	5	2	4	11	3
deer	シカ	8.7	2.16	248	5.51	13747	9.53	2482	7.82	4	1	2	1	1
demand	要求	17.1	2.84	47607	10.77	105817	11.57	31351	10.35	6	2	5	11	1
desk	机	43.9	3.78	4091	8.32	27593	10.23	6507	8.78	4	1	3	1	2

Table 6 (continued)

English name	Japanese name	SUBTLWF	log SBTLWF	AK	logAK	UkWac	logUK	JpWac	logJP	Length	Syll	Mora	ENoS	JNoS
dolphin	イルカ	2.8	1.02	891	6.79	3386	8.13	1755	7.47	7	2	3	2	1
door	ドア	292.1	5.68	4557	8.42	119227	11.69	18755	9.84	4	1	2	5	1
dress	ドレス	87.2	4.47	586	6.37	30358	10.32	3037	8.02	5	1	3	21	1
dresser	たんす	3.6	1.27	869	6.77	1081	6.99	436	6.08	7	2	3	7	1
eagle	ワシ	11.5	2.44	1304	7.17	4151	8.33	1540	7.34	5	2	2	6	1
elephant	象	11.4	2.43	1732	7.46	7564	8.93	2431	7.80	8	3	2	2	1
exit	出口	15.6	2.75	2061	7.63	28353	10.25	3997	8.29	4	2	3	5	1
fail	失敗	24.6	3.20	14314	9.57	46844	10.75	22422	10.02	4	1	4	11	1
find	見つける	831.0	6.72	12877	9.46	669574	13.41	6568	8.79	4	1	4	18	2
firm	会社	35.3	3.56	121162	11.70	93876	11.45	124291	11.73	4	1	3	14	1
fish	魚	83.5	4.42	7205	8.88	100302	11.52	14341	9.57	4	1	3	6	1
flag	旗	17.5	2.86	3398	8.13	26318	10.18	307	5.73	4	1	2	13	4
flute	フルート	2.1	0.75	610	6.41	5704	8.65	1063	6.97	5	1	4	4	1
fool	バカ	89.3	4.49	1727	7.45	9773	9.19	8887	9.09	4	1	2	7	7
foot	足	64.9	4.17	19920	9.90	72861	11.20	34179	10.44	4	1	2	14	10
fork	フォーク	8.8	2.18	1406	7.25	8845	9.09	2010	7.61	4	1	3	9	2
fox	キツネ	21.6	3.07	867	6.77	8476	9.04	701	6.55	3	1	3	10	2
frog	カエル	11.8	2.47	794	6.68	3595	8.19	2234	7.71	4	1	3	4	1
front	前	181.6	5.20	156283	11.96	203471	12.22	240980	12.39	5	1	2	13	15
fuel	燃料	17.2	2.84	7782	8.96	68646	11.14	11055	9.31	4	1	4	5	1
fund	資金	10.6	2.36	59829	11.00	72473	11.19	30818	10.34	4	1	3	9	1
future	将来	103.5	4.64	35007	10.46	383038	12.86	3582	810.49	6	2	4	7	2
genre	ジャン	1.1	0.06	1599	7.38	20766	9.94	9316	9.14	5	2	3	4	1
giraffe	キリン	1.5	0.40	891	6.79	975 6	88	893	6.79	7	2	3	1	2
glass	グラス	60.7	4.11	670	6.51	75896	11.24	3578	8.18	5	1	3	12	3
goal	ゴール	16.8	2.82	6455	8.77	98872	11.50	9685	9.18	4	1	3	4	3
goat	ヤギ	10.5	2.35	414	6.03	4710	8.46	889	6.79	4	1	2	4	1
gorilla	ゴリラ	5.6	1.71	348	5.85	1249	7.13	553	6.32	7	3	3	1	1
grapes	ブドウ	3.9	1.37	1281	7.16	4655	8.45	844	6.74	6	1	3	2	1
guitar	ギター	15.6	2.75	1317	7.18	48378	10.79	8290	9.02	6	2	3	1	1
hammock	ハンモック	1.4	0.33	51	3.93	939	6.84	136	4.91	7	2	5	2	1
hanger	ハンガー	1.4	0.30	225	5.42	1138	7.04	464	6.14	6	2	4	2	1
hate	憎む	214.6	5.37	648	6.47	28358	10.25	581	6.36	4	1	3	2	1
head	頭	371.5	5.92	21291	9.97	236151	12.37	57859	10.97	4	1	3	4	9
heart	ハート	244	2 5.50	424	6.05	188643	12.15	3927	8.28	5	1	3	10	3
helicopter	ヘリコプター	15.8	2.76	4902	8.50	10646	9.27	1498	7.31	10	4	6	1	1
helmet	ヘルメット	9.5	2.25	1135	7.03	9128	9.12	1127	7.03	6	2	5	2	2
hope	希望	320.6	5.77	23915	10.08	215310	12.28	31622	10.36	4	1	3	9	2
ideal	理想	7.3	1.99	4744	8.46	91640	11.43	13474	9.51	5	2	3	5	1
iron	アイロン	17.9	2.89	406	6.01	42295	10.65	791	6.67	4	2	4	6	1
jar	つぼ	8.3	2.12	730	6.59	6000	8.70	703	6.56	6	2	4	7	6
joint	関節	27.6	3.32	1374	7.23	85612	11.36	5100	8.54	5	1	4	13	1
joke	ジョーク	73.0	4.29	584	6.37	15061	9.62	2129	7.66	4	1	4	6	1
jury	陪審	42.8	3.76	921	6.83	13815	9.53	1917	7.56	4	2	4	2	1
kangaroo	カンガルー	2.3	0.84	106	4.66	1065	6.97	453	6.12	8	3	5	1	1
kick	キック	73.4	4.30	394	5.98	28820	10.27	2206	7.70	4	1	3	14	2
kiss	キス	121.2	4.80	647	6.47	8097	9.00	3627	8.20	4	1	2	6	1
ladder	はしご	9.3	2.22	609	6.41	13413	9.50	614	6.42	6	2	3	4	2

Table 6 (continued)

English name	Japanese name	SUBTLWF	log SBTLWF	AK	logAK	UkWac	logUK	JpWac	logJP	Length	Syll	Mora	ENoS	JNoS
learn	習う	118.6	4.78	2022	7.61	164507	12.01	1022	6.93	5	1	3	6	1
left	左	484.5	6.18	13434	9.51	456491	13.03	21651	9.98	4	1	3	24	6
lemon	レモン	12.0	2.49	84	4.43	8657	9.07	1896	7.55	5	2	3	5	1
lesson	レッスン	32.2	3.47	773	6.65	34865	10.46	7638	8.94	6	2	4	4	2
lion	ライオン	15.4	2.73	757	6.63	8053	8.99	2365	7.77	4	2	4	4	1
lips	唇	31.2	3.44	868	6.77	12390	9.42	3214	8.08	4	1	4	5	1
loan	ローン	19.9	2.99	4079	8.31	64240	11.07	9567	9.17	4	1	3	3	1
lobster	ザリガニ	7.3	1.99	128	4.85	2446	7.80	358	5.88	7	2	4	2	2
local	ローカ	41.7	3.73	477	6.17	807598	13.60	4432	8.40	5	2	4	5	1
loose	緩い	41.8	3.73	874	6.77	29939	10.31	466	6.14	5	1	3	18	6
lucky	ラッキー	143.5	4.97	317	5.76	37067	10.52	3772	8.24	5	2	4	3	1
matter	物事	370.6	5.92	1252	7.13	199009	12.20	6597	8.79	6	2	4	7	1
maze	迷路	2.6	0.94	509	6.23	6095	8.72	951	6.86	4	1	3	2	1
morale	モラル	4.1	1.42	1612	7.39	7057	8.86	2857	7.96	6	2	3	2	1
naked	裸	39.3	3.67	1617	7.39	10531	9.26	3419	8.14	5	2	3	5	4
necklace	ネクレス	9.8	2.28	456	6.12	2095	7.65	732	6.60	8	2	5	1	1
normal	普通	70.4	4.25	7667	8.94	136501	11.82	49509	10.81	6	2	3	5	2
nose	鼻	69.8	4.24	2050	7.63	23573	10.07	10735	9.28	4	1	2	10	2
past	過去	123.8	4.82	40726	10.61	288728	12.57	43084	10.67	4	1	2	6	4
peanut	ピーナツ	12.4	2.51	181	5.20	1963	7.58	120	4.79	6	2	5	5	1
pelican	ペリカン	1.8	0.56	95	4.55	614	6.42	176	5.17	7	3	4	1	1
pencil	鉛筆	9.9	2.29	1357	7.21	8138	9.00	1694	7.43	6	2	4	5	1
penguin	ペンギン	2.9	1.06	395	5.98	1597	7.38	1450	7.28	7	2	4	1	1
pig	豚	39.1	3.67	1432	7.27	11464	9.35	5090	8.54	3	1	2	9	2
pipe	パイプ	19.4	2.96	4833	8.48	22290	10.01	2827	7.95	4	1	4	10	4
place	場所	602.7	6.40	26951	10.20	711596	13.48	67227	11.12	5	1	2	30	3
plain	明白	21.8	3.08	515	6.24	38613	10.56	3613	8.19	5	1	4	11	1
pool	プール	47.0	3.85	3298	8.10	65923	11.10	5905	8.68	4	1	3	7	4
prison	刑務所	66.0	4.19	2762	7.92	47075	10.76	4255	8.36	6	2	4	2	1
profit	利益	11.0	2.39	35064	10.46	45239	10.72	32021	10.37	6	2	3	4	2
pyramid	ピラミッド	4.0	1.39	713	6.57	4171	8.34	2057	7.63	7	3	5	7	1
rabbit	ウサギ	20.9	3.04	1535	7.34	10445	9.25	2175	7.68	6	2	3	4	1
race	レース	61.9	4.13	8147	9.01	132707	11.80	10294	9.24	4	1	3	10	2
radio	ラジオ	77.2	4.35	8696	9.07	95220	11.46	13013	9.47	5	3	3	5	1
rain	雨	48.9	3.89	12241	9.41	50650	10.83	24566	10.11	4	1	2	4	3
rank	ランク	8.5	2.14	2963	7.99	20069	9.91	4716	8.46	4	1	3	13	1
real	リアル	442.8	6.09	205	5.32	310064	12.64	7644	8.94	4	1	3	13	1
regular	レギュラー	33.9	3.52	867	6.77	141490	11.86	2115	7.66	7	3	5	17	3
release	リリース	36.3	3.59	72	4.28	126790	11.75	10788	9.29	7	2	4	22	2
rental	レンタル	4.8	1.57	1324	7.19	20916	9.95	4584	8.43	6	2	4	4	1
return	リターン	91.7	4.52	185	5.22	243755	12.40	2153	7.67	6	2	4	29	3
ring	指輪	92.8	4.53	1076	6.98	55053	10.92	2219	7.70	4	1	3	15	4
rocket	ロケット	11.8	2.47	3714	8.22	8802	9.08	4432	8.40	6	2	4	7	1
roll	ロール	63.3	4.15	183	5.21	36851	10.51	3265	8.09	4	1	3	33	4
rule	ルール	48.1	3.87	9481	9.16	81760	11.31	19419	9.87	4	1	3	5	1
sailor	水兵	12.4	2.52	239	5.48	3869	8.26	249	5.52	6	2	4	3	1
scale	スケール	9.5	2.25	1023	6.93	100719	11.52	3747	8.23	5	1	4	18	3
scissors	はさみ	6.7	1.90	1369	7.22	2986	8.00	753	6.62	8	2	3	4	3

Table 6 (continued)

English name	Japanese name	SUBTLWF	log SBTLWF	AK	logAK	UkWac	logUK	JpWac	logJP	Length	Syll	Mora	ENoS	JNoS
score	スコア	30.4	3.42	1257	7.14	59974	11.00	4810	8.48	5	1	3	18	2
screen	スクリーン	23.4	3.15	1305	7.17	117693	11.68	5172	8.55	6	1	5	16	4
screw	ネジ	37.5	3.62	576	6.36	8785	9.08	1384	7.23	5	1	2	10	2
sense	センス	131.8	4.88	1057	6.96	200193	12.21	5735	8.65	5	1	3	9	2
share	シェア	69.5	4.24	0	NA	165513	12.02	5722	8.65	5	1	2	10	1
shark	サメ	15.0	2.71	457	6.12	4491	8.41	760	6.63	5	1	2	5	1
shirt	ワイシャツ	46.4	3.84	697	6.55	14529	9.58	378	5.93	5	1	4	2	1
shock	ショック	28.8	3.36	5825	8.67	30052	10.31	8844	9.09	5	1	3	17	3
show	ショー	488.4	6.19	2643	7.88	375088	12.83	8899	9.09	4	1	3	16	2
shower	シャワー	41.1	3.72	903	6.81	33304	10.41	3608	8.19	6	1	3	11	1
sign	サイン	133.3	4.89	2956	7.99	110620	11.61	7017	8.86	4	1	3	20	2
single	シングル	72.1	4.28	1196	7.09	267442	12.50	4476	8.41	6	2	4	10	9
size	サイズ	46.1	3.83	2164	7.68	190529	12.16	16925	9.74	4	1	3	9	1
ski	スキー	8.1	2.09	5444	8.60	19514	9.88	6912	8.84	3	1	3	2	2
skill	スキル	7.9	2.07	23	3.14	43361	10.68	9451	9.15	5	1	3	2	7
skirt	スカート	10.0	2.30	1124	7.02	4414	8.39	2375	7.77	5	1	4	7	2
slipper	スリッパ	2.2	0.80	191	5.25	660	6.49	31	3.43	7	2	4	2	1
slow	スロー	76.0	4.33	107	4.67	62118	11.04	3386	8.13	4	1	3	11	1
smell	香り	83.1	4.42	3540	8.17	19164	9.86	8796	9.08	5	1	3	10	2
snake	ヘビ	22.4	3.11	1211	7.10	6661	8.80	2462	7.81	5	1	2	8	1
snowman	雪だるま	1.9	0.64	357	5.88	469	6.15	382	5.95	7	2	5	1	1
sock	靴下	9.0	2.19	645	6.47	1793	7.49	872	6.77	4	1	4	3	1
solid	固体	19.6	2.97	364	5.90	52114	10.86	4135	8.33	5	2	3	18	1
spoon	スプーン	7.6	2.03	420	6.04	5199	8.56	1426	7.26	5	1	4	5	2
stroke	なでる	13.1	2.57	592	6.38	25763	10.16	174	5.16	6	1	3	16	3
style	スタイル	30.1	3.40	3962	8.28	160463	11.99	16291	9.70	5	1	4	12	3
swan	白鳥	6.8	1.92	793	6.68	2265	7.73	1223	7.11	4	1	4	4	1
tank	戦車	25.6	3.24	3495	8.16	29573	10.29	2679	7.89	4	1	3	7	1
task	タスク	12.7	2.54	12	2.48	94278	11.45	2342	7.76	4	1	3	4	2
telephone	電話	32.4	3.48	61289	11.02	108445	11.59	72126	11.19	9	3	3	2	2
television	テレビ	33.9	3.52	60636	11.01	83184	11.33	52198	10.86	10	4	3	2	1
tent	テント	17.5	2.86	2719	7.91	14763	9.60	2943	7.99	4	1	3	2	1
tiger	トラ	18.5	2.92	1494	7.31	4744	8.46	4731	8.46	5	2	2	2	2
toaster	トースター	3.9	1.36	56	4.03	1520	7.33	189	5.24	7	2	5	2	1
tomato	トマト	5.9	1.77	1785	7.49	7362	8.90	3901	8.27	6	3	3	2	1
tractor	トラクター	3.7	1.32	493	6.20	4219	8.35	611	6.42	7	2	5	2	1
trap	ワナ	23.8	3.17	867	6.77	13950	9.54	540	6.29	4	1	2	16	2
truck	トラック	72.9	4.29	9543	9.16	12650	9.45	109267	11.60	5	1	4	3	1
trumpet	トランペット	4.1	1.42	317	5.76	5276	8.57	773	6.65	7	2	6	4	1
turtle	カメ	17.0	2.84	399	5.99	2867	7.96	3126	8.05	6	2	2	3	1
umbrella	傘	7.5	2.01	1706	7.44	9722	9.18	3470	8.15	8	3	2	4	2
vest	ベスト	5.6	1.72	4736	8.46	2830	7.95	10586	9.27	4	1	3	7	2
view	眺め	38.5	3.65	294	5.68	340159	12.74	10098	9.22	4	1	3	13	1
violin	バイオリン	4.8	1.56	1613	7.39	7773	8.96	1655	7.41	6	3	5	1	1
wake	覚める	105.2	4.66	1023	6.93	20154	9.91	976	6.88	4	1	3	9	4
warm	暖かい	52.1	3.95	4794	8.48	66980	11.11	3991	8.29	4	1	5	13	6
waste	浪費	53.3	3.97	853	6.75	112286	11.63	1504	7.32	5	1	3	16	1
wolf	オオカミ	20.3	3.01	895	6.80	4996	8.52	1754	7.47	4	1	4	6	2

Table 6 (continued)

English name	Japanese name	SUBTLWF	log SBTLWF	AK	logAK	UkWac	logUK	JpWac	logJP	Length	Syll	Mora	ENoS	JNoS
work	ワーク	798.0	6.68	212	5.36	1615542	14.30	8683	9.07	4	1	3	24	1
youth	ユース	16.8	2.82	520	6.25	65880	11.10	2805	7.94	5	1	3	6	1
zebra	シマウマ	2.5	0.92	74	4.30	1543	7.34	157	5.06	5	2	4	1	1

English name, Japanese name, the item pair in English and Japanese; *SBTLWF,* the word frequency per million of the English item taken from the SUBTLEX corpus (Brysbaert & New, 2009); *logSBTLWF*, the log-transformed SBTLWF word frequency; *AK*, the raw word frequency of the Japanese item taken from a corpus of Japanese newspaper articles (Amano & Kondo, 2000); *logAK*, the log-transformed AK word frequency; *UkWac*, the raw word frequency of the English item taken from thea 1-billion word web-corpus (Kilgariff et al., 2004); *logUK*, the log-transformed UkWac word frequency; *JpWac*, the raw word frequency of the Japanese item taken from a 300-million word web-corpus (Kilgariff et al., 2004); *logJP*, the log-transformed JpWac word frequency; *Length*, the number of letters in the English word; *Syll*, the number of syllables in the English word; *Mora*, the number of mora in the Japanese word; *ENoS*, the total number of senses of the English word taken from WordNet (Princeton, 1990); *JNoS*, the total number of senses of the Japanese word, taken from Meikyou dictionary (2008)

References

Allen, D., & Conklin, K. (2013). Cross-linguistic similarity and task demands for Japanese–English bilingual processing. *PLOS ONE, 8*(8), e72631. doi:10.1371/journal.pone.0072631

Amano, S., & Kondo, T. (2000). *Nihongo-no goi-tokusei [Lexical properties of Japanese] (Vol. 7) [CD-ROM]*. Tokyo, Japan: Sanseido.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of present word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language, 54,* 324–345.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. doi:10.1080/14640748108400805

Dijkstra, T. (2007). The multilingual lexicon. In M. G. Gaskell (Ed.), *Handbook of psycholinguistics* (pp. 251–265). Oxford, UK: Oxford University Press.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496–518.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappeli, M., & Baayen, R. H. (2010). How cross-linguistic similarity affects cognate recognition. *Journal of Memory and Language, 62,* 284–301.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5,* 175–197.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology, 57,* 98–107.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281. doi:10.1037/0096-3445.113.2.256

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1122–1139.

Grosjean, F. (1989). Neurolinguists, beware: The bilingual is not two monolinguals in one person. *Brain and Language, 36,* 31–15.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-linguistic activation survive a change of script? *Cognition, 106,* 501–511.

Igarashi, Y. (2008). *The changing role of katakana in the Japanese writing system: Processing and pedagogical dimensions for native speakers and foreign learners*. Saarbrücken, Germany: VDM Verlag.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. doi:10.1080/01621459.1995.10476572

Kawaguchi, R., & Tsunoda, F. (2005). *Nihongo wa darenomonoka [Whom is Japanese for?]*. Tokyo, Japan: Yoshikawa Koubunkan.

Kess, J. F., & Miyamoto, T. (2000). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji processing*. Baltimore, MD: Benjamins.

Kilgariff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). *The sketch engine*. In Proceedings from EURALEX 2004, pp. 105–116. Available at the.sketchengine.co.uk

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and priming. *Journal of Memory and Language, 49,* 484–499.

Koujien Dictionary, 5th edition. (1998). Tokyo, Japan: Iwanami Shoten.

Koujien Dictionary, 6th edition. (2008). Tokyo, Japan: Iwanami Shoten.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods, 44,* 978–990. doi:10.3758/s13428-012-0210-4

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407. doi:10.1037/0033-295X.88.5.375

Meikyo Japanese dictionary. (2008). Tokyo, Japan: Taishukan.

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese–English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes, 27,* 1563–1583.

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods, 37,* 398–416. doi:10.3758/BF03192709

Prem, M. (1991). Nihongo kyouiku no neck: Gairaigo [Loanwords: the bottleneck in teaching Japanese to foreigners]. *Nihongo Kyouiku, 74*, 28–33.

Princeton University. (2010). WordNet: A lexical database for English. About WordNet [Web page]. Retrieved from wordnet.princeton.edu

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. doi:10.3758/PBR.16.2.225

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55,* 197–212.

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes, 22,* 106–129.

Shinnouchi, M. (2000). Gendai gairaigo jijoo [The situation concerning contemporary loanwords]. *Data and Statistics, 7*(510), 8–13.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215. doi:10.1037/0278-7393.6.2.174

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51,* 247–250. doi:10.1016/j.jml.2004.03.002

Taft, M. (2002). Orthographic processing of polysyllabic words by native and nonnative English speakers. *Brain and Language, 81,* 532–544.

Tamaoka, K., & Makioka, S. (2009). Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on Japanese speech production. *Language and Speech, 52,* 79–112.

Tokowicz, N., Kroll, J. F., de Groot, A. M. B., & van Hell, J. G. (2002). Number-of-translation norms for Dutch–English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, & Computers, 34,* 435–451. doi:10.3758/BF03195472

Tomita, T. (1991). Nihongo kyouiku to gairaigo oyobi sono hyouki [Loanwords in teaching Japanese and its orthography]. *Nihongogaku [Japanese Linguistics], 10*, 37–44.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64,* 88–107. doi:10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science, 20,* 923–927. doi:10.1111/j.1467-9280.2009.02389.x

Van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1,* 193–211.

Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context affects lexical decision and word translation. *Acta Psychologica, 128*, 431–451. doi:10.1016/j.actpsy.2008.03.010

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39,* 458–483.

Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.

Yokokawa, H. (2009). *Nihonjin eigogakushushano eitangoshimitsudo onseipan: Kyouiku kenyuunotameno dainigengogakudeetabeesu [English word familiarity of Japanese learners of English, Audio Edition: Second language research database for education and research]*. Tokyo, Japan: Kuroshio.