



Translation norms for Malay and English words: The effects of word class, semantic variability, lexical characteristics, and language proficiency on translation

Soon Tat Lee¹ · Walter J. B. van Heuven² · Jessica M. Price¹ · Christine Xiang Ru Leong¹

Accepted: 6 September 2022 / Published online: 11 October 2022
© The Psychonomic Society, Inc. 2022

Abstract

Translation equivalents are widely used in bilingual research concerning word processing (e.g., Eddington & Tokowicz, 2013; Jouravlev & Jared, 2020) and second-language vocabulary learning (e.g., Bracken et al., 2017; Degani et al., 2014). Although translation norms exist in several languages, to date there are no Malay-English translation norms. This study presents the first Malay-English translation norms collected with highly proficient Malay-English bilinguals. Furthermore, the study investigates the impact of lexical characteristics on translation ambiguity. The forward translation (FT) task ($N = 30$) collected English translations for 1004 Malay words selected from the Malay Lexicon Project (Yap et al., 2010), and subsequently the backward translation (BT) task ($N = 30$) gathered Malay translations for 845 English words obtained from the FT phase. The data revealed a high prevalence of translation ambiguity in both translation directions. Specifically, verbs, adjectives, and class-ambiguous words were more translation-ambiguous than nouns. Furthermore, within-language semantic variability and word length were positively correlated with translation ambiguity, whereas word frequency only correlated with translation ambiguity in FT. Word length and word frequency of the source words and their translations were positively correlated. Intriguingly, only in FT were bilinguals with higher Malay proficiency more likely to provide accurate and dominant translations for the Malay words. These findings contrast with those reported in translation norming studies involving other language pairs. The translation norms provide a useful resource for bilingual language studies involving Malay-English bilinguals.

Keywords Translation equivalents · Translation ambiguity · Bilingualism

Introduction

Malay is commonly spoken in countries of Southeast Asia such as Malaysia, Brunei, Indonesia, and Singapore (Lee et al., 1998; Lee & Wheldall, 2011; Tan et al., 2009). Being a language from the Austronesian language family, it is commonly investigated in psycholinguistic research for cross-linguistic comparisons with English (e.g., Rusli & Montgomery, 2020). Malay and English share the same 26 letters, but the former has shallower orthography depth, simpler syllable structures, and more transparent affixation than the latter

(Yap et al., 2010). Furthermore, Malay possesses a more complex morphological system, where words can be formed via rule-based affixation (Yap et al., 2010). For instance, a noun (e.g., “*penulis/author*”) can be formed by adding a noun prefix “*peN-*” to a verb “*tulis/write*.” In a similar way, an adjective (e.g., “*bertulis/having writing*”) can be formed by adding a verb prefix “*ber-*” to the same word. In view of these morphological differences, Malay words have more syllables and a wider range in word length than English words (Lee et al., 1998). Taken together, cross-linguistic research involving Malay and English could generate important insights regarding the effects of different language-specific characteristics (e.g., morphological complexity, orthography depth) on bilingual language processing.

Studies investigating bilingual language processing often use translation equivalents, which are words from two different languages that share similar meaning (e.g., Basnight-Brown et al., 2018; Lee et al., 2018). For example, the Malay

✉ Christine Xiang Ru Leong
christine.leong@nottingham.edu.my

¹ School of Psychology, University of Nottingham, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

² University of Nottingham, Nottingham, UK

word “*beras*” and English word “rice grain” are Malay-English translation equivalents, in which both words refer to the seeds of a swamp grass that are cooked and consumed for food. The process of identifying appropriate translation equivalents requires researchers to be proficient in both languages, so that the meaning of the source word can be adequately represented in the translation. However, not all researchers are necessarily proficient in the languages of investigation (e.g., non-native Malay speakers conducting research in Malay). Furthermore, challenges in identifying translation equivalents are complicated by many words that do not have one-to-one corresponding translation from one language to another (Schwieter & Prior, 2020). For instance, the Malay word “*angka*” can be translated into “number,” “digit,” and “figure” in English, and thus have one-to-many mapping from Malay (source language) to English (target language). This one-to-many mapping from a source language to a target language is termed as translation ambiguity (Prior et al., 2007; Tokowicz et al., 2002).

Translation ambiguity could be driven by several reasons (Degani & Tokowicz, 2013; Prior et al., 2011; Schwieter & Prior, 2020). For example, translation ambiguity happens when the meanings of a source word can be represented by different translations in the target language (e.g., the Malay homonym “*mangga*” can be translated into “mango,” a type of fruit, and “lock,” a tool that keeps doors fastened, in English), or when a specific meaning of a source word (e.g., “*batu*” that refers to the solid substance found in the ground) can be translated into several possible English translations that share similar meanings (e.g., the synonyms “rock” and “stone”). In addition, the conceptual and morphological differences (e.g., the use of affixations to signal meaning) between a language pair also contribute to the degree of translation ambiguity between two languages (Degani et al., 2016; Prior et al., 2007). For instance, the English word “thick” covers the meaning of “not thin” for both solid and liquid substances; however, these concepts are distinctly represented by “*tebal* (for solid)” and “*pekat* (for liquid)” in Malay.

There is currently no psycholinguistic database that provides translation ambiguity for every word that exist in any given language pairs (Schwieter & Prior, 2020). Nevertheless, several translation norming studies had been conducted to estimate the prevalence of translation ambiguity for specific language pairs. In these studies, bilinguals were asked to provide translations for words across the two languages they speak (e.g., Prior et al., 2007; Tokowicz et al., 2002; Wen & van Heuven, 2017). Researchers then proceeded to identify the translation-unambiguous (words that have one-to-one mapping) and translation-ambiguous words. Intriguingly, not all possible translations of the translation-ambiguous words share the same status. In particular, the translation that is most frequently provided by bilinguals is identified as

Table 1 Summary of translation ambiguity from past translation norming studies

Translation norms	FT		BT	
	N	%	N	%
Dutch-English (Tokowicz et al., 2002)	562	25.3	562	30.4
Spanish-English (Prior et al., 2007)	762	48.2	670	58.5
Chinese-English (Tseng et al., 2014)	-	-	562	67.3
Chinese-English (Wen & van Heuven, 2017)	-	-	1429	71.2

Tokowicz et al. (2002) and Tseng et al. (2014) normed on the same set of English words

the dominant translation (e.g., the Spanish word “*permitir*” is a more dominant translation choice for the English word “answer” than the word “*dejar*”; Prior et al., 2007). The existence of dominant translations and translation norms enable researchers to further investigate factors affecting word translation and bilingual word processing (Schwieter & Prior, 2020), such as how the consistency of translation choice could be affected by translation ambiguity (e.g., Prior et al., 2011), and how the translation dominance of words (i.e., dominant and subordinate translations) affects bilingual language performance (e.g., Laxén & Lavour, 2010).

Previous translation norming studies have demonstrated high translation ambiguity across several language pairs (Allen & Conklin, 2014; Prior et al., 2007; Tokowicz et al., 2002; Tseng et al., 2014; Wen & van Heuven, 2017). The prevalence of translation ambiguity varies across different language pairs and translation directions (see Table 1 for summary). Forward translation (FT; first language to second language translation, L1-to-L2) consistently resulted in lower translation ambiguity when compared with backward translation (BT; L2-to-L1 translation) (Allen & Conklin, 2014; Prior et al., 2007; Tokowicz et al., 2002)¹. These observed differences in the prevalence of translation ambiguity could partially be attributed to methodological differences across studies (e.g., different sets of word stimuli, different number of participants) and unique language-specific linguistic characteristics (e.g., morphological complexity) (Schwieter & Prior, 2020). However, in view of (a) the current lack of variety in the language pairs being normed, and (b) the fact that all the available norms shared English as one of the languages, the extent to which language-specific linguistic characteristics contribute to translation ambiguity remains speculative.

In the past, many bilingual studies assumed the words they used were translation-unambiguous (Tokowicz et al.,

¹ It is important to note that in all the reviewed translation norming studies, English is consistently being used as the target language in FT, and source language in BT.

2002). Such assumption can be problematic for interpreting research findings because studies have shown that bilinguals' performance on linguistic tasks can be affected by the degree of translation ambiguity (Eddington & Tokowicz, 2013; Jouravlev & Jared, 2020; Laxén & Lavaur, 2010). Specifically, bilinguals were found to recognize translation-unambiguous word pairs faster than translation-ambiguous word pairs, and the dominant translations were recognized faster than the nondominant translations (see Schwieter & Prior, 2020, for a review). In such a scenario, translation norms are crucial for selecting translation equivalents for psycholinguistics studies investigating bilingual language processing.

Despite the growing number of studies investigating cross-linguistic word processing in Malay (e.g., Luniewska et al., 2019; Yap et al., 2010; Yap et al., 2017), there are no translation norms for Malay and English, which are commonly used as language pair in Malay cross-linguistic research. Therefore, the selection of Malay-English translation equivalents is subject to possible unforeseen extraneous variables and biases. Hence, the present Malay-English translation norming project aimed to create the first freely available large database of Malay-English and English-Malay translation norms using the “first translation” method (Allen & Conklin, 2014; Prior et al., 2007; Tokowicz et al., 2002; Tseng et al., 2014; Wen & van Heuven, 2017). This translation norming project started with the FT phase that included 1004 Malay words before these words were translated back from English to Malay in BT phase. Separate groups of proficient Malay-English bilinguals were recruited for each phase. The translations gathered were summarized into ambiguous and unambiguous translation equivalents, supplemented with word class, semantic variability (number of senses), word frequency and word length information. The availability of this information also allows further investigations into how lexical and semantic factors as well as individual differences might affect translation ambiguity and bilinguals' translation choice.

Word class

Past studies suggest that verbs impose greater processing demands than nouns due to the complex relationship between semantics, syntax, and morphology of verbs (see Vigliocco et al., 2011 for a review). In general, nouns refer to discrete entities while verbs refer to actions or events. When comparing nouns and verbs within a language, meaning of verbs is often more context-dependent (Earles & Kersten, 2017; Gentner, 1981) and more polysemous (Miller & Fellbaum, 1991). Nouns across languages also have stronger conceptual overlap and are perceived to be more concrete than verbs in general (Bultena et al., 2013; Gentner, 1981; Laxén & Lavaur, 2010; van Hell & de Groot, 1998). In some

languages, members of a word class can be morphologically more complex than the others. For instance, whereas only English verbs can be inflected with different markers to indicate tenses and direction of actions, both Malay nouns and verbs can be inflected with several forms of affixes to form new words. These common irregularities of verbs could cause behavioral uncertainties and delay processing efficiency during language tasks. Unlike past translation norming studies that mostly focused on nouns (e.g., Tokowicz et al., 2002; Wen & van Heuven, 2017), the present translation norms also include words from other word classes (e.g., verbs and adjectives). The only translation norming study that compared translation ambiguity across different word classes (Spanish-English: Prior et al., 2007) revealed that verbs were significantly more translation-ambiguous than nouns in both translation directions. In addition to nouns and verbs, the present study set out to also compare the translation ambiguity of words from other grammatical classes, namely adjectives and class-ambiguous words.

Within-language semantic variability

Previous translation norming studies also reported that within-language semantic variability, or words with multiple related senses within a language, are likely to be translation-ambiguous (Allen & Conklin, 2014; Degani et al., 2016). In addition, the dominant meaning of the source words was more frequently translated in the translation, compared with their subordinate meanings (Degani et al., 2016). For instance, different Malay translation equivalents are possible for the English word “big” because it has two senses, with “*besar*” referring to the size of an object (i.e., large/not small), and “*penting*” referring to the importance of an event (i.e., important). Taking meaning dominance into account, “*besar*” is expected to be the dominant translation for the English word because it carries the dominant (more common) meaning of the word. Employing senses information from official dictionaries, the present study investigated the effects of within-language semantic variability on translation ambiguity, as well as meaning dominance probability in the translations.

Word length and word frequency

Previous translation norming studies have shown that word length and word frequency affect translation ambiguity. However, the effects were inconsistent across studies and dependent on which particular language pairs were involved (Prior et al., 2007; Tseng et al., 2014; Wen & van Heuven, 2017). For example, Prior et al. (2007) found that low-frequency words were more translation-ambiguous than high-frequency words in both Spanish-English and English-Spanish translation. In contrast, the opposite finding was

observed for English-Chinese translation, such that more frequent English words were inclined to have more Chinese translations (Tseng et al., 2014; Wen & van Heuven, 2017). Furthermore, Wen and van Heuven (2017) also found that word frequency affected translation choice, where high-frequency English words tended to be translated into high-frequency Chinese translations.

Word frequency and word length effects are also predicted by bilingual word processing models that account for speeded translation accuracies and latencies. However, it is important to note that translation in speeded tasks is different from ("offline" or unspeeded) translation production studied in our and other translation norming studies. For instance, in Multilink (Dijkstra et al., 2019) word frequency affects the activation of word candidates in online translation production, where more frequent word candidates are activated faster than the less frequent ones. The activation of word candidates is also expected to be stronger and more effortless if they share orthographic similarity with the source words (Dijkstra et al., 2019). If these findings also apply to offline translation tasks where bilinguals are asked to provide the first translation that come into their mind, translation candidates of high word frequency and with similar word length as the source words should be provided as the translation more readily. However, it is unclear yet whether the predictions of speeded responses in Multilink can be extended to offline tasks that likely involve additional task-related processes.

Negative word length effects, where shorter words tended to be more translation-ambiguous, have been observed when English source words were translated into Spanish, but not when translating in the other direction (i.e., Spanish-English; Prior et al., 2007), nor when a different target language was involved (i.e., English-Chinese; Tseng et al., 2014). This finding is surprising because longer English words with seemingly lower word frequencies should be more translation-ambiguous (Sigurd et al., 2004). Unfortunately, Prior et al. (2007) did not offer any explanation for this negative word length effect. Because Prior et al.'s (2007) study was the only study that showed the negative word length effect, it could be the result of the specific language pair and their unique cross-linguistic interactions. To replicate their findings with other language pairs is therefore important.

Individual differences

In addition to the semantic and lexical effects on translation ambiguity, previous translation norming studies also reported individual differences in the translation word choice. Prior et al. (2007) revealed that more proficient L2 speakers were more consistent at producing the dominant translation (i.e., translation choice made by majority of the participants), but only in FT. Interestingly, L2 proficiency

was shown to be correlated with translation accuracy and translation choice in the English-Chinese BT norms when L2 proficiency was estimated by an objective language proficiency measure, LexTALE (Lemhöfer & Broersma, 2012), but not when subjective self-rated proficiency was employed (Wen & van Heuven, 2017). Taken together, previous studies revealed an influence of L2 proficiency on translation (effect sizes ranged from $r = .39$ to $r = .51$), such that bilinguals with higher L2 proficiency are more likely to achieve greater agreement in the translation choice. Expectations of the language proficiency effects on translation performance can differ depending on the translation direction (Laufer & Aviad-Levitzky, 2017; Laufer & Goldstein, 2004; Schwieter & Prior, 2020). For instance, bilinguals who are less proficient in their L2 might have smaller L2 vocabulary size, or face L2 word retrieval difficulties when translating from their L1 to their L2. This could then result in lower translation accuracy in FT. Conversely, when translating from their L2 to L1, less proficient bilinguals might not have complete semantic representation for the L2 source words, leading them to translate the only meaning that they know (which might not be the dominant meaning), and hence, showing lower agreement on the translation choice.

The present Malay-English translation norms gathered correct translations in forward and backward translation directions. For each source word, the index of translation ambiguity (the number of distinct translations that matched with the meanings in dictionaries) and the dominant translations agreed by the majority were identified. In line with previous translation norming studies (Prior et al., 2007; Wen & van Heuven, 2017), bilinguals with higher L2 proficiency were expected to perform better in the translation tasks and more likely to provide a translation that matches the dominant translations provided by the majority than bilinguals with lower L2 proficiency. Furthermore, factors underlying translation ambiguity, translation choice, and translation accuracy were examined. We expected greater translation ambiguity for verbs, adjectives, and word-class-ambiguous items than for nouns (in line with Prior et al., 2007), and we expected source words with a higher number of senses to be more translation-ambiguous, with a higher tendency for the dominant meaning of the source words to be provided as the dominant translation (in line with Allen & Conklin, 2014; Degani et al., 2016). In addition, we explored the relationship between Malay lexical characteristics (word length and word frequency) and translation ambiguity. Translation equivalents were expected to resemble lexical characteristics of the source words, in which frequent words were expected to yield frequent dominant translations and longer words were expected to yield longer dominant translations (in line with Wen & van Heuven, 2017).

Method

Participants

Sixty proficient Malay-English bilinguals were recruited. Half of the participants (11 male and 19 female) performed the FT from Malay to English, and the other half (9 male and 21 female) performed the BT from English to Malay. The recruitment was conducted in two phases, with FT participants recruited before the BT participants. All participants self-identified themselves as Malay-dominant speakers and were students studying at the University of Nottingham Malaysia. Participants were informed that the dominant language is operationalized as the most frequently used language in daily life and the language that participants find themselves to be most proficient in (Treffers-Daller, 2016). All participants met the English proficiency entry requirement of the university (International English Language Testing System [IELTS] Academic overall score 6.5 or equivalent). They received course credits or monetary compensation for their participation.

Participants completed a language background questionnaire to report their language history, as well as their self-rated language proficiency in Malay and English on a scale from 1 (*very poor*) to 7 (*native-like*). All participants were early bilinguals who reported to have learned Malay prior to English. Paired sample *t*-tests revealed that there was no significant difference between participants' self-rated Malay and English proficiency, $t_s \leq 1.69$, $p_s \geq .10$, suggesting that participants were highly fluent in both languages.

In addition to self-rated proficiency, participants' LexTALE (Lemhöfer & Broersma, 2012) scores confirmed that they were intermediate ($n = 7$ with 60–80% accuracy) to advanced ($n = 53$ with >80% accuracy) English users² (Lemhöfer & Broersma, 2012). Importantly, participants in the FT and BT phases were matched in terms of their self-rated Malay and English proficiency as well as LexTALE score, $t_s \leq 1.63$, $p_s \geq .11$. A summary of the language background questionnaire and the LexTALE scores is presented in Table 2.

Stimuli

The present study used the Malay Lexicon Project (Yap et al., 2010) database as the main corpus for lexical information of Malay words, and SUBTLEX-US (Brysbaert et al., 2012; Brysbaert & New, 2009) for lexical

Table 2 Summary of language background questionnaire and LexTALE data

	FT		BT	
	Mean	SD	Mean	SD
Age (years)	21.33	2.12	21.33	3.74
Age exposed to Malay (L1) (years)	0.43	1.02	0.10	0.54
Age exposed to English (L2) (years)	2.17	2.35	2.08	2.24
Self-rated L1 proficiency				
Reading	6.47	0.67	6.47	0.72
Writing	5.67	0.87	5.53	1.18
Listening	6.53	0.72	6.67	0.65
Speaking	6.23	1.12	6.27	1.03
Average	6.23	0.64	6.23	0.74
Self-rated L2 proficiency				
Reading	6.37	0.66	6.20	0.70
Writing	5.70	1.00	5.73	0.77
Listening	6.20	0.75	6.33	0.60
Speaking	5.73	0.89	5.93	0.73
Average	6.00	0.65	6.05	0.56
LexTALE score	87.21	8.57	90.42	6.26

LexTALE (Lemhöfer & Broersma, 2012); Language background questionnaire measured self-rated proficiency on a 7-point scale (1 = very poor, 7 = native-like)

information of English words. Word concreteness ratings for English words³ were taken from Brysbaert et al. (2014), ranging from 1 (*abstract*) to 5 (*concrete*). Zipf scale (van Heuven et al., 2014) was used as the word frequency measure instead of frequency per million words because it offers a more intuitive interpretation for the users. Zipf values given to lexical items range from 1 (*very low frequency*) to 7 (*very high frequency*), with the boundary between low-frequency and high-frequency words lying between 3 and 4 (van Heuven et al., 2014). These categorization labels allow users to identify lexical items based on their frequency categories. Because the Malay Lexicon Project (Yap et al., 2010) provides only frequency count per million words, the Zipf value for each Malay word was calculated using the equation provided in van Heuven et al. (2014).

The 1004 Malay words involved in FT were selected from the 1520 words used in Yap et al.'s (2010) lexical decision and speeded pronunciation experiments. This subset of words included 570 words that originated from 190 morphemic triplets. Each triplet contained a root word (e.g., “*hidup/live*”), its noun-affixed form

² Currently there is not a freely available objective Malay vocabulary measure (e.g., Malay version of LexTALE). Therefore, only English vocabulary knowledge was measured objectively.

³ Because a concreteness rating database was not available for Malay words, only concreteness ratings for English words were made available for the present study.

(e.g., “*penghidupan*/life”) and verb-affixed form (e.g., “*menghidupkan*/give life”). From these 570 words, 498 words were excluded to ensure every word appeared only once in the word list, either in root word form or affixed form. Root words were retained whenever possible, and affixed words with the highest word frequency were kept in cases where root words were absent. In the example given above, the root word “*hidup*” was kept and its affixed forms “*penghidupan*” and “*menghidupkan*” were removed. Subsequently, these words were checked against a Malay-English dictionary (*Kamus Melayu-Inggeris Dewan*, Dewan Bahasa dan Pustaka, 2012) to identify and exclude words that have sole culture-specific (e.g., “*joget*/a type of Malay dance”) or religious meaning (e.g., “*iblis*/devil”) because they do not have a direct translation in English.

The final word set (1004 Malay words) had a mean word frequency (in Zipf value) of 3.94 ($SD = 0.73$) and a mean word length of 6.80 ($SD = 2.59$) in the Malay Lexicon Project (Yap et al., 2010). Word class information obtained from *Kamus Perdana* (Cheng & Lai, 2019) revealed that the word set comprised of 374 nouns, 228 verbs, 116 adjectives, 278 word-class-ambiguous items (e.g., “*aksi*” can be a noun or an adjective), four adverbs, one classifier, one pronoun, one numeral, and one interjection. The Malay words were randomly split into 10 blocks of 100 words (except for one block that had 104 words). Words in the blocks were matched in Zipf value and word length. One sample *t*-test conducted against the average Zipf value ($M = 3.94$, $SD = 0.73$) and the average word length ($M = 6.80$, $SD = 2.59$) revealed no significant differences with individual word block Zipf values and word lengths, $t_s \leq 1.12$, $p_s \geq .27$.

After English translations for the 1004 Malay words were gathered in the FT task, all correct dominant single-word English translations were used as stimuli for the BT task. For Malay words that received no correct translation, or correct dominant translations that have more than one word in FT, the expected single-word English translations from the reference Malay-English dictionary (*Kamus Melayu-Inggeris Dewan*, Dewan Bahasa dan Pustaka, 2012) were used. Malay words with no single-word English translations according to the Malay-English translation norms and the reference dictionary were excluded ($n = 12$). Furthermore, the English translations that appeared more than once in the FT norms were presented only once in BT task (e.g., “level” was the dominant English translation for the Malay words “*darjat*,” “*paras*,” and “*peres*,” and it was presented only once in BT). The final BT stimuli set consisted of 845 English words.

Overall, the English word stimuli had a mean word frequency (Zipf value) of 4.26 ($SD = 0.91$), mean word length of 6.20 ($SD = 2.27$), and mean concreteness

ratings of 3.25 ($SD = 0.97$). To match with the word class classification of the Malay words in the FT task, we utilized the *all part-of-speech*⁴ information for English word class (Brysbaert et al., 2012). There were 123 nouns, 94 verbs, 46 adjectives, 576 word-class-ambiguous items, four adverbs, one determiner, and one interjection. The English words were randomized into nine blocks of 100 words (except for the final block that had 45 words). One sample *t*-test confirmed that words in the blocks were matched in Zipf value, word length and concreteness, $t_s \leq 1.82$, $p_s > .07$.

Procedure

In the FT phase, participants translated four blocks of words every day and completed the translation task in 3 days within a week. The presentation of word blocks within a day and words within each block was randomized. The word stimuli were presented in lowercase, one word at a time, as black characters on a silver background using PsychoPy (Peirce et al., 2019). Participants were required to enter the first translation that came to their mind. They could skip items by pressing the ENTER key if they could not provide a translation. After finishing each block, participants were prompted to take a short break. On the third day of translation, the LexTALE test (Lemhöfer & Broersma, 2012) and language background questionnaire were administered on Qualtrics (<https://www.qualtrics.com>), after participants had completed the final two blocks of words. The same procedure was adopted for BT, except that the BT participants translated three blocks of words a day and completed the 845 translations in three days within a week. The experiment was approved by the Ethics Committee in the School of Psychology at the University of Nottingham Malaysia. Written consent was acquired from participants before data collection started.

Scoring

Translation accuracy of participants was determined by comparing their translations against the expected translations provided by the Malay-English and English-Malay dictionaries. For the expected Malay-English translations,

⁴ There are two types of word class classification available for the English words (Brysbaert et al., 2012), namely *All part-of-speech* (All_PoS) and *dominant part-of-speech* (dom_PoS). For instance, the word “float” was observed as a noun (105 times) and a verb (276 times) in Brysbaert et al. (2012). Consequently, “float” is classified as word-class-ambiguous under All_PoS because more than one word class is associated with the word. In addition, “verb” is listed as the dom_PoS of “float” in view of its higher frequency of occurrence when compared with the noun form.

Kamus Melayu-Inggeris Dewan (Dewan Bahasa dan Pustaka, 2012) was used as the primary reference source, and *Kamus Perdana* (Cheng & Lai, 2019) was used as the secondary reference. For English-Malay translation, *Kamus Dwibahasa* (Dewan Bahasa dan Pustaka, 2002) was chosen as the primary reference, while the Oxford English-English-Malay Dictionary (Oxford University Press & Oxford Fajar, 2018) was used as the secondary reference. The primary reference dictionaries are widely used by Malay language users as the authoritative dictionary in Malaysia, because they were published by the Institute of Language and Literature, the official government body that monitors Malay language development and usage in the country.

Grammatical affixations that did not transform the word class of a word, such as third person singular “-s” and plural “-s” in English, were collated to its root word and accepted as correct responses if they matched the expected translations. Spelling errors were corrected and accepted on the condition that the errors did not result in another real word in the target language. Two proficient Malay-English coders further examined the translations that did not match with the expected dictionary translations. Synonyms of the expected dictionary translations and colloquial meanings provided were further examined and coded as correct responses only upon agreement achieved from both coders. Some judgment criteria used to accept exceptional translations included: (a) the translations shared similar meaning as the expected translations provided by the dictionaries and both could be used interchangeably (e.g., “*siap*” was accepted as a synonym for “*habis*” and “*selesai*” because both carry the meaning of “finish”), and (b) translations matched with the word choice used colloquially in daily conversations (e.g., “*orang*” as a translation for “human”). Responses that described the meaning of the source words instead of being the direct translation were rejected (e.g., “hairless” for “*botak/bald*”).

Results

The obtained Malay-English bidirectional translation norms and the translation ambiguity index are described in this section. The semantic and lexical information of the words are provided in the supplementary material (see [Database section](#)). The roles of language proficiency, source word frequency, and word length in influencing translation accuracy were explored. Finally, source words that received at least one correct translation were further investigated to determine the roles of word class, within-language semantic variability, word frequency, and word length in translation ambiguity.

Table 3 Proportion of Malay and English words according to their translation ambiguity for the Malay-English and English-Malay translation norms

Type of translation pair	Number of source words	Proportion (%)
Malay-English FT norms ($N = 1004$)		
Translation-unambiguous	325	32.4
Translation-ambiguous	635	63.3
No correct translation	44	4.4
English-Malay BT norms ($N = 845$)		
Translation-unambiguous	180	21.3
Translation-ambiguous	659	78.0
No correct translation	6	0.7

Translation norms

Malay-English forward translation (FT)

Translation accuracy The FT task resulted in a total of 27,130 English translations (90.1%) and 2990 omitted responses (9.9%). A total of 18,378 translations (67.7%) were correct responses. Of the 1004 Malay words, 64.2% (645 words) were correctly translated by at least 50% of the participants, 31.4% (315 words) received correct translations from at least one participant, and 4.0% (44 words) of the stimuli received no correct translations.

Translation ambiguity Translation ambiguity was determined by the number of possible translations provided for each source word. When a source word yielded only one unique correct translation, it was considered translation-unambiguous, and a source word was considered translation-ambiguous when it resulted in more than one correct translation. In the FT norms, the number of possible translations provided for the Malay words ranged from zero to eight. Of the 1004 Malay words, the proportion of translation-ambiguous words were 63.3% (see [Table 3](#)).

Dominant translations For translation-unambiguous words, the unique translation equivalents are the dominant translations. The dominant translations for the translation-ambiguous words were identified by selecting the correct translations that were most frequently provided by the participants. In cases where the translation-ambiguous word had more than one dominant translation, the translation that matched with the dominant meaning from the primary reference dictionary was selected. The results revealed that the dominant English translations of the FT norms covered a wide range of word lengths ($M = 6.20$, $SD = 2.65$, minimum = 2, maximum = 23), word frequencies (Zipf value) ($M = 4.37$, $SD = 0.90$, minimum = 1.59, maximum = 7.62),

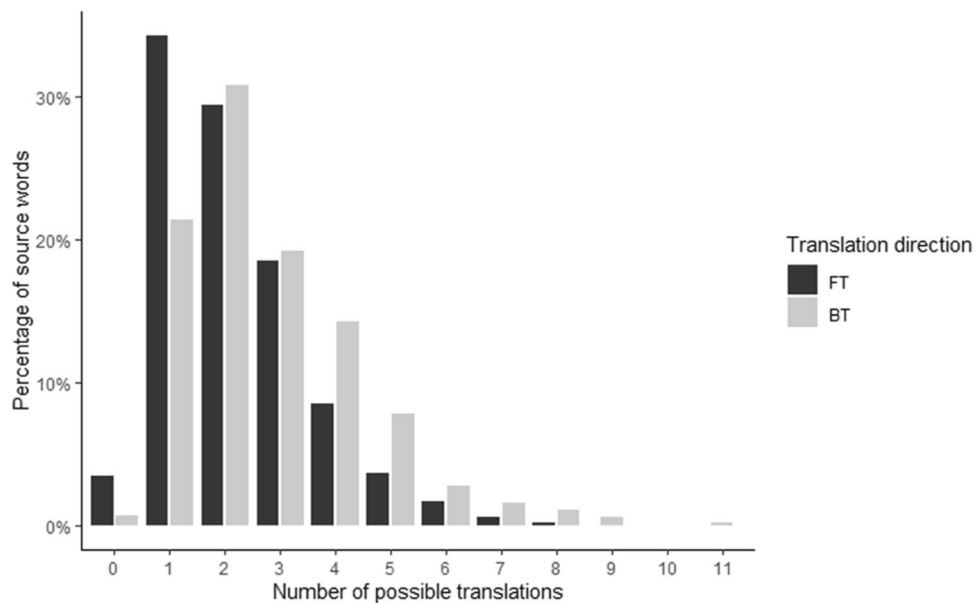


Fig. 1 Distributions of the 845 Malay and English words according to their number of possible translations for the Malay-English FT and English-Malay BT norms

and concreteness ratings ($M = 3.25$, $SD = 0.95$, minimum = 1.19, maximum = 5).

English-Malay backward translation (BT)

Translation accuracy The BT task resulted in 23,813 Malay translations (93.9%) and 1537 omitted responses (6.1%). Of the Malay translations, 20,454 responses were correct translations (85.9%). Overall, 88.4% (747 words) of the 845 English words received correct translations from at least 50% of the participants, 10.9% (92 words) were translated correctly by at least one participant, and six words (0.7%) received no correct translations from the participants.

Translation ambiguity The number of possible translations in the BT norms ranged from 0 to 11, with 78.0% of the 845 English words being translation-ambiguous (see Table 3). The translation ambiguity of BT was compared against the FT norms using the same set of 845 source words used in both translation directions. In FT, 34.2% (289 words) of these words were translation-unambiguous while 62.4% (527 words) were translation-ambiguous. The numerical percentages suggest that BT resulted in more translation ambiguity than FT (see Fig. 1).

Dominant translations In the BT norms, the dominant Malay translations had a mean word length of 6.80 ($SD = 2.49$, minimum = 3, maximum = 19), and mean word frequency (Zipf value) of 4.17 ($SD = 0.75$, minimum = 2.83, maximum = 6.63).

Translation accuracy

This set of analyses assessed the factors that affect bilinguals' translation accuracy. The role of language proficiency was investigated at the participant level, followed by word length and word frequency analyses at both participant and item levels. Dominant translation scores were determined based on the percentage of correct dominant translations each participant provided (participant level) or gathered for each source word (item level), and translation accuracy scores were defined as the percentage of correct translations made in total, independent of whether the translation was dominant or nondominant.

Language proficiency

At the participant level, the influence of language proficiency on the translation performance of proficient Malay-English bilinguals was investigated. Participants' Malay language proficiency was estimated using self-ratings and an English vocabulary test (LexTALE, Lemhöfer & Broersma, 2012), and self-ratings were used to obtain subjective and objective measures of English proficiency.

In FT, Spearman's rho test revealed a statistically significant moderate, positive correlation between self-rated L1 Malay proficiency and participants' dominant translation scores, as well as translation accuracy scores (see Table 4). Participants who perceived themselves as having higher Malay proficiency provided more dominant

Table 4 Spearman’s rho (r_s) for language proficiency and translation accuracy

Variable	Dominant translation score		Translation accuracy score
	r_s	p	
FT ($N = 30$)			
LexTALE score	.09		.13
Self-rated L1 proficiency	.49**		.38*
Self-rated L2 proficiency	.25		.31
BT ($N = 30$)			
LexTALE score	-.08		.08
Self-rated L1 proficiency	.13		.10
Self-rated L2 proficiency	-.04		.09

*Correlation is significant at the 0.05 level (two-tailed)
 **Correlation is significant at the 0.01 level (two-tailed)

Table 5 Proportion of source words according to lexical characteristics

Variable	Forward translation		Backward translation	
	N	Percentage	N	Percentage
Word frequency				
High	428	42.63%	525	62.13%
Low	576	57.37%	320	37.87%
Word length				
Long	300	29.88%	209	24.73%
Short	704	70.12%	636	75.27%

translations and more correct translations. However, L2 proficiency measures (i.e., LexTALE and self-rated English proficiency) did not correlate with these translation scores, $ps > .09$. Interestingly, none of the language proficiency

measures in the BT group correlated with participants’ translation scores, $ps > .50$.

Word length and word frequency

Shapiro-Wilk tests of normality were conducted and revealed non-normal distribution of the translation scores ($ps < .01$); therefore, nonparametric tests were conducted for the subsequent analyses. At the participant level, Wilcoxon signed-rank tests were conducted to compare the translation accuracy of high- and low-frequency words as well as long and short words in both translation directions. Source words with a Zipf value of 4 and above were considered high-frequency words, and source words with Zipf value below 4 were considered low-frequency words. At the same time, the source words from each direction were split into two groups around the mean word length (mean word length for FT = 7.00; BT = 6.21). Table 5 summarizes the proportion of source words in each lexical group.

Wilcoxon signed-rank tests revealed that the translation accuracy of the high-frequency words was significantly higher than that of low-frequency words in both translation directions, $ps < .001$. Also, the translation accuracy for shorter words were significantly higher than that of longer words, $ps \leq .007$ (see Table 6). Overall, participants demonstrated higher translation accuracy and were more likely to provide dominant translation for high-frequency and short words, in contrast to low-frequency and long words.

Spearman’s rho was conducted to assess the subsequent relationships between source words’ lexical characteristics and translation performance. In both translation directions, source words’ frequency positively correlated with dominant translation and translation accuracy scores, while source words’ length negatively correlated

Table 6 Wilcoxon signed-rank tests to compare translation accuracy by lexical characteristics

Variable	Dominant translation score			Translation accuracy score		
	T	z	Effect size (r)	T	z	Effect size (r)
FT ($N = 30$)						
Word frequency	465	-4.78***	.87	465	-4.78***	.87
Word length	102	-2.68**	.49	87	-2.99**	.55
BT ($N = 30$)						
High vs. low frequency words	465	-4.78***	.87	465	-4.78***	.87
Long vs. short words	23	-4.31***	.79	465	-4.78***	.87

Effect size in r was converted from z score (Clark-Carter, 2019)

**Difference was significant at the <.01 level (two-tailed)
 ***Difference was significant at <.001 level (two-tailed)

Table 7 Spearman’s rho (r_s) for source words’ lexical characteristics and translation accuracy

Variable	Dominant translation score	Translation accuracy score
1004 Malay source words		
Word frequency	.32***	.39***
Word length	-.15***	-.17***
845 English source words		
Word frequency	.42***	.62***
Word length	-.30***	-.43***

***Correlation is significant at the 0.001 level (two-tailed)

Table 8 Translation ambiguity index according to word class in FT

Word class	<i>N</i>	TA (%)
Nouns	374	54.8
Verbs	228	70.2
Adjectives	116	68.1
Word-class-ambiguous	278	67.3

996 words retrieved from the 1004 Malay source words. *TA* translation ambiguity

with both dominant translation and translation accuracy scores, $ps < .001$ (see Table 7).

Translation ambiguity

Word class

To investigate if translation ambiguity was affected by word class, source words from each translation direction were grouped by four distinct word classes: nouns, verbs, adjectives, and word-class-ambiguous items. Source words that belong to other word classes (i.e., adverb, classifier,

Table 10 Translation ambiguity index according to word class in English

Word class	<i>N</i>	TA (%)
Nouns	123	63.4
Verbs	94	90.4
Adjectives	46	71.7
Word-class-ambiguous	576	79.5

839 words retrieved from the 845 English source words. *TA* translation ambiguity

determiner, interjection, numeral, and pronoun) were excluded from this analysis because the sample size for each of these word classes was too small to generate meaningful comparisons (see Table 8 for word class distribution). A Kruskal–Wallis analysis of variance (ANOVA) indicated that there are significant differences across translation ambiguity of nouns, verbs, adjectives, and word-class-ambiguous items, H (corrected for ties) = 27.85, $df = 3$, $N = 952$, $p < .001$, Cohen’s $f = .17$. Separate Mann-Whitney U post hoc tests revealed that translation ambiguity for nouns is significantly lower than that of verbs, adjectives, and word-class-ambiguous items, $ps < .005$. There is no significant difference across the translation ambiguity of verbs, adjectives, and word-class-ambiguous items, $ps \geq .18$. Table 9 presents the post hoc tests results.

Similar word class analyses were conducted on the 845 English words in the BT task (see Table 10 for word class distribution). Kruskal-Wallis ANOVA confirmed that there are significant differences across translation ambiguity of nouns, verbs, adjectives, and word-class-ambiguous items, H (corrected for ties) = 36.89, $df = 3$, $N = 833$, $p < .001$, Cohen’s $f = .22$. Mann-Whitney U post hoc tests revealed that verbs are significantly more translation-ambiguous than nouns, adjectives, and word-class-ambiguous items, $ps \leq .02$. At the same time, adjectives and word-class-ambiguous items are significantly more translation-ambiguous than

Table 9 Post hoc Mann-Whitney U tests to compare translation ambiguity across word class in FT

Word class	<i>U</i>	z (corrected for ties)	<i>p</i>	Effect size (<i>r</i>)
Nouns vs. verbs	29,661	-4.80***	.000	.20
Nouns vs. adjectives	16,574	-2.84**	.004	.13
Nouns vs. word-class-ambiguous	39,372	-3.74***	.000	.15
Verbs vs. adjectives	11,570	-.93	.351	
Verbs vs. word-class-ambiguous	27,227	-1.33	.184	
Adjectives vs. word-class-ambiguous	15,066	-.08	.936	

Effect size in r was converted from z score (Clark-Carter, 2019)

**Difference was significant at the <.01 level (two-tailed)

***Difference was significant at <.001 level (two-tailed)

Table 11 Post hoc Mann-Whitney U tests to compare translation ambiguity across word class in BT

Word class	<i>U</i>	<i>z</i> (corrected for ties)	<i>p</i>	Effect size (<i>r</i>)
Nouns vs. verbs	3104	−5.80***	.000	.40
Nouns vs. adjectives	2255	−1.97*	.048	.29
Nouns vs. word-class-ambiguous	24,836	−5.04***	.000	.19
Verbs vs. adjectives	1620	−2.38*	.017	.20
Verbs vs. word-class-ambiguous	22,617	−2.40*	.017	.09
Adjectives vs. word-class-ambiguous	11,936	−1.09	.275	

Effect size in *r* was converted from *z* score (Clark-Carter, 2019)

*Difference was significant at the 0.05 level (two-tailed)

***Difference was significant at < .001 level (two-tailed)

nouns, $p_s \leq .05$. There is no significant difference between translation ambiguity of adjectives and word-class-ambiguous items, $p = .28$ (see Table 11 for summary).

Within-language semantic variability

The relationship between within-language semantic variability and translation ambiguity was further investigated. Semantic variability was defined by the number of senses (meaning) a word has according to the primary reference dictionary. All possible meanings associated with a particular word form were summed up, including meanings of homonyms (words that share the same form but carry distinct meanings; e.g., “*guna*” was considered to have three senses, namely the two related senses “use” and “role,” as well as the [third] unrelated sense “spell”). Nineteen Malay words from FT and eight English words from BT were excluded from the analysis because their number of senses were not provided by the primary reference dictionary. Nonparametric Spearman’s rho tests indicated statistically significant positive correlations between the number of senses of words and number of possible translations in FT, $r_s = .23$, $p < .001$, two-tailed, $N = 951$, and BT, $r_s = .25$, $p < .001$, two-tailed, $N = 833$. Words with higher semantic variability tend to have a higher number of possible translations.

Word length and word frequency

Spearman’s rho conducted indicated weak, yet statistically significant positive correlation between Malay word length and the number of translations provided, $r_s = .08$, $p < .05$, two-tailed, $N = 960$. Similarly, Malay word frequency also correlated weakly and positively with the number of translations provided, $r_s = .09$, $p < .01$, two-tailed, $N = 960$. Malay words with longer strings and of higher frequency were more likely to yield more translations. The same correlation

analyses were conducted for the English words in BT; however, only word length showed a trend towards a positive correlation with the number of translations provided, $r_s = .06$, $p = .06$, two-tailed, $N = 839$.

Translation word choice

The next analyses investigated the effects of meaning dominance, word frequency, and word length on translation word choice. Only translation pairs for which at least 50% of the participants provided the dominant translations were further examined to ensure that the translations under investigation truly represent the translation choice of the majority of the participants.

Meaning dominance

This section focuses on the roles of semantic and lexical characteristics in bilinguals’ translation word choice. The probability of a meaning dominance effect, defined by the likelihood for the dominant meaning of a source word (as indicated by the primary reference dictionary) to also be a dominant translation, was first examined. For instance, the effect was demonstrated when most of the participants translated the English word “direction” into its dominant meaning “*arah*,” rather than its subdominant meaning “*arahan*.”

Of the 502 Malay translation-ambiguous words in FT, 405 Malay source words had their dominant meaning translated by the majority of the participants, and 97 words had their subdominant meaning translated by the majority. A chi-square test for goodness of fit was conducted to assess if the dominant meaning of source words were more frequently translated than the subdominant meaning. The chi-square test revealed that the frequency of the dominant meaning being translated into the dominant translation was significantly higher than that of the

subdominant meaning, $\chi^2(1, N = 502) = 188.97, p < .001$ (Cohen's $w = 0.61$).

For the 576 English translation-ambiguous words in BT, 341 had their dominant meaning translated by the majority of the participants, and 235 had their subdominant meaning translated by the majority. The dominant meanings of English words, when compared with subdominant meanings, were also more frequently translated into the dominant Malay translations, $\chi^2(1, N = 576) = 19.51, p < .001$ (Cohen's $w = 0.18$).

Word length and word frequency

The present study also examined the relationship between word length of the source words and their dominant translations. In FT, Spearman's rho test revealed a relationship between the word length of Malay source words and English translations, $r_s = .31, p < .001$, two-tailed, $N = 502$, indicating that longer Malay words were translated into longer English words. Similarly, there was also a statistically significant correlation between Malay and translated English word frequency, $r_s = .41, p < .001$, two-tailed, $N = 502$, indicating that more frequent Malay words were translated into more frequent English words.

In BT, significant correlation was also found between the word length of English source words and Malay translations, $r_s = .49, p < .001$, two-tailed, $N = 576$. The positive correlation indicates that longer English words were translated into longer Malay words. Before proceeding to the word frequency correlational analysis, an additional 39 English-Malay translation pairs were excluded because the word frequency information was not available for the Malay translations. Spearman's rho indicated a moderate yet statistically significant positive correlation between the word frequency of English source words and Malay translations, $r_s = .46, p < .001$, two-tailed, $N = 537$. Thus, more frequent English words were translated into more frequent Malay translations.

General discussion

The present study aimed at creating the first freely available Malay and English translation norms with proficient Malay-English bilinguals. As a result, a database of Malay-English and English-Malay translation norms for 1004 Malay words and 845 English words is formed. The norms predominantly consist of nouns, verbs, adjectives, and class-ambiguous words that span across a range of semantic variability, word frequencies, and word lengths. We also examined the degree of Malay-English and English-Malay translation ambiguity, and their relationship with the semantic and lexical characteristics of the source words. In addition, factors affecting bilinguals' translation word choice and accuracy were also examined.

Translation ambiguity

The Malay-English FT norms revealed a high proportion of translation-ambiguous Malay words (63.3%). This proportion is higher than that for other translation norms that also involved English as the target translation language (e.g., Dutch-English: 25.3%, Tokowicz et al., 2002; Spanish-English: 48.2%, Prior et al., 2007). The exceptionally low translation ambiguity reported in the Dutch-English norms is likely an underestimation of the proportion of translation-ambiguous words, because the stimuli were chosen and assumed to be translation-unambiguous by previous research (Schwieter & Prior, 2020). In contrast, the Malay source words used in this study were not selected based on being translation-unambiguous. Similarly, the English-Malay BT norms also revealed high translation ambiguity between the two languages (78.0%), which was higher than other translation norms (e.g., Dutch-English: 30.4%, Tokowicz et al., 2002; Spanish-English: 58.5%, Prior et al., 2007), even when compared with the English-Chinese translation norms in which the two languages are differently scripted (67.3% in Tseng et al., 2014; 71.2% in Wen & van Heuven, 2017).

We attributed the high translation ambiguity observed in the present study to the conceptual mapping differences between Malay and English. Malay as an Austronesian language and English as an Indo-European language come from two different language families. In comparison to language pairs that belong to the same language family group (e.g., Dutch and English, which are both varieties of West Germanic languages of the Indo-European language family), Malay and English are likely to have relatively more distinct concepts for words (Schwieter & Prior, 2020; Tseng et al., 2014). Translation ambiguity could emerge when a source language has a wide conceptual space for words (e.g., “thick” for both solid and liquid), whereas the target language provides finer distinctions for the concepts (e.g., “tebal” for solid and “pekat” for liquid). In such case, a single concept carried by a source word can result in two different translations in the target language.

On top of that, we also found translation ambiguity of English-Malay BT norms to be higher than the Malay-English FT norms. This finding is consistent with past translation norming studies (Prior et al., 2007; Tokowicz et al., 2002), in which translation from English as a source language to another target language (e.g., English-Dutch) always resulted in higher translation ambiguity than translation in the other direction (e.g., Dutch-English). Because the higher translation ambiguity has been observed with English as the source language, it is likely that the language-specific properties of English, such as greater within-language semantic variability (Degani et al., 2016), contributed to the higher number of possible translations in the target languages. In addition,

the morphological mapping differences between English and Malay could have added to the variability in translation too, with English being morphologically less complex than Malay. As an example, the English word “need” can be translated into different forms of the Malay word “*perlu*,” including the root word “*perlu*,” verb-affixed form “*memerlukan*,” and noun-affixed form “*keperluan*.”

The higher translation ambiguity and translation accuracy observed in BT compared with FT could also be due to the L2-L1 translation direction because bilinguals were translating from their less dominant language to their more dominant language in BT.⁵ These bilinguals were likely to be more proficient in Malay than English, even though their self-rated language proficiency for the two languages did not differ significantly. If we assume a larger vocabulary size in the bilinguals’ L1 (Rahman et al., 2018), more translation choices would be available for translation equivalents in L1, than when translation was conducted in the other direction. However, as far as we are aware of, all existing BT norms use English as the source language; hence, it is not possible to pinpoint the higher translation ambiguity in BT to language-specific properties (e.g., polysemous English) or a language-universal factor (e.g., better vocabulary knowledge in the target language). Thus, future BT studies could consider (a) employing a source language other than English to provide additional evidence regarding the role of language-specific characteristics of the source language in translation ambiguity (Schwieter & Prior, 2020), and (b) recruiting bilinguals who speak English as their L1 or dominant language to perform the same translation task. If the source language of a BT task has a narrower conceptual space (Schwieter & Prior, 2020) than the target language, and yet still results in higher translation ambiguity than the FT task, the L2-L1 effect explanation on translation ambiguity (language-universal factor) would be supported. If dominant or L1 English speakers performing in an English-Malay translation task (L1-L2 translation) show higher translation ambiguity than the Malay-English translation task, it would suggest that the translation ambiguity observed in the present study is induced by language-specific characteristics of the English language.

With respect to lexical factors affecting translation ambiguity, the present study replicated the findings from Prior et al. (2007) by showing that verbs were more translation-ambiguous than nouns in both translation directions. In addition, adjectives and word-class-ambiguous

items were at least as translation-ambiguous as verbs. Because verbs, adjectives, and word class-ambiguous items were significantly more translation-ambiguous than nouns, it is likely that the higher translation ambiguity found in the present study than in other translation norming studies involving mostly nouns (e.g., Allen & Conklin, 2014; Prior et al., 2007; Tokowicz et al., 2002) could be partly attributed to the additional word classes used. For instance, when the translation ambiguity of words from different word classes were taken into account, English words were more translation-ambiguous with Malay (78.0% in the present study) than Chinese (67.3% in Tseng et al., 2014; 71.2% in Wen & van Heuven, 2017). However, when only the translation ambiguity of nouns was considered, the translation ambiguity index of English-Malay translation became less ambiguous (63.4%) than the English-Chinese translation.

The present findings also replicated the positive relationship between within-language semantic variability and translation ambiguity in both translation directions. In the past, English words with more senses (high semantic variability) tend to produce a greater number of possible translations in Dutch, German, Spanish, and Hebrew (Degani et al., 2016). Although the present study investigated a different language pair in two translation directions, similar effects were found. Allen and Conklin (2014) also reported in their Japanese-English translation norming study a similar effect of semantic variability in both translation directions.

In addition to the impact of the number of senses on translations, longer and frequent Malay words resulted in more translations in English. Surprisingly, only English word length showed a trend towards a positive correlation with translation ambiguity in BT. This difference between FT and BT could be attributed to the difference in source-target language pairing. Previous research found that word frequency and word length effects are inconsistent and sensitive to the source and target language identity. For instance, for Spanish and English, word length effects on translation ambiguity became negligible when the source-target language was changed (i.e., from English-Spanish to Spanish-English) (Prior et al., 2007). Furthermore, the direction of word frequency effects could change when the source language remained the same and only the target language was substituted (e.g., negative correlations for English-Spanish translations but positive correlations for English-Chinese translations; Prior et al., 2007; Tseng et al., 2014; Wen & van Heuven, 2017). In sum, it appears that the relationship between two languages could differ according to language-specific properties of the language pair in question. Because different English word sets were employed across these studies, it is difficult to pinpoint

⁵ We also conducted an additional analysis to investigate the diversity of language use among these bilinguals and how it could potentially affect their dominant translation scores and translation accuracy scores. This language entropy analysis can be found in page 1 of the Supplementary Analyses document.

which factor contributed to the discrepancy. Future studies should consider using same set of source words for meaningful cross-linguistic and cross-study comparisons.

Translation choice

The present study also replicated the meaning dominance effect whereby the dominant meaning of source words provided in the primary dictionary was more likely to become the dominant translation (Degani et al., 2016). Although the dictionary we used provides a brief statement that the meanings of the vocabulary items are arranged according to the commonality of usage, to our knowledge, there is no empirical evidence yet that supports the dominance of the meanings first listed in it. The present findings provide the first preliminary evidence as such. The effect suggests consideration of the semantic overlapping between source words and translations is common during translation (Laxén & Lavour, 2010). Although previous studies only investigated meaning dominance effects in BT, the findings from our study provide empirical evidence that meaning dominance effects occur in both translation directions.

Besides the consideration of meanings, further correlational analyses also revealed that in both translation directions longer source words were translated into longer words, and more frequent source words were translated into words with higher frequencies. These findings are in line with previous translation norming studies that employed different language pairs (Allen & Conklin, 2014; Wen & van Heuven, 2017), indicating that lexical characteristics of the source words have an influence on translation choice for any language pair and translation direction.

Translation accuracy and language proficiency

Only in the FT task, participants who rated themselves with higher Malay (L1) proficiency were more likely to provide correct and dominant translations. Surprisingly, this correlation was not found in BT. One possible explanation is that the overall word frequency of the Malay and English source words differed across the two translation directions. For the 845 source words shared by both translation directions, the mean word frequency (in Zipf values) of Malay source words in FT ($M = 3.98$, $SD = 0.73$) was significantly lower than that of English source words in BT ($M = 4.27$, $SD = 0.91$), $t(1612.88) = -7.25$, $p < .001$. A closer look to the proportion of high- and low-frequency words involved also revealed that more than half of the FT source words (57.37%) were low-frequency words with Zipf value less than 4, while only 37.87% of the BT source words were of low frequency (see Table 5). This high number of low-frequency words in FT could be a potential confound of the L1

proficiency effect observed, whereby high proficiency and vocabulary knowledge in L1 Malay became an important factor for participants to perform well in FT. To investigate whether the L1 proficiency effect on translation accuracy remained when word frequency and word length in both tasks were matched, an additional analysis was conducted using a subset of 709 words that were carefully matched. These results revealed again a significant effect of L1 language proficiency on FT translation performance (see page 2 in the Supplementary Analyses document for more details).

To the best of our knowledge, no past translation norming study has investigated and revealed the impact of L1 proficiency on translation performance, probably because bilinguals' L1 proficiency was always assumed to be homogeneous as a group. The present study provides preliminary evidence to point out that even though most bilingual studies assumed "native-speaker" proficiency (Izura et al., 2014), there could still be potential variation in L1 proficiency within a rather homogeneous group, and it could potentially influence L1 speakers' language performance. Future study should consider extending the investigation of bilingual word processing to also include measures for L1, to account for possible language proficiency effects.

Surprisingly, in contrast to previous research, there was no correlation between L2 proficiency (indicated by objective LexTALE scores and subjective self-ratings) and translation word choice in FT and BT. Prior et al. (2007) found that the Spanish L1 group with higher L2 proficiency were more likely to produce dominant translations, but only in FT. The impact of L2 proficiency was also found in the English-Chinese BT study (Wen & van Heuven, 2017). However, it is important to note that these studies utilized different sets of stimuli and proficiency measures, which complicates direct comparison of findings across studies. Again, future studies should consider using objective L2 proficiency measure (e.g., LexTALE) and similar sets of source words for meaningful cross-study comparisons.

We suspect our bilinguals' high L2 competence to be the reason why we did not find a relationship between L2 proficiency and translation accuracy. Most past translation norming studies recruited unbalanced bilinguals (e.g., Prior et al., 2007; Wen & van Heuven, 2017), who reported to have learned L2 in school and were only later immersed in an L2 environment during tertiary education. The present study, however, involved highly proficient bilinguals who learned the L2 before attending school (< 7 years old). Most of them had rated themselves to be equally proficient in Malay and English too, despite reporting Malay as their dominant language (cf. Duyck & Brysbaert, 2004).

Lastly, the present study also demonstrates that source words with higher word frequency and shorter word length were more likely to be translated correctly in both translation directions. These words seem to be easier items for the translation tasks.

Correspondingly, Wen and van Heuven (2017) also found that their Mandarin-English bilinguals were more reliable in providing the dominant translations for high-frequency English words. As pointed out by one of the reviewers, longer Malay words are likely to be words with affixations, which may or may not share the same word class with the root words (e.g., the root word “*hidup/live*” and one of its affixed forms “*menghidupkan/* give life” are verbs; while another affixed form “*penghidupan/* life” is a noun). The uncertainties in word class of these longer Malay words with affixations could result in a higher chance of making translation mistakes, because participants have to first accurately identify the right meaning and word class form of the affixed words, before performing the translation. Taken together, our study provides evidence that word frequency and word length influence translation accuracy and hence can be used to estimate translation stimuli difficulty level for highly proficient Malay-English bilinguals. Whether this finding can be generalized to other types of bilinguals with varying L2 proficiency remains to be tested.

Conclusion

The present study created the Malay-English and English-Malay translation norms through forward and backward translation tasks. The present translation norms are the first norms collected from balanced bilinguals. Our data analyses showed high prevalence of translation ambiguity between the Malay and English language and replicated some lexical characteristics and semantic variability effects on translation ambiguity. Although attempts to explain the inconsistency in these effects met with challenges due to the inconsistency in word stimuli used in past translation norming studies, we suggest standardizing future norming items to help setting apart the language-specific and language-universal factors towards translation ambiguity.

The present translation norms provide the first database for researchers conducting language research with Malay-English bilinguals. Together with lexical and semantic information of the source and target words, these norms could be good references to aid stimuli selection for future experimental studies (e.g., Jouravlev & Jared, 2020) and computer simulations (e.g., Dijkstra et al., 2019).

Database

The Malay-English and English-Malay translation norms are available at https://osf.io/cnkjq/?view_only=54b5521c763241faa18a5b70963f2550. Two files are available for each translation direction (FT/BT). Translation Norms A of each translation direction contains 17 columns:

1. Word number (nr)
2. Source word (word)
3. Number of letters of the source word (word_length)
4. Word frequency of the source word (word_Zipf)
5. Number of senses of the source word (senses)
6. Concreteness rating of the source word (BT only; concreteness)
7. Word class of the source word (FT: pos_Mal, BT: dom_pos & all_pos)
8. Dominant translation (dom_trans)
9. Number of letters of the dominant translation (dom_trans_word_length)
10. Word frequency of the dominant translation (dom_trans_Zipf)
11. Concreteness rating of the dominant translation (FT only; concreteness_Eng)
12. Word class of the dominant translation (FT: dom_trans_dom_pos & dom_trans_all_pos, BT: dom_trans_pos_Mal)
13. The percentage of the dominant translation (percent_dom_trans)
14. The percentage of correct translations (percent_correct)
15. The percentage of incorrect translations (percent_incorrect)
16. The percentage of omitted responses (percent_omitted)
17. The number of correct possible translations (num_corr_resp)

Translation Norms B contains 19–25 columns, depending on the number of correct possible translations in each translation direction:

1. Word number (nr)
2. Source word (word)
3. The number of correct possible translations (num_corr_resp)
4. The correct possible translations and number of participants who provided the translations (e.g., trans1_count is the number of participants who provided trans1)

Note. The Malay word frequency (Zipf value) were converted from frequency-per-million-word information extracted from Yap et al. (2010). The English word frequency (Zipf value) and word class information were taken from SUBTLEX-US (Brysbaert et al., 2012; Brysbaert & New, 2009). The concreteness ratings for English words were retrieved from Brysbaert et al. (2014).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-022-01977-3>.

Code availability Not applicable.

Funding This work was funded by Fundamental Research Grant Scheme (Grant number: FRGS/1/2019/SS05/UNIM/02/2).

Data availability The datasets generated during the current study are available in the Open Science Framework repository, https://osf.io/cnkjq/?view_only=54b5521c763241faa18a5b70963f2550.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval This study has been reviewed and approved by the University of Nottingham Malaysia Science and Engineering Research Ethics Committee (Application Identification Number: LST050220).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Not applicable.

References

- Allen, D., & Conklin, K. (2014). Cross-linguistic similarity norms for Japanese–English translation equivalents. *Behavior Research Methods*, 46(2), 540–563. <https://doi.org/10.3758/s13428-013-0389-z>
- Basnight-Brown, D., Kazanas, S., & Altarriba, J. (2018). Translation ambiguity in Mandarin-English bilinguals: Translation production differences in concrete, abstract, and emotion words. *Linguistic Approaches to Bilingualism*, 10(4), 559–586. <https://doi.org/10.1075/lab.17037.bas>
- Bracken, J., Degani, T., Eddington, C., & Tokowicz, N. (2017). Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words. *Bilingualism (Cambridge, England)*, 20(4), 783–794. <https://doi.org/10.1017/S1366728916000274>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, 44(4), 991–997. <https://doi.org/10.3758/s13428-012-0190-4>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Bultena, S., Dijkstra, T., & van Hell, J. G. (2013). Cognate and word class ambiguity effects in noun and verb processing. *Language and Cognitive Processes*, 28(9), 1350–1377. <https://doi.org/10.1080/01690965.2012.718353>
- Cheng, S. H., & Lai, C. (Eds.). (2019). *Kamus perdana: Bahasa Melayu-bahasa Cina-bahasa Inggeris* (Edisi Keempat ed.). United Publishing House (M) Sdn. Bhd.
- Clark-Carter, D. (2019). *Quantitative psychological research: the complete student's companion* (4th ed.). Routledge.
- Degani, T., & Tokowicz, N. (2013). Cross-language influences: translation status affects intraword sense relatedness. *Memory & Cognition*, 41(7), 1046–1064. <https://doi.org/10.3758/s13421-013-0322-9>
- Degani, T., Tseng, A. M., & Tokowicz, N. (2014). Together or apart: Learning of translation-ambiguous words. *Bilingualism (Cambridge, England)*, 17(4), 749–765. <https://doi.org/10.1017/S1366728913000837>
- Degani, T., Prior, A., Eddington, C. M., da Luz, A., Fontes, A. B., & Tokowicz, N. (2016). Determinants of translation ambiguity. *Linguistic Approaches to Bilingualism*, 6(3), 290–307. <https://doi.org/10.1075/lab.14013.deg>
- Dijkstra, A. F. J., Wahl, A. R., Buytenhuijs, F., van Halem, N., Al-jibouri, Z. S. H., de Korte, M., & Rekke, S. T. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism (Cambridge, England)*, 22(4), 657–679. <https://doi.org/10.1017/S1366728918000287>
- Duyck, W., & Brysbaert, M. (2004). Forward and Backward Number Translation Requires Conceptual Mediation in Both Balanced and Unbalanced Bilinguals. *Journal of Experimental Psychology. Human Perception and Performance*, 30(5), 889–906. <https://doi.org/10.1037/0096-1523.30.5.889>
- Earles, J. L., & Kersten, A. W. (2017). Why Are Verbs So Hard to Remember? Effects of Semantic Context on Memory for Verbs and Nouns. *Cognitive Science*, 41, 780–807. <https://doi.org/10.1111/cogs.12374>
- Eddington, C. M., & Tokowicz, N. (2013). Examining English–German translation ambiguity using primed translation recognition. *Bilingualism (Cambridge, England)*, 16(2), 442–457. <https://doi.org/10.1017/S1366728912000387>
- Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, 4, 161–178.
- Ibrahim, S. (Ed.). (2002). *Kamus Dwibahasa: Bahasa Inggeris-Bahasa Melayu* (Edisi Kedua ed.). Dewan Bahasa dan Pustaka.
- Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, 35(1), 49–66.
- Jouravlev, O., & Jared, D. (2020). Native language processing is influenced by L2-to-L1 translation ambiguity. *Language, Cognition and Neuroscience*, 35(3), 310–329. <https://doi.org/10.1080/23273798.2019.1652764>
- Kamus Melayu-Inggeris Dewan*. (2012). Dewan Bahasa dan Pustaka.
- Laufer, B., & Aviad-Levitzky, T. (2017). What Type of Vocabulary Knowledge Predicts Reading Comprehension: Word Meaning Recall or Word Meaning Recognition? *The Modern Language Journal (Boulder, Colo.)*, 101(4), 729–741. <https://doi.org/10.1111/modl.12431>
- Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, 54(3), 399–436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Laxén, J., & Lavaur, J.-M. (2010). The role of semantics in translation recognition: effects of number of translations, dominance of translations and semantic relatedness of multiple translations. *Bilingualism (Cambridge, England)*, 13(2), 157–183. <https://doi.org/10.1017/S1366728909990472>
- Lee, L. W., & Wheldall, K. (2011). Acquisition of Malay word recognition skills: lessons from low-progress early readers. *Dyslexia (Chichester, England)*, 17(1), 19–37. <https://doi.org/10.1002/dys.421>
- Lee, L. C., Liow, S. J. R., & Wee, M.-L. O. (1998). Morphological structure of Malay: Using psycholinguistic analyses of rated familiarity. *SEALS*, VIII, 109.
- Lee, Y., Jang, E., & Choi, W. (2018). L2-L1 Translation Priming Effects in a Lexical Decision Task: Evidence From Low Proficient Korean-English Bilinguals. *Frontiers in Psychology*, 9, 267. <https://www.frontiersin.org/article/10.3389/fpsyg.2018.00267>

- Lemhöfer, K. M., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Luniewska, M., Wodniecka, Z., Miller, C. A., Smolik, F., Butcher, M., Chondrogianni, V., Hreich, E. K., Messarra, C., A. Razak, R., Treffers-Daller, J., Yap, N. T., Abboud, L., Talebi, A., Gureghian, M., Tuller, L., & Haman, E. (2019). Age of acquisition of 299 words in seven languages: American English, Czech, Gaelic, Lebanese Arabic, Malay, Persian and Western Armenian. *PLoS One*, *14*(8), e0220611–e0220611. <https://doi.org/10.1371/journal.pone.0220611>
- Miller, G. A., & Fellbaum, C. (1991). Semantic Networks of English. *Cognition*, *41*(1–3), 197–229.
- Oxford English-English-Malay Dictionary* (3rd Ed. Updated Ver.). (2018). Oxford Fajar Sdn. Bhd.
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*(4), 1029–1038. <https://doi.org/10.3758/BF03193001>
- Prior, A., Wintner, S., MacWhinney, B., & Lavie, A. (2011). Translation ambiguity in and out of context. *Applied Psycholinguistics*, *32*(1), 93–111. <https://doi.org/10.1017/S0142716410000305>
- Rahman, A., Yap, N. T., & Darmi, R. (2018). The Association between Vocabulary Size and Language Dominance of Bilingual Malay-English Undergraduates. *3L, Language, Linguistics. Literature the South East Asian Journal of English Language Studies*, *24*(4), 85–101. <https://doi.org/10.17576/3L-2018-2404-07>
- Rusli, Y. A., & Montgomery, J. (2020). Sentence comprehension and working memory in Malay Adults. *GEMA Online® Journal of Language Studies*, *20*(1), 16–34.
- Schwietter, J. W., & Prior, A. (2020). *Translation Ambiguity. In Bilingual Lexical Ambiguity Resolution* (pp. 96–125). Cambridge University Press. <https://doi.org/10.1017/9781316535967.006>
- Sigurd, B., Eeg-Olofsson, M., & van Weijer, J. (2004). Word length, sentence length and frequency - Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. <https://doi.org/10.1111/j.0039-3193.2004.00109.x>
- Tan, T.-P., Xiao, X., Tang, E. K., Chng, E. S., & Li, H. (2009). MASS: A Malay language LVCSR corpus resource. *Oriental COCODA International Conference on Speech Database and Assessments*, *2009*, 25–30.
- 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*(3), 435–451. <https://doi.org/10.3758/BF03195472>
- Treffers-Daller, J. (2016) Language dominance: the construct, its measurement and operationalization. In C. Silva-Corvalán, & J. Treffers-Daller (Eds.), *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (pp. 235–265). Cambridge University Press.
- Tseng, A. M., Chang, L.-Y., & Tokowicz, N. (2014). Translation ambiguity between English and Mandarin Chinese: The roles of proficiency and word characteristics. In: J. W. Schwieter & A. Ferreira (Eds.), *The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science* (pp. 107–165). Cambridge: Cambridge Scholars Publishing.
- 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 (Cambridge, England)*, *1*(3), 193–211. <https://doi.org/10.1017/S1366728998000352>
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* (2006), *67*(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(3), 407–426. <https://doi.org/10.1016/j.neubiorev.2010.04.007>
- Wen, Y., & van Heuven, W. J. B. (2017). Chinese translation norms for 1429 English words. *Behavior Research Methods*, *49*(3), 1006–1019. <https://doi.org/10.3758/s13428-016-0761-x>
- Yap, M. J., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*(4), 992–1003. <https://doi.org/10.3758/BRM.42.4.992>
- Yap, N. T., Razak, R. A., Haman, E., Luniewska, M., & Treffers-Daller, J. (2017). Construction of the Malay cross-linguistic lexical task: A preliminary report. *Language Studies Working Papers*, *8*, 47–61.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.