Bilingualism: Language and Cognition 20 (4), 2017, 783–794 © Cambridge University Press 2016 doi:10.1017/S1366728916000274

Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words*

JENNIFER BRACKEN
University of Pittsburgh
TAMAR DEGANI
University of Haifa, Israel
CHELSEA EDDINGTON
University of Pittsburgh
NATASHA TOKOWICZ
University of Pittsburgh

(Received: June 16, 2015; final revision received: February 8, 2016; accepted: February 8, 2016; first published online 31 March 2016)

Translations often do not align directly across languages, and indirect mappings reduce the accuracy of language learning. To facilitate examination of this issue, we developed a new continuous measure for quantifying the semantic relatedness of words with more than one translation (hereafter translation-ambiguous words). Participants rated the similarity of each translation to every other translation, yielding a Translation Semantic Variability (TSV) score, ranging from 1.00 (unrelated) to 7.00 (related). Then, we determined how relatedness between translations affects translation-ambiguous word learning from German to English. German words with low TSV scores were recognized as translations more slowly and less accurately than German words with high TSV scores. TSV explains unique variance beyond the previously-used dichotomous classification of words as form vs. meaning ambiguous. We propose that the relatedness of the translation alternatives influences learning because it affects the ease with which a one-to-one mapping can be established between form and meaning.

Keywords: Translation ambiguity, Translation Semantic Variability, meaning similarity, translation recognition

Learning a second language (L2) cannot be successful without effective acquisition of the vocabulary that makes it up. However, indirect mappings of translations across languages pose problems for bilingual language processing and language learning. These mis-mappings occur when one word has multiple translations into another language, creating what is referred to as TRANSLATION AMBIGUITY (see review in Tokowicz, 2014b). Here, we examine the role of translation ambiguity in the very beginning stages of L2 vocabulary learning, focusing on the role of the relatedness between the multiple translations of translation-ambiguous words.

Multiple studies have shown that cross-language translation ambiguity causes difficulty in L2 production and/or recognition (Boada, Sánchez-Casas, Gavilán, García-Albea & Tokowicz, 2013; Degani & Tokowicz, 2010; Degani, Tseng & Tokowicz, 2014; Eddington & Tokowicz, 2013; Laxén & Lavaur, 2010; Prior, Kroll & MacWhinney, 2013; Tokowicz & Kroll, 2007), such that words with multiple translations are produced and recognized less accurately and more slowly than words with a single translation by bilingual speakers, across a wide range of proficiencies (Tokowicz, 2014b; Tokowicz & Degani, 2010).

Notably, translation ambiguity is far more common than may be expected: Tokowicz, Kroll, de Groot and van Hell (2002) found that over 25% of a set of English words selected to have a single translation in fact had more than one translation in Dutch. In other normative studies with this same set of English words, they yielded about 45% translation-ambiguous words when translated into German (Eddington, Degani & Tokowicz, 2016) and 65% when translated into Mandarin Chinese (Tseng, Chang & Tokowicz, 2014). Similarly, Prior, MacWhinney and Kroll (2007) found that over 50% of a different set of English words (including some that were part-of-speech ambiguous) had more than one translation in Spanish. Thus, translation ambiguity is an important factor to consider when conducting cross-language research (Degani, Prior, Eddington, Arêas da Luz Fontes & Tokowicz, in press).

Previous research contrasted translation ambiguity that arises from the source language (e.g., semantic/lexical and part-of-speech ambiguity) with ambiguity that is due to near-synonymy in the target language (e.g., Degani & Tokowicz, 2010; Degani et al., 2014; Eddington & Tokowicz, 2013). In particular, translation ambiguity that arises from lexical ambiguity has been labeled MEANING AMBIGUITY. In such cases, multiple translations capture

* Study 2 formed the basis of JB's Honors thesis; we thank Timothy Nokes-Malach and Tessa Warren for their comments on this work. We thank the Office of Experiential Learning at the University of Pittsburgh for providing funding for Study 2. We thank the PLUM Lab members for research assistance. During the writing of this manuscript, TD was supported by EU-FP7 grant CIG-322016 and NT was supported by NIH R01 HD075800.

Address for correspondence:
Natasha Tokowicz, Learning Research & Development Center, 3939 O'Hara St., Room 634, University of Pittsburgh, Pittsburgh, PA 15260 USA
Tokowicz@pitt.edu

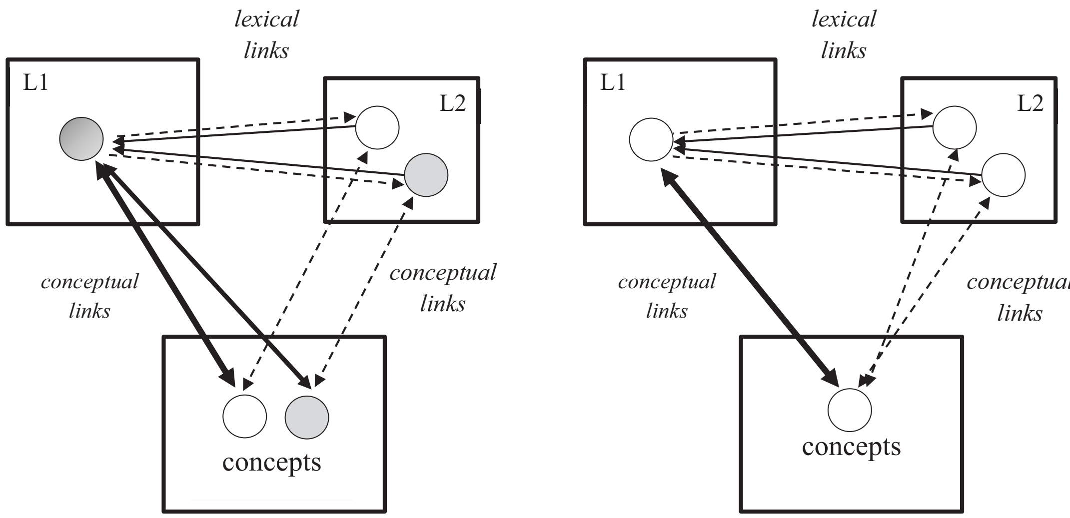

Figure 1. The Revised Hierarchical Model of Translation Ambiguity (adapted from Eddington & Tokowicz, 2013). The left panel depicts a meaning translation ambiguous word; the right panel depicts a form translation ambiguous word. The direction of ambiguity depicted is from L1 to L2 – a single L1 word translates to two L2 words. The opposite situation is also possible but is not shown here.

different meanings of the source word (e.g., *bark* translates into Spanish as *ladrido* to denote the sound a dog makes and *corteza* to denote the outer layer of a tree). By contrast, translation ambiguity that arises from near-synonymy has been labeled FORM AMBIGUITY. In such cases multiple translations have the same or very similar meanings (e.g., *couch* and *sofa* are both translations of the Spanish word *sofá*).

Degani and Tokowicz (2010) explored the effect of both types of translation ambiguity on beginning L2 vocabulary learning by teaching Dutch words to native English speakers. Their critical stimuli were English words with one or two translations into Dutch. Across multiple sessions and tests, translation-ambiguous words were harder to learn than translation-unambiguous words. Ambiguous words were produced and recognized significantly less accurately immediately and after a delay.

Furthermore, Degani and Tokowicz (2010) found that form-ambiguous translations were particularly difficult to learn. They reasoned that the specific difficulty with form-ambiguous words was that a single English word had to be mapped to two Dutch words, whereas for meaning-ambiguous words, learners could use the meanings to keep the two Dutch words separate (see Figure 1). Therefore, a one-to-many mapping (from one meaning to two words) was required for form-ambiguous words, but one-to-one mappings (between a specific meaning and a Dutch word) were possible for the meaning-ambiguous words.

In previous research, form and meaning ambiguity have always been examined as a dichotomy. However, Degani and Tokowicz (2010) noted that their meaning-ambiguous items varied in the extent to which the two meanings were related; some corresponded to homonyms with unrelated meanings (e.g., *change*) and others to polysemes with related senses (e.g., *people*). Classifying translation-ambiguous words as form- or meaning-ambiguous provides only an overall level of semantic ambiguity for the source word. Here, we extend the existing research on this topic by developing a new continuous measure for characterizing translation-ambiguous words. Specifically, we developed the Translation Semantic Variability (TSV) measure, which quantifies the degree of semantic relatedness between the translations of translation-ambiguous words. We further developed a divergence score, which allows researchers to quickly determine whether a given word is 'multiply ambiguous', in that it is not simply form or meaning ambiguous but rather has multiple translations of both types.

Examining a range of semantic variability, rather than simply dichotomizing this dimension, is both theoretically and empirically motivated. In particular, near-synonymous translations may in fact capture different nuances or usages of a word. For example, the words *drunk* and *inebriated* are near-synonyms that differ in their formality, with *drunk* being less formal than *inebriated*. The near-synonyms *foe* and *enemy* share the same core meaning but each word captures a slightly different sense (Inkpen & Hirst, 2006). Moreover, lexical ambiguity may entail two meanings that are dissimilar as for homonyms (e.g., *bark* – dog bark, tree bark) or two senses that are closely related as for polysemes (e.g., *paper* – academic paper, wrapping paper). Importantly, the semantic relatedness of the two meanings of ambiguous words affects processing (e.g., Klepousniotou & Baum, 2007; Klepousniotou, Titone & Romero, 2008; Laxén &

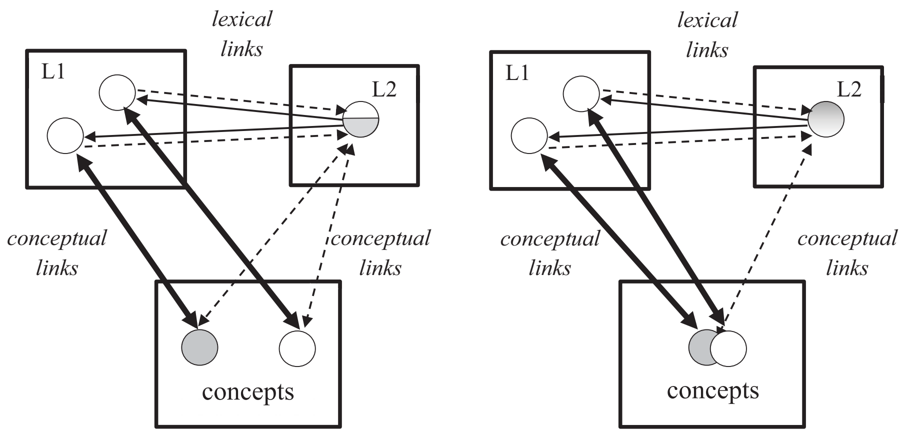

Figure 2. The Revised Hierarchical Model of Translation Ambiguity adapted to capture continuous variations in meaning similarity between translations. The left panel depicts a translation-ambiguous word with less related meanings; the right panel depicts a translation-ambiguous word with more related meanings. The direction of ambiguity depicted is from L2 to L1 – a single L2 word translates to two L1 words. The opposite situation is also possible but is not shown here.

Lavaur, 2010; Rodd, Berriman, Landau, Lee, Ho, Gaskell & Davis, 2012; Rodd, Gaskell & Marslen-Wilson, 2002; for a review, see Eddington & Tokowicz, 2015).

For instance, Rodd et al. (2002) showed that polysemes were processed more quickly than unambiguous words in a lexical decision task, but this advantage did not hold for homonymous words (see also Klepousniotou & Baum, 2007), and some studies using meaning-based tasks (e.g., category judgment) show a disadvantage in processing for homonyms (e.g., Beretta, Fiorentino & Poeppel, 2005). Similarly, in Klepousniotou et al. (2008), polysemes showed reduced effects of context relative to homonymous words. It is therefore useful to examine the semantic similarity of translations that map to a single source word because it will allow a clearer understanding of the continuous dimension of similarity.

The new measure we present here, TSV, allows exactly that, because it takes into account the semantic similarity of all of the pairwise comparisons of translations to a single word, thereby capturing the full range of semantic similarity between translations. For instance, the two English translations for the German word *Folge* – *episode* and *result* – which are not related in meaning (i.e., homonyms), are predicted to receive a low TSV score, whereas the two related translations (i.e., polysemes) to the word *Vererbung* – *inheritance* and *heredity* – are predicted to receive a TSV score in the middle range. Lastly, two English translations of the German word *Versuch* – *try* and *attempt* – which overlap greatly in their meanings (i.e., near-synonyms), are predicted to receive a high TSV score.

After describing the TSV measure in Study 1, we examine the role of this new measure in predicting learning of translation ambiguous words in Study 2. Specifically, in Study 2, native English speakers were asked to learn a set of German translation-ambiguous words that map onto two English translations that varied in their semantic similarity. Degani and Tokowicz (2010) claimed that the inconsistent (one-to-many) mapping of a meaning to two words is the primary reason for the difficulty in learning translation-ambiguous words. If this is correct, we would predict that learners in the present study should have more difficulty learning words with lower TSV scores. That is, it should be harder to learn that a German word maps to two UNRELATED English words than to learn that a German word maps to two RELATED English words. This is because when the two English words are unrelated, the learner will need to create a one-to-many mapping between the German translation and the two meanings. In contrast, when the two English translations are related in meaning, the learner could create a one-to-one mapping between the German translation and a more distributed concept encompassing both English meanings (see Figure 2). Thus, in this vocabulary training study, we test how TSV scores of translation-ambiguous German words influence learners' performance, and predict that words with lower TSV scores will be more difficult to learn than words with higher TSV scores.

Study 1 – Computing TSV scores

Methods

Participants

One hundred and nineteen native English speaking students at the University of Pittsburgh participated in the study for credit toward an Introduction to Psychology course requirement. Participants had no knowledge of German, and had not been exposed to any language other than English before the age of 10. Data from

21 additional participants were excluded following data collection because they did not meet these criteria.

Materials

Materials for the TSV task were obtained from the correct responses provided in the German to English number-of-translation norms that had been collected by Eddington et al. (2016) from a group of six proficient English–German bilinguals (three native English speakers and three native German speakers). These participants were recruited from the University of Pittsburgh and the local German language club, and had indicated relatively balanced proficiencies in English ($M = 8.83$) and German ($M = 8.44$), measured on a 10-point self-rated proficiency scale on which 10 indicated the highest level of proficiency. The authors calculated the number of translations for a word as the number of correct translations given across participants (see also Prior et al., 2007; Tokowicz et al., 2002); the maximum number of translations was six. In the present study, all words that received two or more translations were paired, such that each English translation for a given German word was paired with every other English translation for that word. For example the German word *Spannung* was translated correctly into the English words *tension*, *suspense*, and *tense*. We therefore obtained semantic similarity ratings for all three pairs: tension-suspense, tension-tense, and suspense-tense. Nine lists were created such that no word was repeated within the same list. Moreover, translation pairs were intermixed with filler pairs that were expected to vary in their relatedness rating.

Procedure

Participants rated each word pair in terms of its meaning similarity and then separately in terms of its combined spelling and sound similarity, both on a 1 (completely different) to 7 (exactly the same) scale (see also Tokowicz et al., 2002). They completed both ratings together to reduce the influence of form similarity on the meaning-similarity ratings. Each participant rated between 137 and 154 pairs. To derive TSV scores, we averaged the semantic similarity ratings across all participants for all combinations of English word pairs that had been provided as correct translations of a given German word. A score of 1 indicates that the translations are very different in meaning, whereas a score of 7 indicates that the translations are very similar in meaning. Each word pair was rated by a minimum of 12 participants.

To elucidate whether the derived TSV score for a given word with three or more translations is the result of some of the translations being very similar in meaning and others being less related in meaning, we also computed a DIVERGENCE SCORE for each word. For example, the German word *Gleich* translates as *immediately*, *same*, and *equal*; notably, *same* and *equal* may be more similar

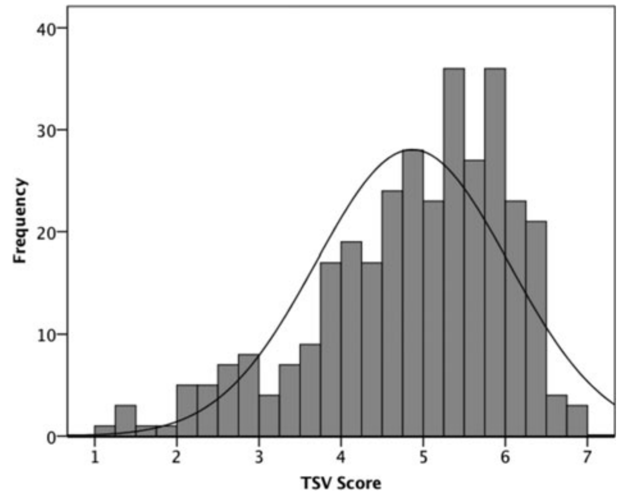

Figure 3. Distribution of TSV scores.

to each other than are *immediately* and *equal*. This divergence score was computed as the standard deviation (*SD*) of the averaged semantic similarity ratings given to the different English word pairs that had been provided as translations for each German word. A low divergence score indicates a lack of variance in the ratings of all translations given, and a high divergence score indicates a combination of some translation pairs being semantically related and others being unrelated (i.e., a combination of form and meaning translation ambiguous).

Results and discussion

TSV scores

The TSV ratings can be downloaded at: http://plumlab.pitt.edu/norms/ listed as a function of the German translation. TSV scores ranged from 1.16 to 7.00 ($M = 4.85$, $SD = 1.17$). Words that were previously classified by Eddington and Tokowicz (2013) as form ambiguous received a higher average TSV score ($M = 5.43$, $SD = 0.78$, range: 2.33 to 7.00) than words that were previously classified as meaning ambiguous ($M = 3.5$, $SD = 1.03$; range: 1.16 to 5.46), $t\ (328) = 17.06$, $p < .001$. As can be seen in Figure 3, the TSV scores do not appear to be bimodally distributed, as would be expected if the form vs. meaning classification were psychologically valid. Rather, many words received intermediate TSV scores. For instance, the German word *Wahl,* which was translated in the norms as *election* and *choice*, received a TSV score of 4.0. Such a word would have been artificially classified as either 'form-ambiguous' or 'meaning-ambiguous' when in fact the semantic similarity of the two translations falls somewhere in between. To directly examine whether the distribution of TSV scores was unimodal, we conducted Hartigan's dip test as implemented in the 'diptest' package (Maechler,

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press

2015) in R (R core team, 2013). According to this test, TSV scores are unimodally distributed, $D = 0.0146$, $p = .9446$. This suggests that obtaining norms to determine semantic variability may be superior to the form/meaning dichotomy. Further, based on the Shapiro-Wilks test for normality (shapiro.test) in R (R Core Team, 2013), the TSV data are not normally distributed, but rather are negatively skewed, $W = 0.95$, $p < .0001$. Thus, most translation pairs are considered relatively similar in meaning.

Divergence scores

To reiterate, the divergence score is computed for words that have three or more translations as the *SD* of the averaged semantic similarity ratings for a word. Thus, a lower divergence score indicates closer similarity in the translations given. The divergence scores ranged from 0.03 to 2.45 ($M = .85$, $SD = .51$). To illustrate what the divergence score reflects, consider the example given above of the German word *Gleich*, which translates into English as *immediately*, *same*, and *equal*. *Equal* and *same* are indeed considered very similar in meaning (6.31), whereas *immediately* is considered relatively unrelated to either of these words (2.08 and 2.07, respectively). Therefore, *Gleich* has a high divergence score because two of its translation pairs have low TSVs and one of its translation pairs has a high TSV, yielding a high divergence score of 2.45. By contrast, German words that elicited only translations of the same type (e.g., only synonymous translations or only unrelated translations) received low divergence scores. For example, *Leichtigkeit* translates into English as *ease*, *easiness*, and *simplicity*. These words are all highly related to each other (averaged TSV 5.82; minimum TSV 5.69), therefore there is little variability in their ratings yielding a low divergence score of .12. Similarly, the German word *Karte* translates into English as *card*, *ticket*, and *map*. These words are all minimally related to each other (averaged TSV 2.95; minimum 2.23), and this lack of variability in their ratings yields a low divergence score of 0.87. Thus, by examining both the averaged TSV score and its corresponding divergence score, researchers can select words that have translations that are more or less related, and that are more or less varied in their semantic relation to each other.

Conclusion

In Study 1, we demonstrated that TSV is a useful metric for quantifying the meaning relatedness between the translations of translation-ambiguous words. Furthermore, this rating can be used in lieu of the previously-used form vs. meaning ambiguous dichotomous classification. For example, form-ambiguous words such as *bag*/*sack* (*Tüte*), received high TSV scores (actual mean rating 6.5), whereas meaning-ambiguous words such as *jaw*/*pine* (*Kiefer*) received low TSV scores (actual mean rating 1.25). And, these scores were not bimodally distributed as would have been expected if a true dichotomy were to underlie this continuum. Note also that the overall mean TSV rating was above the midpoint of the scale (mean overall rating 4.87), which reflects the fact that more translation-ambiguous words are form ambiguous than meaning ambiguous (e.g., Tokowicz et al., 2002; Tseng et al., 2014).

The divergence scores are computed as the standard deviation of the TSV scores for words with three or more translations. The divergence score is a metric of the relative similarity of the translations of a given word, and can be used to quickly determine whether a word has multiple translations of varied types. This metric can help researchers select particular types of words for their research. As mentioned above, the word *Gleich* has two translations that are similar to each other, and another that is quite different. Interestingly, words with this particular profile are relatively over-represented in the Mandarin Chinese–English language pair (see Tseng et al., 2014).

In Study 2, we examine the consequences of the range of meaning similarity between translations for the initial exposure to vocabulary in a group of individuals who had not previously been exposed to the target language (German) or similar languages (in this case, Dutch). Because this is a first investigation of the continuous nature of the semantic similarity of translations as reflected by the TSV score, we opted to focus on words with only two translations, and to leave the investigation of the divergence score and its influence to future studies.

Study 2 – Learning of Translation Ambiguous Words

Study 2 examines the relationship between TSV and the difficulty in learning translation-ambiguous words. Previous studies have explored the consequences of ambiguity type for language learning and processing only in terms of the form/meaning dichotomy; here, we investigate the difficulties in learning ambiguous translations across a continuum of similarity. We tested learning using the translation-recognition task (e.g., de Groot, 1992). We chose this task because the use of translation tasks is an accepted way of assessing vocabulary knowledge (see de Groot, 2011, Chapter 3). Furthermore, this task is particularly useful in the case of less-proficient learners because it is less difficult than the translation-production task, and has been used in similar translation-ambiguity learning studies in the past (see Degani & Tokowicz, 2010; Degani et al., 2014).

We predicted that words with low TSV scores would be more difficult to learn (i.e., recognize as translations) than words with high TSV scores. As noted above and depicted in Figure 2, this is because words with low TSV scores necessitate ambiguous one-to-many mappings between the L2 lexical node and two L1 lexical nodes and/or two (unrelated) semantic representations. In contrast, words with a high TSV score may allow the learner to override the ambiguous mapping by creating a one-to-one mapping between the L2 lexical node and a single, more distributed semantic representation.

Methods

Participants

Eighteen right-handed native English speakers with no prior knowledge of the target language, German, participated in the study for $20. Individuals with prior exposure to Dutch were not qualified to participate due to the language's similarity to German. Data from one participant were excluded due to technical error and data from another participant were excluded to maintain counterbalancing; the final analyses are based on data from 16 participants.

Design

This study used a session (first vs. second) within-subjects design. TSV was then used as a continuous predictor of the reaction time and accuracy data.

Stimuli

The German stimuli and their English translations were taken from the Eddington et al. (2016) norms. The 34 critical German words had two translations into English that had been normed for TSV in Study 1. Thirty-four items with only one translation into English were also included as fillers. Words with two translations were shown with both translations; a total of 102 German–English word pairs were trained.

Two training lists were created and counterbalanced across participants. Each list included 34 translation-unambiguous German words and their English translations, and 34 translation-ambiguous words, shown on consecutive trials, once with each translation (e.g., Degani et al., 2014). The order of the two translations was counterbalanced across participants. The two translations were matched in terms of a number of important dimensions (see Table 1 for stimulus characteristics and matching details).

Four counterbalancing list versions of a translation recognition test were created. Each participant was tested in two list versions of this task, one on each day; the particular versions were counterbalanced across participants. Across the four counterbalancing list versions, ambiguous German words appeared twice, paired with a different English word each time. One quarter of the ambiguous words was paired correctly both times, one quarter was paired incorrectly both times, one quarter was paired correctly with the words' first translation but incorrectly with the words' second translation, and one quarter was paired incorrectly with the words' first translation but correctly with the words' second translation. The counterbalancing of lists for this task removed any contingency across words, so that participants could not predict if the second pairing of an ambiguous word was correct or incorrect. An equal number of participants saw each of these versions crossed with each of the training orders.

On incorrect trials, a German word was paired with the English translation of a different German word in the training set. During the pairing process, care was taken to ensure that related pairings were not accidentally created.

Session 1 procedure

The experiment consisted of two sessions, held one week apart. During the first session, participants completed three training cycles in which they learned the 102 German–English translation pairs; thus, there were a total of 306 training trials. Each trial began with a fixation cross that remained on the screen until the participant pressed a button. This was then replaced with a German word shown simultaneously with its English translation. The German–English translation pair remained on the screen for 800 ms (following Degani & Tokowicz, 2010). For translation-ambiguous words, the second translation always appeared on the trial immediately following the first (Degani et al., 2014). Directly following training, participants completed a Stroop task (Stroop, 1935) as a distracter between training and testing.

Participants were then tested using a translation-recognition task. In this task, upon seeing the fixation cross, participants initiated a trial via button press. The English and German words were presented on the screen at the same time. The participant pressed the 'yes' button with the right index finger to indicate that the English word was a correct translation of the German word, and the 'no' button with the left index finger to indicate that the English word was not a correct translation. The German–English pair remained on the screen until the participant responded or up to 3500 ms. There was an inter-stimulus interval of 100 ms. German words with multiple translations were shown twice, with one translation at a time, in a random order determined by the computer program (E-Prime software, Psychology Software Tools, Pittsburgh, PA).

Session 2 procedure

Participants returned one week later to complete the second session, which consisted of a different version of the translation-recognition task, an Operation Span task (Turner & Engle, 1989), and a language history questionnaire (Tokowicz, Michael & Kroll, 2004).

Table 1. *Stimulus characteristics.*

	Translation 1	Translation 2	*p*-value
English length (number of letters)	5.53 (.36)	5.44 (.37)	.829
SUBTLEXUS frequency	2.98 (.45)	3.04 (.76)	.685
Number of orthographic neighbors	4.66 (1.17)	6.50 (1.20)	.266
Frequency of orthographic neighbors	24.65 (7.11)	27.77 (6.86)	.757
Concreteness	3.65 (.22)	3.53 (.19)	.391

Note. Standard deviations are shown in parentheses. SUBTLEXUS frequency is from Brysbaert and New (2009). Number of orthographic neighbors and frequency of orthographic neighbors uses Coltheart's N measure (from the MCWord database; Medler & Binder, 2005). Concreteness values are from Brysbaert, Warriner and Kuperman (2014).

Table 2. *Summary of intercorrelations between TSV, accuracy, and reaction times on the translation-recognition task as a function of session.*

Measure	1	2	3	4	5
TSV	—	.04	.34*	−.11	−.52**
Yes Accuracy 1	.04	—	.31*	−.47**	−.26
Yes Accuracy 2	.34*	.31*	—	−.29*	−.24
Yes RT 1	−.11	−.47**	−.29*	—	.19
Yes RT 2	−.52**	−.26	−.24	.19	—

*$p < .01$, two-tailed. **$p < .01$, two-tailed.

Results

The data for ambiguous words were first analyzed using repeated measures Analyses of Variance (ANOVAs) to explore effects of session. In these analyses, session (first, second) served as a within-participants variable. The dependent variables were accuracy and reaction time; reaction time analyses were performed only on correct trials. Following the ANOVAs, TSV was examined as a predictor of accuracy and reaction time. Trial type ('yes', 'no') was not included as a factor because the 'no' trials were created by pairing a German word with the English translation of another word. Therefore, these trials were not of theoretical relevance.

Accuracy

Participants were more accurate in the translation-recognition task during Session 1 immediately following training ($M = 90.6\%$) than during Session 2 after a delay ($M = 82.4\%$), F (1, 15) = 6.00, MSE = .009, $p < .05$, $\eta p^2 = 29$.

Reaction time

The effect of session was not significant in the reaction time analysis, $F < 1, p = .42$.

Correlations

To examine the effects of meaning similarity directly, TSV scores were correlated with accuracy and reaction time; the eight items with 0 accuracy were treated as missing. Because English and German word length and English word frequency are known to affect performance but were not matched a priori, they were controlled for using partial correlations. Similarly, the form similarity ratings from Study 1 were used as a control variable. See Table 2 for the correlation scores.

TSV correlated positively with accuracy on 'yes' trials in Session 2, r (50) = .34, $p < .05$, see Figure 4. TSV correlated negatively with reaction time for 'yes' trials in Session 2, r (50) = -.52, $p < .01$, see Figure 5. These findings demonstrate that for ambiguous stimuli, participants responded more quickly and accurately to words that are higher on the TSV scale (i.e., that had translations more similar in meaning), but that these effects were statistically significant only after a delay. No other correlations with TSV were significant ($|r|$s ≤ .17).

To directly assess whether this new continuous measure better predicts performance than the dichotomous form vs. meaning distinction, we used a hierarchical linear regression analysis. In this type of hierarchical analysis, variables are entered on separate steps to determine whether a subsequently-entered variable explains additional variance beyond a previously-entered

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press

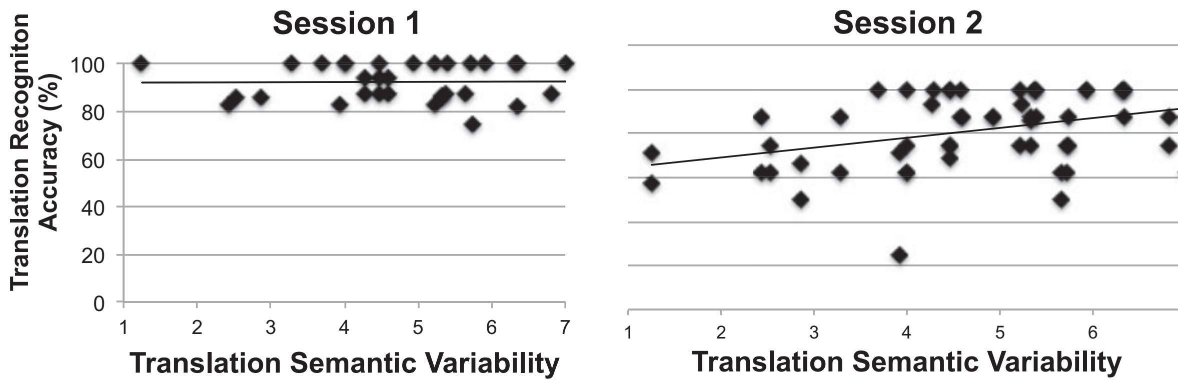

Figure 4. Correlation between TSV scores and accuracy for 'yes' trials in Sessions 1 and 2 in the translation-recognition task.

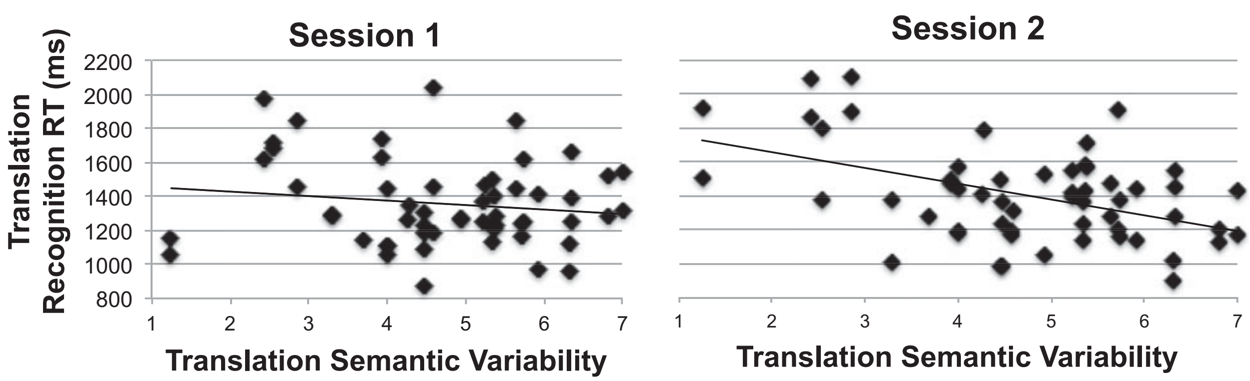

Figure 5. Correlation between TSV scores and reaction time for 'yes' trials in Sessions 1 and 2 in the translation-recognition task.

variable. Here, we ask whether the continuous TSV factor explains additional variance beyond the categorical form vs. meaning variable and vice versa. We classified the 10 words in our sample with the lowest TSV scores as meaning ambiguous and the 10 words with the highest TSV scores as form ambiguous. Although this is necessarily an arbitrary number of stimuli, note that this stacks the deck against us because it is likely that most samples of form vs. meaning ambiguous items would not be at the extreme ends of the continuum. We then ran separate regressions for each of the four dependent measures (Session 1 'yes' reaction time and accuracy; Session 2 'yes' reaction time and accuracy). We did not include the 'no' trials because these pair incorrect translations and therefore a valid TSV score cannot be computed. As with the correlations, because English and German word length and English word frequency are known to affect performance but were not matched in this sample of items, they were entered on the first step of the analysis along with form overlap ratings.

For Session 2 'yes' reaction time, if TSV is entered first, TSV is significant, $\Delta R^2 = .61$, F (1, 14) = 32.10, $p < .01$, but form vs. meaning does not predict additional variance, $\Delta R^2 = .02$, F (1, 13) = 1.20, $p = .29$ (total R^2 for all three steps = .76). By contrast, if form vs. meaning is entered first, it predicts significant variance, $\Delta R^2 = .42$, F (1, 14) = 12.82, $p < .01$, but TSV still predicts additional variance beyond that, $\Delta R^2 = .21$, F (1, 13) = 11.40, $p < .01$ (total R^2 for all three steps = .76). TSV and form vs. meaning did not predict significant variance for any of the other dependent measure; the full statistics are reported in Table 3. Thus, with as extreme examples as possible of form vs. meaning ambiguous items in our sample, the continuous TSV measure predicted additional variance beyond the categorical measure in reaction time in Session 2.

General Discussion

The results of the current study underscore the importance of semantic relatedness for learning of translation-ambiguous words. In previous research, Degani and Tokowicz (2010) demonstrated that translation-ambiguous words were harder to learn than translation-unambiguous words, and that learning of form-ambiguous items, which require a one-to-many mapping, is especially difficult. We extended these findings by using a new measure called Translation Semantic Variability to investigate a spectrum of similarity, and exemplified its predictive power for learning translation-ambiguous words.

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press

Table 3. *Summary of regression analyses testing the ability of TSV and ambiguity type to predict variance in translation recognition reaction time and accuracy for 'yes' trials.*

	TSV first				Type first		
Step	ΔR^2	ΔF	p	Step	ΔR^2	ΔF	p
			Session 1 Reaction Time				
1-Control	.43	2.87	.06	1-Control	.43	2.87	.06
2-TSV	.000	.001	.99	2-Type	.000	.002	.97
3-Type	.000	.02	.90	3-TSV	.001	.01	.91
			Session 2 Reaction Time				
1-Control	.12	.53	.72	1-Control	.12	.53	.72
2-TSV	.61	32.10	<.01**	2-Type	.42	12.82	.003**
3-Type	.02	1.20	.29	3-TSV	.21	11.40	.005**
			Session 1 Accuracy				
1-Control	.26	1.30	.32	1-Control	.26	1.30	.32
2-TSV	.000	.001	.97	2-Type	.004	.07	.80
3-Type	.02	.29	.60	3-TSV	.01	.22	.65
			Session 2 Accuracy				
1-Control	.19	.88	.50	1-Control	.19	.88	.50
2-TSV	.05	.82	.38	2-Type	.009	.15	.70
3-Type	.06	1.04	.33	3-TSV	.09	1.70	.22

Note. Step 1 includes form rating from Study 1, SUBTLEXUS Lg10 English frequency (Brysbaert & New, 2009), and German and English length in number of letters. Steps 2 and 3 include TSV or type as denoted by the column heading. Degrees of freedom vary by step (Step 1 = 4, 15; Step 2 = 1, 14; Step 3 = 1, 13). Type refers to the form vs. meaning categorization.
*$p < .01$, two-tailed. **$p < .01$, two-tailed.

First, our paper details a new method for capturing meaning similarity between translations, and extends the previously-used form vs. meaning ambiguity dichotomy (Degani, Prior & Tokowicz, 2011; Degani & Tokowicz, 2010; see also Laxén & Lavaur, 2010), by capturing the full range of semantic similarity between translations. Critically, TSV scores are not bimodally distributed at the two extreme points of relatedness, highlighting that a dichotomous classification does not adequately capture this variable. Moreover, the semantic similarity ratings collected in Study 1 can be used by researchers interested in the similarity of pairs of English words because the norms provide ratings of meaning and form similarity for 630 English word pairs. Further, in another study, we collected a comparable set of TSV norms for a set of English words, by asking German–English bilinguals to rate the semantic-similarity of all pairwise German translations of each English word (Eddington et al., 2016). These latter TSV data can be used to estimate the semantic ambiguity of the English source word itself because English words with lower TSV scores are likely to be more semantically ambiguous than English words with higher TSV scores.

Second, in a translation-recognition task, TSV scores correlated with learners' performance on Session 2 such that words with lower TSV scores were harder to learn than words with higher TSV scores. This difficulty was present in the accuracy and latency data alike. Interestingly, our regression analyses demonstrated that the continuous TSV measure was a better predictor of learners' performance than the form vs. meaning dichotomy used in previous translation ambiguity studies. TSV scores, which better align with participants' perception of semantic relatedness, may prove to be a highly valuable tool in L2 acquisition research more broadly.

Accuracy and rapidity of responses in the translation-recognition task decreased as the relatedness of the translations decreased. We suggest that the difficulty associated with unrelated translations stems from the one-to-many mapping (from form to meaning) required during learning. For translations with lower TSV scores, a one-to-many mapping needs to be established between a single German translation and two meanings that are unrelated for native English speakers. This is arguably more difficult than mapping a single German translation onto two related meanings, because the two meanings are conceivably highly connected or closer in semantic space and may be treated as a single more distributed concept. For translation-ambiguous words with medium TSV scores, a one-to-many mapping may still be required but their meanings will be more closely related and may facilitate learning relative to words with lower TSV scores.

In a recent study, Rodd et al. (2012) showed that adult learners were better at learning new meanings to familiar

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press

words when these meanings were semantically related to the known meaning of the word than when it was a new unrelated meaning. Recall of the newly acquired meanings was significantly better, both immediately and following a delay, when the new meanings were related to the familiar meaning. The authors explain the RELATEDNESS EFFECT by proposing that the semantic relatedness of the new and old meanings makes it easier for learners to create the proper form-to-meaning link, or that during recall unrelated meanings compete with the novel meaning and hinder performance (e.g., Eddington, 2015).

The relatedness effect observed in the current study may be explained similarly as arising during learning or during the recognition task. Specifically, as suggested above, it may be easier to map a new German label to two related meanings because these may be treated as a single more distributed concept, thus allowing a direct one-to-one mapping. Thus, for instance, learners may be able to link the new German word *Versuch* to a more distributed concept including both *attempt* and *try*. Such a mapping is not suitable, however, when the two meanings are unrelated because these are more distant from each other in semantic space and therefore require a split in the mapping from (German) form to meaning. Conversely, the relatedness effect we observe may stem from competition arising during the translation recognition test from the unrelated meaning. When asked to determine if *Kiefer* and *jaw* are correct translations, the simultaneous activation of the alternative unrelated meaning *pine* interferes with participants' decision that the two words (*Kiefer* and *jaw*) are correct translations of each other. Conversely, during recognition there is a lack of competition for the more related translations, which is consistent with within-language studies of semantic ambiguity (e.g., Rodd et al., 2002).

In the current study, the relation between the semantic relatedness of the translations, as reflected by the TSV score, and translation recognition performance was evident following a delay, in the accuracy and latency data of Session 2 but not immediately after learning (in Session 1). Several sources may explain this delayed effect. First, because the performance in Session 1 was higher than that in Session 2 (see Figure 4), a restricted range may have limited the potential for a significant correlation between TSV and performance. Notably, however, such an explanation is less suitable for the latency data, in which performance similarly ranged in the two sessions (see Figure 5). Second, as discussed above, the relatedness effect may be rooted in mapping processes that take place during learning or in competition processes during the recognition test. The delayed nature of the relatedness effect may suggest that it is more heavily based on competition during testing, because representational processes should have affected performance both immediately and following a delay. Finally, the delayed nature of the relatedness effect may reflect a change in the weights of lexical and semantic representations. In particular, all translation-ambiguous words entail a one-to-many mapping between L2 lexical nodes and L1 lexical nodes (see Figure 2), but words with a higher TSV score may enjoy a one-to-one mapping between L2 lexical nodes and semantic representations. It is possible that immediately after learning, during Session 1, learners rely more on lexical links, and thus regardless of the TSV score, all translation-ambiguous items suffer from a one-to-many mapping. Following a delay, however, learners may begin to rely more on semantic links and thus can benefit from the one-to-one mapping afforded by items with higher TSV scores. Thus, the influence of the TSV measure may be more prominent as learners begin to weight the mapping of lexical to semantic representation over lexical-translational links. It is interesting to note that this finding is consistent with our past research in which the difference between form and meaning-ambiguous words was observed only after a delay (Degani & Tokowicz, 2010).

The second important aspect of our findings relates to the direction of ambiguity. In a previous learning study (Degani & Tokowicz, 2010), native English speakers were asked to map a single English word onto two Dutch words. When the English word was already ambiguous in its own right (e.g., *change*, which can either refer to the result of modification, or to a collection of coins), it was easier for participants to link it to two Dutch words, each connected to one of the meanings. This creates an advantage for meaning-ambiguous words. In contrast, when the ambiguous words are L2 words, as in the current study, the more similar the two English (L1) translations, the easier they were to learn. In this case, two English words converge onto one German translation. The greater the semantic relation between the English words to begin with, the easier it will be to link them to one shared translation.

Previous research with proficient bilinguals has demonstrated that a shared translation in a bilingual's other language may increase the semantic similarity of its two translations (Degani et al., 2011). Specifically, in a semantic relatedness rating task with Hebrew–English and English–Hebrew bilinguals, Degani et al. (2011) showed that two English words that share a translation in Hebrew were rated by bilinguals as more similar in meaning than two words that do not share a Hebrew translation. This effect was observed both when Hebrew was the first language of the participants and when it was the L2. Moreover, the shared translation effect, reflected in the increase in semantic relatedness for bilinguals compared to monolinguals, was evident both when the two English words were related in meaning (e.g., *house* and *home*, which share the translation *bait* in Hebrew; see also Jiang, 2002, 2004) and when the two English

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press

words were unrelated in meaning (e.g., *beak* and *source* share the Hebrew homonym translation *makor*). The authors suggested that co-activation of the two English words and meanings with their corresponding shared (Hebrew) translation creates convergence in semantic space, resulting in an increase in the semantic relatedness of the two words. The present study taps the earliest stages of learning a shared translation in an L2, and suggests that this convergence process is easier if the two words are initially related in meaning.

One potential limitation of our study is that we tested the efficacy of our new measure with learners who were not studying the target language outside of the laboratory setting and who had not previously been exposed to the language in question. We chose this population so that we would have strict experimental control over exposure to the target vocabulary (for discussion, see Tokowicz & Degani, 2015). In addition, we are interested in the first exposure to a language because all adult learners begin at this point, and because learning at this stage may set the trajectory for how words are represented. Nonetheless, because the learners in our study were naïve learners, the findings regarding the influence of the TSV measure may not directly generalize to more experienced learners. Specifically, more proficient bilinguals add vocabulary items to already established semantic networks, and it remains to be examined whether the advantage for learning more semantically-similar translation-ambiguous words holds when one of the translations is already known to the learner and the additional translation must be learned. In a previous study, we found no interaction between meaning similarity and learning two translations on different days of training vs. on the same day. Instead, for both more- and less-similar translations, learning on the same day led to better learning (Degani et al., 2014). However, although the translation learned on the first day enjoyed a substantial advantage, it was not truly known prior to training, as would be the case for more experienced learners.

In future research, it will also be important to extend our findings with the TSV measure to other pairs of languages. This is particularly important because language pairs vary in their form relatedness, with some pairs having somewhat high form overlap and large numbers of cognates (e.g., Portuguese and Spanish) and others having relatively fewer (e.g., Portuguese and English). Notably, in the current study the effect of semantic relatedness (TSV) was established after statistically controlling for the influence of form overlap.

To summarize, given the importance of semantic relatedness for processing of ambiguous words (Eddington & Tokowicz, 2015; Rodd et al., 2012), the current study presented a new continuous measure for semantic relatedness of translation-ambiguous words. Such translation ambiguity is a common characteristic of the mapping between languages and has previously been shown to influence performance of bilinguals and L2 learners (for a review, see Tokowicz, 2014a, Chapter 6; Tokowicz, 2014b). The new measure we created, Translation Semantic Variability, captures the full range of semantic similarity of multiple translations and as such goes beyond the previously-used dichotomy between form-ambiguous and meaning-ambiguous translations. In a training study with native English speakers we demonstrated that the TSV score predicts learning of translation-ambiguous German words. In particular, learners were less accurate and slower to respond in a translation-recognition task as the relatedness of the two English translations decreased. This relatedness effect likely stems from the indirect one-to-many mapping between form and meaning that is unavoidable when the translations are unrelated. The findings highlight the importance of ambiguity in the mapping from form-to-meaning at the earliest stages of vocabulary learning.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1366728916000274

References

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24*(*1*), 57–65.

Boada, R., Sánchez-Casas, R. M., Gavilán, J. M., García-Albea, J. E., & Tokowicz, N. (2013). Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition, 16,* 183–197.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(*5*), 1001–1018.

de Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. New York: Psychology Press.

Degani, T., Prior, A., Eddington, C. M., Arêas da Luz Fontes, A., & Tokowicz, N. (in press). Determinants of translation ambiguity: A within and cross-language comparison. *Linguistic Approaches to Bilingualism*.

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology, 23,* 18–28.

Degani, T., & Tokowicz, N. (2010). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition, 13,* 299–314.

Degani, T., Tseng, A. M., & Tokowicz, N. (2014). Together or apart? Learning of ambiguous words. *Bilingualism: Language and Cognition, 17,* 749–765.

Eddington, C. M. (2015). *Effects of within- and cross-language semantic ambiguity in learning and processing*. Unpublished Doctoral Dissertation. University of Pittsburgh, Pittsburgh, PA.

Eddington, C. M., Degani, T., & Tokowicz, N. (2016). *English and German translation norms: The role of proficiency in translation ambiguity*. Manuscript in revision.

Eddington, C. M., & Tokowicz, N. (2013). Examining English-German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition, 16,* 442–457.

Eddington, C. M., & Tokowicz, N. (2015). How context and meaning similarity influence ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review, 22,* 13–37.

Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics, 32*(*2*)*,* 223–262.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24,* 617–637.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *Modern Language Journal, 88,* 416–432.

Klepousniotou, E., & Baum, S. (2007). Clarifying further the ambiguity advantage effect in word recognition: Effects of aging and left-hemisphere damage on the processing of homonymy and polysemy. *Brain and Language, 103,* 148–149.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1534–1543.

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: Language and Cognition, 13*(*2*)*,* 157–183.

Maechler, M. (2015), *Hartigan's dip test statistic for unimodality-corrected code*. R package version 0.75-6.

Prior, A., Kroll, J. F., & MacWhinney, B. (2013). Translation ambiguity but not word class predicts translation performance. *Bilingualism: Language and Cognition, 16* (*Special Issue 02*)*,* 458–474. doi: 10.1017/S1366728912000272

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.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition, 40,* 1095–1108.

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*, 245–266.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(*6*)*,* 643–661.

Tokowicz, N. (2014a). *Lexical processing and second language acquisition*. New York: Routledge.

Tokowicz, N. (2014b). Translation ambiguity affects language processing, learning, and representation. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, A. Tuninetti & D. Walter (Eds.), *Selected proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines* (pp. 170–180). Somerville, MA: Cascadilla Press.

Tokowicz, N., & Degani, T. (2010). Translation ambiguity: Consequences for learning and processing. In B. Van Patten & J. Jegerski (Eds.), *Research on second language processing and parsing* (pp. 281–293). Amsterdam: John Benjamins.

Tokowicz, N., & Degani, T. (2015). Learning second language vocabulary: Insights from laboratory studies. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 216–233). Cambridge: Cambridge University Press.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*, 727–779.

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.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*(*3*)*,* 255–272.

Tseng, A. M., Chang, L.-Y., & Tokowicz, N. (2014). Translation ambiguity between English and Mandarin Chinese: The role of proficiency. In J. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 107–165). Cambridge: Cambridge Scholars Publishing.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

https://doi.org/10.1017/S1366728916000274 Published online by Cambridge University Press