Bilingualism: Language and Cognition 16 (2), 2013, 458–474 © Cambridge University Press 2012 doi:10.1017/S1366728912000272

Translation ambiguity but not word class predicts translation performance*

ANAT PRIOR
Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel
JUDITH F. KROLL
Department of Psychology, Center for Language Science, The Pennsylvania State University, USA
BRIAN MACWHINNEY
Department of Psychology, Carnegie Mellon University, USA

(Received: September 8, 2011; final revision received: February 13, 2012; accepted: May 10, 2012; first published online 13 July 2012)

We investigated the influence of word class and translation ambiguity on cross-linguistic representation and processing. Bilingual speakers of English and Spanish performed translation production and translation recognition tasks on nouns and verbs in both languages. Words either had a single translation or more than one translation. Translation probability, as determined by normative data, was the strongest predictor of translation production and translation recognition, after controlling for psycholinguistic variables. Word class did not explain additional variability in translation performance, raising the possibility that previous findings of differences between nouns and verbs might be attributed to the greater translation ambiguity of verbs relative to nouns. Proficiency in the second language was associated with quicker and more successful production of translations for ambiguous words, and with more accurate recognition of translations for ambiguous words. Working memory capacity was related to the speed of recognizing low probability translations for ambiguous words. These results underscore the importance of considering translation ambiguity in research on bilingual lexical and conceptual knowledge.

Keywords: translation ambiguity, word class, bilingual lexicon

The conceptual and lexical organization of the two languages of bilingual speakers has received much attention in the psycholinguistic literature (for a review see Kroll & Tokowicz, 2005). One topic of particular concern has been the extent to which the bilingual's two languages share a common set of conceptual representations. Past studies have generally demonstrated that words in each of the bilingual's two languages are probably linked to a single shared conceptual store (Francis, 2005). However, much of this research has focused on translation pairs that have a unique one-to-one mapping across languages. Recently, researchers have begun to pay attention to the widespread occurrence of translation ambiguity (Degani, Eddington, Tokowicz & Prior, 2009; Prior, MacWhinney & Kroll, 2007; Prior, Wintner, Lavie & MacWhinney, 2011; Tokowicz, Kroll, De Groot & Van Hell, 2002), and the implications of this ambiguity for bilingual lexical and conceptual representation (Degani, Prior & Tokowicz, 2011; Gathercole & Moawad, 2010; Morford, Wilkinson, Villwock, Pinar & Kroll, 2010). In particular, words that have multiple translations are more difficult to translate (Boada, Sánchez-Casas, García-Albea & Ferré, 2011; Laxen & Lavaur, 2010; Tokowicz & Kroll, 2007) and are harder to learn (Degani & Tokowicz, 2010).

The majority of studies of the bilingual lexicon have focused on the nominal lexicon. As we begin to extend our scope to include verbs and other grammatical word classes, we find additional, often more complicated, patterns of cross-linguistic conceptual overlap (Gentner, 1981; Van Hell & De Groot, 1998). Recent work examining both these issues in concert revealed that verbs were more ambiguous in translation than nouns, both in an offline translation task (Prior et al., 2007) and in large parallel language corpora (Prior et al., 2011).

The present study extends this line of research to examine the implications of word class and translation ambiguity in online, time-sensitive translation production and translation recognition tasks. Specifically, we investigated the possibility that some of the differences in the cross-language overlap between nouns and verbs noted in previous studies might be attributed to the word

* Anat Prior was supported by post-doctoral NRSA grant F32HD049255 and by EU-FP7 grant IRG-249163. The writing of this article was also supported in part by NIH Grant HD053146 and NSF Grants BCS-0955090 and OISE-0968369 to Judith F. Kroll and by the NSF Pittsburgh Science of Learning Center. The authors thank Mercedes Farrell, Anna Guitchounts, Sofia Murra, Shayna Watson and Huiling Yu for data collection and coding. The authors also gratefully thank Prof. Ana I. Schwartz and members of the PROBAR Lab from the University of Texas, El Paso, for help in recruiting participants. Finally, the authors wish to thank two anonymous reviewers who made helpful comments and suggestions on a previous version of the paper.

Address for correspondence:
Anat Prior, Department of Learning Disabilities, University of Haifa, Mount Carmel, Haifa, Israel, 31905
aprior@edu.haifa.ac.il

class differences in translation ambiguity. We adopted the method of hierarchical regression to determine whether word class differences in translation performance remain after removing variance related to differences in other psycholinguistic characteristics, most importantly translation ambiguity.

The current study focuses on the two major grammatical classes of nouns and verbs, that tend to encode different types of meaning (Levin & Rappaport Hovav, 1996; Wierzbicka, 1988). Nouns typically denote entities, are usually perceptually grounded, with meanings that are generally less dependent on the specific linguistic context (but see Barsalou, 1982). Verbs, on the other hand, usually encode relations (Ferretti, McRae & Hatherell, 2001), have more senses (Miller & Fellbaum, 1991), and can be more easily adjusted by contextual demands (for example, metaphorical readings). These word class differences in meaning are reflected in the finding that verbs in general are less concrete and imageable than nouns (Bird, Franklin & Howard, 2001; Chiarello, Shears & Lund, 1999).

Several studies have probed the neural representation of nouns and verbs (Black & Chiat, 2003, Cappa & Perani, 2003; Caramazza & Hillis, 1991). While most have reported dissociations between the two word classes within a language (Damasio & Tranel, 1993; Federmeier, Segal, Lombrozo & Kutas, 2000; Mestres-Missé, Rodriguez-Fornells & Münte, 2010; Perani, Cappa, Schnur, Tettamanti, Collina, Rosa & Fazio, 1999; Pulvermuller, Lutzenberger & Preissl, 1999) and cross-linguistically (Willms, Shapiro, Peelen, Pajtas, Costa, Moo & Caramazza, 2011), others report a large degree of overlap in the areas processing nouns and verbs (e.g., Tyler, Russel, Fadili & Moss, 2001). However, even for those studies claiming distinct brain involvement for the two word classes, it is hard to pinpoint these differences to a specific level of linguistic representation and they might reflect differences not only at the conceptual level, but differences in syntax, morphology, phonology and the manner in which nouns and verbs are used in context. An additional line of research focuses on the order in which lexical items belonging to different word classes are acquired. Concrete nouns are the dominant lexical category in the early speech of children learning a variety of languages, and verbs tend to appear later in child speech (Gentner, 1982; Gentner & Boroditsky, 2001). Similar tendencies have been identified in the acquisition of a second language (Dietrich, 1990), which might be a consequence of the lesser degree of overlap in verb meanings cross-linguistically.

Of specific interest in the current context are studies examining the implications of word class differences for cross-language representation. Gentner (1981) showed that, in a within-language paraphrase task, more nouns than verbs were preserved in the new formulation. Further, higher stability was found for nouns when compared with verbs following a double translation procedure, in which sentences were translated by one bilingual to a different language, and then translated back to the original language by a different bilingual. These findings were interpreted as demonstrating that verbs are more easily adjusted to fit the linguistic context, and that they are characterized by greater cross-linguistic variability. Van Hell and De Groot (1998) examined word class, as well as concreteness, through the use of a word association task, both within and across languages. They found that nouns elicited more associations than verbs, and that concrete words elicited more associations than abstract words. These patterns held in all within- and across-language association conditions of the study. Interestingly, the study also found greater cross-linguistic associative similarity for nouns than for verbs, hinting that nominal translation equivalents share more conceptual features than do verbal concepts. Finally, a direct comparison of the number of different translations given to nouns and verbs in English and in Spanish (Prior et al., 2007) found that, after controlling for effects of frequency and imageability, verbs generated a higher number of possible translations than nouns, confirming and extending the previous findings. This last finding also raises the possibility that cross-language ambiguity in lexical-conceptual mapping might play a major role in the observed performance differences between nouns and verbs. Such increased ambiguity for verbs over nouns could also offer an alternative account for the findings described above, beyond the role played by concreteness *per se*. This issue is addressed directly in the current study.

Concrete words enjoy processing advantages in various linguistic tasks, including lexical decision (Schwanenflugel, Harnishfeger & Stowe, 1988) and translation (e.g., De Groot, Dannenberg & Van Hell, 1994). These advantages might be due to the fact that concrete words can be processed by both imagery and verbal codes (Paivio, 1971), whereas abstract words need to rely on the latter route only, or might reflect the ease with which world knowledge can be instantiated for various words (Schwanenflugel et al., 1988). A first investigation into the consequences of concreteness and translation ambiguity on performance of an online translation production task was reported by Tokowicz and Kroll (2007). That study demonstrated an interaction between the two factors: nouns with only one translation did not show a concreteness advantage whereas nouns with more than one translation were translated more rapidly when they were concrete. Recently, Laxen and Lavaur (2010) examined the consequences of multiple translations and concreteness in a translation recognition task. In one study they found a concreteness effect that did not interact with number of translations (Experiment 1) and in a second study they found concreteness effects only in forward translation (L1–L2) and only when the target word was presented with a low-probability translation. Thus, the present investigation of word class differences and the role of translation ambiguity must also take into account the concurrent differences between nouns and verbs in the types of concepts they denote. In the present study, we chose to operationalize these differences by using a measure of context availability, which is highly correlated with concreteness.

Ambiguity in translation can have different reasons (Prior et al., 2007). In some cases, within-language homonymy might lead to multiple translations. For example, the English word *glass* has two distinct meanings – the material and the drinking vessel. Each of these translates onto a different Spanish word – *vidrio* for the former and *vaso* for the latter. Within-language synonymy can also lead to multiple translations – the Spanish word *sofá* may be translated into English as either *sofa* or *couch*. Word class ambiguity also often results in multiple translations. The English word *cook* can mean either the action, i.e. the verb, in which case it translated into the Spanish *cocinar*, or the person, i.e. the noun, in which case it translates to the Spanish *cocinero*. Finally, there are cases where multiple translations are a result of the differences in the lexical-conceptual mappings of the two languages. The Spanish noun *reloj* covers the concepts denoted by both *clock* and *watch* in English, each of which is a correct translation. In the same way, the meaning of the English verb *know*, which covers both knowing facts and knowing people, is carried by two distinct verbs in Spanish – *saber* for the former and *conocer* for the latter. Although ambiguity is obviously not a uniform phenomenon, the current study does not distinguish between the different sources of translation ambiguity.

Translation ambiguity can be represented in existing models of bilingual language representation, especially those models that distinguish between lexical and conceptual levels of representation. Models such as BIA (Dijkstra, Van Heuven & Grainger, 1998; Grainger, Midgley & Holcomb, 2010), the Revised Hierarchical Model (Kroll & Stewart, 1994), or the Distributed Feature Model (De Groot, 1992) allow for a lexical item in one language to be linked to two different concepts or groups of concepts, which are in turn linked to two different words in the other language, as in the case of within-language homonymy. Within-language synonymy is represented in these models by a concept being linked to only one lexical item in L1, for example, but having links to two lexical items in the L2, or vice versa. The slowing of recognition in translation ambiguity arises in these models from the fact that spreading activation leads to the activation of two or more competing interpretations. The preferred interpretation is then selected during a post-access competition resolution process.

Systematic estimates of the degree of ambiguity in translation across an entire bilingual lexicon are very difficult to achieve, but recently the prevalence of translation ambiguity has been assessed in several bilingual samples. A set of 400 Dutch and English nouns, which had previously been used in a large number of bilingual experiments (e.g., De Groot, 1992) and were assumed to have a single, or a clearly dominant translation, were found to be ambiguous up to 25% of the time (Tokowicz et al., 2002). Due to the nature of the materials, it is likely that this is an underestimate of the actual degree of ambiguity. Supporting this possibility, 60% of a less constrained set of 700 Spanish and English nouns and verbs were found to be ambiguous in translation (Prior et al., 2007; Prior et al., 2011), and 40% of English and German words were found to be ambiguous in translation. Overall, the extent to which translation ambiguity has been underrepresented in previous studies suggests that we need to examine carefully the implications of this factor for bilingual language processing.

Of the few studies that have examined the consequences of translation ambiguity for bilingual performance, all have found it to have a significant impact. In one study, intermediate English–Spanish bilinguals translated ambiguous words more slowly and less accurately than single translation words (Tokowicz & Kroll, 2007). Further, in a training study, Degani and Tokowicz (2010) found that native English speakers had more difficulty learning translation-ambiguous as opposed to translation-unambiguous words. Finally, Laxen and Lavaur (2010) reported slower and more error-prone performance for translation-ambiguous words in a translation recognition task. These authors also distinguished between dominant and non-dominant translations for the translation-ambiguous words, and found a recognition advantage for the former over the latter. The explanation offered was that the task allowed participants to anticipate the translation that would be presented after the target word to be translated, and that they tended to activate dominant translations more often, incurring a cost when the non-dominant translation was then presented. Boada et al. (2011) report similar findings for a sample of translation-ambiguous words in Spanish and Catalan.

The current study examines the issue of translation ambiguity in a more graded fashion, by examining the probability of different translations as a continuous variable. Translation probability, a finer grained measure of ambiguity, is calculated from norming studies, in which bilinguals each provide a single translation for the target word, allowing a distinction between less and more probable translation choices. Importantly, translation probability does not make a dichotomous distinction between dominant and non-dominant translations, allowing a more complete description of cross-language lexical mappings. Thus, the word *carne* in Spanish is translated as *meat* with a probability of .9, and as *flesh* with a probability of .1, a situation that lends itself easily to distinguishing dominant and non-dominant translations. However, many words do not fall easily into this pattern. For example, the Spanish verb *asistir* is translated as *attend* with a .44 probability, and as *assist* or *help* with a probability of .28 each. Further, even in cases of only two possible translations, the degree of dominance or the difference in the probabilities of these options can vary greatly. We therefore use translation probability, as derived from the Prior et al. (2007) norms, as opposed to just the binary dominance dimension, as a predictor of translation recognition and translation production in the current study.

Predictions regarding the possible consequences of translation probability for translation production are less straightforward than the dominance effects demonstrated for translation recognition, because the paradigm does not include a stage during which possibly incorrect expectations can be generated. In the production task, bilinguals are likely instead to accept and produce the first plausible translation that reaches activation threshold. Here the question is whether, when bilinguals choose to produce a less-probable translation for a certain word, as determined by normative data, they suffer a cost for this choice. Such a cost could be seen as reflecting the need to overcome increased competition from a more probable translation during the activation and selection process. Similar processes have been described in the general case where unbalanced bilinguals produce a word from the weaker L2 and need to overcome competition from the more dominant L1 alternative (Costa, 2005; Kroll, Bobb & Wodniecka, 2006; La Heij, 2005). Thus, the question arises whether similar mechanisms underlie the resolution of within-language competition between alternative translations and the across-language competition between translation equivalents. Alternatively, the cost of producing a less-probable translation might be a result of retrieval difficulties or imperfect representation. If bilinguals are searching for the high-probability translation but are unable to access it in real time, they may then resort to activating a less probable translation, resulting in reduced performance. We will return to this issue in the discussion.

Previous studies directly comparing translation production and recognition (e.g. De Groot, 1992) have shown that these tasks are sensitive to different aspects of stimuli. The translation recognition task as implemented here consists of two words from the two languages presented sequentially and the participant needs to determine whether the second word is an accurate translation of the first. In the current framework, as was the case in Laxen and Lavaur (2010), translation recognition should show consequences of the activation of multiple possibilities, because of the delay between the presentation of the target word and its translation. During the final stages of this task, the decision process must compare the suggested translation displayed on the computer screen with the projected translations generated from the target word. If the most highly activated projected translation matches the visual input, then reaction times should be relatively fast. However, if the most highly activated projected translation is a mismatch, if there is close competition between projected translations, or if no translation is activated, then reaction times will be slower.

Translation production, on the other hand, can provide a purer measure of competition for selection, because a single lexical form has to be identified and spoken, unlike the case for the recognition task. Moreover, the translation production task does not involve additional stages of matching to a visual target and does not include variability produced in trials when no anticipatory translation is generated or when close competitors are generated. Further, we expect production but not recognition to show effects of direction of translation (slower and more error-prone forward than backward translation), as has been demonstrated previously (Kroll, Michael, Tokowicz & Dufour, 2002: Kroll & Stewart, 1994 but see La Heij, Kerling & Van Der Velden, 1996).

Finally, previous research has demonstrated that different stimulus properties and experimental manipulations do not influence the performance of all bilinguals in the same way (Talamas, Kroll & Dufour, 1999, Tokowicz, Michael & Kroll, 2004), and this might be the case for translation ambiguity as well. Bilinguals who are more proficient in the L2 might have learned more efficient ways of coping with ambiguity (see Blumenfeld & Marian, 2011, for evidence that proficient bilinguals are better able than monolinguals to inhibit competing lexical alternatives in spoken word recognition). Similarly, individual differences in cognitive resources such as working memory capacity might also prove important in this context. Thus, bilinguals who have higher working memory spans might have better ability to navigate and successfully resolve ambiguity than lower-span individuals, because they have more resources at their disposal for resolving competition and selecting one of several options (Gernsbacher & Faust, 1991). However, the opposite pattern is also possible, namely that greater resources might enable high-span bilinguals to maintain activation for two competing translations more easily than low-span bilinguals, thus leading to stronger effects of ambiguity and reduced performance for ambiguous as opposed to non-ambiguous target words (Miyake, Just & Carpenter, 1994). This pattern of larger ambiguity effects for high-span individuals also receives some support in the literature (Just & Carpenter, 1992; Kroll et al., 2002; Tokowicz et al., 2004). To investigate these possibilities we also included an analysis of ambiguity effects in translation recognition and translation production, and examined to what degree they were correlated with

Table 1. *Participant characteristics, means and SDs.*

	English dominant (N = 34, 19 females)	Spanish dominant (N = 30, 17 females)
Age	25.5 (9.6)	30.7 (10.8)
Years studying L2	14.3 (10.2)	13.8 (8.9)
L1 Self-rated proficiency	9.5 (0.8)	9.7 (0.8)
L2 Age of Acquisition*	10.9 (7.3)	16.6 (10.5)
L2 Self-rated proficiency	7.5 (1.3)	7.9 (1.2)
L1 Use**	4.8 (0.3)	3.7 (0.7)
L2 Use**	3.2 (0.9)	4.6 (0.5)
L2 Lexical Decision D′**	1.5 (0.4)	2.3 (0.7)
O-span math (accuracy, max = 60)	44.1 (16.7)	40.3 (14.5)
O-span word (accuracy, max = 60)	53.3 (7.2)	52.8 (5.1)

* Groups significantly different, $p < .05$
** Groups significantly different, $p < .001$
Notes: (i) Proficiency self-ratings are on a scale from 1 (not at all) to 10 (perfect command) and are averaged across oral and written comprehension and expression. (ii) Language use reports are on a scale from 1 (almost never) to 5 (every day), and are averaged across oral and written comprehension and expression, and media exposure.

individual differences in L2 proficiency and working memory.

To summarize, the current study assessed the independent contributions of ambiguity and word class to online translation performance. These influences are examined in two frequently used bilingual tasks – translation recognition and translation production. Importantly, ambiguity is treated as a continuous variable, and not dichotomized as in previous research. Based on prior findings, we predicted that translation ambiguity will lead to decreased performance, both in accuracy and in speed, in both the experimental tasks. We further predicted that the impact of ambiguity will remain significant even after statistically controlling the influence of other psycholinguistic word characteristics that have been found to influence translation performance.

Based on the extensive literature on differences between nouns and verbs in learning, processing, and neural representation, a straightforward prediction would be that bilinguals would encounter greater difficulties in the cross-linguistic processing of verbs relative to nouns. Thus, one can expect slower reaction times and lower accuracy for verbs relative to nouns in both translation tasks. However, none of the previous studies controlled for the significant differences in degree of ambiguity between nouns and verbs. Therefore, it might be the case that previous findings of superior performance and greater cross-language overlap for nouns over verbs (Gentner, 1981; Van Hell & De Groot, 1998) might actually be a consequence of greater ambiguity in translation for verbs over nouns, and not a correlate of word class *per se*. This issue will be addressed directly in the present study by examining whether word class remains a significant predictor of performance even after controlling for translation ambiguity.

Finally, with respect to individual differences, we hypothesize that higher L2 proficiency will lead to more efficient ambiguity resolution. In the production task, higher L2 proficiency bilinguals will find it easier to activate target forms. In the translation recognition task, proficiency levels may not be as important as working memory capacity, or the ability to quickly process and inhibit incorrectly anticipated words and then to boost the activation of secondary targets.

Method

Participants

Sixty-four bilingual speakers of Spanish and English participated in the study (26 males), ages 18 to 56 years, mean age of 28 years.[1] All listed Spanish and English as their two strongest languages, even if they had some knowledge of other languages. We recruited highly proficient bilinguals – selection criteria included studying the second language for a minimum of 5–6 college semesters or having commensurate language experience. Table 1 describes participant characteristics, separately for Spanish-dominant and English-dominant participants. For both groups, the average age of acquisition (AoA) of L2 was well past early childhood. Sixty of the participants completed the experiment in Pittsburgh, PA, and four participants completed the study in El Paso, TX.

[1] The population of the current study spanned a wider range of adult bilinguals and was not limited to young, college-age students.

Table 2. *Word characteristics, means and SDs.*

	English			Spanish	
	Nouns (N = 87)	Verbs (N = 76)	Nouns/Verbs (N = 237)	Nouns (N = 218)	Verbs (N = 182)
Word frequency	104.7 (112.7)	189.8 (747.2)	142.4 (308.9)	99.7 (145.6)	62.5 (161.2)
Word length	6.14 (2.3)	5.3 (1.8)	4.7 (1.4)	6.28 (2.0)	6.5 (1.6)
Number of translations	1.9 (1.2)	2.0 (1.2)	2.8 (1.6)	1.7 (1.1)	2.2 (1.6)
Context availability	5.3 (0.7)	5.2 (0.6)	5.3 (0.6)	5.4 (0.7)	5.2 (0.6)
Concreteness	4.6 (1.3)	3.6 (0.7)	4.5 (1.1)	4.8 (1.1)	4.1 (0.9)
Form similarity	2.8 (3.7)	1.0 (2.5)	1.5 (2.9)	2.3 (3.4)	1.0 (2.4)

Notes: Word frequency is measured in occurrences per million. Word length in letters. Context availability was rated on a scale of 1–10, and concreteness on a scale of 1–7; cognate rating was on a scale of 1–10.

Participants signed an informed consent, and were paid for their participation.

Materials

Four hundred translation pairs in Spanish and in English were used. Table 2 presents word frequencies (Kucera & Francis, 1967, for English; Pérez, Alameda & Cuetos, 2003, for Spanish), word length, grammatical class, number of translations (Prior et al., 2007) and context availability ratings. Table 2 also provides descriptive information on concreteness ratings, although we used context availability, rather than concreteness, as a variable in our design. In addition, Table 2 shows that over half of the words in English (N = 237) were ambiguously nouns or verbs. The cognate ratings given in Table 2 were taken from the Prior et al., 2007 norms. These ratings were generated by monolingual English speakers who performed a translation elicitation task (Dufour & Kroll, 1995; Kroll & Stewart, 1994). Such ratings are comparable to ratings obtained from bilingual speakers (Friel & Kennison, 2001). These cognate ratings ranged from 0 (no cross-language overlap) to 10 (highly similar lexical forms). For more details see Prior et al., 2007.

For the translation recognition task, two of the possible translations were selected for each of the ambiguous words. Words were selected from the set normed by Prior et al. 2007. A word was deemed ambiguous if it was given at least two correct translations by different participants in the norming study. When the word had two possible translations, both were included in the recognition list. When there were more than two possible translations, we selected the highest and the lowest probability translations for each item. Thus, in the recognition task the unambiguous items were always paired with the same translation. Two versions of the materials were constructed for the ambiguous items, such that each item was paired with a high probability translation in one version and with a low probability translation in the other version. Unrelated word pairs were constructed by sampling words matched in length, frequency, and grammatical class to the target words, but unrelated in meaning to the cue words.

To control for the presence of cognates among the correct translations in the translation recognition task (roughly 20% of the trials across the different lists), a similar percentage of form-related pairs were included among the unrelated controls (roughly 18% across the experimental lists). For example, the Spanish word *playa*, which means "beach", was presented as the target for the English cue word *play*. Thus, participants had to ignore the form similarity of the words and correctly respond "no", since the Spanish word is not an accurate translation of the English target word. The inclusion of such unrelated controls meant that form similarity was rendered an unreliable cue for the translation accuracy decision. As mentioned above, unrelated controls were always selected from the same grammatical class as the correct translations, since it has been demonstrated that bilinguals can utilize grammatical class in performing translation recognition (Sunderman & Kroll, 2006).

In the translation production task the first word presented in translation recognition was designated as the target to be translated. Four lists of 100 items each were created so that one list was presented in each language within each of the two experimental tasks. In translation recognition, half of the 100 words in a given list were paired with their correct translation and the other half were paired with the matched unrelated control. Overall, the lists were matched on frequency, length, the percentage of nouns, verbs and class-ambiguous words, the percentage of cognates and the percentage of translation-ambiguous words. A full set of the materials used in both tasks can be found in the online supplementary materials.

Procedure

In the first stage of the experiment, participants completed the translation recognition and translation production tasks. The order of these tasks was counterbalanced across participants, as was the order of the languages. In each of the tasks, half of the participants performed Spanish-to-English translation first, and the remainder started with English-to-Spanish translation. The order of translation directions was held constant for each participant across two tasks so that participants who first translated from Spanish to English also first performed translation recognition in which the Spanish word appeared prior to the English word within a given pair.

After completing the translation tasks, participants were given a short break and then completed a language history questionnaire (LHQ), a lexical decision task in the L2 and a working memory task (operation-span), which are described below.

Translation production

On each trial a word in one language appeared on the screen. The word remained on the screen until the participant triggered the voice key by producing the translation or by saying “no”, “I don’t know” or their Spanish equivalents. When the voice key was triggered or after a maximum duration of 4 seconds, the word was replaced by a fixation point. Participants pressed a button to initiate the following trial. Participants completed one block of 100 words in one translation direction, and following a brief rest, went on to produce translations for 100 words in the other direction. Responses were recorded and coded for accuracy and the specific translation produced offline.

Translation recognition

Each trial started with a 500 ms fixation point, followed by a 50 ms blank screen and then by a word in one language. The single word remained on the screen for 150 ms, and then a word in the other language appeared below it. Both words remained on the screen until participants made a response, or for a maximum of 5 seconds, to preserve similarity with the production task. Participants pressed a button to indicate whether the bottom word was an accurate translation of the original target word. Once again, trials were blocked by language. There were 100 trials in each language, 50 that were correct translations and 50 that were unrelated words. Of the correct translation trials, 1/3 were single translation-unambiguous words, 1/3 were ambiguous words paired with a high probability translation and 1/3 were ambiguous words paired with a low probability translation. Items were rotated across conditions for different participants, so that no participant viewed a given item in more than one condition.

Language history questionnaire

Participants provided information about their language learning experience and rated their self-perceived proficiency in each language (Table 1). Language dominance was assessed as follows: If there were differences in the self-ratings of proficiency in the two languages on the LHQ scales, the language rated by the bilingual as his or her stronger language was assumed to be the dominant language. Participants who rated themselves equally proficient in English and in Spanish were questioned orally, and asked which language they would select as being their stronger language. If they were able to make such a choice, their assigned dominance reflected this choice. Finally, there were three participants who were unable to make the determination and we assumed them to be English dominant, by virtue of currently residing in a predominantly English-speaking environment. This procedure resulted in 34 participants being designated as English dominant, and 30 being designated as Spanish dominant.

Lexical decision

participants performed a lexical decision task in their L2, as determined above. Two versions of a lexical decision task were developed, one in English and one in Spanish. The procedure for selecting the words was based on that described by Kempe and MacWhinney (1996). Each list included 168 words and 168 orthographically and phonotactically legal non-words. Each trial started with a fixation point presented for 500 ms, followed by the target string which remained on the screen until the participant made a response, by button press, using a response box. The task took approximately 15 minutes to complete. D-prime (d′) measures of performance reflected participants’ ability to discriminate between actual words and non-words in the language. The d′ score was used as an added online measure of L2 proficiency, to augment self-ratings.

Operation-span task

This is a working memory task adapted from the operations-words task (O-Span) introduced by Turner and Engle (1989). Participants solved mathematical expressions, while maintaining sets of words in memory. In each trial, a fixation cross appeared in the middle of the screen for 1000 ms, followed by a single mathematical expression, which remained on the screen for 2500 ms, and was replaced by a question mark appearing for 1250 ms. While the question mark remained on the screen, participants had to push a button indicating whether the mathematical expression was correct or incorrect. Upon response, or time out, the question mark was replaced with a word appearing for 1250 ms. Participants had to retain the words in memory until the end of the set. At the end of a set, a recall prompt appeared on the screen, at which point participants wrote down as many words as they recalled from that set in a booklet, and pressed a button to initiate the following set. Sets ranged in size from two to six operation-word pairs per set, and were presented in ascending order, with three sets of each size, for a total of 15 sets. Each set included approximately equal numbers of correct and incorrect mathematical expressions. Before completing the experimental sets, participants performed two practice sets (one with four items and one with six items). Two versions of this task were constructed, one in English and one in Spanish. Participants completed the operation-span task in their dominant language.

Results

The data analysis approach we adopted in this study was to use hierarchical regression to assess the impact of predictor variables on the dependent measures. Thus, we allowed many lexical properties of the stimuli to vary simultaneously, and did not attempt to create orthogonal conditions by matching all variables of interest. We believe this approach allows for the creation of more representative and natural stimulus lists, and has the added advantage of avoiding the problems associated with dichotomizing continuous variables (Cohen, 1983; MacCallum, Zhang, Preacher & Rucker, 2002).

In all cases, predictor variables were entered into the regression models based on knowledge gained from previous research. Namely, variables that have demonstrated more stable influences on performance were always entered in the early steps of the analysis, followed by less well-established variables. Finally, the novel variables introduced in this study were always entered last into the regression model, so as to explore any effect remaining after accounting for the variance attributed to all other variables.

Item analysis

Translation production

Upon examination of the data, it became apparent that several of the items were rarely translated accurately and therefore all the responses given to six English words (*aim*, *crawl*, *curse*, *drill*, *hesitate*, *lead*) and one Spanish word (*reto*) were eliminated from all further analysis. This resulted in 6256 English-to-Spanish and 6368 Spanish-to-English translation trials.

Reaction time analysis

For the reaction time (RT) analysis, the data were further trimmed by eliminating the following trials types for the English-to-Spanish and Spanish-to-English translation directions, respectively: error trials (8.0% and 6.7%), no response trials (3.4% and 1.9%), trials on which the participant stated he or she did not know the translation (5.2% and 3.7%), voice key failures, including RTs shorter than 300 ms and longer than four seconds (3.5% and 4.0%), and trials on which participants produced a correct translation but one which did not appear in the original translation norming (2.6% and 1.5%). Thus, for English-to-Spanish translation 22.9% of the data were removed and for Spanish-to-English translation 18.0% of the data were removed. Because of our approach to data analysis, we did not follow common procedures for eliminating responses that differed from individual participant means. Instead, absolute reaction time criteria were applied to all participants (less than 300 ms or more than 4000 ms). Differences across participants with respect to proficiency and/or dominance were later examined within the context of the regression analyses.

The average RT across participants was calculated for each unique combination of target word and translation. For example, if a given target word elicited three different correct translations across participants, RTs for each of the three pairings of targets and translations contributed separate data to the analyses. Separate averages were calculated for participants performing forward and backward translation. Thus, when examining Spanish-to-English translation, the reaction times of Spanish-dominant and English-dominant participants were averaged separately. This procedure resulted in a total of 2031 translation pairs, each one produced by between one and thirteen participants, with an average of 4.7 responses per pair, an SD of 2.9.

Seven predictor variables were entered into the regression model, in the following order: the average length and frequency (Kucera & Francis, 1967, for English, and Davis & Perea, 2005, for Spanish) of the target and translation words, the direction of translation (forward or backward) and the cognate rating of the pair (Prior et al., 2007). Because the stimuli materials included both nouns and verbs, including many class-ambiguous words in English, we deliberated whether or not to use previously collected concreteness, familiarity or imageability ratings. This is because we were not certain whether these constructs apply to both word classes in the same manner. In addition, previous studies in which such norms were collected did not always distinguish between the different readings of ambiguous words. Finally, De Groot (1992) identified context availability as a better predictor of translation production and recognition than imageability. Therefore, we decided to collect context availability ratings for the experimental materials. Thus, 10 native English speakers rated how easy it was for them to imagine a context in which each of the English words would be used on a seven-point scale (using the original instructions appearing in Schwanenflugel et al., 1988). In this procedure nouns and verbs were clearly marked (nouns were preceded by an indefinite article, and verbs were presented in the infinitive preceded by *to*). None of the bilinguals who participated in the main experiment completed the context availability ratings. The mean value of these ratings across the word sample was 5.4, with an SD of 0.6. Thus, context availability was entered as the fifth predictor variable in the regression model, followed by the probability of the translation (Prior et al.,

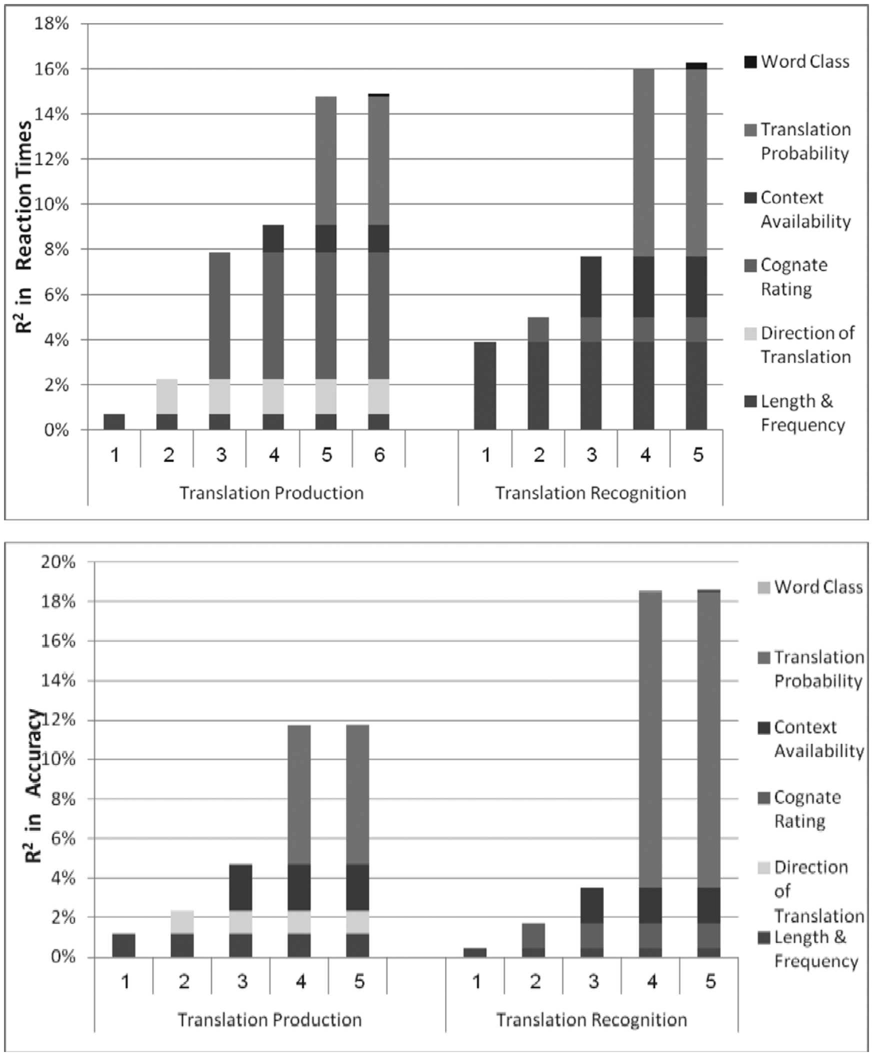

Figure 1. Explained variance in reaction times (top) and accuracy (bottom) for translation production and translation recognition for each of the predictor variables entered in the hierarchical regression models.

2007), and finally the word class (noun or verb). English word class ambiguous items were coded depending on the specific translations that they received in Spanish. Because Spanish morphology allows easy distinction between nouns and verbs, each target-translation pair was coded on the basis of whether the Spanish translation was a noun or a verb. Because of the extremely large size of the sample, we adopted a conservative level of statistical significance of $p < .001$, in an attempt to avoid a preponderance of significant effects.

The average length and frequency of the target and translation words significantly predicted RT, $R^2 = .007$, $F(2,2028) = 7.13$, $p < .001$ (Step 1). Translation from L2 to L1 was faster than translation from L1 to L2 (1316 ms vs. 1420 ms) $\Delta R^2 = .016$, $F(1,2027) = 32.5$, $p < .001$ (Step 2). Cognate translations were produced more quickly than translations with low cognate ratings $\Delta R^2 = .056$, $F(1,2026) = 122.56$, $p < .001$ (Step 3), as predicted. Translation pairs high in context availability were produced more quickly than words low on this variable, $\Delta R^2 = .013$, $F(1,2025) = 27.95$, $p < .001$ (Step 4). Again as predicted, and replicating previous findings of the role of translation probability in translation production, highly probable translations were produced more quickly than less probable translation options $\Delta R^2 = .057$, $F(1,2024) = 135.42$, $p < .001$ (Step 5). Finally, after accounting for the role of all previous variables, the effect of word class was not significant, although noun translations were produced somewhat more quickly than verb translations, $\Delta R^2 = .001$, $F(1,2023) = 2.98$ $p = .08$ (Step 6) (see Figure 1).

We probed the additional impact of three two-way interactions of specific theoretical significant. Thus, we wished to see whether the effect of translation probability differed for nouns and verbs (Probability × Word Class interaction) or across the two translation directions (Probability × Direction). The third question we addressed was whether context availability influenced nouns and verbs similarly (Context Availability × Word Class). The three interaction terms were entered into the model as the final step of the model, but were not found to contribute significantly to the explained variance in RT, $\Delta R^2 = .002$, $F(3,2021) = 1.43$ $p = .23$ (Step 7).

To summarize, the two most significant predictors of translation reaction time, each accounting for approximately 6% of the variance, were cognate status and translation probability. All other variables examined accounted for less than 2% of the variance. The model overall accounted for 15% of the variability in translation production reaction time.

Accuracy analysis

The percent of accurate translations given to each target word was calculated, again separately for each direction of translation. Specifically, for each word we counted how many of the participants that were required to translate it were able to come up with a correct translation. In this analysis translations that did not appear in the norming study but were correct translations of the target word were considered correct responses. Thus, each target word received a single accuracy value, regardless of whether different participants produced a single translation or one of several options. Each target word was translated by between three and thirteen participants, with an average of eight responses per item, per direction of translation, with an SD of 2.6. Several differences were introduced in the variables used in this regression model, to account for the fact that only properties of the target word could be used, because accuracy was collapsed across different possible translations. Thus, only the length and frequency of the target word were used. Cognate rating was not entered into the model, as it requires a unique translation to be identified. Finally, the translation probability measure, which depends again on a unique cue-translation pair, was replaced by an alternative measure of ambiguity, namely the number of different translations identified for the target word in the translation norms (Prior et al., 2007).

Overall, the pattern of results emerging from the accuracy analysis was quite closely aligned with the pattern of the RT results, therefore reducing the likelihood of speed-accuracy tradeoffs. The average length and frequency of the target words significantly predicted accuracy, $R^2 = .012$, $F(2,1567) = 9.17$, $p < .001$ (Step 1). Translation from L1 to L2 was less accurate than translation from L2 to L1, $R^2 = .012$, $F(1,1566) = 19.01$, $p < .001$ (Step 2). Words high in context availability were translated more accurately than words low on this variable, $\Delta R^2 = .023$, $F(1,1565) = 37.99$, $p < .001$ (Step 3). Number of possible translations was negatively correlated with accuracy, such that ambiguous words were translated less accurately than unambiguous words $\Delta R^2 = .071$, $F(1,1564) = 124.97$, $p < .001$ (Step 4) (see Figure 1). Again, as in the RT analysis, there were no significant effects of word class on the accuracy of translation $\Delta R^2 = .00$, $F(1,1563) < 1$, $p = .5$ (Step 5).

As in the analysis of reaction times, we examined the added explanatory power of three two-way interactions (Probability × Word Class, Probability × Direction of translation and Context Availability × Word Class). The three interaction terms were entered as the final step of the model, but as in the analysis or RTs, were not found to contribute significantly to the explained variance in accuracy, $\Delta R^2 = .002$, $F(3,1561) = .98$, $p = .39$ (Step 6).

Translation recognition

Reaction time analysis

Invalid recognition trials were excluded, including 56 trials with reaction times faster than 400 ms and 95 trials with reaction times longer than 4000 ms. Average reaction times to correct "yes" responses were calculated separately for forward and backward translation, and each item received correct responses from between one and seven participants, with an average of 2.79. Initial analyses demonstrated that direction of translation was not a significant predictor of translation recognition, for either reaction time or accuracy. In light of this, and due to the relatively low number of responses per item, we decided to collapse our analysis across forward and backward translation, and recalculated the average RT and accuracy rates for each item. Each item now received between one and eight correct responses, with an average of 4.6 and an SD of 2.4.

The average RT across participants was calculated for each unique combination of target word and translation. Thus, ambiguous translation words were analyzed both with their high- and with their low-probability translations. This procedure resulted in a total of 1031 translation pairs. The predictor variables included in the model were identical to those described in the translation production RT analysis, and the resulting pattern of results was overall similar across the two tasks, with several differences that will be highlighted below.

The average length and frequency of the target and translation words significantly predicted recognition latency, $R^2 = .039$, $F(2,1029) = 20.84$, $p < .001$ (Step 1). Cognate translations were recognized more quickly than translations with low cognate ratings $\Delta R^2 = .011$, $F(1,1028) = 11087$, $p < .001$ (Step 2). Translation pairs high in context availability were recognized more quickly than words low on this variable, $\Delta R^2 = .027$, $F(1,1027) = 30.14$, $p < .001$ (Step 3). As predicted, and extending previous findings regarding the role of translation probability beyond the translation production task, highly probable translations were recognized more quickly than less probable translation options $\Delta R^2 = .083$, $F(1,1026) = 101.02$, $p < .001$ (Step 4). Indeed, translation probability proved to be the strongest predictor variable in the model. Finally, as in the translation production task, word class did not significantly affect translation recognition latency, after all other variables had been accounted for $\Delta R^2 = .003$, $F(1,1025) = 3.6$, $p = .06$ (Step 5) (see Figure 1). The two-way interaction terms between Translation Probability × Word Class and between Context Availability × Word Class were added on the final step of the model (Step 6), but did not add any explained variance ($F < 1$).

Accuracy analysis
The percent of accurate responses was calculated for each target-translation pair, as presented to participants in the recognition task. The predictors included in the accuracy analysis were exactly the same as those included in the RT analysis, since the unique target-translation pairings were preserved, because both words were presented to the participants. Each such pair was presented to between one and eight participants, with an average of 5.1 participants per pair and an SD of 2.5. Thus, 1047 translation pairs were included in the accuracy analysis.

The average length and frequency of the target and translation words were not significant predictors of recognition accuracy, $R^2 = .004$, $F(2,1044) = 2.25$, $p = .11$ (Step 1). Cognate translations were recognized more accurately than translations with low cognate ratings $\Delta R^2 = .013$, $F(1,1043) = 13.66$, $p < .001$ (Step 2). Translation pairs high in context availability were recognized more accurately than pairs with low context availability, $\Delta R^2 = .018$, $F(1,1042) = 19.11$, $p < .001$ (Step 3). Strikingly, once again, translation probability was a significant and strong predictor of recognition accuracy, with highly probable translations being recognized more accurately than less probable translations, $\Delta R^2 = .15$, $F(1,1041) = 191.96$, $p < .001$ (Step 4). Word class did not significantly predict recognition accuracy, $\Delta R^2 = .00$, $F(1,1040) < 1$, $p = .78$ (Step 5) (Figure 1). Finally, as in previous analyses, the two-way interaction terms between Translation Probability × Word Class and between Context Availability × Word Class were added on the final step of the model (Step 6), but did no add any explained variance $\Delta R^2 = .004$, $F(2,1038) = 2.8$, $p = .06$.

To summarize the findings of the item analyses, although there were some task-dependent differences between translation production and recognition, translation probability emerged as the strongest predictor of performance across both translation tasks. Direction of translation was a significant factor only in the production tasks and even there it was not as important as translation probability and cognate status. In contrast to these predicted effects, the translation performance for nouns and verbs seems to be highly similar, after differences in other psycholinguistic variables (such as context availability and most notably translation ambiguity) were accounted for. Finally, the effects of translation probability and context availability were stable across word class, further demonstrating the similarity of noun and verb translation.

Individual differences analyses

Average performance was calculated separately for ambiguous and unambiguous words for each participant within each task. The effect of ambiguity was measured by subtracting the performance on ambiguous words from the performance for unambiguous items, for both RT and accuracy. Thus, for the translation production task, two scores were calculated for each participant – ambiguity-driven RT differences and accuracy differences. For the translation recognition task, these same measures were computed, and in addition, performance was compared for the low- and high-probability translations, within each participant. Thus, overall four difference scores were calculated for each participant, two for the difference between ambiguous and unambiguous words (RT and accuracy) and two for the difference between high-probability and low-probability translations for the ambiguous items (RT and accuracy).

Participants' self-rated L2 proficiency and their reported daily use of the L2 were examined as possible predictors of the influence of ambiguity on translation performance. Additionally, the d′ score from the L2 lexical decision task was taken as an objective measure of L2 vocabulary knowledge. This score indicates the participant's ability to distinguish between real words and pseudo-words, and thus reflects vocabulary size and word knowledge. Self-rated proficiency was moderately correlated with reported use ($r = .357$, $p < .005$), and both measures were also significantly correlated with performance in the lexical decision task in the L2, as expressed in the d′ measure ($r = .466$, $p < .001$ for proficiency, and $r = .533$, $p < .001$ for use).

In these analyses we also examined the putative role of working memory capacity on ambiguity effects in translation production and recognition. Memory for words and correct responses to the mathematical equations in the operation-span task were averaged and used as a single measure of working memory capacity (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005).

For the translation production task, correlation analyses showed that all three measures of L2 proficiency were significantly negatively correlated with the decrease in accuracy for producing translations for ambiguous as opposed to unambiguous words. ($r = -.39$, $p < .01$ for the lexical decision d′, $r = -.37$, $p < .01$ for L2 self-rated proficiency, and $r = -.25$, $p < .05$ for self-reported L2 use). Thus, participants who were less proficient in the L2 had a greater disparity in accuracy rates between ambiguous and unambiguous words than more proficient participants. All three proficiency measures were also negatively correlated with the differences in mean translation time between ambiguous and unambiguous words. The cost in RT for ambiguous as opposed to unambiguous stimuli was not significantly correlated with the lexical decision d′ ($r = -.12$, $p = .31$), was marginally correlated with L2 self-rated proficiency ($r = -.21$, $p = .09$) and was significantly correlated with L2 reported use ($r = -.33$, $p < .01$). Thus, participants who were more proficient in the L2 suffered from a smaller delay in reaction times when producing translations for ambiguous as opposed to unambiguous words, when compared with less proficient speakers of the L2. Working memory span was not related to ambiguity driven differences in performance in the translation production task ($p > .4$ for accuracy and RT).

In the translation recognition task, some similarities and some differences emerged. When examining performance differences driven by ambiguity, namely a comparison between ambiguous and unambiguous items, we again found a role for L2 proficiency. Thus, L2 self-reported proficiency was negatively correlated with the ambiguity-driven cost in accuracy ($r = -.32$, $p < .01$). Thus, participants who were less proficient in the L2 showed a greater decrease in their ability to correctly recognize translation for ambiguous relative to unambiguous words, when compared with more proficient L2 speakers. However, none of the proficiency measures were related to the ambiguity-driven cost in RT in the translation recognition task (all *p*s >.25). Finally, similar to what was found in the translation production task, working memory span was not correlated with ambiguity-driven costs in either RT or accuracy (both *p*s >.2).

When we examined performance costs driven by the probability of the specific translation presented for the ambiguous words, an interesting and distinct pattern emerged. For the first time, we found working memory span to be related to performance, as there was a significant negative correlation between working memory span and the probability-driven cost in RTs ($r = -.27$, $p < .05$). Thus, participants with higher working memory span incurred less of a cost in reaction time, when recognizing low- as opposed to high-probability translations. We hypothesize that this pattern reflects the enhanced ability of the high-span participants to negotiate and successfully resolve the ambiguity, a process that the lower-span participants found more taxing. Additionally, there was a significant positive correlation between self-reported L2 use and the probability-driven cost in accuracy ($r = .31$, $p < .05$). This indicates that participants who reported using the L2 a greater percent of the time actually made more mistakes of incorrectly rejecting low probability translations of ambiguous words than did participants who reported using the L2 less often. One tentative interpretation of this admittedly surprising pattern might be that the more time spent using the L2 the stronger becomes the tendency to activate the high-probability translation of ambiguous words, and consequently the mismatch upon the presentation of the low-probability alternative in the translation recognition task leading to errors. We shall return to this issue in the discussion.

Finally, in the translation recognition task, there were significant correlations between the cost incurred for ambiguous as opposed to unambiguous words and the cost incurred for low- versus high-probability translations ($r = .43$, $p < .001$ for RT, and $r = .39$, $p < .01$ for accuracy). This finding might hint that similar cognitive and/or linguistic mechanisms are recruited for resolving these two types of competition in translation recognition, though of course a correlation is only suggestive in this case.

Discussion

The present study investigated the cross-linguistic representation of nouns and verbs, as reflected in the performance of translation recognition and translation production tasks. The translation of words from different grammatical classes was examined after controlling for the differences in other psycholinguistic properties that correlate with word class, most importantly ambiguity in translation. Our results suggest that, after such differences are accounted for, any differences in the processing of nouns and verbs are greatly reduced if not eliminated altogether. Additionally, the current study demonstrates the impact of ambiguity in translation on the performance of bilinguals, in almost all the tasks and measures examined. The effect of ambiguity also interacted with participant working memory and L2 proficiency. In the remainder of the discussion we will address each of these findings, as well as several methodological issues and directions for future research.

As described in the introduction, most research on lexical processing in bilinguals has focused on nouns, and to the best of our knowledge the only studies comparing the representation and processing of nouns and verbs and demonstrating word class differences used offline measures of performance (Gentner, 1981; Prior et al., 2007; Van Hell & De Groot, 1998). Importantly, these studies all support the idea that verbs are more ambiguous in translation than nouns, and tend to have a less direct meaning overlap across languages. Additional within-language research has shown word class differences in the characteristic meanings of nouns and verbs, with the former tending to be more concrete and enjoying higher context availability than the latter (Bird et al., 2001). Therefore, in the current study, we jointly examined the influence of context availability, translation ambiguity and word class on translation performance.

The results across the two translation tasks that we examined are consistent for both nouns and verbs. Thus, within-language psycholinguistic variables including word length, frequency and context availability, seemed to influence translation performance similarly across both word classes – shorter, more frequent and higher available words and concepts were translated and recognized more quickly and accurately, regardless of whether they were nominal or verbal. The same was true for across-language word characteristics – cognate nouns and verbs tended to be translated and recognized more quickly and accurately. We also found faster and more accurate translation production (but not recognition) in the backward direction (from L2 to L1) than in the forward direction (from L1 to L2), as predicted by the Revised Hierarchical Model (Kroll & Stewart, 1994; for recent discussion see Brysbaert & Duyck, 2010; Kroll, Van Hell, Tokowicz & Green, 2010). Finally, translation pairs with high translation probability, or fewer alternative translations, were translated and recognized more quickly and accurately, again regardless of their word class. In fact, translation ambiguity proved to be the strongest predictor of performance across both tasks and both word classes. Surprisingly, after accounting for these variables, we found no residual influence of word class, for either RT or accuracy, in translation production and recognition. Thus, bilinguals were equally fast to translate and to recognize the accurate translations of nouns and verbs, once other factors had been controlled.

This finding underlines the importance of considering translation ambiguity in any investigation of word class differences in cross-language representation and processing. It also raises the possibility that previously reported word class differences might in fact be reflecting differences in translation ambiguity, and not necessarily differences between the conceptual representation of nouns and verbs. To illustrate, a finding of greater cross-language similarity in the associative networks for nouns than for verbs (Van Hell & De Groot, 1998) might be a reflection of the fact that verbs tend to have more alternative translations, and therefore their meaning would overlap to a lesser degree with each of the alternatives (see also Tokowicz et al., 2002).

These results are of special importance in light of recent findings that word class information can constrain lexical access and modulate cross-language interference in bilingual settings (Baten, Hofman & Loeys, 2010; Sunderman & Kroll, 2006). Therefore, despite the fact that bilinguals are able to utilize word class information to reach more efficient processing of interlingual homographs and to overcome the competition that is produced by lexical form neighbors, the current study suggests that the basic mechanisms for translating nouns and verbs are fundamentally the same. The regression approach adopted in our work also implies that some performance differences in bilingual word processing across word classes can most likely be ascribed to differences in word and concept characteristics that tend to be associated or confounded with different word classes, most notably the degree of translation ambiguity.

An important question regarding the current finding, namely that translation ambiguity played a stronger role in translation performance than word class, is the extent to which it is characteristic of lexical processing out of sentential context. Along similar lines, Sunderman and Kroll (2006) found that conflicting word class information eliminated lexical interference but not semantic interference in out-of-context translation recognition, again showing a strong role for semantics in this paradigm. However, it is possible that different results might emerge when investigating the translation of words in context or of larger segments of text. Under these conditions, where syntactic processes are actively engaged, word class may exert a stronger or at least equal influence to that found for semantic factors, such as translation ambiguity. This is an area ripe for further investigation and future studies of bilingual language processing should address the issue of determinants of lexical processing when words are embedded in wider linguistic context. Future research should also extend the present findings to other lexical classes, most notably adjectives.

The second striking finding of the current study is that, across almost all tasks and measures, translation probability was the strongest predictor of bilingual performance. Thus, a word that tends to be chosen often as a translation for a target word according to normative and corpus data tends to be produced more quickly and accurately in a translation task, although in this task cognate rating was an equally strong predictor, at least for reaction times. High translation probability also facilitates performance in a translation recognition task, aligning well with the results of Laxen and Lavaur (2010) and Boada et al. (2011), who report similar findings using dichotomous measures of translation ambiguity. Our analysis first examined the effects of word frequency and length, direction of translation, context availability and form similarity and replicated previous research demonstrating the influence of these variables on translation performance. Frequency, context availability and cognate status have also been shown to be positively correlated with translation probability (Prior et al., 2007), though they jointly explain only a relatively low proportion of variability in translation probability. However, even after variance in translation performance attributed to these variables was controlled, translation probability emerged as a strong predictor of translation production and recognition.

Laxen and Lavaur (2010) found that translation recognition of words that were unambiguous in translation was faster and more accurate than for words that had two possible translations. They further compared the recognition of dominant and non-dominant translations for the ambiguous words, and found superior performance for the former. The explanation they offer is that in the translation recognition task the translation is presented at a certain delay after the target word (as was the case in the present study) allowing participants to generate a possible translation, most likely the dominant translation in the case of ambiguous words. Then, when they are presented with the non-dominant translation, it does not match the expectation, thus leading to a delay in reaction time and more errors.

The current results raise the possibility that this is not the only mechanism at play, because translation probability was found to be a significant predictor of translation recognition performance across the entire range of probabilities. Thus, even among translations with a probability above .5 (which are by definition dominant translations) higher probabilities lead to faster and more accurate recognition. When examining the lower half of the distribution, namely translations with a probability of less than .5 (which are mostly non-dominant, though might in certain cases be equi-dominant with other possible translations), again translation probability was the strongest predictor of RT (along with frequency) and the only significant predictor of accuracy. Thus, across the range of translation probabilities, higher values allow for easier and more expedient access from a word to its translation. This finding complements the explanation offered by Laxen and Lavaur (2010), as most likely both mechanisms operate in concert.

As described briefly in the introduction, the possible influence of translation probability on translation production is theoretically less straight-forward. On the one hand, because translation production is a task that is mostly internally guided it might be hypothesized that any translation that a bilingual chooses to produce must be the item most highly available at the moment of production. This description would then lead to a finding that translation production might be influenced by normative probability distributions to a lesser degree than translation recognition, or indeed not at all. The present data show otherwise – normative translation probability was a strong predictor of translation production performance. Importantly, this relationship emerged after controlling for other variables known to facilitate translation production (e.g., frequency and cognate status). Therefore, it seems that in order to produce a less probable translation bilinguals need to overcome interference from other possible translations, most specifically those with a higher normative probability. Why then would a bilingual produce a lower probability translation? This might happen as a result of a momentary difficulty in retrieving the higher probability option, or it may be due to stochastic properties of the lexical representation network in the bilingual lexicon that allow for different possible responses to be produced, similar to the case in monolingual naming contexts (Peterson & Savoy, 1998). In this regard it is also interesting to note that unbalanced bilinguals produce lower probability alternatives constantly, whenever they speak in the L2 as opposed to the L1 (e.g., in naming a picture in L2 rather than L1). Thus, although it is not clear whether the ability to produce a lower probability translation in the current design is more analogous to producing less dominant alternatives within a language or across languages, this ability is crucial in enabling L2 use more generally.

In this regard, the differential modulatory influence we demonstrated for L2 proficiency and working memory capacity on ambiguity effects in the two tasks can be informative. In translation production, L2 proficiency was negatively correlated with the magnitude of ambiguity effects in accuracy, and to a lesser degree in RT. Specifically, participants who were less proficient in the L2 displayed a larger disparity in their ability to come up with correct translations for ambiguous as opposed to unambiguous items than did more proficient L2 speakers. Less proficient participant were also slower to produce translation for ambiguous as opposed to unambiguous items. The effect of L2 proficiency was less pronounced in the translation recognition task, but again supported the notion that increased proficiency is linked to higher facility in managing ambiguity. Specifically, we found that less proficient L2 speakers made more errors in recognizing translations for ambiguous as opposed to unambiguous words. However, in this task proficiency was not related to reaction time.

These findings hints at the greater difficulty in retrieving a correct translation when more than one exists, perhaps due to interference between the competing possible translations. Alternatively, it might be the case that ambiguous words are learned more slowly in the process of second language acquisition and undergo a longer period of partial knowledge until stable connections are established to L1 words and concepts. This possibility is also supported by a recent training study showing that ambiguous words were harder to learn (Degani & Tokowicz, 2010). An additional possibility in this regard is that the same mechanisms that allow bilinguals to overcome competition from the dominant L1 when producing a word in L2 are at the basis of selecting one of several possible translations. From this perspective, bilinguals who are less proficient in the L2 are less adept at overcoming this type of interference in production, and thus also exhibit a disadvantage in resolving ambiguity in translation.

Conversely, we found that working memory capacity was only significantly correlated with the RT cost associated with recognizing low as opposed to high-probability translations in the translation recognition task. Returning to the two possible mechanisms described earlier for explaining probability effects in translation recognition, high-span individuals might be better able to activate several possible translations for ambiguous words, improving performance for lower probability items. Additionally, high-span individuals might have an improved ability to inhibit a higher-probability translation that has been generated upon expectation when the lower-probability translation is then presented. Along similar lines, Tokowicz et al. (2004) report that among L2 learners who had been immersed in the L2, high-span individuals were more likely to generate meaning errors, suggesting that they may be better able to tolerate ambiguity and error for the sake of communication. These processes are reminiscent of work describing the effects of memory span on the ability of monolingual readers to resolve word ambiguity in sentential context (Gernsbacher & Faust, 1991). Finally, the two explanations need not be mutually exclusive, and indeed most probably operate in concert.

A final intriguing result in the individual differences analyses demonstrated a significant association between reported degree of L2 use and error rates for low-probability translations in the translation recognition task. Specifically, bilinguals who reported using their L2 a larger percent of the time were more likely to erroneously reject low-probability translations for ambiguous words. A possible explanation is that increased L2 use results in a finer tuned probability distribution for ambiguous words, following the notion of frequency matching (Gibson, 1986; Prior et al., 2011). Thus, bilinguals who use the L2 more extensively might be more likely to activate the high-probability translation of ambiguous words, and consequently the mismatch upon the presentation of the low-probability alternative in the translation recognition task is greater, leading to errors. This suggestion is mostly speculative at this point, and should be further investigated in future research.

When examined together, this pattern of results highlights the task differences between translation production and recognition, and demonstrates how their performance might rely on various linguistic and cognitive resources of bilinguals. Translation production is internally driven, and successful retrieval of an accurate translation, especially for difficult ambiguous items, is strongly dependent on language proficiency. The nature of the task requires the bilingual to initiate a search of the lexicon to identify an appropriate response, and if the information is not represented in a stable manner, or if the retrieval process encounters interference from other possible translations the end result might be a failure. Translation recognition, on the other hand, requires less of a retrieval effort. Instead, it places an emphasis on decision-making processes that evaluate the match between internally-generated anticipatory translations and visually-presented translations. Successful performance of the task requires the bilingual to manage possible competition between these two information sources, and to rely on the ability to inhibit information, when it is no longer relevant.

Finally, from a methodological perspective, we believe that the approach adopted in this research, namely using a graded measure of probability and regression analysis, allowed us to bring to light results that might have been more difficult to identify using dichotomous variables and standard orthogonal stimuli construction. Thus, the finding that word class *per se* does not seem to impact translation processes after controlling for the influence of psycholinguistic variable that are confounded with word class is an important outcome of the current design and analytic method. Similarly, our demonstration that the effect of translation probability on translation recognition is continuous suggests that more than a single cognitive process is at play, in addition to the expectation based explanation offered previously (Laxen & Lavaur, 2010).

To conclude, the present study demonstrated the pervasive influence of translation ambiguity on translation recognition and translation production, while at the same time highlighting important differences between the two tasks and the cognitive and linguistic resources they rely on. As such, translation ambiguity is emerging as an important psycholinguistic factor that should be considered in future studies of bilingual lexical representation and processing (for a recent review see Tokowicz & Degani, 2010). In contrast, word class *per se* did not exert a significant influence on translation performance in either task, raising the possibility that previous findings of word class differences in bilingual performance might in fact be a reflection of differences in translation ambiguity Thus, although bilinguals are able to utilize word class information to constrain activation and to improve performance in translation tasks, it seems that the lexical and conceptual mechanisms supporting the translation of single nouns and verbs are fundamentally similar, at least in lexical tasks performed without wider linguistic context.

References

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10,* 82–93.

Baten, K., Hofman, F., & Loeys, T. (2010). Cross-linguistic activation in bilingual sentence processing: The role of word class meaning. *Bilingualism: Language and Cognition, 14,* 351–359.

Bird, H., Franklin, S., & Howard, D. (2001). Noun–verb differences? A question of semantics: A response to Shapiro and Caramazza. *Brain & Language,76,* 213–222.

Black, M., & Chiat, S. (2003). Noun–verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics, 16,* 231–250.

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influence inhibitory control in auditory comprehension. *Cognition, 118,* 245–257.

Boada, R., Sánchez-Casas, R., García-Albea, J. E., & Ferré, P. (2011). Translation dominance and semantic relatedness in multiple translation words: A study with highly proficient bilinguals. Poster presented at the 8th International Symposium of Bilingualism, Oslo, Norway.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual memory after 15 years of service? *Bilingualism: Language and Cognition, 13,* 359–371.

Cappa, S. F., & Perani, D. (2003). The neural correlates of noun and verb processing. *Journal of Neurolinguistics, 16,* 183–189.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349,* 788–790.

Chiarello, C., Shears, C., & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, & Computers, 31,* 603–637.

Cohen, J. (1983) The cost of dichotomization. *Applied Psychological Measurement,7,* 249–253.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and a user's guide. *Psychonomic Bulletin and Review, 12,* 769–786.

Costa, A. (2005). Lexical access in bilingual production. In Kroll & De Groot (eds.), pp. 308–325.

Damasio, A. R., & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Science, 90,* 4957–4960.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37,* 665–671.

Degani, T., Eddington, C. M., Tokowicz, N., & Prior, A. (November, 2009). Comparing within-language and across-language word ambiguity. Poster presented at the meeting of the Psychonomic Society, Boston, MA.

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.

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

De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation. *Journal of Memory and Language, 33,* 600–629.

Dietrich, R. (1990). Nouns and verbs in the learner's lexicon. In H. W. Dechert (ed.), *Current trends in European second language acquisition research*, pp. 13–22. Clevedon: Multilingual Matters.

Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica, 38,* 177–196.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23,* 166–180.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123,* 2552–2566.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas and thematic role concepts. *Journal of Memory and Language, 44,* 516–547.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In Kroll & De Groot (eds.), pp. 251–267.

Friel, B. M., & Kennison, S. M. (2001). Identifying German–English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language & Cognition, 4,* 249–274.

Gathercole, V. C. M., & Moawad, R. A. (2010). Semantic interaction in early and late bilinguals: All words are not created equally. *Bilingualism: Language and Cognition, 13,* 385–408.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4,* 161–177.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (ed.), *Language development: Language, cognition and culture*, pp. 301–334. Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (eds.), *Language acquisition and conceptual development*, pp. 215–256. Cambridge: Cambridge University Press.

Gernsbacher, M. A., & Faust, M. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 245–262.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum. [Original work published 1979.]

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (eds.), *Language acquisition across linguistic and cognitive systems*, pp. 267–284. Philadelphia, PA: John Benjamins.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kempe, V., & MacWhinney, B. (1996). The crosslinguistic assessment of foreign language vocabulary learning. *Applied Psycholinguistics, 17,* 149–183.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition, 9,* 119–135.

Kroll, J. F., & De Groot, A. M. B. (eds.). (2005). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18,* 137–171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for symmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149–174.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In Kroll & De Groot (eds.), pp. 531–554.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13,* 373–381.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In Kroll & De Groot (eds.), pp. 289–301.

La Heij, W., Kerling, R., & Van Der Velden, E. (1996). Nonverbal context effects in forward and backward translation: Evidence for concept mediation. *Journal of Memory and Language, 35,* 648–665.

Laxen, 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,* 157–183.

Levin, B., & Rappaport Hovav, M. (1996). From lexical semantics to argument realization, Ms., Northwestern University & Bar Ilan University.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2010). Neural differences in the mapping of verb and noun concepts onto novel words. *NeuroImage, 49,* 2826–2835.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition, 41,* 197–229.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33,* 175–202.

Morford, J. P., Wilkinson, E., Villwock, A., Pinar, P., & Kroll, J. F. (2010). When deaf signers read English: Do written words activate their sign translations? *Cognition, 118,* 286–292.

Paivio, A. (1971). *Imagery and verbal process*. New York: Holt, Rinehart, & Winston.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of verb and noun processing: A PET study. *Brain, 122,* 2337–2344.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 539–557.

Pérez, M. A., Alameda, J. R., & Cuetos, F. (2003). Frecuencia, longitud y vecinidad ortografica de las palabras de 3 a 16 letras del diccionario de la lengua Española (RAE, 1992). *Revista Electrónica de Metodología Aplicada, 8,* 1–10.

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,* 1029–1038.

Prior, A., Wintner, S., MacWhinney, B., & Lavie, A. (2011). Translation ambiguity in and out of context. *Applied Psycholinguistics, 32,* 93–111.

Pulvermuller, T., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex, 9,* 497–506.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27,* 499–520.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28,* 387–422.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2,* 45–58.

Tokowicz, N., & Degani, T. (2010). Translation ambiguity: Consequences for learning and processing. In B. VanPatten & J. Jegerski (eds.), *Research on second language processing and parsing*, pp. 281–294. Amsterdam: John Benjamins.

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., & 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,* 255–272.

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

Tyler, L. K., Russell, R., Fadili, J., & Moss, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain, 124,* 1619–1634.

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

Wierzbicka, A. (1988). *The semantics of grammar*. Amsterdam: John Benjamins.

Willms., J. L., Shapiro, K. A., Peelen, M. V., Pajtas, P. E., Costa, A., Moo, L. R., & Caramazza, A. (2011). Language-invariant verb processing regions in Spanish–English bilinguals. *Neuroimage, 57,* 251–261.