Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 4, 843–858

0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.4.843

Investigating Linguistic Relativity Through Bilingualism: The Case of Grammatical Gender

Stavroula-Thaleia Kousta, David P. Vinson, and Gabriella Vigliocco
University College London

The authors investigated linguistic relativity effects by examining the semantic effects of grammatical gender (present in Italian but absent in English) in fluent bilingual speakers as compared with monolingual speakers. In an error-induction experiment, they used responses by monolingual speakers to establish a baseline for bilingual speakers and show that gender affects the semantic substitution errors made by monolingual Italian speakers compared with monolingual English speakers. They then showed that Italian–English bilingual speakers behave like monolingual English speakers when the task is in English and like monolingual Italian speakers when the task is in Italian, hence exhibiting appropriate semantic representations for each language. These results show that for bilingual speakers there is intraspeaker relativity in semantic representations and, therefore, that gender does not have a conceptual, nonlinguistic effect. The results also have implications for models of bilingual semantic memory and processing.

Keywords: linguistic relativity, bilingualism, grammatical gender, speech errors

In order to speak any language, one has to pay attention to the distinctions obligatorily expressed in the language, and in this way every linguistic community differs from every other. In other words, there is linguistic relativity because speakers of typologically different languages are required to verbalize different aspects of reality when constructing linguistic messages. What are the cognitive consequences of linguistic relativity? For Whorf (1956), to whom the linguistic relativity hypothesis is attributed, linguistic relativity is closely linked to linguistic determinism, that is, the proposal that language determines the way the external world is perceived, categorized, and acted on.

Nowadays virtually no one would claim that language has such a deterministic role on cognition as Whorf originally envisioned it. It is, however, still a matter of intense debate to what extent language affects (rather than determines) cognition. According to one hypothesis, the linguistic classifications imposed by language affect only other linguistic processes and representations (Brysbaert, Fias, & Noël, 1998; Finkbeiner, Nicol, Greth, & Nakamura, 2002; Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Massey, and Gleitman, 2002; Slobin, 1996; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). A stronger hypothesis argues that language also affects nonlinguistic cognition, above and beyond the processes involved in speaking and comprehending a particular language. Data in support of this claim have been provided in a number of studies (Imai & Gentner, 1997; Kita & Özyürek, 2003; Levinson, 1996, 1997; Levinson, Kita, Haun, & Rasch, 2002; Lucy & Gaskins, 2001; Lucy, 1992; McDonough, Choi, & Mandler, 2003; Pederson et al., 1998; Yoshida & Smith, 2005), but there is also a substantial body of evidence that argues against this claim (Gennari et al., 2002; Huntley-Fenner, Carey, & Solimando, 2002; Li & Gleitman, 2002; Mazuka & Friedman, 2000; Munnich, Landau, & Dosher, 2001; Soja, Carey, & Spelke, 1991; Vigliocco et al., 2005). In other words, the main debate in the field focuses on the extent and pervasiveness of the effect of language on cognition.

Until now, standard investigations of the link between language and thought have involved testing two or more monolingual populations on some nonlinguistic task (involving perception, categorization, or action) to determine whether a particular linguistic difference between the two populations affects cognitive performance. Most of the time, this type of work is motivated by a linguistic task in which it is demonstrated that the relevant linguistic dimension does produce the predicted effect in tasks involving language. In this study, we adopted a different approach: We tested both monolingual and proficient bilingual populations on the same linguistic task in two languages and then drew inferences about the strength of the effect of language on cognition on the basis of the extent to which the behavior of bilingual speakers patterns with that of monolingual speakers. If the results show that an aspect of grammar (e.g., grammatical gender) affects monolingual performance in a task that taps into semantic representations, this would

Stavroula-Thaleia Kousta, Department of Psychology, University College London; David P. Vinson and Gabriella Vigliocco, Department of Psychology and Deafness Cognition and Language Centre, University College London.

This research was supported by Economic and Social Research Council Grant (ESRC) RES000230038 and European Union (FP6-2004-NEST-PATH) Grant 028714 to Gabriella Vigliocco and by ESRC Grant RES620286001 to the Deafness, Cognition, and Language Research Centre. We thank Tamar Gollan, Susie Flett, and David Green for their help with constructing the bilingual language background questionnaire. We are also grateful to Lara Pelizzon and Raffaella Delbello for their help with collecting the monolingual Italian data.

Correspondence concerning this article should be addressed to Stavroula-Thaleia Kousta, Department of Psychology, University College London, London WC1H 0AP, United Kingdom. E-mail: s.kousta@ucl.ac.uk

establish that there is a cross-linguistic difference in grammar that affects semantic representations. That is, we would have evidence of an effect of language on language, in a similar vein to studies that compare performance of two monolingual populations on a linguistic task. However, bilingual performance allows us to go beyond the effect of language on language. If the results show a difference between the two languages of proficient bilingual speakers who acquired a second language after their first was largely in place (i.e., intraspeaker linguistic relativity) and this difference is parallel to the difference found between groups of monolingual speakers, then the conclusion could be only that language has a truly limited effect: It affects only representations and processes relevant to a specific language and does not extend to conceptual, nonlinguistic representations. Our reasoning is as follows. Normally developing children successfully acquire nearly all aspects of linguistic competence by the age of 5 or 6. If language affects thought, then by that time nonlinguistic conceptual representations should bear the mark of the specific language that has been learned. When a second language is acquired after that age, at advanced levels of second-language proficiency, second-language semantics is conceptually mediated rather than being mediated by the first language (de Groot, 1992a; Potter, So, von Eckardt, & Feldman, 1984). If conceptual representations have been affected by first-language linguistic distinctions, we should be able to see bilingual speakers behaving the same way as monolingual speakers of their mother tongue, regardless of the language involved. If, on the other hand, we replicate the difference found between monolingual speakers within the bilingual individual, we would have evidence that argues against a significant role of language on shaping nonlinguistic thought.

This approach has two main advantages. First, the same task and the same stimuli can be used to make inferences about the role of cross-linguistic grammatical differences on both semantic and conceptual representations, unlike monolingual studies, in which different tasks or different stimuli are used to probe linguistic and nonlinguistic representations. In these latter studies, obtaining an effect of language on language but not on nonlinguistic cognition can be attributed to changes in the stimuli or task, rather than providing strong evidence that grammatical differences affect only the semantic level and do not extend to conceptual representations. This, however, cannot be the case when the stimuli and task are held constant. Second, this approach allows us to ask even more interesting questions about the link between language and cognition. If language affects cognition, is it possible that a second language modifies cognitive dispositions established by a first language? This would be the conclusion suggested if bilingual speakers are found to behave in the same way in both their languages, but their behavior patterns with that of monolingual speakers of their second language rather than their first language.

We do not believe that this approach can completely replace traditional investigations of the linguistic relativity question. If no differences are found between the languages of bilingual speakers, it would then be necessary to test the bilingual speakers on a nonlinguistic task to determine whether the effect can be attributed to nonlinguistic thought. However, we do believe that finding intraspeaker relativity effects with bilingual speakers is strong evidence against a significant role of language in shaping cognition. We also argue that this approach makes it possible to ask a broader range of questions than traditional approaches have allowed.

It is also important to note that we maintain a distinction between semantic and conceptual representations (see Jackendoff, 2002; Levinson, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004), in which semantic representations are linguistic whereas conceptual representations are nonlinguistic. More precisely, following Vigliocco, Vinson, Lewis, and Garrett (2004), we assume that conceptual representations are based on distributed modality-specific featural representations: The concept "dog," for instance, may be composed of features such as *has legs*, *barks*, *has fur*, *is sexuated*, and so on. Semantic representations are an intermediate level of representation, binding conceptual features to linguistic (lexico-syntactic, phonological, and orthographic) information. As such, semantic representations can be shaped by both conceptual and linguistic properties. Maintaining a distinction between conceptual and semantic representations in these terms makes it possible to account for cross-linguistic differences in word meaning that may or may not be reflected in conceptual structures. If, on the other hand, there is no distinction between these two levels, the question of whether language affects cognition cannot be meaningfully asked: Cross-linguistic differences in patterns of lexicalization are a fact; if identity of semantic and conceptual representations is assumed, one has to conclude that Whorf was right (i.e., that language determines thought). We now know that this absolute determinism is wrong (see Levinson, 2003, and Vigliocco & Filipovic Kleiner, 2004, for a discussion).

The few studies on bilingualism and thought that are available have focused on a number of different linguistic domains but have not come to definitive conclusions. There is comparatively more work carried out with lexical rather than grammatical information. For instance, in the domain of color research, Caskey-Sirmons and Hickerson (1977), Ervin (1961), and Jameson and Alvarado (2003) have shown that bilingual speakers identify and categorize colors differently from monolingual speakers of their first language, because they are influenced by color naming patterns in their second language. Effects of bilingual speakers' second language on their first language have also been reported in other domains; for example, Wolff and Ventura (2003) have shown that the first-language semantic representations of causal verbs of Russian–English speakers who are equally proficient in their two languages are affected by their second language. Also, Ameel, Storms, Malt, and Sloman (2005) found that the naming patterns of compound Dutch–French speakers (i.e., bilingual speakers from birth, who use both their languages in the same contexts) for different types of bottles and dishes diverged from monolingual naming patterns but also differed within bilingual speakers for each of their languages (even though this difference was significantly smaller than that between monolingual speakers).

Research on the effects of grammar on cognition with bilingual speakers is even scarcer. Athanasopoulos (2006) extended Lucy's (1992) work on grammatical number to a different language (Japanese) and to bilingual populations. Japanese, like Yucatec Maya in Lucy's original study, does not obligatorily mark plurality. Plural forms, when used, refer to humans and less commonly animals but are never used with words having inanimate referents. As a result Japanese monolingual speakers are shown to be more sensitive to changes in the number of humans and animals but not in the number of artifacts. In English, however, plural morphology

is used with both animate and inanimate referents, with substances being the only exception to pluralization. Monolingual English speakers are therefore shown to be more sensitive to changes in the number of both animate and discrete inanimate entities. Athanasopoulos asked monolingual English, monolingual Japanese, and bilingual Japanese–English speakers at two levels of proficiency (intermediate and advanced) to match pictures according to similarity and found that intermediate Japanese–English bilingual speakers tended to follow the pattern of monolingual Japanese speakers whereas advanced bilingual speakers behaved similarly to the monolingual English speakers. Although Athanasopoulos replicated Lucy's (1992) findings in a nonlinguistic task comparing monolingual populations who speak languages that either mark (English) or do not mark (Japanese) inanimate discrete entities for plurality and showed that advanced Japanese–English bilingual speakers resemble monolingual English speakers in their preferences, he was very cautious about interpreting the results as evidence for a role of language on nonlinguistic cognition. He argued that it is entirely possible that language was used to mediate the nonlinguistic task because the language in which instructions are given seems to play a role in the type of responses given in seemingly nonlinguistic tasks: Athanasopoulos (2001, quoted in Athanasopoulos, 2006) and Cook, Bassetti, Kasai, Sasaki, and Takahasi (2006) have found that bilingual speakers who are instructed in their first language deviate from either monolingual norm, suggesting that the language of instruction plays a crucial (but not straightforward) role in so-called nonlinguistic tasks. Furthermore, if it is true that the task was linguistically mediated, there is doubt as to whether it is informative about anything else apart from the ability to use language as a strategy to resolve the task at hand.

Scheutz and Eberhard (2004) tested bilingual German–English speakers and monolingual English speakers on a task that involved the interpretation of the *–er* agentive morpheme in English sentences. In German this morpheme has the same form and similar function as the English equivalent but is predominantly associated with male referents. Scheutz and Eberhard found that German bilingual speakers showed a slight tendency to carry this male association over to the English morpheme, which according to the authors does not have any male associations. However, the *–er* morpheme in English, although not predominantly male associated, is one of the morphemes used to mark masculine referents on nouns that have both masculine and feminine forms (e.g., *widower–widow*, *waiter–waitress*, *master–mistress*). So the English and German *–er* morphemes differ only in terms of degree of association with male referents. Given also the fact that the morphemes are cognates, it is unsurprising that German bilingual speakers exhibit a trend for male association for this morpheme.

More relevant to the phenomenon under investigation in this article are studies of grammatical gender. Boroditsky and colleagues (Boroditsky & Schmidt, 2000; Phillips & Boroditsky, 2003; see also Boroditsky, Schmidt, & Phillips, 2003) showed that first-language grammatical gender affected bilingual Spanish–English and German–English speakers' performance in a number of linguistic and nonlinguistic tasks carried out in English or with English instructions, although these speakers had a minimum of 14 years of "experience" with English. They further showed that behavior is predicted by proficiency rather than one's native tongue by investigating Spanish–German–English trilingual speakers with varying degrees of proficiency in their different languages. Boroditsky and colleagues used these results to argue for a pervasive role of language in shaping cognition. The tasks they used to support their arguments involved comparisons of inanimate objects to humans (e.g., *toaster–girl*) or memory for the gender of an arbitrarily assigned name to an object (e.g., *apple–Patrick/Patricia*). These types of tasks, however, are open to methodological and interpretation criticisms. When asked to rate a toaster and a girl for similarity, one may strategically use grammatical gender to perform the novel, rather puzzling, task; when asked to remember an arbitrary name for an object with no affective value, one may fall back on grammatical gender as a strategy. Eberhard, Scheutz, and Heilman (2005), in a series of computer simulations, provided evidence toward such an explanation of the results obtained by Boroditsky and colleagues.

In short, the limited literature on bilingualism and cognition has produced mixed results. One observation that emerges from these studies is that proficiency in each of a bilingual speaker's languages, age of acquisition of the second language, the extent of overlap between languages with respect to the phenomenon under study, and the nature of the experimental task play an important role, accounting for the differences in the obtained results.

Grammatical Gender

In this article, we investigate bilingual cognition in relation to a phenomenon, grammatical gender, that has attracted considerable interest in the literature on linguistic relativity–determinism. In many languages, grammatical gender is a formal category: Nouns are marked masculine, feminine, neuter, and so on. Italian, for instance, like Arabic, French, Spanish, and many other languages, has a two-gender system, and all nouns are marked either feminine or masculine. In English and other languages, such as Chinese, Estonian, and Hungarian, gender is not a formal category. Does the grammatical gender of words affect their semantic representation and the conceptual representation of their referents? Early work with monolingual speakers using ratings of words on a semantic differential scale (Konishi, 1993), male or female voice assignment to pictured objects (Sera, Berge, & del Castillo-Pintado, 1994), and picture sorting performed by children (Martinez & Shatz, 1996) suggested that grammatical gender affected both the semantic representation of words and the conceptual representation of the referents of those words. More recent research, however, has suggested that grammatical gender effects are substantially constrained. Sera et al. (2002) again used a voice assignment task and found that gender effects are obtained with speakers of languages such as French and Spanish but not German, even though all three languages have formal gender. They proposed an explanation based on the fact that Spanish and French have a two-gender system (masculine–feminine) that exhibits a high correlation between grammatical and natural gender and gender is morphologically marked across several grammatical categories, whereas German has a three-gender system (masculine–feminine–neuter) with a less straightforward relationship between grammatical and natural gender and gender is marked in fewer grammatical categories. Nevertheless, it is possible that the difference between Spanish and French, on the one hand, and German, on the other, arose because Spanish and French speakers could successfully use gender as a strategy to assign male and female voices to pictures,

whereas in German speakers the strategy fails with nouns that are marked neuter, hence leading to a null result. Vigliocco et al. (2005) used elicitation of semantic substitution errors and triadic judgment tasks to investigate the difference between two- and three-gender languages and showed that gender effects are even further constrained: Apart from being limited to languages with two genders and with a close correspondence between grammatical and natural gender, grammatical gender was shown to affect the semantic representation of words referring to sexuated entities (e.g., animals) but not the representation of words referring to entities that lack natural gender (e.g., artifacts). Moreover, they found that this constrained semantic effect did not generalize to a nonlinguistic, conceptual task: triadic judgments with pictures. However, the triadic judgment task is quite explicit and may have also invited the use of strategies.

In summary, in studies with monolingual speakers, grammatical gender effects are obtained with languages possessing only two gender categories (masculine–feminine) and with a regular correspondence between grammatical and natural gender, such as Italian, French, and Spanish. However, there is still no agreement as to whether the language-specific effects generalize to entities that do not possess natural gender or, importantly, to nonverbal tasks.

Bilingual Semantic Memory and Processing

Research in bilingual representation and processing was dominated for a long time by one central question: Are a bilingual speaker's two languages stored and accessed together or independently? When it comes to semantic representation and processing, there is general consensus in the field that bilingual speakers possess a single semantic store[1] and that semantic representations are accessed in a language-independent manner. Evidence for such a position comes from semantic priming studies (e.g. Chen & Ng, 1989; de Groot & Nas, 1991), studies based on variations of the Stroop interference paradigm (e.g., Costa, Miozzo, & Caramazza, 1999; Ehri & Ryan, 1980), and studies involving same- and different-language semantic comparisons of words (e.g., Caramazza & Brones, 1980; Potter et al., 1984). There is also neuroimaging evidence supporting the shared-store, language-independent access account (Illes et al., 1999). However, it has become increasingly apparent that translation-equivalent words more often than not do not have identical semantic representations in bilingual semantic memory, especially for proficient bilingual speakers (de Groot, 1992a, 1992b; Van Hell & de Groot, 1998a, 1998b). The semantic content of translation-equivalent words is only partially overlapping. For instance, the word *family* in English typically refers to the nuclear family, whereas the Spanish translation-equivalent *familia* commonly refers to both the nuclear and extended family (Tokowicz, 2000). A model that combines a shared semantic memory with only partially overlapping semantics for translation-equivalent words has been proposed by de Groot and colleagues (de Groot, 1992a, 1992b; de Groot, Dannenburg, & van Hell, 1994; Van Hell & de Groot, 1998a, 1998b). The distributed feature model assumes that there is a common pool of semantic features that underlies both of a bilingual speaker's languages. These features, however, combine in a language-specific manner, and this is what determines to what extent the meaning of translation-equivalent words overlap. The model was originally meant to capture differences in the semantic content of translation-equivalent words referring to abstract concepts, assuming that the semantic content of translation-equivalent words referring to concrete entities was mostly the same across languages. However, it is true that even the referents of translation-equivalent concrete words are very commonly not identical, suggesting that their semantic representations are different (Paradis, 1997).

Although there is relative consensus in the literature that, at least for proficient bilingual speakers, the semantic representations of translation-equivalent words diverge in accord with the different referential function of these words, it is an open question whether bilingual speakers maintain distinct representations for words that do not differ in their extension. For instance, the words *tiger* in English and *tigre* in Italian refer to the same feline mammal. However, the research on gender outlined above has shown that the semantic representation of the word *tigre* in Italian differs from the representation of the word *tiger* because grammatical gender plays a role in determining meaning. Do bilingual Italian–English speakers form both of these distinct semantic representations, even though the words *tigre* and *tiger* denote the same creature?

Objectives

In the present study, we test the strength of language effects on cognition by investigating to what extent bilingual speakers develop semantic representations that are appropriate for their second language and to what extent semantic representations in the first language are affected by learning a second language. When it comes to grammatical gender, evidence of its effect on conceptual representations is ambiguous and complicated by the fact that the vast majority of tasks used to investigate its effects are explicit and subject to the use of strategies. Here we focus on semantic errors arising during continuous naming, thus using an online methodology that is not subject to the use of strategies.

We asked monolingual Italian and monolingual English speakers to name pictures of common land animals presented at a fast rate. Bilingual Italian–English speakers were also asked to carry out the same task in both languages. The aim of the task was to elicit semantic substitution errors. Semantic substitution errors (saying "eye" when "ear" is intended) are assumed to arise as a result of competition between semantically related lexical candidates in a conceptually driven lexical selection process during production (Garrett, 1984). If grammatical gender increases semantic similarity between nouns that share gender, then monolingual Italian speakers should produce more gender-preserving errors than monolingual English speakers for whom Italian gender is irrelevant. The comparison of Italian to English speakers makes it possible to exclude the contribution of visual and language-independent semantic similarity to the effects of gender. In other words, the English data act as a baseline for Italian performance.

[1] Very frequently in the bilingual literature, the terms *semantic store* and *conceptual store* are used interchangeably (see, for instance, Francis, 1999). We assume that in most discussions of language-dependent or independent access to the bilingual lexicon, it is semantic store that is intended rather than conceptual store. When conceptual store is intended, the claim that there might be two such stores equals extreme Whorfianism: If two separate nonlinguistic stores are maintained in the bilingual mind–brain, each resulting from and interfacing with one of the bilingual speaker's languages, then language completely determines thought.

Vigliocco et al. (2005) indeed found that for animals, intruders in Italian preserved the gender of targets above the English rate, hence obtaining an effect of grammar on semantics. We replicate this study with monolingual speakers in order to establish a baseline for comparison for bilingual speakers.[2] The crucial question is how bilingual speakers are going to behave. If language affects nonlinguistic cognition, then one would expect bilingual speakers to behave in the same way in both their languages and not to differ from monolingual Italian speakers. If, however, the effect of language is limited to the semantic representations of that same language, then the expectation is that there will be a significant difference in performance between the two languages for bilingual speakers, with their errors patterning with monolingual Italian errors when they do the task in Italian and with monolingual English errors when they do the task in English. As a final possibility, because grammatical gender has only a limited conceptual motivation (the grammatical gender of words referring to humans nearly always corresponds to conceptual gender, but gender assignment for other words is largely arbitrary), it is possible that learning an ungendered second language foregrounds the arbitrary nature of gender assignment in the bilingual speakers' mother tongue and leads to a restructuring of semantic representations. This is plausible also because for the bilinguals in the present experiment, the first language was the lesser used language. In this scenario, it might turn out that bilingual speakers' error patterns in both their languages are more similar to the English monolingual norm rather than the Italian monolingual norm.

In the literature on language production, it is commonly assumed that grammatical gender affects the production of phrases but not the production of bare nouns (La Heij, Mark, Sander, & Willeboordse, 1998; Paganelli, Vigliocco, Vinson, Siri, & Cappa, 2003; Pechmann & Zerbst, 2002; Schriefers, Jescheniak, & Hantsch, 2002; Vigliocco, Vinson, Indefrey, Levelt, & Hellwig, 2004; but see Cubelli, Lotto, Paolieri, Girelli, & Job, 2005). This is because grammatical gender is a lexico-syntactic property that is necessarily recruited when sentential frames are being constructed, but not being semantic in nature, it is not mandatory for the production of bare nouns. However, for languages like Italian and semantic fields like animals (which have natural gender), Vigliocco et al. (2005) have shown that gender is also a lexico-semantic property, affecting the production of bare nouns. In order to investigate whether grammatical gender effects are due to only the activation of syntactic information necessary for computing agreement, errors in single word naming were compared with errors in determiner plus noun naming (Italian determiners are marked for gender).

Experiment 1: Monolingual Error Induction

Method

Participants. Twenty-six native speakers of English (14 men, 12 women; mean age: 22.0 ± 2.6 years) were recruited for the experiment through the University College London psychology subject pool. Four participants were replaced because of language backgrounds inappropriate for this study (they reported moderate or higher proficiency in any gendered language). Two further participants were replaced (1 because he reported he was dyslexic and found the task unmanageably difficult and 1 because of incomprehensible responses). Participants were paid at a rate of £6 per hour. Twenty-five native speakers of Italian (8 men, 17 women; mean age: 23.2 ± 2.7), all of whom were psychology students at the Università degli Studi di Trieste, Italy, also volunteered for the experiment. Participants were paid at a rate of €9 per hour for their participation or received course credit. Twelve participants were replaced because of language backgrounds inappropriate for the study (they reported moderate or higher proficiency in English or another second language).

Materials and design. The set of items consisted of pictures depicting common land animals. They were the same items used for Experiment 2 in Vigliocco et al. (2005), namely, the pictures of 27 animals (11 marked feminine; 16 marked masculine—see Appendix A for the full list of the picture names). Each picture was presented in black on a white background and scaled to fit within a 240 × 240 pixel area.

Forty-one blocks of 10 pictures were created for each condition (bare noun vs. noun phrase) by randomly selecting pictures but observing the following constraints: First, a picture appeared no more than once within a block, and second, a picture never appeared as the last item in one block and the first item in the next. Each picture appeared either 15 or 16 times in each condition, and each participant saw a different random order of blocks for each part of a session. Stimuli presentation was controlled using the E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002).

Procedure. Participants were told that they were taking part in a study of speech patterns under time pressure and that they would be asked to name pictures depicting animals as they appeared on screen. The experiment was carried out in two parts: In one part participants were asked to name the pictures using only a bare noun, and in the other part they were asked to produce noun phrases consisting of a definite article plus a noun. The order of the two parts was counterbalanced across participants.

Instructions emphasized that speakers should attempt to keep up with the rate of presentation, that is, naming pictures as they appeared, rather than retaining them in memory. All participants gave consent to have their responses recorded; all spoken responses were digitally recorded and later transcribed and scored.

The experiment began with an untimed name-agreement phase, in which each picture was presented, and participants were asked to name them without time pressure. The experimenter noted any variation from the intended names and also provided prompts if the participants were not able to produce a label for the picture. The name-agreement phase allowed us to ensure that responses that mismatched our intended target but were nonetheless used by participants to refer to a given picture were not considered to be lexical errors.

After the name-agreement phase, participants performed a set of six practice blocks. In each practice block, 10 pictures in a row

[2] The data in Vigliocco et al. (2005, Experiment 2) cannot be used as a baseline because that experiment investigated errors with pictures of both animals and artifacts (we use only animals); bare noun naming was involved (we also included a condition in which participants had to name the pictures using a determiner + noun—see below for details); and fewer blocks of trials were included (we increased the number from 27 to 41 blocks of 10 pictures per condition).

were presented in one of four possible locations on the screen, and participants were instructed to name each aloud. These practice blocks were intended not only to familiarize participants with the task but also to allow the experimenter to adjust the rate of presentation according to each participant's performance. After each block of 10 pictures, the experimenter altered the rate of presentation to accommodate each speaker's speech rate. Initial presentation parameters were 1,000-ms display time for each picture. The display rate was altered by the experimenter (minimum step size = 100 ms) during the practice session in order to make the task difficult but manageable for each speaker. Final presentation rates ranged between 400 and 900 ms of display time for English (average = 633 ms) and between 500 and 1,200 ms for Italian (average = 688 ms). Italian speakers may have needed longer presentation times because the Italian words in the experiment were significantly longer than English words, both in terms of phonemes, $t(26) = 7.573$, $p < .001$, and in terms of syllables, $t(26) = 9.067$, $p < .001$.

Once the practice session was completed, the experimental blocks were presented. Participants pressed a key to begin each block, and then the 10 items in the block appeared in sequence, with time parameters as determined in the practice session. Pictures appeared at randomly selected positions on the screen (a pilot study indicated that there were noticeably more naming errors when pictures were presented with "location uncertainty"; presentation at a single location, instead, resulted in an elevated rate of omission errors, presumably because masking-type effects led to greater likelihood of participants not seeing one of the pictures in a sequence). After each block, participants were given the opportunity to take a break if necessary.

Results

The data collected were transcribed and scored as belonging to one of the following categories. *Correct* responses were those for which the participant produced the intended noun or noun phrase in entirety. *Different label* responses were those for which the participant produced an alternative label for one of the animals (e.g., *badger* instead of *skunk*) but used the label consistently to name the specific animal and did not use the label consistently for another animal in the experiment. Different labels were identified in any of the following ways: The participant used the different label in the initial untimed naming phase, the participant used the different label three or more times in the experiment itself without self-correcting, and/or we judged that the response word was also an acceptable label for the given picture. These items were treated as acceptable responses and were not included in the error analyses below. *Lexical errors* were those for which the participant produced a label that did not qualify as a different label. *Omissions* were cases in which participants failed to provide a response. *Self-corrections* were instances in which participants started producing an incorrect word but changed their response to the correct target before it was complete. Self-corrections were scored as lexical errors if the incorrect word was produced completely before being corrected. *Miscellanea* were other responses not included above, such as dysfluencies, incomplete or inaudible responses, agreement errors (in the production of noun phrases in Italian), and nonword responses. Only responses classified as lexical errors were included in the analyses reported below. All lexical errors were semantically related words referring to animals (91.32% of the English errors and 84.31% of the Italian observations were within-set errors, i.e., words referring to other animals used in the experiment). Table 1 presents the number of observations in each of the above response categories, broken down by language and response type. Note that the error rate is so high because participants were under time pressure. Figure 1 presents the average proportion of target–error pairs sharing gender in the two languages and response types. To compare English and Italian, we assigned English words the gender that translation-equivalent items have in Italian. In other words, an error in English was classified as a gender-preserving error if the translation-equivalent error in Italian shared the same gender with the translation of the target. The English monolingual data in this case function as a baseline, so that we could determine whether gender affects error rates in Italian: Because gender is not a formal category in English, English semantic errors are driven by visual and conceptual similarity (e.g., shared habitat, size, behavior, membership in the same genus). Do the errors produced by Italian speakers reflect only these types of similarity or does grammatical gender contribute as well?

An inspection of the proportion of gender-preserving errors for English participants indicates that (Italian) gender was preserved below chance, suggesting a negative correlation between visual–

Table 1
Number and Percent Occurrence of Responses as a Function of Language and Response Required for Experiment 1

	Italian				English			
	Bare noun		Noun phrase		Bare noun		Noun phrase	
Response type	No.	%	No.	%	No.	%	No.	%
Acceptable								
Correct	8,419	82.14	8,579	83.70	9,244	86.72	9,287	87.12
Different label	560	5.46	531	5.18	330	3.10	379	3.56
Error								
Lexical	272	2.65	309	3.01	284	2.66	284	2.66
Omission	765	7.46	665	6.49	684	6.42	623	5.84
Self-correction	44	0.43	35	0.34	36	0.34	42	0.39
Miscellanea	190	1.85	131	1.28	82	0.77	45	0.42

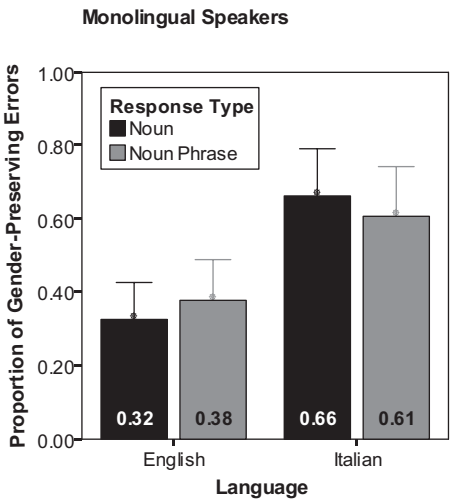

Figure 1. Mean proportion of gender-preserving errors by monolingual Italian and English speakers. Error bars represent 95% confidence intervals.

conceptual similarity and gender. Vigliocco et al. (2005) also found a negative correlation between visual–conceptual similarity and gender for both animals and artifacts in English. Although it is not clear why this is the case, it does not in any way affect the validity of the English data as a baseline for Italian data: If only visual and language-independent conceptual similarity governs Italian semantic errors, there should be no difference in the proportion of gender-preserving errors in the two languages.

The likelihood of substituting one word with another is increased if target and intruder share phonological in addition to semantic similarity (i.e., one is more likely to say "oyster" when "lobster" is intended than to say "crab"; the mixed error effect; Dell & Reich, 1981; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996). It is important to exclude such errors from the data set, because there are clear phonological correlates of grammatical gender in Italian. Moreover, Vigliocco et al. (2005) demonstrated that target–error pairs in Italian that shared gender were characterized by greater phonological overlap, which could not be attributed only to the inflectional suffix carrying the gender information. In order to eliminate the confounding effect of phonological similarity, we carried out the following procedure. First, we phonologically transcribed all the animal nouns using standard British English and standard Italian pronunciations using the online version of the *Oxford English Dictionary* (Oxford University Press, n.d.) and the *Grande Dizionario Italiano* (2006), respectively, as sources. We then excluded all target–intruder pairs that shared more than 33% of their phonemes in the same position (onset, nucleus, coda of the first syllable, middle syllable[s], and the final syllable). This value was determined by setting a threshold at the mean, plus one standard deviation, of the phonological similarity proportion of all targets and errors. For participants whose accents differed from the standard, we phonologically transcribed their responses in the name-agreement phase and recalculated phonological similarity. We then excluded for each participant errors that exceeded the 33% similarity cutoff as before. After applying these procedures, there remained 393 English errors (197 for the bare noun condition and 196 for the noun phrase condition) and 344 Italian errors (162 and 181 for the bare noun and noun phrase conditions, respectively). These errors can be considered semantic, rather than mixed (i.e., semantic and phonological), errors.

For each participant and item, we calculated the proportion of errors for which the target and the intruder shared Italian gender (see Figure 1). Four participants were excluded from subsequent analyses (2 English, 2 Italian) because they produced no errors in one of the conditions.

We performed 2 × 2 analyses of variance (ANOVAs) with participants and items as random factors, investigating the effects of language (English and Italian, between participants but within items) and response type (bare noun and noun phrase, within participants and items). There was a significant main effect of language in both analyses, $F_1(1, 45) = 19.009$, $p < .001$, $MSE = 0.065$, and $F_2(1, 19^3) = 13.300$, $p < .01$, $MSE = 0.057$, with Italian errors sharing gender more often than English errors ($M_{\text{difference}} = 0.28$, 95% confidence interval [CI] = 0.13). There was no main effect of response type or interaction. These results show that, as expected, grammatical gender affects the perceived semantic similarity between different animals for Italian speakers as compared with English speakers (replicating Vigliocco et al., 2005).

However, it is possible that the selection of an arbitrary criterion of phonological similarity, based only on shared segments all treated equally, may not rule out all aspects of phonological similarity that are correlated with shared gender among the items tested. Therefore we conducted an additional regression analysis, using maximum likelihood estimation, to consider whether gender actually affects the probability of error produced by Italian speakers once a number of potential confounding phonological factors are taken into account (all regression analyses reported in this article were carried out using the software R; R Development Core Team, 2007). Because the majority of lexical errors were also from within the response set, we began by creating a confusion matrix by pairing all targets with all possible errors. In each cell of the matrix ($N = 702$), we entered the number of errors observed in Italian (the dependent variable in this analysis) expressed as a proportion of the total number of opportunities for the specific error to occur in the experiment.

The model included the following predictors. The proportion of errors for each target–error pair in the English data set was used to predict the probability of error in Italian. This takes into account those aspects of similarity among animals that are common between English and Italian, such as conceptual similarity and visual similarity among the pictures used in the experiment. A variety of separate predictors related to phonological similarity in Italian were also included: whether the two words shared the same number of syllables, stress pattern, onset phoneme, final phoneme, and also the continuous measure of form similarity described above (% of phonemes shared in the same position between target and error). Finally, shared grammatical gender was used as a predictor. Taking such an approach, common variance due to correlations between gender and morphophonological similarity, other semantic similarity,

[3] There are 19 degrees of freedom in this analysis because there were no errors for some of the items.

or visual similarity is partialed out. Therefore, if gender is a significant predictor when visual and other semantic and phonological variables are taken into account, it can be said that gender per se is predicting the probability of error.

Unsurprisingly, the presence or absence of errors in English was a strong predictor of Italian errors, $\chi^2(1, N = 702) = 1233.81$, $p < .001$. With the exception of shared onset phoneme, all other phonological variables significantly predicted probability of error: % segments shared by position, $\chi^2(1, N = 702) = 34.42$, $p < .001$; shared number of syllables, $\chi^2(1, N = 702) = 44.35$, $p < .001$; shared stress pattern, $\chi^2(1, N = 702) = 17.41$, $p < .001$; and shared final phoneme: $\chi^2(1, N = 702) = 7.77$, $p < .01$. Crucially, gender was also a significant predictor of probability of error: $\chi^2(1, N = 702) = 68.51$, $p < .001$. This shows that the gender effect observed in the ANOVA is not simply a product of the arbitrary cutoff at which errors were decided to be "phonologically similar." A plot of the partial effect of gender (i.e., when the effect of all other variables has been removed) appears in Figure 2.

Here we argue that this effect is semantic in nature, because it has been shown that grammatical gender does not affect the production of bare nouns (see Objectives above). One might still object, however, that this difference is due to syntactic differences between the two languages (syntactic gender is encoded at the lexical level in Italian but not in English) without any involvement of semantics: Gender preservation in errors may simply reflect the coactivation of words sharing grammatical gender. If this is the case, one would expect gender to have the same effect on errors regardless of semantic category (e.g., for both animals and artifacts). In order to confirm that the effect is not syntactic, we reanalyzed the Italian–English tool error data in Vigliocco et al. (2005), who found that there was no gender effect for tools in Italian as opposed to animals but used a somewhat arbitrary threshold value to determine phonological similarity. Here we used the logistic regression procedure detailed above to determine whether gender has an effect on tool errors in that data set when a continuous measure of phonological similarity is used and, crucially, when the effect of the morphological marking of gender (i.e., the final phoneme in Italian words) is also partialed out ($N = 2,450$). The presence or absence of errors in English was a strong predictor of Italian tool errors, $\chi^2(1, N = 2,450) = 420.81$, $p < .001$. With the exception of shared onset phoneme, all other phonological variables were significant predictors: % segments shared by position, $\chi^2(1, N = 2,450) = 50.56$, $p < .001$; shared number of syllables, $\chi^2(1, N = 2,450) = 33.49$, $p < .001$; shared stress pattern, $\chi^2(1, N = 2,450) = 5.85$, $p < .05$; and shared final phoneme, $\chi^2(1, N = 2,450) = 19.09$, $p < .001$. In contrast to Vigliocco et al. (2005) and because we are using a more sensitive analysis here, we also found that gender significantly predicted probability of error, $\chi^2(1, N = 2,450) = 24.85$, $p < .001$. Crucially, however, the size of this gender effect on tool errors is several orders of magnitude smaller than the gender effect on animal errors: This result can be clearly seen in the plots of the partial effects of gender in the two different analyses (see Figure 2).

It is not clear at this stage whether the small gender effect on tools is syntactic or semantic in origin. The most important finding of this reanalysis, however, is that the magnitude of the gender effect is not invariant across semantic fields, as would be expected if it was purely syntactic in origin, therefore confirming the (uniquely or at least predominantly) semantic origin of the gender effect on animals.

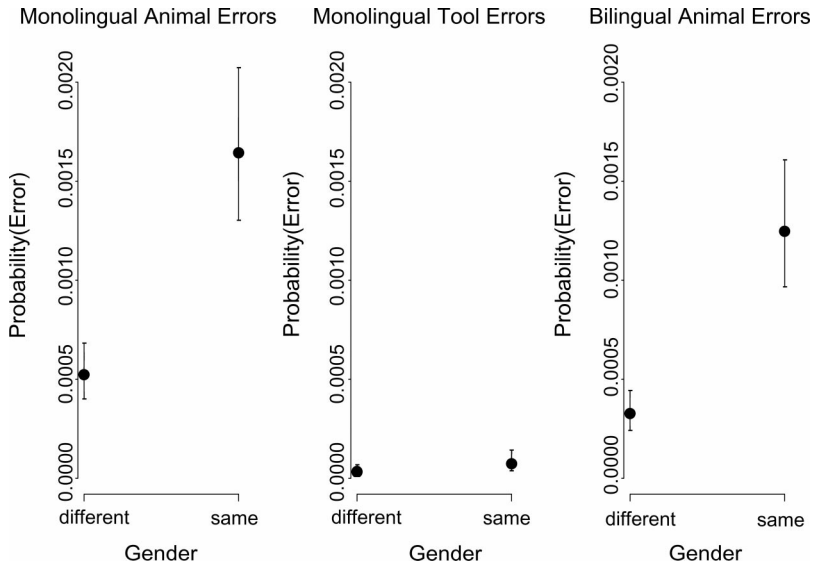

Figure 2. Plots of the partial effect of gender (i.e., the effect of gender when the effect of all other predictors in the regression model is held constant) on probability of error in the monolingual Italian animal data set (left), the monolingual Italian tool data set from Vigliocco et al., 2005 (middle), and the bilingual Italian animal data set (right). All other categorical predictors in each model are held constant at the level *different* (i.e., the effect of gender is calibrated for different onset phonemes, etc.); all numerical predictors in each model are held constant at the median. Error bars represent 95% confidence intervals.

Discussion

Experiment 1 demonstrates that grammatical gender affects perceived semantic similarity in Italian as compared with English (a language with no formal gender), replicating the results in Vigliocco et al. (2005). The monolingual Italian speakers' pattern of errors was different than that of the monolingual English speakers, exhibiting a significantly higher proportion of gender-preserving errors. Grammatical gender, essentially a lexico-syntactic property, has acquired semantic status in Italian, such that the meaning of words is affected by their grammatical specification (masculine vs. feminine). Although this was expected, there was one result that requires explanation: The proportion of gender-preserving errors in Italian was the same for both the bare noun and the noun phrase conditions. However, on the basis of previous work (Paganelli et al., 2003; Vigliocco, Vinson, Indefrey, et al., 2004; Vigliocco et al., 2005), one might have expected gender-preservation rates to have been higher for the noun phrase condition than for the single word condition in Italian. This is because, apart from a semantic effect with these specific nouns, gender has been shown to primarily affect the retrieval of syntactic frames in noun phrase production. Therefore, one might expect the semantic and syntactic effects of gender to be additive. This, however, need not be the case. There is no consensus in the literature as to when and under which circumstances grammatical gender effects arise in noun phrase production, which makes the interpretation of these results more complex. According to one account (Alario & Caramazza, 2002; Caramazza & Miozzo, 1997; Costa & Caramazza, 2002), grammatical gender effects in noun phrase production are due to determiner selection. In languages like Italian, in which determiners are selected on the basis of both syntactic and phonological information, determiner selection occurs late, during the phonological encoding of the noun phrase. Because determiner selection occurs so late it has little effect on lemma selection. Then presumably, gender-preservation rates in our experiment are not significantly different in the bare noun and the noun phrase condition because their source is purely semantic in both cases. According to an alternative account (Vigliocco, Vinson, Indefrey, et al., 2004), gender-preservation effects arise as a result of monitoring at the morphophonological level. Lemmas (abstract forms containing semantic and syntactic but no phonological information) that compete for selection also activate their corresponding syntactic frames, which are specified for gender. In cases of high competition, the most active frame may be retrieved even before a lemma is selected. When a lemma is selected, it is inserted in the frame for morphophonological encoding. Frames with incompatible gender features are more likely to be detected by a monitoring system at this level than are frames with compatible gender features, hence leading to a gender-preservation effect. In this framework, the lack of a difference in gender-preservation rates in the present experiment could be attributed to the fact that fewer ill-formed frames are generated in the first place, because lemma competition is stronger among same-gender lemmas (activated because they are semantically more similar as a result of gender). At any rate, resolving this issue is beyond the scope of this article, and it does not affect the main purpose of this data set, which is to serve as a baseline for bilingual speakers.

Experiment 2: Bilingual Error Induction

Experiment 1 confirmed that monolingual Italian and English speakers differ in the type of errors they make: Italian speakers' spontaneous semantic substitution errors are affected by the grammatical gender of the target and the intruder. This result demonstrates that grammatical gender increases semantic similarity between words that share the same gender in comparison to those that do not. In other words, an arbitrary grammatical property of nouns affects the meaning representation of those nouns in Italian: an effect of language on language. The main question addressed in this article, however, is whether this effect extends beyond first-language semantics to conceptual representations and, hence, second-language semantics. If the effect of gender is limited to semantic representations and does not affect conceptual representations, we expect bilingual speakers to behave differently in each of their languages, following monolingual patterns. If the first language shapes aspects of nonlinguistic cognitive representation, we expect bilingual speakers to behave similarly to Italian monolingual speakers, regardless of the language in which the task is carried out. Finally, it may be the case that learning a second language alters cognitive dispositions established by means of the first language; in this case, grammatical gender should not have any effect on either of a bilingual speaker's languages.

Method

Participants. Twenty-nine native speakers of Italian (10 men, 19 women; mean age: 28.1 $\pm$ 5.1) who were also highly proficient in English were recruited from the London and Cambridge areas. All had learned English after the age of 6 and had been resident in the United Kingdom for at least 9 months prior to the time of testing. Detailed information about the linguistic background of the bilingual speakers appears in Appendix B. Volunteers received monetary compensation at a rate of £12 per hour for their participation.

Materials and design. The materials and design were the same as for Experiment 1.

Procedure. The procedure was the same as that used with monolingual speakers. The only difference was that bilingual speakers did the same experiment in English and Italian (on different days, with at least 1 day intervening between sessions). The order of the two language sessions was counterbalanced across participants to avoid any order effects. The order of the bare noun and the noun phrase conditions was also counterbalanced across participants, but each participant did the two conditions in the same order for the two language sessions.

Results

The data were recorded and transcribed as in Experiment 1. Responses broken down by category appear in Table 2. Once again, to compare English and Italian, we assigned English words the gender that translation-equivalent items have in Italian. As in the monolingual data, all lexical errors were semantically related, with 90.02% of errors in the Italian version and 93.74% of errors in the English version of the task being within-set errors. The proportion of gender-preserving errors as a function of language and response type appears in Figure 3.

Table 2
Number and Percent Occurrence of Responses as a Function of Language and Response Required for Experiment

	Italian				English			
	Bare noun		Noun phrase		Bare noun		Noun phrase	
Response type	No.	%	No.	%	No.	%	No.	%
Acceptable								
Correct	9,856	82.89	10,143	85.31	10,163	88.53	10,237	89.17
Different label	492	4.14	492	4.14	319	2.78	320	2.79
Error								
Lexical	282	2.37	286	2.41	235	2.05	229	1.99
Omission	956	8.04	759	6.38	616	5.37	582	5.07
Self-correction	77	0.65	79	0.66	53	0.46	43	0.37
Miscellanea	227	1.91	131	1.10	94	0.82	69	0.60

Note. One participant did not return for the second (English) session, so the Italian version contains data from 29 particpants, whereas the English version contains data from 28 participants.

We analyzed all lexical errors in the same manner as in Experiment 1. We first excluded mixed errors (errors that shared both semantic and phonological similarity), using the same procedures we used for monolingual speakers. Additionally, we applied a third procedure: We excluded errors from both languages in which the target and the translation of the error shared 50% or more of their phonemes in the same serial order, but disregarding syllable boundaries and structure. These errors resemble mediated errors in aphasic speech, in which, for instance, unicorn is spoken as house, with horse being the mediator (Martin, Dell, Saffran, & Schwartz, 1994). The reason for excluding these errors was because form similarity between languages affects bilingual lexical processing (Colomè, 2001). Indeed, we observed that one particular error involving the same animals in English (target: *bear*; error: *horse*; translation of target: *orso*) and in Italian (target: *cavallo*; error: *orso* (*bear*); translation of the target: *horse*) was particularly frequent in the bilingual data (38 instances in English; 19 instances in Italian) but infrequent in the monolingual data (1 instance in English; 2 instances in Italian[4]). These were the only errors that were excluded (from both the analyses of variance and the regression analyses reported subsequently) on the basis of this procedure.

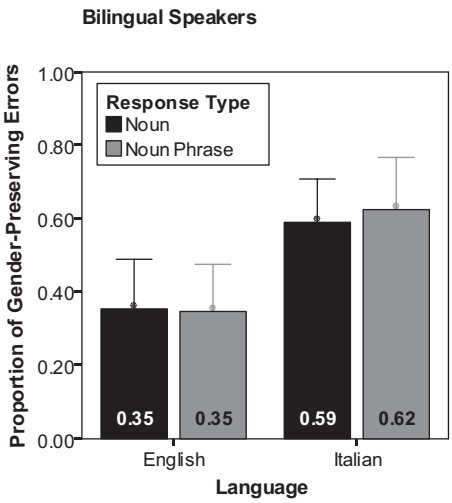

Figure 3. Mean proportion of gender-preserving errors by bilingual Italian and English speakers. Error bars represent 95% confidence intervals.

After mixed errors had been excluded, there remained 347 Italian errors (169 in the bare noun condition and 178 in the noun phrase condition) and 283 English errors (155 in the bare noun condition and 129 in the noun phrase condition). For each participant and item, we calculated the mean proportion of errors that preserved gender, and we submitted the means to two ANOVAs to test generality over participants (F_1) and items (F_2). Only participants who contributed observations to both conditions in both languages were included in the analyses. The analyses yielded the same pattern of results as for the monolingual speakers: There was a significant main effect of language, $F_1(1, 23) = 13.292$, $p < .001$, $MSE = 0.070$, and $F_2(1, 20^5) = 7.499$, $p < .05$, $MSE = 0.042$, with more gender-preserving errors in Italian rather than in English ($M_{\text{difference}} = 0.25$, 95% CI = 0.14). There was no main effect of response type or interaction (Fs < 1). In other words, the bilingual speakers in this experiment were significantly more likely to produce same-gender intruders when they were carrying out the task in Italian than when they were carrying out the same task in English.

As in Experiment 1, we explored this grammatical gender effect with regression analyses. We analyzed the proportion of Italian bilingual errors using the proportion of English bilingual errors, Italian form similarity, and gender as predictors ($N = 700$). Errors in English were a strong predictor of probability of error in Italian, $\chi^2(1, N = 700) = 1{,}298.03$, $p < .001$. All form-related variables were significant predictors, with the exception of shared final phoneme: % shared segments, $\chi^2(1, N = 700) = 25.32$, $p < .001$; shared number of syllables, $\chi^2(1, N = 700) = 24.75$, $p < .001$; shared stress pattern, $\chi^2(1, N = 700) = 25.63$, $p < .001$; and shared onset phoneme, $\chi^2(1, N = 700) = 7.67$, $p < .01$. Gender was also a significant predictor, $\chi^2(1, N = 700) = 68.53$, $p < .001$, showing that bilingual speakers' errors preserved grammatical gender when they were responding in Italian.

[4] These observations were excluded from the monolingual data when comparisons with the bilingual data were carried out.

[5] There are 20 degrees of freedom in this analysis because there were no errors for some of the items.

It was also important to determine to what extent the performance of the bilingual speakers was equivalent to that of the monolingual speakers. We therefore carried out two further sets of ANOVAs assessing the effects of language (bilingual and monolingual, between participants but within items) and response type (bare noun and noun phrase, within participants and items) for each of the two languages. In order to perform these analyses, we used exactly the same criteria for the exclusion of phonologically related errors for monolingual speakers as we did for bilingual speakers. When the bilingual Italian data were compared with the monolingual Italian data, there was no significant main effect or interaction (Fs < 1). For these analyses, as before, participants who had not contributed any observations to one of the two response type conditions (bare nouns vs. noun phrases) were excluded. In order to increase power and investigate further any differences between the two language groups, we collapsed these two conditions, also including observations by participants who had contributed data only to one of the conditions. Once again, there was no significant main effect or interaction ($Fs < 1$). We followed the same procedures for comparing the bilingual English and monolingual English data. There were no significant main effects or interactions: language, $F_1 < 1$, and $F_2(1, 16) = 2.334$, $p = .146$; response type, $F_1(1, 45) = 1.237$, $p = .272$, and $F_2 < 1$; Language $\times$ Response Type, $F_1(1, 45) = 1.581$, $p = .215$, and $F_2 < 1$. After collapsing the two response type conditions, there was again no significant difference between the two participant groups: $F_1 < 1$, and $F_2(1, 23) = 2.227$, $p = .149$.

We followed up these analyses with regression analyses. In a first set of analyses, we wanted to determine whether the effect of gender was reduced in bilingual speakers' first language as compared with monolingual Italian speakers. We carried out two logistic regression models: one in which the proportion of bilingual Italian errors for each picture pair in the experiment, gender, and form-related variables were used to predict the proportion of monolingual Italian errors for each picture pair in the experiment and one in which the proportion of monolingual Italian errors, gender, and form-related variables were used to predict the proportion of bilingual Italian errors for each picture pair. In the model with monolingual errors as the dependent variable, gender was a significant predictor of probability of error, even though the effect of all other predictors had been removed, $\chi^2(1, N = 700) = 17.48$, $p < .001$. However, the same was true for the model with bilingual errors as the dependent variable, $\chi^2(1, N = 700) = 9.21$, $p < .001$. Moreover, the size of the gender effect in the two analyses was extremely similar (see Figure 4): Even errors that were not shared between the two groups of participants were affected by gender in extremely similar ways for both bilingual and monolingual speakers. In other words, gender did not seem to have a diminished effect in the bilingual Italian data. In a further analysis, we asked whether gender predicts some variance in the bilingual English data when the effect of other variables is removed, in order to test whether bilingual speakers' responses in their second language were affected by first-language gender. With proportion of bilingual English errors for each pair of pictures as the dependent variable, and proportion of monolingual English errors, gender, and form-related variables as predictors, we found that gender did not have a significant effect on probability of error, $\chi^2(1, N = 700) = 2.04$, *ns*.

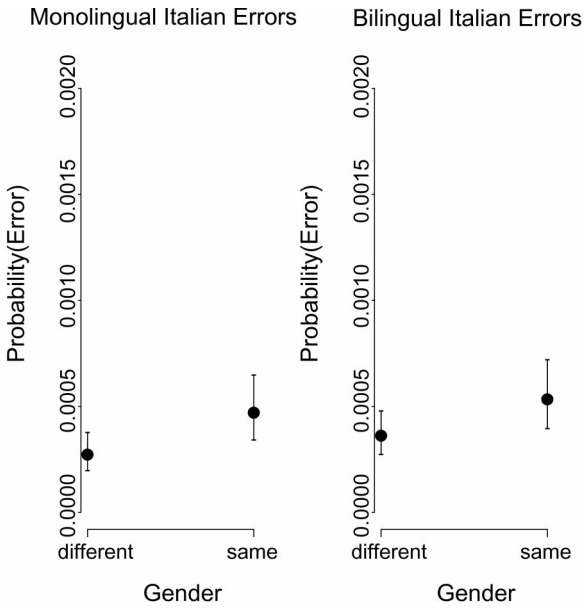

Figure 4. Plots of the partial effect of gender on monolingual Italian errors (when the effect of bilingual Italian errors and other phonological variables is held constant) and on bilingual Italian errors (when the effect of monolingual Italian errors and other phonological variables is held constant). All other categorical predictors in each model are held constant at the level *different* (i.e., the effect of gender is calibrated for different onset phonemes, etc.); all numerical predictors in each model are held constant at the median. Error bars represent 95% confidence intervals.

Discussion

Experiment 2 showed that bilingual Italian–English speakers produced more gender-preserving errors when they were carrying out the task in Italian than when they were carrying out the task in English. Moreover, the proportion of gender-preserving errors for bilingual speakers did not differ significantly from either the monolingual English data (in the comparison with bilingual English data) or from the monolingual Italian data (in the comparison with bilingual Italian data). This cannot be due to a lack of sensitivity in the between-subjects comparison, because there was a significant difference between the two monolingual groups. We interpret these results as evidence for the fact that the semantic representations of bilingual speakers in these studies are appropriate for each of their languages. In other words, there is intraspeaker linguistic relativity and the behavior of bilingual speakers is predicted by the behavior of monolingual speakers.

Evidence for linguistic relativity based on comparisons of the behavior of monolingual speakers has clearly different consequences for claims of linguistic determinism than evidence for linguistic relativity with bilingual speakers. In the monolingual literature, differences between linguistic groups on a specific semantic variable have been interpreted as indicative of the potentially central role language has on shaping cognition, which is normally followed up in a nonlinguistic task. Evidence for linguistic relativity within the same speaker (the bilingual individual) can have only the reverse interpretation: If proficient bilinguals who acquired their second language after their first show evidence of adaptation to monolingual norms for their second language, then

their first language has a very limited effect, that is, an effect on just that language. Therefore, these results obviate the need to perform a nonlinguistic task.

General Discussion

Our aim in this article has been to investigate the strength and pervasiveness of language-specific effects on cognition through the comparison of monolingual and bilingual speakers using the same task and identical stimuli. We asked to what extent a grammatically motivated semantic distinction in a first language would be transferred to a second language or, as a completely opposite alternative, to what extent second-language semantic representations may affect first language representations. We have shown that bilingual Italian–English speakers' performance was significantly different in each of their languages, thus demonstrating relativity in their semantic representations. Moreover, we did not find any evidence of transfer from first language: There was no significant difference between the monolingual and the bilingual English data. Finally, learning a second language without grammatical gender does not seem to affect semantic representations in the gendered first language, because there was no significant difference between the monolingual Italian and bilingual Italian data.

Evidence of transfer from one's native or most proficient language into a second language has been interpreted as evidence for the pervasive effect of language on cognition, especially when highly proficient bilingual speakers are involved (Boroditsky, 2001; Boroditsky et al., 2003). We found no evidence of transfer from Italian to English of the semantic effects of gender and interpret this lack of transfer as evidence for the constrained role grammatical gender has on bilingual cognition. If bilingual speakers are able to develop semantic representations that are appropriate for a second language, even when they are not as proficient in their second language as in their first, this constitutes evidence for a limited cognitive effect of the particular variable. This pattern of results is in line with earlier work that has found no significant difference between monolingual Italian and English speakers in a nonlinguistic triadic judgment task (Vigliocco et al., 2005, Experiment 4).

An effect of grammar on semantic representations that does not extend to conceptual representations can be accommodated within approaches to meaning that make a distinction between lexico-semantic representations, on the one hand, and nonlinguistic conceptual representations, on the other (Jackendoff, 2002; Levinson, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). In the featural and unitary semantic space framework (Vigliocco, Vinson, Lewis, & Garrett, 2004), for instance, conceptual representations are decompositional and semantic representations bind conceptual features to lexico-syntactic and phonological–orthographic representations. Because semantic representations interface with both conceptual and lexico-syntactic information, syntactic features, if highly correlated with conceptual features, can affect the development of semantic representations. Regarding the semantic but not conceptual effect of grammatical gender in particular, it could arise in the following way (Vigliocco et al., 2005; Vigliocco & Kita, 2006). In languages such as Italian, there is a transparent link between grammatical gender and conceptual gender for words referring to human entities. During linguistic development, language learners notice this relationship between the grammatical gender of words referring to humans and the conceptual gender (i.e., sex) of their referents. This transparent relationship increases the probability that words with this lexico-syntactic property will be mapped on broadly similar lexico-semantic space. Once this association is established, it is generalized to other words for which there is an opaque (e.g., animals) or no correspondence (e.g., artifacts) between grammatical and conceptual gender. The strength of the generalized association depends crucially on the degree to which there is conceptual support for the grammatical features: The conceptual representation of animal concepts includes the information that they are sexuated, hence facilitating the association; the conceptual representation of artifacts, on the other hand, does not include any sex-related features, and hence the generalization of the association is extremely weak (if present at all). In other words, lexico-semantic representations develop as an intermediate level that binds conceptual features to lexico-syntactic/phonological forms and as such are shaped by both conceptual and linguistic information. Although this architecture does not preclude an effect of linguistic on conceptual representations, it can also accommodate its absence.

Boroditsky and colleagues (Boroditsky & Schmidt, 2000; Phillips & Boroditsky, 2003), who also investigated gender effects with bilingual and multilingual speakers, reached a strikingly different conclusion than we did. As we mentioned in the introduction, the locus of the discrepancy most likely lies in the types of tasks used. We used a methodology that made the variable of interest immune to the use of strategies: speech error induction. Inadvertently produced semantic substitution errors reflect automatic, online processes and are much more revealing of the effects of gender on semantic representations than unfamiliar tasks that may induce any number of strategies in order to be resolved.

How general are the results we obtained? Could it be reasonably claimed that they are representative of all different linguistic phenomena and all different levels of bilingual proficiency postintermediate level? Such conclusions are not warranted at this stage for a variety of reasons. First, there is very limited research on language and thought carried out with bilingual speakers; this research has investigated an extremely limited set of linguistic phenomena. There is no a priori reason to assume that all differences in grammar would produce exactly the same degree of cognitive effect (given that even the same grammatical property, i.e., gender, seems to have different effects for animals and artifacts as we report here). Second, there is substantial behavioral (Chen & Leung, 1989; Chen, 1990; Kroll & Curley, 1988; Mägiste, 1984; Poulisse & Bongaerts, 1994; Talamas, Kroll, & Dufour, 1999) and neuropsychological (Kotz & Elston-Güttler, 2004; Ojima, Nakata, & Kakigi, 2005; Perani et al., 1998; Tatsuno & Sakai, 2005) evidence to show that proficiency plays a major role in the type of conclusions that can be drawn about bilingual memory and processing. For instance, the fact that we did not obtain any evidence of restructuring (Pavlenko, 1999, 2005) of our bilingual speakers' mother tongue representations may be because our bilingual speakers were not proficient enough in English or had not been immersed in an English-speaking environment long enough for such restructuring to take place. Some evidence for restructuring has been obtained with bilingual speakers with different linguistic profiles than the ones in the present study. For instance, Wolff and Ventura (2003) have shown that Russian–

English speakers' first-language semantic representations of causal verbs are affected by their second language. However, the bilingual speakers in their Experiment 2 seem to be almost equally proficient in their two languages (they reported average proficiency of 8.6 for Russian and 8.2 for English). In other words, it may be possible to show that gender effects on semantic representations are significantly reduced in bilingual speakers' mother tongues by testing bilingual speakers whose proficiency in their two languages is very similar or for whom their second language has displaced their first as the dominant language.

Regarding the implications of our results for models of bilingual semantic memory and lexical access, one might argue that the difference in our bilingual speakers' performance in their two languages indicates that they access each of their languages independently and that two separate semantic stores are involved. However, there is abundant evidence in the literature (summarized in the introduction) to suggest that this is not the case. How, then, is it possible to reconcile the language-independent semantic access, single semantic store account with the observed difference in Experiment 2 above? Under the language-independent, single semantic store account, the intention to produce a verbal label upon seeing a picture of a skunk in bilingual Italian–English speakers activates conceptual features such as *animal*, *lives in North America*, *produces nasty smell when in danger*, *is sexuated*, and so on, which activate the semantic nodes for both *skunk* and *puzzola* (its translation equivalent in Italian), as well as other semantically related words in both languages. Experiment 1 and earlier research show that the semantic competitors of *skunk* are not exactly the same as the semantic competitors of *puzzola*, even though the two words do not differ in reference. The semantic representation of the word *puzzola* binds the conceptual feature "female" (and possibly other features related to female entities) to the linguistic form by virtue of the latter's grammatical gender and is hence more semantically similar to words that share the same gender in Italian. For instance, *puzzola* shares more of its semantic representation with *giraffa* ($giraffe_{fem}$) than *skunk* does with *giraffe*. If competition for selection is language independent, then Italian–English bilingual speakers should show the same pattern of errors in both their languages (which, presumably, should also differ from that of monolingual speakers of either language). This is clearly not the case. One account for how the current pattern of results can be obtained is in terms of Green's inhibitory control model (Green, 1998b, 2003). According to this model, the intention to carry out a linguistic action (e.g., utter a word in a specific language) gives rise to competing language schemas. A supervisory attentional system specifies the intended language (e.g., Italian or English) and transmits it to the competing language schemas. The schemas regulate the activation levels of representations in the lexico-semantic system and inhibit outputs from the system. Selection of the language-appropriate alternative occurs at the lexico-semantic level: Lexico-semantic representations are language tagged, and those with inappropriate tags are suppressed, but only after they have been activated. This is just one account of how the nonspecific access proposal is compatible with our data. There are numerous other proposals that can account for our results (e.g., accounts that do not involve inhibition; De Bot & Schreuder, 1993; Poulisse & Bongaerts, 1994). The important point is to show that our results are not incompatible with the single-storage, language-independent access account of bilingual semantic memory.

The data presented above are also relevant to questions regarding the representation of concrete words in the bilingual lexicon. Despite claims that the representation of cognates and translation-equivalent concrete words are identical, it has become apparent that translation-equivalent concrete words very often are not coextensive (Paradis, 1997) and hence not identical in meaning. Ameel et al. (2005) and Malt and Sloman (2003) have provided some evidence showing that the representations of translation-equivalent words with concrete referents, which, crucially, are not coextensive (i.e., they do not refer to the same entities), are only partially overlapping. Here we show that words that share extension can also have nonidentical semantic representations. This concept is important in demonstrating that appropriate second-language lexical knowledge is not acquired purely through mismatches in extension as previously thought (de Groot, 1992a, 1992b; Van Hell & de Groot, 1998a, 1998b). In other words, for a second-language learner to observe that the translation equivalent of the word *bottle* does not refer to the same type of things in Mandarin as it does in English is not the only way of fine tuning second-language semantics. Subtler, intralinguistic, nonreferential information (e.g., the absence of grammatical gender) can shape the representation of meaning in the second language.

Conclusion

In this article, we demonstrate the validity of comparing the performance of monolingual and bilingual populations on a purely linguistic task as a method of investigating the strength and pervasiveness of language effects on cognition. We also show how the investigation of linguistic relativity through bilingualism makes it possible to ask a wider range of questions concerning the link between language and thought. We found that bilingual speakers performed differently in each of their languages, and their behavior followed that of monolingual speakers of each language. These results demonstrate that Italian grammatical gender cannot logically have an effect on the nonlinguistic, conceptual representations of bilingual speakers and lend support to work that has failed to find a gender effect in a conceptual task with monolingual speakers.

Bilingualism and multilingualism seem to be the norm rather than the exception around the world (Harris & McGhee Nelson, 1992). Despite this, research on language-specific effects on cognition has focused on monolingual populations, not addressing how two or more different ways of talking about the world are accommodated within a single mind–brain. This type of research is essential not only in order to determine how language may affect cognition but also in order to understand the cognitive consequences of bilingualism.

References

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition, 82,* 179–223.

Ameel, E., Storms, G., Malt, B., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 52,* 309–329.

Athanasopoulos, P. (2001). *L2 acquisition and bilingual conceptual struc-*

ture. Unpublished master's thesis, University of Essex, Colchester, United Kingdom.

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition, 9,* 89–96.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22.

Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, and semantics. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 42–47). Mahwah, NJ: Erlbaum.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.

Brysbaert, M., Fias, W., & Noël, M.-P. (1998). The Whorfian hypothesis and numerical cognition: Is "twenty-four" processed in the same way as "four-and-twenty"? *Cognition, 66,* 51–77.

Caramazza, A., & Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology, 34,* 77–81.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition, 64,* 309–343.

Caskey-Sirmons, L. A., & Hickerson, N. P. (1977). Semantic shift and bilingualism: Variation in the color terms of five languages. *Anthropological Linguistics, 19,* 358–367.

Chen, H. C. (1990). Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory and Cognition, 18,* 279–288.

Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 316–325.

Chen, H. C., & Ng, N. L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition, 17,* 454–462.

Colomè, A. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45,* 721–736.

Cook, V. J., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism, 10,* 137–152.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding during speech production. *Journal of Memory and Language, 46,* 153–177.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41,* 365–397.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language, 53,* 42–59.

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 191–214). Amsterdam: John Benjamins.

de Groot, A. M. B. (1992a). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 389–412). Amsterdam: North Holland.

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

de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30,* 90–123.

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

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Eberhard, K. M., Scheutz, M., & Heilman, M. (2005). An empirical and computational test of linguistic relativity. In B. B. Bara, L. Barsalou, & M. Buccarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 618–623). Mahwah, NJ: Erlbaum.

Ehri, L. C., & Ryan, E. B. (1980). Performance of bilinguals in a picture-word interference task. *Journal of Psycholinguistic Research, 9,* 285–302.

Ervin, S. E. (1961) The connotations of gender. *Word, 18,* 249–261.

Finkbeiner, M., Nicol, J., Greth, D., & Nakamura, K. (2002). The role of language in memory for actions. *Journal of Psycholinguistic Research, 31,* 447–457.

Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin, 125,* 193–222.

Garrett, M. (1984). The organization of processing structure for language production. In D. Caplan, A. Lecourse, & A. Smith (Eds.), *Biological perspectives on language* (pp. 172–193). Cambridge, MA: MIT Press.

Gennari, S. P., Sloman, S., Malt, B., & Fitch, T. (2002). Motion events in language and cognition. *Cognition, 83,* 49–79.

Grande dizionario Italiano. (2006). Milano: Garzanti Linguistica.

Green, D. W. (1998b). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1,* 67–81.

Green, D. W. (2003). The neural basis of the lexicon and the grammar in L2 acquisition: The convergence hypothesis. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and the lexicon in second language acquisition* (pp. 197–218). Amsterdam: John Benjamins.

Harris, R. J., & McGhee Nelson, E. M. (1992). Bilingualism: Not the exception any more. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 3–14). Amsterdam: North-Holland.

Huntley-Fenner, G., Carey, S., & Solimando, A. (2002). Objects are individuals but stuff doesn't count: Perceived rigidity and cohesiveness influence infants' representations of small numbers of discrete entities. *Cognition, 85,* 203–221.

Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Gover, G. H., Poldrack, R., et al. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language, 70,* 347–363.

Imai, M., & Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition, 62,* 169–200.

Jackendoff, R. (2002). Foundations of language. Oxford, England: Oxford University Press.

Jameson, K. A., & Alvarado, N. (2003). Differences in color naming and color salience in Vietnamese and English. *Color Research and Application, 28,* 113–138.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48,* 16–32.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research, 22,* 519–534.

Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics, 17,* 215–235.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 2, pp. 389–395). London: Wiley.

La Heij, W., Mark, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research, 61,* 209–219.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. Oederson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109–169). Cambridge, MA: MIT Press.

Levinson, S. (1997). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology, 7,* 98–131.

Levinson, S. C. (2003). Language and mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 25–46). Cambridge, MA: MIT Press.

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the tables: Language affects spatial cognition. *Cognition, 84,* 155–188.

Li, P., & Gleitman, L. (2002). Turning the tables: Spatial language and spatial cognition. *Cognition, 83,* 265–294.

Lucy, J. A. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis.* Cambridge, England: Cambridge University Press.

Lucy, J., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, England: Cambridge University Press.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 304–315.

Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition, 6,* 47–67.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. E. (1994). Origins of paraphasia in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and language, 47,* 609–660.

Martin, N., Gagnon, D. A., Schwartz, M. E, Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11,* 257–282.

Martinez, I. M., & Shatz, M. (1996). Linguistic influences on categorization in preschool children: A crosslinguistic study. *Journal of Child Language, 23,* 529–545.

Mazuka, R., & Friedman, R. (2000). Linguistic relativity in Japanese and English: Is language the primary determinant in object classification? *Journal of East Asian Linguistics, 9,* 325–351.

McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46,* 229–259.

Munnich, E., Landau, B., & Dosher, B.A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81,* 171–207.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17,* 1212–1228.

Oxford University Press. (n.d.). *Oxford English dictionary online*. Retrieved June 1, 2005, from http://www.oed.com/

Paganelli, F., Vigliocco, G., Vinson, D. P., Siri, S., & Cappa, S. (2003). An investigation of semantic errors in unimpaired and Alzheimer's speakers of Italian. *Cortex, 39,* 419–439.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition, 84,* 189–219.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 331–354). Mahwah, NJ: Erlbaum.

Pavlenko, A. (1999). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition, 2,* 209–230.

Pavlenko, A. (2005). Bilingualism and thought. In A. de Groot & J. Kroll (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 433–453). Oxford, England: Oxford University Press.

Pechmann, T., & Zerbst, D. (2002). The activation of word class information during speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 233–243.

Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language, 74,* 557–589.

Perani, D., Paulesu, E., Sebastian-Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121,* 1841–1852.

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15,* 36–57.

R Development Core Team. (2007). *R: A Language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Scheutz, M. J., & Eberhard, K. M. (2004). Effects of morphosyntactic gender features in bilingual language processing. *Cognitive Science, 28,* 559–588.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Schriefers, H. J., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 941–950.

Sera, M. D., Berge, C., & del Castillo-Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development, 9,* 261–292.

Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodriguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General, 131,* 377–397.

Slobin, D. (1996). From "thought and language" to "thinking for speaking." In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, England: Cambridge University Press.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's induction of word meaning: Object terms and substance terms. *Cognition, 38,* 179–211.

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.

Tatsuno, Y., & Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *Journal of Neuroscience, 25,* 1637–1644.

Tokowicz, N. (2000). *Meaning representation within and across languages.* Unpublished doctoral dissertation, Pennsylvania State University, State College, PA.

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

Van Hell, J. G., & de Groot, A. M. B. (1998b). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology, 51*(A)*,* 41–63.

Vigliocco, G., & Filipovic Kleiner, L. (2004). From mind in the mouth to language in the mind. *Trends in Cognitive Sciences, 8,* 5–7.

Vigliocco, G. & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes, 21,* 790–816.

Vigliocco, G., Vinson, D., Indefrey, P., Levelt, W., & Hellwig, F. (2004). Role of grammatical gender and semantics in German word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 483–497.

Vigliocco, G., Vinson, D., Lewis, W., & Garrett, M. F. (2004). Representing the meaning of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48,* 422–488.

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General, 134,* 501–520.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* [J. Carroll, Ed.]. Cambridge, MA: MIT Press.

Wolff, P., & Ventura, T. (2003). When Russians learn English: How the meaning of causal verbs may change. In B. Beachley, A. Brown, & F. Conlin (Eds.). *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 822–833). Boston: Cascadilla Press.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16, 90–*95.

Appendix A

Experimental Items

Italian	Gender	English
Asino	M	Donkey
Cammello	M	Camel
Cane	M	Dog
Capra	F	Goat
Cavallo	M	Horse
Cervo	M	Deer
Coniglio	M	Rabbit
Elefante	M	Elephant
Gatto	M	Cat
Giraffa	F	Giraffe
Leone	M	Lion
Leopardo	M	Leopard
Lupo	M	Wolf
Maiale	M	Pig
Mucca	F	Cow
Orso	M	Bear
Pantera	F	Panther
Pecora	F	Sheep
Puzzola	F	Skunk
Riccio	M	Hedgehog
Scimmia	F	Monkey
Scoiattolo	M	Squirrel
Talpa	F	Mole
Tigre	F	Tiger
Topo	M	Mouse
Volpe	F	Fox
Zebra	F	Zebra

Note. F = feminine; M = masculine.

Appendix B

Linguistic (English–Italian) Background Information for Bilingual Participants

Mean age of acquisition: 10.1 years (*SD* = 2.9)
Mean period of instruction: 7.4 years (*SD* = 2.0)
Mean length of stay in an English-speaking country: 4.6 years (*SD* = 3.5)
Mean % language usage per day: English, 70.8% (*SD* = 19.8%); Italian, 29.2% (*SD* = 19.8%)

Self-rated mean proficiency, from 1 (*almost none*) to 7 (*native*)

	English	Italian
Speaking	5.7 (*SD*=0.9)	6.9 (*SD* = 0.3)
Reading	6.1 (*SD*=0.7)	7
Writing	5.6 (*SD*=0.9)	7
Listening	6.0 (*SD*=0.9)	7

Received July 12, 2007
Revision received February 12, 2008
Accepted February 13, 2008 ■