Journal of Experimental Psychology: General
2005, Vol. 134, No. 4, 501–520

0096-3445/05/$12.00 DOI: 10.1037/0096-3445.134.4.501

Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use

Gabriella Vigliocco, David P. Vinson, Federica Paganelli, and Katharina Dworzynski
University College London

In 4 experiments, the authors addressed the mechanisms by which grammatical gender (in Italian and German) may come to affect meaning. In Experiments 1 (similarity judgments) and 2 (semantic substitution errors), the authors found Italian gender effects for animals but not for artifacts; Experiment 3 revealed no comparable effects in German. These results suggest that gender effects arise as a generalization from an established association between gender of nouns and sex of human referents, extending to nouns referring to sexuated entities. Across languages, such effects are found when the language allows for easy mapping between gender of nouns and sex of human referents (Italian) but not when the mapping is less transparent (German). A final experiment provided further constraints: These effects during processing arise at a lexical–semantic level rather than at a conceptual level.

Keywords: semantic similarity, linguistic relativity, cross-linguistic research, grammatical gender, language development

There are nearly 7,000 spoken languages and more than 100 sign languages documented to date in the world (Grimes, 2004). Languages differ in many ways, beyond the obvious differences in, for example, the repertoire of phonemes, phonotactic rules, word forms, and sentence-level syntax. Of interest here is that, as Roman Jakobson (1959) put it, "languages differ essentially in what they *must* convey and not in what they *may* convey" (p. 236). That is, languages differ in which conceptual properties must be realized in sentential forms, and such cross-linguistic differences in obligatory expression may lead speakers of different languages to pay more attention to those dimensions of meaning that are obligatorily expressed in their language and less attention to those dimensions that are optional (Lucy, 1992). A wide body of evidence from a variety of domains suggests that such differences do have consequences on cognition in general (e.g., Boroditsky, 2001; Davidoff, Davies, & Roberson, 1999; Levinson, 1996; Lucy, 1992; Sapir, 1921; Sera, Elieff, Forbes, Burch, Rodriguez, & Dubois, 2002; Whorf, 1956) or at least on cognition pertaining to language use (Brysbaert, Fias, & Noel, 1998; Gennari, Sloman, Malt, & Fitch, 2002; Slobin, 1982, 1996). Here we investigate the source of such effects as well as how strong and pervasive they can be.

Gabriella Vigliocco, David P. Vinson, Federica Paganelli, and Katharina Dworzynski, Department of Psychology, University College London, London, England.

The research reported here was supported by Economic and Social Research Council Grant RES000230038. We thank Stavroula Kousta for her comments on drafts of this article.

Correspondence concerning this article should be addressed to Gabriella Vigliocco, Department of Psychology, University College London, 26 Bedford Way, London WC1H 0AP, England. E-mail: g.vigliocco@ucl.ac.uk

Regardless of whether any effect of language-specific properties is limited to the verbal domain, if such an effect is present in adulthood, it must arise as a consequence, or by-product, of language-learning mechanisms. By assessing whether language-specific properties affect meaning representations of adult speakers, and the strength and breadth of such effects, we can gain some insight into the mechanisms that support language development. We focus on grammatical gender, which, by virtue of its (largely) arbitrary link to conceptual properties of objects, provides us with a strong, conservative test of language-specific effects on cognition and of hypotheses concerning the mechanisms. Broadly stated, we test two classes of hypotheses by which gender effects may arise: In one, gender effects arise as a consequence of similarity in linguistic contexts (similarity and gender); in the other, such effects arise as a generalization from the transparent relationship between sex of human referents and gender of nouns (sex and gender). We test these hypotheses by investigating the gender systems of two languages: Italian and German (summarized in Appendix A).

Mechanisms by Which Grammatical Gender Might Affect Cognition

If effects of grammatical gender on meaning can be found for adult speakers, these effects must arise as a consequence (or by-product) of language-learning mechanisms. Here we consider some alternative mechanisms by which such effects could come about during language development. One possibility, which we call the similarity and gender hypothesis, is that these effects arise as a by-product of a very general learning mechanism used by children to bootstrap aspects of meaning from the linguistic input (Fisher, 1994; Fisher & Gleitman, 2002; Landau & Gleitman,

1985; Landauer & Dumais, 1997).[1] The basic idea is that words that have similar syntactic and morphophonological properties also tend to have similar meanings. Nouns that share the same gender are used in the same linguistic contexts, which differ from those contexts in which nouns of a different gender are used. These differences in linguistic context can be observed at the syntactic level, because words with the same gender require gender agreement with determiners, adjectives, and pronouns in sentences. These differences in context can also be observed at the morphophonological level, because there are correlations in languages between the syntactic specification of gender and how it is realized in morphological and phonological forms for determiners, adjectives, and pronouns as well as inflectional affixes of nouns. This line of reasoning has been proposed in the literature for different syntax-to-meaning mappings: For example, the distinction between count and mass nouns in English has been argued to help children learn the distinction between entities and stuff (Bloom, 1994), and subcategorization information (i.e., whether verbs can take objects and the kinds of syntactic structures that are permitted for a particular verb) has been argued to help children learn aspects of meaning of these verbs (Fisher, 1994; Fisher, Gleitman, & Gleitman, 1991). In this view, gender effects would not depend on establishing associations between grammatical gender of nouns and sex of referents.

The similarity and gender hypothesis predicts that, across languages, these effects should be found for languages with two genders (like Italian) as well as for other languages with three or more genders (like German), because aspects of similarity in linguistic context are present across languages regardless of whether the genders of nouns can be mapped into the two sexes as long as the languages are morphologically rich, thus providing a sufficient number of gender-marked sentence contexts. Within a language, gender effects should be found for all words regardless of the type of referent, because they are based on general aspects of similarity rather than only on whether sex of referents can be associated with gender of words.

Alternatively, effects of grammatical gender could be based on establishing associations between gender of nouns and sex, which we call the sex and gender hypothesis. Across languages there is a core correspondence between genders of nouns and sex of their referents for humans. Children who learn a gendered language could notice this correspondence between gender and sex (male or female features) for nouns referring to humans; they could then extend this principle to encompass other nouns for which there is no direct correspondence but which still refer to sexuated entities (animals). This generalization would lead to similarity effects: Words of the same gender would be more similar among themselves than words of different gender by virtue of sharing male- or female-like semantic properties. In other words, an association would first be established between the gender of nouns referring to humans and the sex of referents because of the co-occurrence of linguistic features (gender: masculine or feminine) and conceptual features (sex: male or female). Once this association is established for words referring to humans, it is then generalized to other nouns: those that possess both the linguistic (gender of nouns) and conceptual (sex) features.

The sex and gender hypothesis predicts differential gender effects across languages. Such effects will be strongest for languages with the greatest correspondence between the gender of nouns referring to humans and the sex of the referents (as is the case for Romance languages in which there are only two genders and few exceptions to the consistent mapping between the gender of nouns referring to humans and the sex of the referents), because the high degree of transparent correspondence greatly simplifies the learning task. They will be weaker (or absent) in languages with multiple genders and/or in which nouns referring to humans fall into more than two classes (like German); in these cases, it is harder for the language-learning child to establish the association between sex and gender of human referents.

Within a language, under this formulation of the sex and gender hypothesis, whether gender effects will be observed depends on the content domain: Such effects should only be found for sexuated entities and not for other entities that lack the relevant conceptual features (those related to sex). However, it is important to note that a less constrained version of the sex and gender hypothesis is also a possibility: The strong association between the gender of nouns and the sex of human referents could go beyond simply enhancing male- or female-like conceptual properties of other sexuated referents, leading to the assignment of male- or female-like conceptual properties, even to those entities for which sex is not a relevant conceptual dimension, simply by virtue of masculine or feminine gender marking on the corresponding nouns. In this less constrained version of the sex and gender hypothesis, gender effects within a language should extend to all words.

Under both the similarity and gender and the sex and gender hypotheses, effects of gender are predicted in verbal tasks, in which conceptual and lexical (lexicosemantic) information is retrieved. Both hypotheses are silent regarding whether these effects are limited to verbal tasks or whether they extend to nonverbal tasks in which conceptual, but not necessarily lexical, information is retrieved. Nonetheless, if we were to find that the effects only arise when lexical information is retrieved, this finding would provide further important constraints on the mechanisms.

When Grammatical Gender Affects Cognition and When It Does Not

A number of previous studies have investigated whether grammatical gender of nouns has consequences for the mental representation of the corresponding entities in the world. These studies investigated whether gender effects can be observed in gendered languages rather than the mechanisms that could give rise to any such effect if present. In general, these previous studies (explicitly or implicitly) assumed an association between grammatical gender and male or female properties of referents and investigated whether concepts for masculine nouns are male-like and concepts for feminine nouns are female-like regardless of semantic category. Thus, getting back to the hypotheses we presented previ-

[1] There are fundamental differences between views such as those of Fisher and Gleitman (2002) and Landauer and Dumais (1997) in the assumptions and in the type of probabilistic information that is taken to be central in the process of deriving aspects of meaning from linguistic context. Regardless of whether syntactic information, co-occurrence, or a combination of the two are used by the child, meaning is inferred from linguistic context in both views.

ously, they may be considered as a test of a less constrained version of the sex and gender hypothesis.

Konishi (1993) asked Spanish and German speakers to rate words on a semantic differential scale (Osgood, Suci, & Tannenbaum, 1957) and found that grammatically masculine words were rated higher on semantic dimensions that have masculine connotations, such as *power*. Crucially, speakers of Spanish and German differed in their ratings for words that had a different gender in the two languages; for example, the word *fork* is masculine in Spanish (*tenedor*) and feminine in German (*Gabel*). On the basis of these results, Konishi argued that words' grammatical genders and conceptual representations are intimately related. Because Konishi used words referring to inanimate objects, his results appear to be problematic for the sex and gender hypothesis. However, no effect was found for feminine nouns, which should have differed along the nurture dimension, which has very strong feminine connotations. Further, in a similar test contrasting English and German speakers, Mills (1986) found that German and English speakers' ratings of nouns did not differ on the basis of German gender.

In a more explicit test of the link between male and female properties and grammatical gender, Sera, Berge, and del Castillo-Pintado (1994) asked Spanish and English speakers to assign a male or female voice to pictured objects and words. Unlike English speakers, Spanish speakers tended to follow the Spanish gender differences in their assignments of voice. Sera et al. (2002) used the same method to investigate two other gendered languages, French and German, in addition to Spanish and English. They replicated the difference between Spanish and English and reported that speakers of French showed gender-specific effects similar to the Spanish speakers. However, German speakers did not perform differently from English speakers. Thus, using the same methodology, effects were reported for Spanish and French but not for German. The finding of an effect in Spanish and French (both Romance languages with two gender classes, like Italian) but not in German is compatible with the sex and gender hypothesis. The finding of gender effects for inanimate objects is compatible with the similarity and gender hypothesis or with an unconstrained version of the sex and gender hypothesis according to which effects of gender are not limited to sexuated entities. Finally, the finding that the effect of gender was the same for words and pictures suggests an effect that goes beyond the domain of language. Note, however, that the studies by Sera et al. are susceptible to criticisms related to the use of strategies in the task, because participants were explicitly asked to classify words according to male–female properties; thus, speakers could use grammatical gender in a conscious manner as a way to solve the puzzling task. This would give rise to the observed effect in Spanish and French. For German, even if participants strategically assigned a female or male voice to characters on the basis of gender for words with masculine or feminine gender, this could not be done for the neuter words, thus leading to a null result.

In a developmental study, Martinez and Shatz (1996) found effects of (Spanish) grammatical gender in 3- to 4-year-old children on a sorting task involving only pictures. In the crucial condition, children were instructed to sort pictures of people and objects into groups "that go together." Unlike any of the English-speaking children, a number of the Spanish-speaking children (6 of 18) sorted items on the basis of grammatical gender. This finding is compatible with both the similarity and gender hypothesis as well as with a relatively unconstrained version of the sex and gender hypothesis. However, again, it is unclear whether these results were obtained not because of broad conceptual consequences of grammatical gender but because some children overtly sorted according to grammatical gender (an acceptable pattern of performance under the vague instructions, which did not explicitly specify that the sorting be done on the basis of meaning). Furthermore, these results conflict with those obtained by Sera et al. (2002), who instead found that the use of gender in assigning a male or female voice to an object was not observed before the age of 6.

In language production, Vigliocco, Vinson, Indefrey, Levelt, and Hellwig (2004) investigated semantically related substitution errors (e.g., saying "arm" when "leg" is intended) in German. The question addressed in the relevant part of the study was whether erroneously produced nouns shared grammatical gender with the intended word. Semantically related substitution errors are assumed to arise as a consequence of competition among semantically similar meaning representations that guide lexical retrieval in speech. Therefore, if words that share the same gender are semantically more similar than words that do not share gender, the likelihood of substituting one word with another of the same gender should be higher than that of substituting one word with another of a different gender. No language-specific effects of grammatical gender were found in this study (even for words referring to animals), and it was, therefore, concluded that grammatical gender of German nouns does not affect meaning similarity among the words. This finding is consistent with the sex and gender hypothesis, which predicts no gender effects in German because of the lack of transparency of gender marking. However, the null effect of gender cannot be considered strong evidence in favor of the sex and gender hypothesis because this study did not include a language in which this hypothesis would predict a gender effect.

Thus, taken together, previous studies suggest that gender of words can sometimes affect meaning representations, at least in some languages. The reported results appear compatible with the similarity and gender hypothesis in some cases and the sex and gender hypothesis in others (e.g., effects in German are found by Konishi, 1993, but not by Mills, 1986, Sera et al., 2002, or Vigliocco, Vinson, Indefrey, et al., 2004); methodological issues (particularly the possibility that participants could have used gender in a strategic manner) only serve to obfuscate the matter. Here we present experiments that are, instead, precisely developed to test the predictions of the alternative hypotheses we have described, using tasks that minimize the possibility that participants can use grammatical gender in a strategic manner.

Plan of the Study

We present a series of four experiments investigating Italian and German using tasks that are sensitive to meaning similarity. To ensure that any gender effect we observed was language specific (rather than simply reflecting general correspondences between grammatical gender and meaning as demonstrated by Foundalis, 2002), we conducted parallel experiments with speakers of gendered languages (Italian, German) and with English speakers.

Because English does not have grammatical gender, the performance of English speakers provides us with a baseline against which to assess any language-specific gender effects above and beyond aspects of semantic and visual similarity, which may be correlated with grammatical gender. In all experiments, we contrasted words referring to animals with words referring to artifacts. The latter category provides a crucial contrast between the similarity and gender hypothesis and the sex and gender hypothesis; only the former predicts a language-specific effect. For animals, both hypotheses predict that language-specific gender effects should be observed either from the link between grammatical gender of nouns and sex of referents (sex and gender) or by the same general mechanisms that apply to all nouns (similarity and gender). Experiments 1 and 2 tested these predictions for Italian speakers. Experiment 3 replicates Experiment 1 in German as an additional test of the hypotheses (gender effects should be observed only in Italian under the sex and gender hypothesis but in both languages under the similarity and gender hypothesis). Finally, Experiment 4 used pictures as materials to test the extent of gender effects, assessing whether gender effects are only observed when lexicosemantic representations are recruited in a task or whether gender effects extend to cases in which the task could be carried out entirely on the basis of conceptual information.

Experiment 1: Triadic Similarity Judgments With Italian and English Words

In Experiment 1 speakers of Italian and English were presented with triplets of words (translation equivalents in the two languages), and their task was to judge which two of the three were most similar in meaning. This similarity judgment task has been successfully used in previous studies investigating semantic organization and its impairments (Fisher, 1994; Fisher et al., 1991; Garrard, Carroll, Vinson, & Vigliocco, 2004; Grossman, Mickanin, Onishi, & Hughes, 1996; Romney, Moore, & Rusch, 1997). Moreover, this task has been shown to be sensitive to linguistic variables at the interface between meaning and syntax. For example, Fisher (1994) showed that English speakers' judgments reflected differences in the kinds of syntactic structures that are permitted for a particular verb; Garrard et al. (2004) showed that English speakers' judgments reflected the distinction between count and mass nouns for food items (for which the semantic divide between countable entities and substances is less obvious). Thus, if grammatical gender of Italian nouns renders nouns sharing the same gender more semantically similar than words not sharing gender, we should observe language differences in this task. Different groups of participants were presented with words referring to animals and to artifacts. Whereas the similarity and gender hypothesis predicts language-specific effects of gender in both semantic categories, the sex and gender hypothesis predicts that gender effects should either be limited to words referring to animals or at least should be stronger for words referring to animals than for those referring to artifacts.

Method

Participants. Participants were 36 native Italian speakers (24 women, 12 men) and 36 native English speakers (27 women, 9 men) from the University College London participant pool. None of the Italian participants reported having fluent knowledge of English: They were Italian undergraduate and postgraduate students visiting London. None of the English speakers reported moderate or better competence in any Romance language. Educational levels were comparable across the two groups. All were paid £3 ($5.30) for participating.

Materials. Words referring to 20 animals (medium–large mammals) and 24 artifacts (tools and implements) were selected for the experiment. We selected only words that were easily recognizable in pictorial format (by both Italian and English speakers) because we wanted to use the same items in all experiments. We also ensured that words of both genders were represented in the item sets. Note here that, for animals in Italian, in the majority of cases the noun (masculine or feminine) refers to both male and female animals. However, because the inflectional process for marking gender for human referents is productive, this can also apply to animals. In particular, a base masculine noun can almost always be turned into a feminine noun for words referring to animals (e.g., *lup-o*, *lup-a* [male wolf, female wolf]; see Dardano & Trifone, 1985; Lepschy & Lepschy, 1981). The situation is different for feminine words referring to animals that, instead, cannot be morphologically changed.[2] The items used in this experiment are listed in Appendix B.

Triads for the Italian and English conditions were created by assembling all possible three-word combinations from the 20 words referring to animals in the experimental set (a total of 1,140 triads) and separately for the 24 words referring to artifacts (a total of 2,024 triads). The order of words in each triad was randomized, and then the order of triads was randomized across participants. The 1,140 animal triads in each language were divided into three lists, each containing 380 word triads; the 2,024 artifact triads were divided into six lists, each containing 337 or 338 word triads. Word triads were printed on size A4 sheets in 10-point (capitalized) Times New Roman font and were presented in a two-column format (eight pages per participant).

Procedure. Each participant judged only a subset of the triads (see Fisher, 1994). Four participants completed each list of 380 animal items; thus, each triad was judged by four different speakers of a language (12 participants per language judged animal triad). Four different participants completed each list of 337 or 338 artifact items; again, each triad was judged by four different speakers (24 participants per language judged artifact triad). All participants were told that the experiment concerned participants' judgments of meaning similarity among triplets of words, and that their task was to choose the two words of the three that were most similar in meaning and to delete the odd one. Instructions emphasized that the decision was to be made on the basis of meaning and not other types of similarity between the words (e.g., phonological or visual similarity). Participants typically completed the task in about 30 min, after which they were asked to describe the strategies that they used in performing the task and to list the easiest and most difficult decisions. The most important aspect of these questions was whether any Italian participants mentioned grammatical gender as an overt basis for making their decisions.

[2] This difference between masculine and feminine nouns referring to animals may introduce a confound. Masculine nouns, although referring to animals of both sexes, may be considered as more male-like by Italian speakers because a feminine version of these nouns is possible. Feminine nouns (also referring to both sexes), instead, may be considered as more neutral because they cannot be changed into masculine. We assessed the role of this potential confound in a pretest in which we included the words we used in Experiments 1 and 2. In the pretest, we asked speakers to judge how often they thought each word could be used to refer to a male or female referent. No difference between the masculine and feminine words was found.

Results

No Italian participants indicated that they used grammatical gender in their similarity judgments in the postexperimental questionnaire. Word pairs selected by each participant were classified into same-gender and different-gender pairs. The dependent variable was the proportion of same-gender word pairs selected as similar, considering only those triads that offered the opportunity for selecting different-gender pairs (i.e., those triads containing two words with masculine [Italian] gender and one word with feminine gender, and those triads containing two feminine words and one masculine word). The average percentage of same-gender pairs selected in the critical triads is presented in Figure 1 and Table 1.

A 2 (category: animals, artifacts) × 2 (language: Italian, English) between-subjects analysis of variance (ANOVA) revealed a main effect of language, $F(1, 68) = 6.97$, $p = .01$, $MSE = .00129$; Italian speakers tended to select word pairs sharing gender more often than did English speakers. The main effect of category was not significant, $F(1, 68) = 1.93$, $p = .17$, $MSE = .00129$. The Language × Category interaction, however, was significant, $F(1, 68) = 5.63$, $p = .02$, $MSE = .00129$. The language-specific gender effect was limited to the category of animals and was not observed for artifacts. This was confirmed by tests of simple main effects: animals, $t(22) = 2.83$, $p = .005$, one-tailed; artifacts, $t(46) = 0.24$, $p = .406$.

Discussion

Meaning similarity judgments by Italian speakers were affected by the grammatical gender of the noun for words referring to animals but not for words referring to artifacts.

Table 1
Average Percentage of Same-Gender (Italian) Word Pairs Selected in Triadic Similarity Judgments by Language and Category.[a]

	Category	
Language	Animals	Artifacts
Italian		
M	33.7	30.4
SEM	0.9	0.7
English		
M	29.2	30.1
SEM	0.9	0.6

[a] Considering only triads with two words of one gender and one word of the other gender.

Before discussing the theoretical implications of these findings for the similarity and gender and sex and gender hypotheses, we need to address the possibility that the difference between Italian and English speakers we reported could have come about because speakers of Italian and English used different semantic strategies in carrying out their judgments and not because of grammatical gender. For example, Italian speakers might have used size to group animals, whereas English speakers could have used ferociousness. If these semantic dimensions were to be correlated to differences in Italian gender, this could introduce a confound. Although there was no suggestion that speakers of the two languages consistently used different semantic grouping on the basis of the postexperimental questionnaire, we further addressed this potential issue by carrying out multidimensional scaling analyses

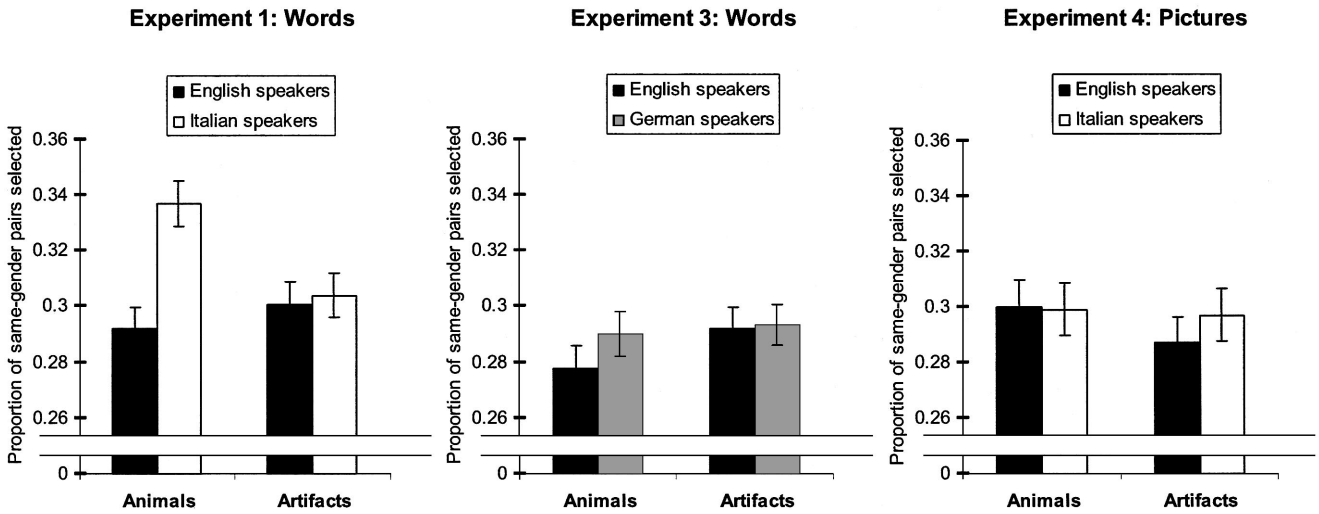

Figure 1. Average proportion of same-gender word pairs selected in triadic similarity judgments (Experiments 1, 3, and 4), considering only triads with two words of one gender and one word of the other gender. Error bars reflect standard error of the mean by participants. The left panel depicts judgments for English and Italian participants to word stimuli (Experiment 1), the center panel depicts judgments for English and German participants to word stimuli (Experiment 3), and the right panel depicts judgments for English and Italian participants to picture stimuli (Experiment 4).

to assess the semantic similarity among the items in each language and category. These analyses (reported in Appendix C) showed a high degree of consistency in the grouping properties used by speakers of the two languages, thus reassuring us that the difference we observed cannot be ascribed to this potential confound.

The results of Experiment 1 are compatible with the sex and gender hypothesis: The gender effect does not generalize beyond entities for which sex is a semantically relevant property, consistent with the notion that the effect is mediated by an association between gender of nouns and male- or female-like aspects of meaning. Had this effect been mediated by more general mechanisms according to which increased semantic similarity is observed for words that are similar on any linguistic dimension, as stated by the similarity and gender hypothesis, we would have expected to observe it both for animals and artifacts. The lack of effects for words referring to artifacts is also in conflict with a less constrained version of the sex and gender hypothesis, according to which grammatical gender influences not only semantic representations for entities for which sex is relevant (i.e., animals) but also semantic representations for which sex-related conceptual properties are not relevant.

In Experiment 2 we seek converging evidence, moving to a task in which meaning representations are accessed in an automatic, online manner: elicitation of semantic substitution errors.

Experiment 2: Elicitation of Semantic Substitution Errors in Italian

Semantic substitution errors (e.g., saying "tiger" when "leopard" is intended) are among the most common spontaneously occurring slips of the tongue (Dell, 1995). It is generally agreed that semantically related lexical retrieval errors reflect coactivation of semantically related lexical candidates during a conceptually driven retrieval process (e.g., Garrett, 1984, 1992; Levelt, Roelofs, & Meyer, 1999). It has also been shown that the likelihood of substitution reflects fine-grained semantic similarity between target and intruding words (Vigliocco, Vinson, Lewis, & Garrett, 2004). Here we use the continuous picture-naming paradigm we have introduced in previous work (Paganelli, Vigliocco, Vinson, Siri, & Cappa, 2003; Vigliocco, Vinson, Indefrey, et al., 2004) to elicit semantic substitution errors. We test whether Italian substitution errors are influenced by grammatical gender once other factors affecting the likelihood to produce a lexical error are taken into account. Predictions from the similarity and gender and sex and gender hypotheses are the same as in Experiment 1.

To take into account other uncontrolled factors that can affect substitution errors, we first compare the errors from Italian participants with those from English participants in the same task and for the same pictures. As in Experiment 1, the comparison with English data allows us to take into account influences of meaning similarity that are unrelated to Italian gender (e.g., speakers may be more likely to substitute *pecora* [sheep, fem.] with *capra* [goat, fem.] than with *asino* [donkey, masc.] because of greater semantic similarity between sheep and goats rather than because of anything to do with grammatical gender).

Moreover, previous work in the language production literature has established that 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). Phonological overlap between target and intruding words cannot be factored out just by comparing data from Italian speakers with data from English speakers. To take this factor into account, we carried out analyses excluding errors in which target and intruder are strongly phonologically similar. This is important because there are clear phonological correlates of grammatical gender in Italian. If we did not exclude mixed errors, we would not know whether any putative gender effect in Italian arises as a consequence of greater meaning similarity between words sharing gender or, rather, general production mechanisms applying to all languages.

Method

Participants. Participants were 28 native speakers of Italian (19 women, 9 men) and 22 native English speakers from the University College London participant pool (15 women, 7 men). All were paid £3 ($5.30) for participating. As in Experiment 1, the Italian participants were students visiting London without fluent knowledge of English, and none of the English speakers reported moderate or better competence in any Romance language.

Materials. The set of items included pictures corresponding to the animal and artifact words used in Experiment 1, plus additional items, listed in Appendix D (7 additional animal items, mainly small mammals, and 26 additional artifact items). Most pictures were selected from Snodgrass and Vanderwart (1980), with some additional exemplars created for the purpose in similar style. Each picture was presented in black on a white background and scaled to fit within a 240 × 240-pixel area.

Seventy-seven blocks of 10 pictures each were constructed by randomly selecting pictures, with several constraints. First, all pictures from a given block were from one category (animals or artifacts). Second, a picture appeared no more than once within a block. Third, a picture never appeared as the last item in one block and the first item in the next. Each picture appeared 10 times in the course of the experiment. Each participant received a different random order of blocks in the course of the experiment (27 blocks of animals and 50 blocks of artifacts[3]). Experimental lists were prepared and stimuli presented using the E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002).

Procedure. Participants were instructed that they were participating in a study of speech patterns under time pressure, and that they would be asked to name pictures aloud, using a single word, as they appeared on the screen. Instructions emphasized that speakers should attempt to keep up with the rate of presentation (i.e., naming pictures as they appeared rather than retaining them in memory). This was accomplished by skipping items if necessary to recover from difficulty. All participants gave consent to have their responses recorded; all spoken responses were tape-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 experimenters 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

[3] In previous studies using this paradigm (Paganelli et al., 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004), we have found that errors almost never cross semantic categories. Thus, we presented items referring to animals and to artifacts in different blocks.

nonetheless used by participants to refer to a given picture were not considered to be lexical errors.

After this, participants performed a set of 12 practice blocks (6 blocks of animals, 6 of artifacts). In each practice block, 10 pictures were presented sequentially and randomly in one of four locations on the screen, and the participant was 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 if this was necessary to accommodate each speaker's speech rate. Initially, each picture was displayed for 1,000 ms, and, for a given participant, all pictures were displayed at the same rate. This rate was altered by the experimenter (minimum step size = 100 ms) during the practice session to make the task difficult but manageable for each speaker. Final display times ranged from 500 ms to 1,000 ms in English (M = 710 ms) and from 700 ms to 1,100 ms in Italian (M = 863 ms). Display times differed across languages because the Italian words were longer than the English words (M = 2.7 syllables and 6.4 phonemes vs. 1.5 syllables and 4.4 phonemes, respectively); therefore, slightly longer presentation times were necessary for Italian participants to name the pictures at a suitable accuracy level.

Once the practice session was completed, the experimental blocks were presented. Participants pressed a key to begin each block. A fixation cross briefly appeared in the center of the screen to indicate that the block was beginning, and then the 10 items in the block appeared in sequence at randomly selected positions on the screen, with time parameters as determined in the practice session. After each block, participants were given the opportunity to take a break if necessary, and at the halfway point a short break was provided. Each picture was presented 10 times for naming in the course of the experiment.

Results

Participants' responses were transcribed and scored in the following categories:

Correct responses: Participants uttered the correct target word completely.

Different label: Participants used a different word than our intended target (e.g., "stag" for "deer"), but this different label was consistently used by that participant and did not refer to another item 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 or used the different label three or more times in the experiment itself without self-correcting, 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.

Lexical errors: Participants produced a word that differed from the target and did not qualify as a different label.

Omissions: Participants did not produce any response for a target picture.

Self-corrections: 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: Other responses not included in the previous scoring categories, such as dysfluencies, incomplete utterances, and inaudible responses. The breakdown of responses by language and category is reported in Table 2. Only responses scored as lexical errors were considered for further analyses.

Table 2
Frequencies (and Percentages) of Different Response Types in the Error Elicitation Experiment (Experiment 2) as a Function of Language and Category

Response type	Italian	English
Animals		
Acceptable		
Correct	5,825 (77.1%)	5,184 (87.3%)
Different label	824 (10.9%)	245 (4.1%)
Error		
Lexical	193 (2.5%)	121 (2.0%)
Omission	650 (8.6%)	238 (4.0%)
Self-correction	45 (0.6%)	73 (1.2%)
Miscellanea	23 (0.3%)	79 (1.3%)
Artifacts		
Acceptable		
Correct	11,882 (84.9%)	8,988 (81.7%)
Different label	378 (2.7%)	1,088 (9.9%)
Error		
Lexical	265 (1.9%)	233 (2.1%)
Omission	1,398 (10.0%)	605 (5.5%)
Self-correction	41 (0.3%)	44 (0.4%)
Miscellanea	36 (0.3%)	42 (0.4%)

As mentioned, effects of Italian gender need to be assessed unconfounded with phonological overlap. In Appendix E we report analyses demonstrating that, indeed, phonological overlap affects these semantic substitutions. Crucially, errors sharing gender with the target also tended to be more phonologically similar than those not sharing gender. Surprisingly, this greater phonological overlap was not simply related to obvious phonological correlates of Italian gender (see Appendix E). Phonological overlap renders words more prone to substitution, thus introducing a confound. To take this into account, we excluded from analyses all errors that exhibited the greatest phonological similarity to the targets (those sharing 33% or more of the target word's phonemes in either language[4]). After applying this criterion, there remained 93 English and 142 Italian errors for animals and 178 English and 181 Italian errors for artifacts, all of which can be considered semantic, rather than mixed (i.e., semantic and phonological), errors.

On this restricted set of errors, for each participant we calculated the proportion of errors for which the target and error words shared (Italian) gender separately for each category (Figure 2). Two (Italian) participants were excluded from analysis because they produced no errors in one of the categories (in both cases artifacts).

We performed 2 × 2 ANOVAs with participants and items as random factors, investigating the effects of language (English and Italian between subjects and between items) and category (animals, artifacts; within subjects, between items) were performed on the proportion of errors sharing gender. The main effect of language was significant: participants, $F(1, 46) = 19.06$, $p < .001$, $MSE = .0655$; items, $F(1, 122) = 7.75$, $p = .006$, $MSE = .152$. Errors made by Italian speakers were more likely to be of the same

[4] This value was determined by setting a threshold at the mean, plus 1 standard deviation, of the phonological similarity proportion of all targets and errors in a given category; comparable thresholds were obtained for animal and tool categories even though they were separately analyzed.

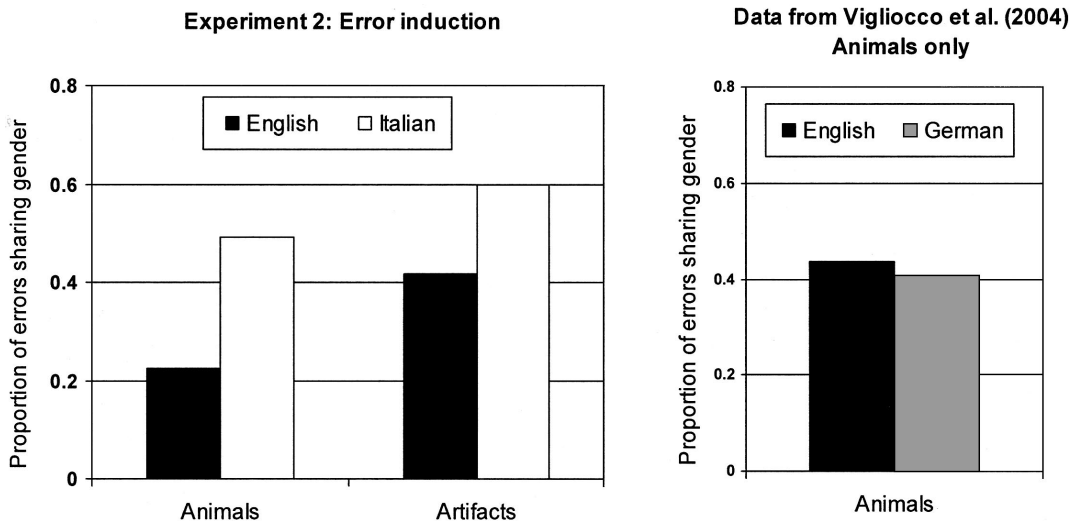

Figure 2. Proportion of Italian gender preservation between target and intruding words for animals and artifacts (excluding form-related errors). The English proportions are computed by assigning English words the gender of the translation-equivalent Italian words. Data in right panel is from Vigliocco, Vinson, Indefrey, Levelt, & Hellwig (2004).

gender as the target than those made by English speakers. The main effect of category was only significant by participants, $F(1, 46) = 19.90$, $p < .001$, $MSE = .0604$; for items, $F(1, 122) = 0.34$, $p = .561$, $MSE = .152$. Errors made for artifacts were more likely to share (Italian) gender with the target than errors made for animals. Crucially, the main effects were qualified by a Language × Category interaction, which was significant by participants, $F(1, 46) = 6.16$, $p = .017$, $MSE = .0604$, but not significant by items, $F(1, 122) = 3.24$, $p = .074$, $MSE = .152$. The difference between Italian and English was greater for the category of animals than it was for artifacts. We assessed the language difference on the proportions of errors sharing gender for each category using planned contrasts on the simple main effects of language. The difference between Italian and English participants was significant for animals: participants, $t(46) = 4.80$, $p < .001$; items, $t(43) = 3.05$, $p = .004$. The difference did not reach significance for artifacts: participants, $t(46) = 1.45$, $p = .077$ one-tailed; items, $t(79) = 0.80$, $p = .428$.

Discussion

By comparing the target–intruder pairs produced by Italian speakers, which could be affected by grammatical gender, with those produced by English speakers, which cannot be affected by Italian grammatical gender, and by limiting our analysis to cases in which phonological similarity was low, we found language-specific effects of Italian gender for animals, convergent with the results of Experiment 1. Together, these results suggest that, for animals, Italian speakers tend to attribute male- or female-like properties to the referents following the grammatical gender of the nouns, consistent with the sex and gender hypothesis. For artifacts, however, the tendency for the intruder to preserve the gender of the target word did not reach significance, in continuity with the results from the similarity judgment task, in which we observed no language-specific effect of Italian gender for the artifacts.

As we already described, another test of the two hypotheses involves cross-linguistic comparisons. The sex and gender hypothesis predicts that effects of gender should be observed (or at least should be strongest) only for languages with two gender classes and a core set of words (nouns referring to humans) for which there is a highly consistent and transparent mapping between sex of referents and gender of nouns. The similarity and gender hypothesis, in contrast, predicts that gender effects should be observed regardless of the number of genders in the language and of the transparency of the mapping between sex of referents and gender of nouns as long as the linguistic context provides sufficient cues to distinguish words of different genders.

Experiment 3: Triadic Similarity Judgments With Words in German

If the strength of associations between the gender of words and male- and female-like attributes of referents is mediated by the consistency in a language between gender of nouns referring to humans and sex of the referents, German provides us with an important test case. In contrast to Romance languages, in which there is a clear and consistent mapping between sex of referents and the genders assigned to nouns referring to humans, the three-gender system of German renders this mapping less consistent, especially because all nouns in the diminutive form are marked as neuter regardless of whether the referent is male or female.

Experiment 3 involved the triadic judgment task in German with words referring to animals and artifacts. As in Experiment 1, English judgments on translation equivalent words are used as a baseline with which to compare results for the German speakers.

Method

Participants. Participants were 28 native German speakers recruited from German cultural clubs in London (16 women, 12 men) and 28 native English speakers (14 women, 14 men) from the University College London participant pool. All were paid £3 ($5.30) for participating. None of the German speakers described themselves as fluent in English; none of the English speakers reported any competence in German or any other language with formal gender systems.

Materials. Words referring to 20 land animals and 22 artifacts were selected in a manner parallel to Experiment 1. Because the distribution of the three genders in these semantic categories can substantially differ, it was not possible to exactly match the number of exemplars from each gender class. For animals we included exemplars from all three genders (5 feminine, 7 masculine, 8 neuter); however, because there are very few artifacts with neuter gender we were forced to use only masculine (11) and feminine (11) items (see Appendix F for a list of items). Triads were created as in Experiment 1: 1,140 animal triads, divided into three lists of 380 triads each, and 1,540 artifact triads, divided into four lists of 385 triads each.

Procedure. As in Experiment 1, participants were told that the experiment concerned judgments of meaning similarity among triplets of words, and that their task was to choose the two words of the three that were most similar in meaning and to delete the odd one.

Results

No German participants indicated that they used grammatical gender in their similarity judgments in the postexperimental questionnaire. As in Experiment 1, the dependent variable was the proportion of same-gender word pairs selected as similar, considering only those triads that contained two words of one (German) gender and only one word with a different gender. Effects of category (animals, artifacts) and language (English, German) were investigated using between-subjects ANOVA; average percentages of same-gender word pairs are presented in Figure 1 and Table 3.

A 2 × 2 ANOVA yielded no main effect of language, $F(1, 52) = 0.45$, $p = .507$, $MSE = .00123$; there was no overall language-specific difference related to German gender. There was no significant effect of category, $F(1, 52) = 0.72$, $p = .715$, $MSE = .00123$, and the interaction was also not significant, $F(1, 52) = 0.36$, $p = .552$; $MSE = .00123$; triadic similarity judgments by German speakers did not appear to be affected by gender in either category. Nonetheless, there remained a high level of similarity between the performance of English and German speakers; when considering the frequency of selection of all possible word pairs from all triads, responses to both categories were extremely highly correlated across language (animals $r = .951$; tools $r = .926$). This fact provides us with an important assurance that participants in both languages were using similar semantic criteria to make their judgments, a crucial point in this instance in which no significant language-specific gender effects were observed.

Table 3
Average Percentage of Same-Gender (German) Word Pairs Selected in Triadic Similarity Judgments, by Language and Category[a]

	Category	
Language	Artifacts	Animals
German		
M	29.3	29.0
SEM	0.9	1.0
English		
M	29.2	27.8
SEM	0.9	1.0

[a] Considering only triads with two words of one gender and one word of a different gender.

Discussion

In contrast to Italian, we observed no effect of gender in German, a pattern of results predicted by the sex and gender hypothesis but not by the similarity and gender hypothesis. Because of the lack of any gender effect in this experiment, we decided it was not necessary to conduct the corresponding error induction experiment in German to obtain converging evidence. Nonetheless, we carried out a reanalysis of a set of semantically related substitution errors by German and English speakers previously reported in Vigliocco, Vinson, Indefrey, et al. (2004).[5] We considered only errors for animals (errors for body parts were also reported), and, as in Experiment 2, we excluded all those instances in which the error was phonologically similar to the target (sharing 33% or more of the target's phonemes in the same positions) before comparing responses in the two languages. In the remaining set, no difference was observed between the gender effects in German and English; 39 of 96 (40.6%) errors by German speakers were of the same gender, and 41 of 94 (43.6%) errors by English speakers were of the same (German) gender. This is in contrast to the comparison between Italian and English reported in Experiment 2 and suggests that language-specific gender effects for words from the same semantic category may vary as a consequence of structural differences between languages, as predicted by the sex and gender hypothesis.[6]

Experiments 1 to 3 lend support to the constrained version of the sex and gender hypothesis, according to which effects of gender can emerge only if the language affords a high degree of correspondence between gender of nouns and sex of human referents, and such effects only arise for referents for which sex is relevant (but not transparently associated to the gender of the nouns). In a final experiment, we assess whether the effects are further constrained by whether lexical information needs to be retrieved in the task.

[5] This reanalysis considered the German errors from Experiment 1 of Vigliocco, Vinson, Indefrey, et al. (2004) and the English errors reported in the Discussion of the same source.

[6] A possible alternative explanation of the difference between Italian and German (Experiment 1 vs. 3) may lie in the participants' extent of experience with English. Namely, our German speakers might have been somewhat more proficient in English than our Italian speakers. Because we did not assess participants' competence in English but relied only on their self-reports, we cannot completely rule out this possibility.

Experiment 4: Triadic Similarity Judgments in Italian With Pictures

The tasks used in Experiments 1 to 3 recruited both conceptual and linguistic information, because they used words as stimuli or required verbalization of pictures. In a number of theories, lexicosemantic representations (roughly corresponding to the meanings of words and thus being linguistic in nature) are clearly distinguished from conceptual representations (roughly corresponding to mental representations for things, events, and so on, which subserve language and other cognitive functions and thus being nonlinguistic in nature; e.g., Garrett, 1984; Jackendoff, 2002; Levelt et al., 1999; Levinson, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). To this point, the reported experiments have assessed whether, and to what extent, the mental representations consulted for the purpose of using language (thinking for speaking; Slobin, 1991, 1996) differ among Italian, German, and English speakers for words with grammatical gender.

The next step is to assess whether the observed effects of gender are strictly dependent on the engagement of lexical knowledge or whether they also arise when the task only requires engagement of conceptual knowledge, a finding that could provide support for a linguistic relativity hypothesis (Whorf, 1956).

In Experiment 4 we replicate Experiment 1 (contrasting Italian and English speakers' similarity judgments for animals and artifacts) using pictures, instead of words, as materials. The use of pictures does not exclude per se the possibility that speakers label the items; however, whereas the task in Experiment 1 cannot be completed without involving lexical knowledge (because only words were presented to participants), the pictorial task does not require verbalization. The sex and gender hypothesis (as well as the similarity and gender hypothesis) is silent with respect to whether effects should be observed with pictures. Establishing whether the effects are limited to the verbal domain, however, provides further constraints on how these effects can arise.

Method

Participants. Participants were 36 native Italian speakers (22 women, 14 men) and 36 native English speakers from the University College London participant pool (24 women, 12 men) who did not participate in Experiments 1 or 2. As in Experiments 1 and 2, the Italian speakers were Italian students visiting London without fluent knowledge of English, and none of the English speakers reported moderate or better competence in any Romance language. All were paid £3 ($5.30) for participating.

Materials. Pictures corresponding to the 20 words referring to animals and the 24 words referring to artifacts in Experiment 1 were used (all of these pictures were also used in Experiment 2). Using the lists from Experiment 1, triads for both language conditions were created by replacing each word with its corresponding picture. Thus, all details of list composition are exactly the same as in Experiment 1, except for differences related to picture presentation. Three or four triads were presented per size A4 page: Each picture was printed in a square space with a maximum dimension of 4 × 4 cm. The three pictures in each triad were centered on a line and were separated by three dashes (approximately 1 cm between pictures); triads were printed on separate lines with at least 4.5 cm of blank space between them.

Procedure. The procedure was the same as in Experiment 1, except that the instructions were changed to refer to pictures instead of words. Participants were explicitly instructed to make their judgments on the basis of meaning similarity between the concepts expressed by the pictures and not their visual similarity.

Results

No Italian participants indicated that they used grammatical gender in their similarity judgments in the postexperimental questionnaire. As in Experiment 1, the dependent variable was the proportion of same-gender picture pairs selected as similar, considering only those triads that offered the opportunity for selecting different-gender pairs (i.e., those triads containing two pictures whose label had a masculine [Italian] gender and one whose label had a feminine gender, and those triads containing two feminine pictures and one masculine picture). Effects of category (animals, artifacts) and language (English, Italian) were assessed using between-subjects ANOVA (see Figure 1 and Table 4 for a summary).

The 2 × 2 ANOVA yielded no main effect of language, $F(1, 68) = 0.58$, $p = .447$, $MSE = .00068$; there were no language differences in the effects of gender overall. There was also no effect of category, $F(1, 68) = 1.44$, $p = .234$, $MSE = .00068$, and no significant interaction, $F(1, 68) = 0.82$, $p = .367$, $MSE = .00068$. Nonetheless, similarity judgments were highly consistent in both languages; considering the frequency of selection of all possible word pairings from all triads (for animals, cross-linguistic $r = .911$; for tools, $r = .915$).

In contrast to Experiment 1, differences were not observed between Italian speakers' similarity judgments for pictures and English speakers' judgments for the same pictures, suggesting that language-specific effects may be strictly limited to verbal tasks. We tested this difference between Experiments 1 and 4 directly using a three-way ANOVA to assess the effects of Modality of Presentation (word, picture) × Language (Italian, English) × Category (animals, artifacts). The main effect of modality was significant, $F(1, 136) = 5.26$, $p = .023$, $MSE = .00098$, as was the main effect of language, $F(1, 136) = 6.70$, $p = .011$, $MSE = .00098$. The main effect of category was not significant, $F(1, 136) = 3.34$, $p = .070$, $MSE = .00098$. None of the two-way interactions was significant: Category × Modality, $F < 1$; Language × Category, $F(1, 136) = 1.94$, $p = .166$; Language ×

Table 4
Average Percentage of Same-Gender (Italian) Picture Pairs Selected in Triadic Similarity Judgments by Language and Category[a]

	Category	
Language	Animals	Artifacts
Italian		
M	29.9	29.7
SEM	0.9	0.6
English		
M	30.0	28.7
SEM	0.9	0.6

[a] Considering only triads with two words of one gender and one word of a different gender

Modality, $F(1, 136) = 2.87$, $p = .093$; all *MSE*s = .00098. Crucially, all these main effects and interactions were qualified by a significant three-way interaction, $F(1, 136) = 6.03$, $p = .015$, $MSE = .00098$: Language-specific effects of gender were observed but only for words referring to animals. No language-specific effects were found for pictures referring to animals or for artifacts in either modality.

Discussion

No effect of Italian grammatical gender was observed in picture judgments for either animals or artifacts, in contrast to Experiment 1 in which we observed a language-specific effect of Italian gender for animals. This finding suggests that grammatical gender effects in Italian are tightly constrained: limited to those semantic categories in which a generalization from conceptual gender is most intuitively plausible (Experiments 1 and 2) but also limited to verbal tasks and not generalizing to a task that does not require verbal mediation. Thus, these effects are strictly thinking for speaking effects (Slobin, 1991, 1996) and suggest that the mechanisms assumed by the sex and gender hypothesis apply to language development but do not extend to conceptual structures. Results from Experiment 4 also pose important constraints to models of lexical representation, which we consider in the General Discussion.

General Discussion

The four experiments we reported were designed to address the strength and pervasiveness of language-specific grammatical gender effects on semantic representations for the corresponding objects. We found language-specific effects that are highly constrained: limited to a language with a two-gender system (Italian). Limited to tasks that require verbalization and in Italian, they were observed in certain semantic categories (animals) and not others (artifacts).

Before discussing the implications of these results for learning and processing mechanisms, let us consider the role of form (phonology–orthography) overlap between words sharing the same gender as a potential confounding factor in our Italian data. There are strong correlations between gender and form regardless of whether the nouns have conceptual or grammatical gender (see Appendix A). In particular, the word ending "-a" marks a majority of feminine nouns, and the word ending "-o" marks a majority of masculine nouns. These same endings also mark other words in sentences that agree in gender with the noun (e.g., adjectives modifying the nouns). These form correlates might have been used by Italian speakers in our verbal tasks. Although this is possible, form correlates of gender cannot be the sole factor at work in our study for two reasons. First, form correlates are present for both animals and artifacts; nonetheless, in Experiment 1 we only observed an effect of gender for animals; the same is true for Experiment 2, in which phonologically related errors were excluded using a strict criterion.

It is, however, possible (if not likely) that form correlates of gender (within the word or in phrasal contexts) play some role in establishing or strengthening the association between grammatical gender of the words and conceptual features (such as male or female properties) during language development (as also argued by Sera et al., 2002). In contrast to Italian, there are fewer form correlates of gender in German (see Appendix A). This lack of form correlates, combined with the less transparent link between gender of nouns and sex of human referents, may explain the different results in the two languages. To assess the role of form correlates, future research could extend the current studies to languages that have two-gender systems (like Italian) but that do not, however, have such clear form correlates.

Regardless of the role played by form during language learning, our results impose constraints on the type of mechanisms that operate during language learning and from which these effects arise as a by-product. Moreover, our results have novel implications for models of the lexical system.

Sex, Similarity, and Grammatical Gender: Constraints on Learning Mechanisms

Early in this article, we presented two hypotheses regarding the manner in which the gender of nouns may cause those entities that share this linguistic property to be more semantically similar. According to the sex and gender hypothesis, the effect would be strictly mediated by and dependent on establishing associations between genders of nouns and male- or female-like properties of referents. According to the similarity and gender hypothesis, the effect would instead come about as a by-product of inferring meaning similarity from use in the same linguistic context. In the sex and gender hypothesis, the association between gender of the nouns and male- or female-like properties requires that children notice the relation between nouns referring to humans and sex of referents (an association that is present to varying degrees across languages; Corbett, 1991), which they can then generalize to other entities. In contrast, the similarity and gender hypothesis requires no association between grammatical and biological gender.

The two hypotheses make different predictions regarding when language-specific effects of grammatical gender should be observed, depending on whether establishing an association between gender of nouns referring to humans and sex of referents is a necessary prerequisite for the effects to arise. The sex and gender hypothesis predicts that the strength (or presence) of gender effects will differ across semantic categories within a language and will differ across languages. In particular, effects of grammatical gender on meaning similarity should be greater within a language such as Italian for sexuated entities than for other types of entities. Moreover, effects of grammatical gender should be greater for languages, such as Italian and (presumably) other Romance languages, with only two genders (for which discovery of the association between sex and gender should be easiest); the greatest association should be between the gender of nouns referring to humans and the sex of referents (aiding in discovery but also strengthening the association and rendering it more generalizable) than for languages (such as German) with more than two genders and with less association between gender and sex of referents. Some previous categorization studies provide evidence compatible with the prediction that the effect of gender may be found in languages with only two genders (Arabic: Clarke, Losoff, McCracken, & Rood, 1984; Clarke, Losoff, McCracken, & Still, 1981; Italian: Ervin, 1962; French and Spanish: Sera et al., 2002) and not in languages with more than two genders (German: Sera et al., 2002).

We also noted that all previous studies that have reported effects of grammatical gender on meaning similarity have assumed (implicitly or explicitly) one version of the sex and gender hypothesis that predicts general and highly pervasive effects of gender, in contrast to the more constrained version we have discussed here. In this less constrained version, the grammatical gender of nouns would enhance the salience of male- or female-like properties of sexuated referents and would allow for the development of male- or female-like properties for referents for which sex is irrelevant (e.g., artifacts). The link between these previous studies and this less constrained version of the sex and gender hypothesis is directly reflected in the choice of a categorization task in which participants are asked to categorize words, pictures, or nonwords according to sex in some studies (e.g., Clarke, et al., 1981, 1984; Ervin, 1962; Sera et al., 1994, 2002). In other cases, this fact is reflected in what are considered to be the semantic repercussions of gender. For example, in the study by Konishi (1993), effects of grammatical gender were predicted to affect those dimensions of the Semantic Differential Scale that are typically associated with male and female entities (e.g., "power" as a male dimension, "nurture" as a female dimension). As we have discussed earlier, however, these studies suffered from a number of methodological problems related to the explicit mention of gender in the task (e.g., Sera et al., 1994, 2002) or the unclear pattern of results (Konishi, 1993). Our studies, which used tasks less susceptible to the use of strategies, favor our more constrained version of the sex and gender hypothesis, in which the semantic effects of grammatical gender do not extend beyond semantic domains for which properties related to sex are relevant (animals).

Regarding the similarity and gender hypothesis, even though such a mechanism does not appear to support the association between grammatical gender and meaning, we certainly allow that it may provide a viable account of language learning in other domains (e.g., in extracting properties of events from syntactic properties of verbs; Fisher, 1994). A cross-linguistic approach provides a novel manner to investigate which mechanisms are at play across domains. It remains an important question for future studies to establish when such a similarity-based approach may be the main force driving the learning process and when, instead, it may not be (as in the case of grammatical gender).

Sex and Gender: Implications for Models of the Lexical System

Can we integrate the sex and gender hypothesis into models of lexical representation and use? There are three novel constraints to such models arising from our work. First, the system must allow for language-specific effects to arise at a lexicosemantic level but not at a conceptual level of representation (to account for the difference between Experiments 1 and 4). That is, conceptual knowledge may not be affected by linguistic knowledge. Second, during development, the system needs to be sensitive to the degree of correlation between conceptual and linguistic features. Differences in the degree of such correlation give rise to the different results we obtained in Italian and German. Finally, during learning, the system must be able to make (limited) generalizations from the correlation between conceptual (sex) and linguistic (gender) properties, extending to cases for which no such transparent relationship is present (animals in Italian).

As mentioned, a distinction between lexicosemantic representations and conceptual representations that subserve linguistic and nonlinguistic cognition is a design feature of many theories of lexical representation (e.g., Garrett, 1984; Jackendoff, 2002; Levelt et al., 1999; Levinson, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). Distinguishing between these two levels allows us to account for the difference between linguistic and nonlinguistic tasks in terms of the level of representation that is consulted in a given task: Lexicosemantic representations must be consulted during linguistic tasks, but only conceptual representations need to be consulted during nonlinguistic tasks. In a related line of investigation, Bowers, Vigliocco, Stadthagen-Gonzalez, and Vinson (1999) asked Spanish speakers to make semantic and gender categorizations for a set of pictures and their corresponding names. They found similar latencies for semantic categorization of pictures and words. However, participants were faster to make gender decisions to words compared with pictures. It was suggested that the difference emerges because picture recognition requires conceptual mediation (accessing conceptual representations before lexical representations), whereas word recognition processes permit direct access to lexical representations. These findings converge with our results in showing that effects of grammatical gender are mediated by linguistic representations.

Assuming a distinction between lexicosemantic and conceptual representations, the developmental mechanisms of the sex and gender hypothesis must also be implemented. The featural and unitary semantic space (FUSS) hypothesis put forward by Vigliocco, Vinson, Lewis, and Garrett (2004) may provide a starting point. In FUSS, conceptual representations are conceived of as distributed featural representations. The motivation for assuming distributed conceptual representations comes primarily from considerations related to neurological plausibility of the system (see Vigliocco, Vinson, Lewis, & Garrett, 2004; Vinson, Vigliocco, Cappa, & Siri, 2003). Lexicosemantic representations are conceived as intermediate representations that bind conceptual information to other types of linguistic information. It is argued that this lexicosemantic level develops by taking properties of the conceptual featural representations into account. Specifically, in the current implementation of FUSS, we simulated the binding of conceptual features into lexicosemantic representations by using self-organizing maps (Kohonen, 1997) to reduce the dimensionality of a similarity space based on speaker-generated semantic features (see Vinson & Vigliocco, 2002, for details). The resulting lexicosemantic space developed an organization solely on the basis of the properties of the input (i.e., aspects of similarity among semantic features for different concepts, including shared features, distinctive features, and properties of correlation among features) but not on other linguistic information such as lexicosyntactic and phonological properties. However, the sex and gender hypothesis suggests that such other linguistic information must also play a role in shaping representations at this level: The lexicosemantic level develops under the joint influence of conceptual and linguistic properties.

Within the framework of FUSS, this implies that the system needs input not only from conceptual properties but also from lexicosyntactic properties such as grammatical gender. The greater the transpar-

ency and consistency of the mapping between lexicosyntactic properties (like gender) and conceptual properties (like sex), the more likely it is that these lexicosyntactic properties will be (implicitly) mapped onto similar lexicosemantic representations. This would occur as a consequence of reducing the dimensionality of a conceptual feature space, under which the system is sensitive to correlations among dimensions of input (in this case, correlations between properties related to sex and grammatical gender). In this manner, language-specific effects could reflect the manner by which the lexicosemantic space develops an organization as a function of conceptual properties and linguistic properties correlated to conceptual properties (see also Vigliocco & Kita, 2005).

Grammatical Gender and Linguistic Relativity

The presence of an effect only for linguistic tasks is in line with the view that language affects only thinking for speaking (Slobin, 1991, 1996) or, more generally, thinking for using language, namely tasks that engage linguistic coding (whether overtly or covertly). Thinking for using language must differ across languages when we consider aspects such as conceptual gender: Italian speakers must pay more attention to the sex of a friend, professor, child, and so forth to produce the correct words in sentences, in contrast to English speakers for whom conceptual gender is less obligatorily marked. Here we showed that differences in thinking for using language extend beyond conceptual gender for nouns referring to humans to other semantic categories such as animals in languages such as Italian. Moving from gender to other language-specific properties, other studies have found language-specific effects limited to linguistic tasks (or tasks in which verbal encoding could have been plausibly engaged; Brysbaert et al., 1998; Finkbeiner, Nicol, Greth, & Nakamura, 2002; Gennari et al., 2002; Malt, Sloman, Gennari, Shi, & Wang, 1999). For example, Brysbaert et al. (1998) have shown that language differences in the ordering of number words affect manipulation of number words but not number symbols. Finkbeiner et al. (2002) and Gennari et al. (2002) showed that language differences in expression of path and manner in the verb stem (see Slobin, 1996, for a discussion) affect tasks that require linguistic encoding but not tasks that do not require linguistic encoding.

Does this mean that we should abandon our quest for language effects on nonlinguistic aspects of cognition (Sapir, 1921; Whorf, 1956) altogether? We believe not. Grammatical gender of nouns is but one of many language-specific properties, and there is no strong a priori reason to believe that findings concerning one type of property will necessarily generalize to other properties.

For example, there are some domains in which researchers have provided some (more or less controversial) evidence of language influence beyond thinking for using language, including color categorization (Davidoff et al., 1999; Roberson, 2005; Roberson, Davies, & Davidoff, 2000, 2002), space representation (Levinson, 1996, 1997; but see Li & Gleitman, 2002), and conception of time (Boroditsky, 2001).

There are many important differences between domains such as color categorization, spatial language, metaphors used to describe time, and the grammatical gender of nouns that may determine how pervasive any effect of language might be. First, as discussed by Boroditsky (2001), language may be a more powerful tool (or teacher) in abstract domains of knowledge compared with concrete domains. This is because abstract concepts are learned predominantly by means of language, whereas concrete concepts may be learned at first predominantly by means of our direct experience in the world. Second, language may have a more pervasive effect for perceptually and conceptually continuous domains (e.g., color, time, and space) than for discrete domains (e.g., objects or sex). It is only by investigating a broad range of languages and cross-linguistic differences that the role of these more general factors can be better understood.

References

Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua, 92,* 297–329.

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

Boroditsky, L., & Schmidt, L. (2000, August). *Sex, syntax, and semantics.* Paper presented at the 22nd Annual Conference of the Cognitive Science Society, Philadelphia, PA.

Bowers, J. S., Vigliocco, G., Stadthagen-Gonzalez, H., & Vinson, D. P. (1999). Distinguishing language from thought: Experimental evidence that syntax is lexically rather than conceptually represented. *Psychological Science, 10,* 310–315.

Brysbaert, M., Fias, W., & Noel, 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.

Clarke, M. A., Losoff, A., McCracken, M. D., & Rood, D. S. (1984). Linguistic relativity and sex/gender studies: Epistemological and methodological considerations. *Language Learning, 34,* 47–67.

Clarke, M. A., Losoff, A., McCracken, M. D., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning, 31,* 159–169.

Corbett, G. (1991). *Gender.* Cambridge, UK: Cambridge University Press.

Dardano, M., & Trifone, P. (1985). *La lingua Italiana, una grammatica completa e rigorosa* [*The Italian language: A complete and rigorous grammar*]. Bologna, Italy: Zanichelli.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398,* 203–204.

Dell, G. S. (1995). Speaking and misspeaking. In L. R. Gleitman & M. Liberman (Eds.), *Language: An invitation to cognitive science* (2nd ed., Vol. 1, pp. 183–208). Cambridge, MA: MIT Press.

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.

Ervin, S. M. (1962). 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.

Fisher, C. (1994). Structure and meaning in the verb lexicon: Input for a syntax-aided verb learning procedure. *Language and Cognitive Processes, 9,* 473–517.

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology, 23,* 331–392.

Fisher, C., & Gleitman, L. R. (2002). Language acquisition. In H. F. Pashler (Series Ed.) and C. R. Gallistel (Volume Ed.), *Steven's handbook of experimental psychology, Vol. 1: Learning and motivation* (pp. 445–496). New York: Wiley

Foundalis, H. E. (2002, August). *Evolution of gender in Indo-European languages.* Paper presented at the 24th Annual Conference of the Cognitive Science Society, Fairfax, VA.

Garrard, P., Carroll, E., Vinson, D. P., & Vigliocco, G. (2004). Dissociation of lexical syntax and semantics: Evidence from focal cortical degeneration. *Neurocase, 10,* 353–362.

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

Garrett, M. F. (1992). Lexical retrieval processes: Semantic field effects. In E. Kittay & A. Lehrer (Eds.), *Frames, fields and contrasts: New essays in semantic and lexical organization* (pp. 377–395). Hillsdale, NJ: Erlbaum.

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

Grimes, B. F. (2004). *Ethnologue: Languages of the world* (14th ed.). Dallas, TX: SIL International.

Grossman, M., Mickanin, J., Onishi, K., & Hughes, E. (1996). Verb comprehension deficits in probable Alzheimer's disease. *Brain and Language, 53,* 369–389.

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

Jakobson, R. (1959). On linguistic aspects of translation. In R. A. Brower (Ed.), *On translation* (pp. 232–239). Cambridge, MA: Harvard University Press.

Kohonen, T. (1997). *Self-organizing maps.* New York: Springer.

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

Köpcke, K.-M., & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache [The cognitive organization of gender assignment for monosyllabic words in the modern German language]. *Zeitschrift für germanistische Linguistik, 11,* 166–182.

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104,* 211–240.

Lepschy, G., & Lepschy, A. L. (1981). *La lingua Italiana* [*The Italian language*]. Milan: Bompiani.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

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

Levinson, S. C. (1997). From outer to inner space: Linguistic categories and non-linguistic thinking. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 13–45). Cambridge, UK: Cambridge University Press

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

Li, P., & Gleitman, L. R. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83,* 265–294.

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

Malt, B. C. Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40,* 230–262.

Martin, N., Gagnon, D. A., Schwartz, M. F., 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.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German.* London: Springer-Verlag.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University Illinois Press.

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

Roberson, D. (2005). Color categories are culturally diverse in cognition as well as in language. *Cross-Cultural Research, 39,* 56–71.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a Stone-Age culture. *Journal of Experimental Psychology: General, 129,* 369–398.

Roberson, D., Davies, I., & Davidoff, J. (2002). Color categories are not universal: Replications and new evidence. In B. Saunders & J. van Brakel (Eds.), *Theories, technologies, instrumentalities of color: Anthropological and historical perspectives* (pp. 25–36). Lanham, MD: University Press of America.

Romney, A. K., Moore, C. C., & Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Proceedings of the National Academy of Sciences, USA, 94,* 5489–5494.

Sapir, E. (1921). *Language.* New York: Harcourt, Brace, and World.

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

Sera, M., 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., 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. I. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 128–172). Cambridge, England: Cambridge University Press.

Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics, 1,* 7–26.

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

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

Twain, M. (1880). The awful German language (Appendix D). In M. Twain (Ed.), *A tramp abroad* (pp. 390–402). Hartford, CT: American Publishing.

Vigliocco, G., & Kita, S. (2005). Language-specific effects of meaning, sound and syntax: Implications for models of lexical retrieval in production. *Language and Cognitive Processes.*

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

Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: Semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics, 15,* 317–351.

Vinson, D. P., Vigliocco, G., Cappa, S. F., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language, 86,* 347–442.

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

Appendix A

Conceptual and Grammatical Gender in Italian and German

Languages differ in whether they include gender distinctions for words referring to entities without biological sex. Corbett (1991) distinguishes between languages that have semantic gender systems and those with formal gender systems. In the first category are languages such as English and Mandarin, in which gender is encoded in linguistic elements only for referents having biological sex. For example, in English, gender distinctions are lexically encoded for some nouns (e.g., actor–actress; uncle–aunt) and are maintained in the pronominal system as a distinction between clearly male and female entities (e.g., he–she) and everything else (it). For languages with formal gender systems, however, gender distinctions apply to all types of nouns regardless of whether their referents have a biological sex or not. Although gender assignment seems to be largely arbitrary in such instances, there is some degree of convergence, particularly among closely related languages (see Boroditsky & Schmidt, 2000; Foundalis, 2002).

Italian and German are among the languages with formal gender systems. In Italian (like other Romance languages such as Spanish and French), all nouns fall into one of two gender classes (masculine or feminine). For nouns referring to humans, there is nearly always a transparent relation between the gender of the noun and the sex of the referent (we refer to this relation as "conceptual gender"). This transparent relation is realized not only through the use of different words (e.g., *uomo–donna* [man–woman]) but also by the use of derivational and inflectional affixes that turn masculine nouns into feminine (e.g., *attore–attrice* [actor–actress]; *bambino–bambina* [male child–female child]). All other nouns (e.g., animals, artifacts, substances, abstract entities, nouns depicting actions and events) are marked for gender as well, although in most of these cases (to which we refer as "grammatical gender") gender bears little if any conceptual force.

There are important form correlates of gender marking for Italian nouns.[A1] For approximately 80% of all nouns (whether marked for conceptual or grammatical gender), the word ending "-o" marks masculine nouns, and the word ending "-a" marks feminine nouns. These same inflections are used to mark other words in sentences to agree with the noun (e.g., adjectives, different types of pronouns, and past participles).

German, alternatively, has three genders (masculine, feminine, and neuter) and a less transparent mapping between sex of referents and the genders assigned to nouns referring to humans. For example, all nouns in the diminutive form take neuter gender regardless of whether the referent is male or female. This fact was captured by Mark Twain (1880), who, in *A Tramp Abroad,* noted that "in German a young lady has no sex, while a turnip has. Think what overwrought reverence that shows for the turnip, and what callous disrespect for the girl" (Twain, 1880, Appendix D, "The Awful German Language").

Nouns in German are also marked for case; some combinations of case and gender also result in less transparent gender marking. In the nominative case (for singular nouns), gender is marked in phrases with different definite determiners (*der* = masculine; *die* = feminine; *das* = neuter), whereas only feminine nouns are distinguished in indefinite phrases (*ein* = masculine–neuter, *eine* = feminine). In the accusative case, all genders are marked differently (whether definite or indefinite), and in the dative case only feminine nouns take different determiners (again, whether definite or indefinite). Although there are few transparent morphophonological correlates of gender in the form of nouns (at least compared with the extent to which gender is marked on the noun in Italian), many fine-grained correspondences between gender and form have been described (Köpcke & Zubin, 1983).

[A1] Here, we prefer to use "form" rather than "phonology" to encompass both phonological and orthographic information because both of them provide cues to gender in Italian (because of its shallow orthography).

(*Appendixes continue*)

Appendix B

Experimental Items Used in Experiment 1 (as Words) and 4 (as Pictures)

Italian	Gender	English
	Animals	
asino	M	donkey
cammello	M	camel
cane	M	dog
capra	F	goat
cavallo	M	horse
cervo	M	deer
elefante	M	elephant
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
scimmia	F	monkey
tigre	F	tiger
volpe	F	fox
zebra	F	zebra
	Artifacts	
ago	M	needle
bottiglia	F	bottle
cacciavite	M	screwdriver
chiave	F	key
coltello	M	knife
cucchiaio	M	spoon
forbici	F	scissors
forchetta	F	fork
martello	M	hammer
matita	F	pencil
pala	F	shovel
penna	F	pen
pennello	M	paintbrush
pettine	M	comb
pinze	F	pliers
rasoio	M	razor
rastrello	M	rake
scala	F	ladder
scopa	F	broom
secchio	M	bucket
sega	F	saw
spremiagrumi	M	juicer
trapano	M	drill
vite	F	screw

Note. M = masculine; F = feminine.

Appendix C

Fine-Grained Similarity Patterns in the Similarity Ratings in Experiment 1

We conducted multidimensional scaling analyses to investigate the properties of similarity among items in the different languages (Italian, English) and categories (animals, artifacts). For each combination of language–category, we calculated pairwise similarity ratios for each possible pairing of words in the set by counting the number of times each given word pair was selected (across all triads), divided by the total opportunity of selection for that word pair (i.e., the number of times those two words appeared in a given triad). These pairwise similarity values were highly correlated across languages for both semantic categories; for animal word pairs, $r = .835$; for artifact word pairs, $r = .882$.

Distance values were then calculated by subtracting each similarity ratio from 1; low distance corresponds to the greatest similarity. For example, in the animal triadic task, "lion" and "tiger" were chosen together by English speakers 63 times of the 72 triads in which those two words occurred, yielding a similarity ratio of 0.88 for this word pair and a distance of 0.12. We fit the resulting distances into a two-dimensional representation space using ALSCAL, with a nonmetric, Euclidean distance model, assuming interval data. This space illustrates the overall patterns of similarity among the items in both semantic categories and across languages (Figure C1).

Speakers of both languages converged in distinguishing animals into four general groupings, roughly corresponding to domestic animals, exotic animals, big cats, and smaller canines–felines; overall commonalities were also observed for artifacts, with the main distinction being between hand tools versus other types of implements. The effects of gender are not apparent at this level of description. Thus, the gender effect we observed in Experiment 1 cannot be attributed to different semantic strategies used by Italian and English speakers.

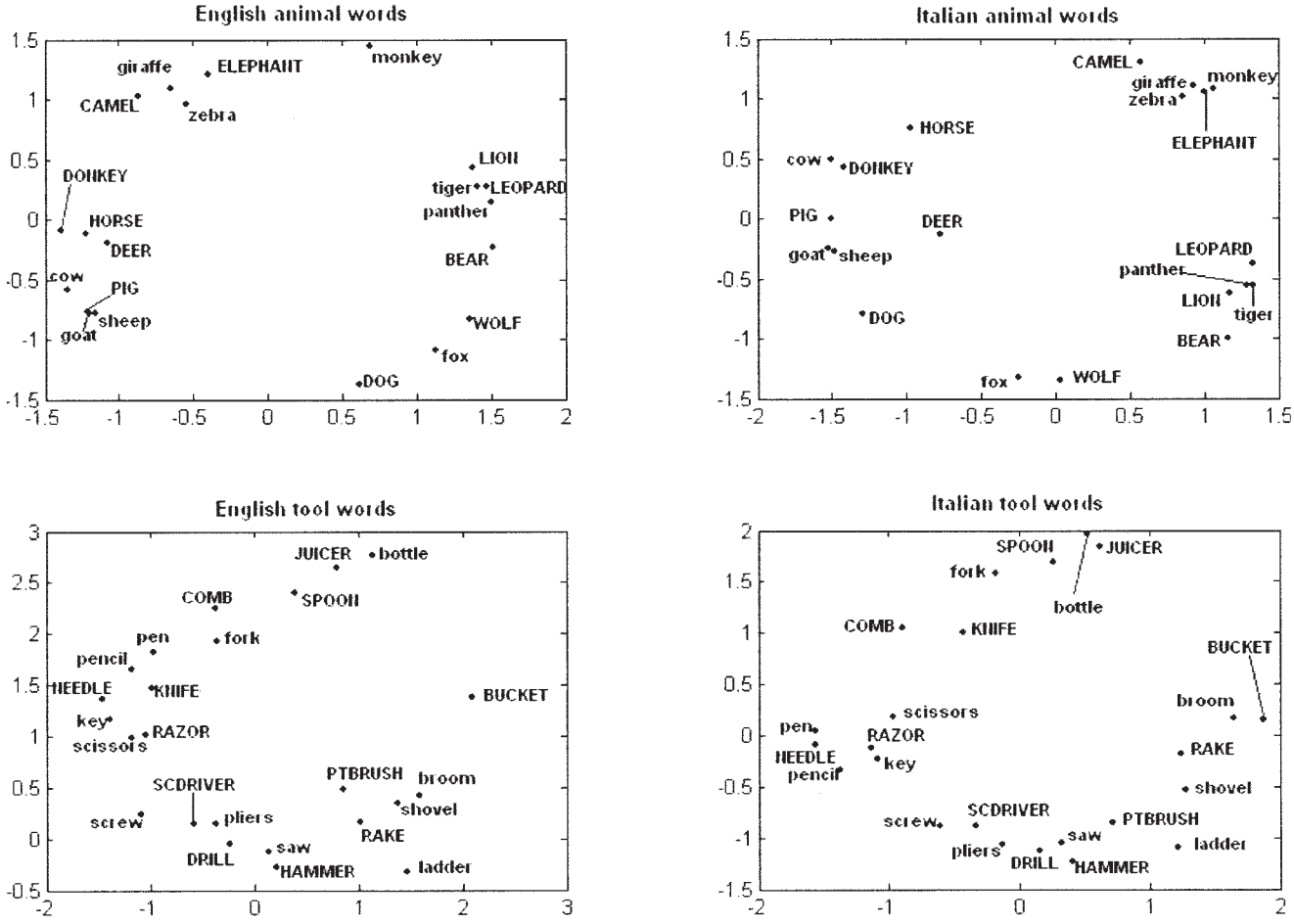

Figure C1. Two-dimensional similarity scaling solution reflecting overall patterns of similarity among words referring to animals and tools in English and Italian. Words with masculine gender in Italian are set in uppercase, and words with feminine gender in lowercase.

(*Appendixes continue*)

Appendix D

Additional Items Used in Experiment 2

Italian	Gender	English
	Animals	
coniglio	M	rabbit
gatto	M	cat
puzzola	F	skunk
riccio	M	hedgehog
scoiattolo	M	squirrel
talpa	F	mole
topo	M	mouse
	Artifacts	
ascia	F	axe
bicchiere	M	glass
bilancia	F	scale
cavatappi	M	corkscrew
chiodo	M	nail
clessidra	F	hourglass
coperchio	M	lid
ditale	M	thimble
falce	F	sickle
frullatore	M	blender
grattugia	F	grater
imbuto	M	funnel
lucchetto	M	lock
maniglia	F	doorknob
mestolo	M	ladle
ombrello	M	umbrella
pentola	F	pot
posacenere	M	ashtray
radio	F	radio
righello	M	ruler
rubinetto	M	tap
sveglia	F	clock
tazza	F	cup
teiera	F	teapot
tostapane	M	toaster
zappa	F	hoe

Note. M = masculine; F = feminine.

Appendix E

Analysis of Phonological Overlap Between Targets and Errors

We first assessed whether there was a tendency in the observed substitution errors for the target and intruding word to share phonological overlap. We considered phonological overlap among all lexical errors using as a measure of phonological similarity the proportion of the target word's phonemes also present in the error word (in the same position). We calculated this by counting the number of phonemes shared between the target and error words in the same positions (i.e., onset–nucleus–coda of first syllable, middle syllable[s], last syllable), divided by the number of phonemes in the target word. Italian target–error pairs sharing gender exhibited greater phonological similarity than those target–error pairs with different gender, for animals $t(191) = 4.91$, $p < .001$; for artifacts, $t(263) = 4.75$, $p < .001$, an unsurprising result given the strong gender cues provided by the final phoneme in most Italian words (Table E1).

Italian gender is reflected in the word-final phoneme; however, phonological overlap in other positions has been shown to be present in mixed errors (especially in word-initial position; Dell & Reich, 1981). To assess whether the Italian target–error pairs sharing gender also showed greater phonological overlap than the target–error pairs that did not share gender, we carried out a follow-up analysis. We calculated a phonological overlap measure (for Italian only) comparable to the previous one, but we considered all phonemes except the last (see Table E2). Italian target–error pairs sharing gender exhibited greater phonological similarity on this measure as well—for animals, $t(191) = 1.99$, $p = .048$; for artifacts, $t(263) = 2.23$, $p = .027$— suggesting that the correspondence between gender and phonology in our error set extends beyond the final phoneme alone. Moreover, as can be seen from Tables E1 and E2, phonological overlap is somewhat greater for the artifacts than the animals.

The finding of greater phonological overlap for Italian words that share gender (even when the final phoneme is not taken into account) justifies our decision to exclude from our analyses of gender effects those errors that exhibited the greatest phonological similarity to the targets to avoid the confounding effect of phonological similarity in our analyses of language-specific gender effects on semantics.

Table E1
Average Proportion of Target Words' Phonemes Shared by Error Words in the Same Position, as a Function of Language, Category, and Grammatical Gender: Experiment 2

		Category	
Language	Italian gender	Animals	Artifacts
English	Different	.180	.170
	Same	.080	.270
Italian	Different	.128	.112
	Same	.260	.244

Table E2
Average Proportion of Target Words' Phonemes Shared by Error Words in the Same Position, Excluding the Final Phoneme, as a Function of Category and Grammatical Gender (for Italian Only): Experiment 2

	Category	
Italian gender	Animals	Artifacts
Different	.113	.104
Same	.143	.157

(*Appendixes continue*)

Appendix F

Items Used in Experiment 3 (German–English Triadic Judgments)

Italian	Gender	English
	Animals	
Affe	M	monkey
Eichhörnchen	N	squirrel
Elefant	M	elephant
Esel	M	donkey
Giraffe	F	giraffe
Hase	M	rabbit
Hund	M	dog
Igel	M	hedgedog
Kamel	N	camel
Känguruh	N	kangaroo
Katze	F	cat
Kuh	F	cow
Maulwurf	M	mole
Maus	F	mouse
Pferd	N	horse
Schaf	N	sheep
Schwein	N	pig
Stinktier	N	skunk
Zebra	N	zebra
Ziege	F	goat
	Artifacts	
Axt	F	axe
Bohrer	M	drill
Deckel	M	lid
Flasche	F	bottle
Gabel	F	fork
Hammer	M	hammer
Korkenzieher	M	corkscrew
Löffel	M	spoon
Meiβel	M	chisel
Mixer	M	blender
Nagel	M	nail
Reibe	F	grater
Säge	F	saw
Schere	F	scissors
Schraube	F	screw
Schraubenzieher	M	screwdriver
Suppenkelle	F	ladle
Tasse	F	cup
Teekanne	F	teapot
Toaster	M	toaster
Topf	M	pot
Zange	F	pliers

Note. M = masculine; F = feminine; N = neuter.

Received May 10, 2004
Revision received May 24, 2005
Accepted May 31, 2005 ■