

# Role of Grammatical Gender and Semantics in German Word Production

Gabriella Vigliocco and David P. Vinson  
University College London

Peter Indefrey, Willem J. M. Levelt, and  
Frauke Hellwig  
Max Planck Institute for Psycholinguistics

Semantic substitution errors (e.g., saying “arm” when “leg” is intended) are among the most common types of errors occurring during spontaneous speech. It has been shown that grammatical gender of German target nouns is preserved in the errors (E. Marx, 1999). In 3 experiments, the authors explored different accounts of the grammatical gender preservation effect in German. In all experiments, semantic substitution errors were induced using a continuous naming paradigm. In Experiment 1, it was found that gender preservation disappeared when speakers produced bare nouns. Gender preservation was found when speakers produced phrases with determiners marked for gender (Experiment 2) but not when the produced determiners were not marked for gender (Experiment 3). These results are discussed in the context of models of lexical retrieval during production.

Speaking involves the retrieval of lexical representations that correspond to our intentions and the development of a syntactically and morphophonologically well-formed frame for the sentence to be uttered. The development of such a frame is, in part, guided by syntactic information specific to each word, for example, its grammatical category, subcategorization requirements of verbs, and, for languages such as German, the gender of nouns. This article focuses on the relationship between the retrieval of a lexical representation that specifies the meaning the speaker wants to convey and the retrieval of its associated lexico-syntactic information.

It has been reported that syntactic properties of words, such as grammatical class (e.g., Garrett, 1980), but also grammatical gender in languages such as German (Marx, 1999), affect lexical errors that spontaneously occur during speech, in which the intended word and the intruder are similar in meaning (e.g., saying “arm” when “leg” is intended). This is an intriguing observation that suggests *prima facie* a very close relationship between retrieval of a lexical representation that corresponds to the meaning the speaker wants to convey and retrieval of syntactic information.

It is generally agreed that semantically related lexical retrieval errors (hereafter *semantic substitutions*) reflect coactivation of semantically related lexical candidates during a conceptually driven retrieval process (e.g., Garrett, 1984, 1992; Levelt, Roelofs, & Meyer, 1999). In most cases, competition in the retrieval process

resulting from this coactivation simply slows the retrieval process, as demonstrated by studies investigating naming latencies using, for example, the picture–word interference paradigm (see Levelt et al., 1999). However, on some occasions the competition is resolved in favor of another semantically related word, leading to semantic substitutions. Furthermore, it has been shown that semantic effects, as reflected both in naming latencies and in substitution errors, are *graded*, namely, modulated by the semantic similarity between words (Vigliocco, Vinson, Damian, & Levelt, 2002; Vigliocco, Vinson, Lewis, & Garrett, in press). Crucial in the present context is the observation that for spontaneously occurring semantic substitutions, the intruders also tend to preserve syntactic properties of targets: Target and intruding words share the same grammatical class (i.e., nouns substitute for other nouns, verbs substitute for other verbs, etc.). In the same vein, Marx (1999) reported that spontaneously occurring semantic substitutions for German nouns tend to preserve not only grammatical class but also the grammatical gender of the target noun. Our goal in the present study was to experimentally explore alternative accounts of these syntactic effects on semantic substitution errors.

How do these effects of syntactic properties on semantic substitutions come about? A first plausible account is that words sharing syntactic properties tend to be semantically more similar than words not sharing those syntactic properties. On this account, syntactic preservation effects would actually be semantic in nature and would occur regardless of whether speakers are producing single words or connected speech. Alternatively, the tendency of targets and intruders to share syntactic properties may reflect the use of these properties for building sentential frames during production and would not occur when words are produced in isolation. Because the reported syntactic preservation effects on semantic substitutions are based on errors in connected speech, these two accounts cannot be teased apart with corpus data; the reason is that lexical retrieval processes and frame building processes are both engaged. Finally, it might be the case that syntactic preservation effects are due to the realization of syntactic properties at the form

---

Gabriella Vigliocco and David P. Vinson, Department of Psychology, University College London, London, England; Peter Indefrey, Willem J. M. Levelt, and Frauke Hellwig, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

Support for this research was provided by National Science Foundation Grant SBR9711829, Human Frontier Science Program Grant HFSP148/2000, and Economic and Social Research Council Grant RES000230038.

Correspondence concerning this article should be addressed to Gabriella Vigliocco, Department of Psychology, University College London, 26 Bedford Way, London WC1, United Kingdom. E-mail: g.vigliocco@ucl.ac.uk

level. These different accounts are discussed in the context of language production models in the next section.

### Syntactic Preservation Effects in Theories of Language Production

Researchers in language production agree in distinguishing processes devoted to retrieving lexical representations and processes devoted to building frames for sentences. Researchers further agree in distinguishing between semantically driven and form-driven lexical retrieval processes. There is ample evidence compatible with these distinctions, as reviewed by Vigliocco and Hartsuiker (2002).

In *WEAVER<sup>++</sup>* (Levelt et al., 1999; Roelofs, 1992) a lexical concept is activated on the basis of the speaker's intentions. In this model of lexical retrieval, developed by Levelt and colleagues, activation from the lexical concept spreads to other concepts that are semantically related to the target and to their corresponding abstract lexical representations (referred to as *lemmas*) via two-way connections. Lemmas corresponding to activated concepts will compete for selection. In this theory, lemmas are considered to be intermediate representations between lexical concepts and word form representations, as well as between lexical concepts and syntactic properties. Syntactic properties such as gender or grammatical class would be retrieved during grammatical encoding, to allow the building of the syntactic frame for the sentence to be uttered. Retrieval of these properties is assumed to depend on selection of the corresponding lemma. Active syntactic frames marked for gender, however, would not feed back activation to the lexical selection process. Hence, retrieval of lexico-semantic representations and retrieval of the corresponding syntactic features are sequential processes, for which the flow of activation is strictly feed-forward: The frame, which is built on the basis of the lexical-syntactic properties, cannot affect the process of lexical (lemma) selection. Thus, in the current formulation of this model, the gender preservation effect reported for spontaneously occurring semantic substitutions can only be accounted for as resulting from greater semantic similarity between target and intruding words that share the same syntactic properties. Because frames cannot be retrieved before lexical selection is completed, the gender of the nonselected target lemma cannot affect the selected intruder. Nonetheless, an intruder sharing the target's gender may be semantically more similar, thus allowing it to be a stronger competitor during lexical retrieval.

The model developed by Dell (1986) and Dell, Schwartz, Martin, Saffran, and Gagnon (1997), as well as the proposal by Stemberger (1985), can also account for the gender preservation effect in terms of greater semantic similarity between target and intruder. These models, however, also allow for different mechanisms. In these proposals, activation is assumed to spread in both directions among lexico-semantic, lexico-syntactic, and word form information. With respect to the interface between lexico-semantic and lexico-syntactic information, interactive activation models are quite underspecified, and clear predictions may strictly depend on details of parameter setting in the models; however, a first conceivable scenario is that the cascading nature of the process allows for the activation (or generation) of gender-marked syntactic frames for highly activated, but not yet selected, lexico-semantic representations. Feedback from a frame to lexical selection could

entail a biasing effect of the syntactic specification of the frame to lexical selection, enhancing the activation of competitors that could fit into the same frame. This is precisely the mechanism described by Stemberger (1985) to account for the grammatical class preservation effect observed in spontaneously occurring semantic substitutions. In this scenario, the semantically driven lexical retrieval process may be affected not only by properties of word form (see Dell, 1986, and Levelt, 1989, for a discussion) but also by the syntactic properties of highly activated lexico-semantic representations. Under this view, the gender preservation effect would arise because the activated (or built-on-the-fly) sentential frames feed back to the lexical selection process, increasing the likelihood of retrieving a semantically related competitor that also shares grammatical gender with the target. It is crucial to note that, in this scenario, the frame feeding back activation to the lexical selection process is specified for abstract syntactic properties but is unspecified with respect to the morphophonological form of the grammatical markers.

An alternative, more complex scenario within spreading activation models would involve feedback to the lexical selection process from both the sentential frame and the morphophonological specification for the phrase. In this scenario, it makes a difference whether the gender of a noun is transparently realized in the determiner's word form; the gender preservation effect could be the product of the combined biasing effect of activated frames (syntactically specified for gender) and transparent morphophonological marking in the word form.

### Semantic Representations and Language-Specific Syntactic Properties

The semantic account just described assumes that syntactic effects on semantic errors arise because words sharing syntactic properties are semantically more similar than words not sharing those syntactic properties. Because syntactic properties can be language specific, as is the case for the grammatical gender of most German nouns, this account has additional implications for whether semantic representations can be affected by language-specific syntactic properties (thus differing among speakers of different languages).

It is important to distinguish cases in which syntactic properties of words are linked to semantics in a manner that is common across languages from those cases in which syntactic properties and meaning are linked in language-specific manners. Across languages, lexico-syntactic properties of words are often closely linked to the meanings of the words. For example, in the case of grammatical class, the semantic distinction among objects, actions, and properties corresponds to the syntactic distinction among nouns, verbs, and adjectives. The semantic distinction between countable entities and substances is captured by the syntactic distinction between count and mass nouns in some languages, and the semantic dimension of biological gender corresponds to a syntactic distinction between masculine and feminine nouns in many languages. In all of these cases, the syntactic distinctions have clear conceptual foundations, as reflected in the fact that languages largely agree in their assignment of syntactic properties (e.g., noun and verb) to semantically distinct referents (e.g., objects and actions).

More intriguingly, correlations across languages have also been found when the conceptual basis of a syntactic distinction is less transparent. In many languages, gender is a property not only of nouns referring to humans and (some) animals but also of nouns referring to objects and abstract entities for which gender has no obvious conceptual foundation. Nonetheless, even in this latter case, some evidence exists that gender attribution across languages is not completely arbitrary (Boroditsky & Schmidt, 2000; Foundalis, 2002). For example, Foundalis (2002) showed that there are significant correlations among languages in terms of the gender assigned to translation-equivalent words; significant correlations were also observed between gender attributions made by speakers of gendered languages and those made by speakers of English. The existence of such correspondences between the semantic and the syntactic dimensions has been taken by some authors to indicate a conceptual foundation for syntactic distinctions (e.g., Bates & MacWhinney, 1982; Boroditsky & Schmidt, 2000; Langacker, 1987). Hence, in such views, the semantic specification of words and their syntactic properties can be considered to be part of the same representation or, if not, to be represented in a strongly interdependent manner.

Syntactic preservation effects on semantic substitutions such as the grammatical class effect can be plausibly accounted for as semantic: arising as a consequence of the conceptual foundation of grammatical class. Gender preservation effects may also reflect cross-linguistic conceptual similarity among entities (despite the fact that the conceptual foundation of grammatical gender is far more modest than the conceptual foundation of grammatical class). Under a semantic account, gender preservation may, in addition to cross-linguistic conceptual similarity, reflect increased semantic similarity among words that happen to share the same syntactic properties only in that particular language. It has been argued that even when gender is language specific and idiosyncratic, it becomes part of the semantic specification of the corresponding words once learned, thus rendering words that share the same language-specific gender semantically more similar (Sera, Berge, & del Castillo-Pintado, 1994). This possibility has implications for linguistic relativity claims.

Within the cognitive psychology tradition, the received view is that we share the mental (conceptual) representations we have for things and events in the world. Although there certainly is some degree of flexibility related to differences in the physical and cultural environments we live in (see Clark, 1996), differences in the language we speak do not affect our conceptual representations of things and events. For example, the fact that Spanish has a smaller repertoire of verbs expressing manner of motion (i.e., the way in which a movement is carried out, e.g., “crawling”) than English (Slobin, 1996) does not imply that Spanish speakers cannot conceptualize different types of motion, just that they verbalize them differently (i.e., with a phrase instead of a single word). The fact that “knowledge” in English cannot be pluralized, whereas “opinion” can, and the fact that in Italian both can be pluralized do not necessarily imply that English speakers conceive of “knowledge” and “opinion” in a way that differs from Italian speakers’ concepts of “conoscenza” and “opinione.”

Alternative views arguing for linguistic relativity according to which language-specific differences can affect our conceptual representations for things and events in the world are also represented in the literature. There are various versions of linguistic relativity

positions, from strong formulations, according to which the language spoken by a community could affect cognition beyond language (Sapir, 1921; Whorf, 1956), to views according to which language-specific properties would affect only cognition specifically related to linguistic expression (“thinking for speaking”) in that specific language (Slobin, 1996). Some recent evidence has been argued to provide support for linguistic relativity in its strong formulation, even for language-specific syntactic properties such as gender. For example, Konishi (1993) asked speakers of Spanish and German to rate words on the 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 of words for which gender differed (for example, the word *fork* is masculine in Spanish [*tenedor*] and feminine in German [*Gabel*]). On the basis of this correspondence between speakers’ semantic ratings and the grammatical gender of nouns, Konishi argued that grammatical gender is intimately related to the conceptual representations of words.

Along similar lines, Sera et al. (1994) asked speakers of Spanish and English to assign male or female voices to pictured objects. Unlike English speakers, Spanish speakers tended to follow the Spanish gender differences in their assignments of voice. In a more recent study involving the same procedure, Sera et al. (2002) extended the investigation to two other gendered languages: French and German. They replicated the difference between Spanish and English and reported that speakers of French (another Romance language) showed gender-specific effects similar to the Spanish speakers. German speakers, however, did not perform differently from English speakers, in contrast to the results of the study by Konishi (1993) just described. Finally, Boroditsky, Schmidt, and Phillips (2003) cited experiments (performed entirely in English) in which bilingual speakers (Spanish–English or German–English) were asked to generate adjectives referring to English nouns whose translation equivalents had different genders in Spanish and German. Independent judges rated the adjectives with respect to whether they reflected more masculine or feminine properties. Speakers tended to report more female-like adjectives for words that had feminine gender in their native language than words that had masculine gender (and vice versa for male-like adjectives), even though the experiment was conducted entirely in English. These findings are also taken to reflect language-specific gender effects on conceptual representations beyond “thinking for speaking” in a given language.

Finding effects of language-specific gender on linguistic tasks indicates differences in thinking for speaking across languages, differences that are a prerequisite for observing effects of language-specific properties beyond linguistic tasks. In our work, we assess whether language-specific (German) gender renders words that share the same gender more semantically similar; thus, we assess whether grammatical gender affects thinking for speaking among German speakers.

In summary, syntactic effects such as the gender preservation effect on semantic substitutions provide us with a window into how semantic and lexico-syntactic properties are represented and how their retrieval is orchestrated in on-line production. On the other hand, the finding of gender preservation in semantic substitutions is compatible with an effect of language-specific syntactic

properties (once learned) on semantic representations. However, it is also compatible with accounts in which these gender preservation effects come about as a consequence of the interplay between retrieving lexico-semantic representations and building sentential frames (at either the syntactic or morphophonological level). The present set of experiments was designed to explore these alternatives.

### Plan of the Study and Relevant Properties of German

We report a series of three experiments, conducted in German, in which we induced semantic substitutions using a continuous picture naming paradigm. In this paradigm, speakers are presented with pictures in quick succession, and their task is to name the pictures. The same pictures are repeated in different blocks. In previous work we established that, when this paradigm is used, speakers' errors tend to involve other words in the response set. This allowed us to control the composition of the semantic neighborhoods. We further established that the produced errors show graded semantic effects. Within semantic fields, the likelihood of observing a given target–intruder pair is modulated by the semantic similarity between the two words (Vigliocco et al., in press). In this previous work as in the current experiment, semantic similarity was operationalized as item-by-item semantic distance measures derived from speaker-generated feature norms (Vinson & Vigliocco, 2002). Thus, this methodology allowed us to assess semantic similarity differences at a fine-grained level. This was crucial given our goal to establish whether words that share the same syntactic property (gender) are more semantically similar in a given language.

Our target language was German, in continuity with previous work (Marx, 1999). German has three genders (masculine, feminine, and neuter), and, in contrast to English, all nouns in the language are marked for gender. In the nominative case, gender is marked in phrases with different definite determiners (“der” = masculine; “die” = feminine; “das” = neuter). However, in phrases including indefinite determiners, both masculine and neuter nouns are introduced by the same determiner (“ein”), whereas feminine nouns are introduced by a different determiner (“eine”). Although there are very few transparent morphophonological correlates of gender in the form of nouns (at least relative to the extent to which gender is marked on the noun in languages such as Italian), many fine-grained correspondences between gender and form do exist (Köpcke & Zubin, 1983). Therefore, to minimize the possible confounds between gender and form, we selected items with a minimum of form similarity within gender classes; furthermore, we took form similarity into account in all of the analyses we performed.

### Experiment 1

If the gender preservation effect reported for spontaneously occurring semantic substitutions arises as a consequence of greater semantic similarity among words sharing the same gender, the effect should be present when speakers are producing bare nouns. By contrast, if the effect arises as a result of the activation of gender-marked frames, it should not be present in bare noun production, because sentence-level frame building mechanisms should not be triggered. Relevant here is the finding, in picture

naming experiments conducted in Dutch and German, that a word presented at the same time as (or shortly after) a target picture speeds up naming if the presented word and the picture name share the same gender; crucially, however, this effect arises only when speakers are asked to produce phrases (hence, when sentence generation mechanisms are triggered) and not when they are asked to produce bare nouns (La Heij, Mak, Sander, & Willeboordse, 1998; van Berkum, 1997). This suggests that grammatical gender is retrieved only during connected speech, when gender-marked frames for the to-be-uttered sentence are built, and that grammatical gender does not affect lexical retrieval per se.

In the continuous picture naming task, lexical substitution errors produced by speakers can, in addition to semantic similarity, reflect phonological similarity between the target and the intruder. That is, speakers may produce errors in which target and intruder are solely phonologically related (e.g., saying “shovel” when “shuttle” is intended), a type of error well documented in spontaneously occurring slips of the tongue (e.g., Garrett, 1984). Furthermore, speakers may also show a tendency to produce more errors when the target and intruder words are both semantically and phonologically related, namely a tendency to produce mixed errors, which are also documented in spontaneously occurring slips (e.g., Dell & Reich, 1981). To minimize the possible confounding effects of phonological similarity, we favored items with no obvious phonological overlap within semantic fields. Because it was not possible to completely avoid phonological similarity among items, we also took phonological similarity into account in all of the analyses we conducted. The presence or absence of errors in all possible within-set target–intruder pairs was entered into a logistic regression model in which we considered semantic similarity (measured in English for translation equivalents of the German words used in the present study), phonological similarity, and gender as predictors. As we describe in more detail subsequently, gender was entered into the regression model as a predictor only after the effects of meaning similarity and form similarity had been partialled out. The crucial question is whether language-specific effects of grammatical gender significantly predict the occurrence of errors once meaning similarity (including cross-linguistic, conceptual contributions to gender) and form similarity are taken into account.

### Method

*Participants.* Eighteen native German speakers ranging from 18 to 34 years of age were recruited from the Nijmegen community and from the Heinrich Heine Universität Düsseldorf. Participants volunteered or received 8 DM (U.S. \$10.24) for taking part; all reported normal or corrected-to-normal vision.

*Materials.* To allow the opportunity for within-field substitution errors to potentially exhibit variation in gender of intruding words, it was necessary to select semantic fields that contained picturable entities varying across masculine, feminine, and neuter gender. However, we could not avoid having different numbers of masculine, feminine, and neuter words in each field, owing to the different frequencies of the three genders in German. We selected the semantic fields of *body parts* and *animals*. Within these fields, items were initially selected on the basis of their picturability. Pictures (line drawings) were derived from Snodgrass and Vanderwart (1980); some additional pictures were prepared specifically for this experiment. An item was included in the experimental set only if its label (and gender) was clearly and unambiguously known to an informal panel of four native German speakers. On this basis, 19 body parts and 28 animals were

selected (see Table 1 for a breakdown of gender by semantic field, and see the Appendix for the specific items and their English translations).

Each picture was presented in black on a white background and scaled to fit within a 240 × 240 pixel space. Blocks of 12–14 pictures were constructed by randomly selecting pictures, with the following constraints. First, each block consisted of only one semantic field (body parts or animals). This was done to increase the number of within-field errors produced. Second, a picture could occur no more than once within a block and could not occur as the last item in one block and the first item in the next. Finally, each picture was to appear 14 times in the course of the experiment (a total of 53 experimental blocks). The order of blocks was randomized for each participant. Stimulus presentation was carried out, with the Nijmegen Experiment Set-Up program, on IBM-compatible computers.

*Procedure.* Participants were instructed that they were taking part in a study of speech patterns under time pressure and that they would be asked to name aloud pictures using a single word as they appeared on the screen. Instructions emphasized that speakers should attempt to keep up with the rate of presentation, skipping items if necessary to recover from difficulty (i.e., naming pictures as they appeared rather than retaining them in memory). 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 the pictures without time pressure. The experimenters noted any variation from the intended names, and they also provided prompts if the participants were unable to produce a label for the picture. After this, participants performed a set of practice trials (in which each target picture appeared once). In the practice trials, 12–13 pictures in a row were presented in one of six possible locations on the screen, and the participant was instructed to name each aloud. After each block of 12–13 pictures, the experimenter altered the rate of presentation if necessary to accommodate each speaker's speech rate. Initial presentation parameters (display times) were 700 ms for each picture, with a blank interstimulus interval (ISI) of 200 ms before the next picture appeared in a block. This rate was altered by the experimenter during the practice session to make the task difficult but manageable for each speaker; final display times ranged from 400 ms to 800 ms ( $M = 617$  ms), always with a 200-ms ISI. If the participant was not comfortable and ready to begin at the end of the practice session, or if the experimenter needed more trials to appropriately adjust the presentation rate, a second practice session was conducted; this was the case for only 1 participant in this experiment.

Once the practice session was complete, the experimental blocks were presented. Participants verbally signaled their readiness, and the experimenter pressed a key to begin each trial. A fixation cross briefly appeared in the center of the screen to indicate that the block was beginning, and then the 12–14 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, the participant was given the opportunity to take a break if necessary, and at the halfway point a short break was enforced. Each picture was presented 14 times for naming in the course of the

experiment, which lasted between 20 and 30 min. At the end of the experiment, participants were debriefed; in particular, they were asked to indicate what they thought the purpose of the experiment might be. Most participants' responses focused on the phonological qualities of their utterances; those participants who correctly guessed that the focus was on errors ( $n = 4$ ) were noted for subsequent analysis. Because the latter participants' pattern of response did not differ in any consistent manner from the other participants, all participants were collapsed into a single group for analysis.

*Scoring criteria.* Participants' responses were transcribed and scored in the following categories. Responses were scored as *correct responses* when the participant uttered the correct target word completely. Responses were scored as *omissions* when participant did not attempt to produce the target word. *Different label* referred to the participant using a different word than our intended target but doing so consistently. This was identified in either of two ways: The participant used the different label in the initial untimed naming phase, or the participant used the different label twice or more in the experiment itself without self-correcting. These items were treated as correct in the analyses. *Lexical errors* were classified as instances in which the participant produced a word that differed from the target and that did not qualify as a "different label." Lexical errors were further classified as "out of set" (intruding words that were not in the response set of the present experiment) and "within set" (items from within the present response set). These latter errors were the crucial focus of the present study. *Corrections* were categorized as instances in which the participant started to produce an incorrect word but changed his or her response to the correct target before it was complete. *Miscellaneous errors* referred to responses other than those just described, such as dysfluencies and unintelligible responses.

*Data analysis.* All of the analyses reported subsequently concern within-set lexical errors (i.e., cases in which the word produced for a target was another word in the experimental set), which were the most common type of error. The presence of lexical errors across participants was recorded as entries in within-field confusion matrices. For example, if any participant mistakenly produced "Pferd" (horse) for the picture "zebra," a 1 was entered into the Pferd–zebra matrix cell; 0 was entered if that error never occurred. Confusion matrices were prepared only within semantic fields (separate confusion matrices for body parts and animals), because there were no between-fields errors in the present study. The complete set of entries corresponding to the presence or absence of errors between each possible within-field target–error pairing was submitted to a sequential logistic regression model considering meaning similarity and form similarity (as detailed subsequently) as predictors in a first step and gender in a second step. This analysis allowed us to assess the unique contribution of gender beyond those aspects that are correlated with cross-linguistic, conceptual contributions to gender.

The first predictor was *meaning similarity*. This consisted, for each target–intruder pair, of a value between 0 and 30 reflecting meaning proximity between the two corresponding words in English (lower values corresponded to greater similarity). In a previous study, we developed these empirically based measures of semantic distance for a set of 456 English words (words used in the present experiment were German translations of a subset of these items). The measures of semantic distance were obtained in the following manner (for detailed descriptions of the methods, see Vigliocco et al., in press; Vinson & Vigliocco, 2002). First, feature norms for the whole set of words were obtained by asking speakers of English to generate features that define and describe a given word. Second, ensemble averages from multiple self-organizing maps (Kohonen, 1997) were used as a statistical tool to reduce the dimensionality of the featural space so as to obtain a semantic space in which each word was represented as the unit best responding to an input vector in the resulting map after training. Semantic distance was operationalized as the average Euclidean distance between any two such units in the space. Development of this model of the semantic space provided us with fine-grained measures of

Table 1  
*Grammatical Gender of Items Used in the Present Experiments*

Semantic field	Feminine		Masculine		Neuter	
	No.	%	No.	%	No.	%
Animals	8	29	13	46	7	25
Body parts	5	26	9	48	5	26

*Note.* Percentages indicate gender distribution within a given semantic field.

semantic similarity. We have previously shown that these measures successfully predict graded semantic similarity effects (Vigliocco, Vinson, et al., 2002; Vigliocco et al., in press), and using them allowed us to assess the role of fine-grained semantic similarity in determining target–intruder pairing in our set.

Crucially, our semantic similarity measures were taken from English. The experiment assessed whether *language-specific* grammatical gender assignment (adding to the baseline of cross-linguistic correlations between gender and semantics) increases the semantic similarity of words sharing the same gender in German, thus rendering them better lexical competitors for retrieval. Using semantic similarity in English, we partialled out any cross-linguistic conceptual contribution to the gender preservation effect, leaving only the contribution of gender specific for German. This is important because, as mentioned earlier, gender is somewhat correlated with meaning across languages. In particular, German gender has been found to be correlated with English speakers' gender categorizations (Foundalis, 2002). Our English semantic similarity measures reflect the correlation between meaning and gender due to conceptual factors, a correlation that determines some degree of agreement in gender assignment between English speakers' intuitions and German grammatical gender. When this predictor is entered in the regression model before gender, a gender effect will be factored out to the extent that English speakers will have correctly predicted German gender. Under the semantic account of syntactic preservation effects, however, German-specific effects of gender on meaning similarity should still be present after the effect of cross-linguistic, conceptual determinants of gender has been partialled out.

The second predictor was *word form similarity*. As mentioned earlier, word form has also been shown to affect the likelihood of substituting one word with another, at least in English and Spanish (del Viso, Igoa, & García-Albea, 1991; Dell & Reich, 1981). The measure of word form similarity among the words in our set was computed as in Damian and Martin (1998). This metric is a linear combination of the following contributions to form similarity: (a) shared number of syllables (0 for different numbers of syllables, 1 for the same number of syllables), (b) syllabic location of main stress (0 for different, 1 for the same), (c) the proportion of shared phonological segments in the two words regardless of their positions (a value ranging from 0 to .8 within this set of words), and (d) the proportion of shared phonological segments in the same positions (also ranging from 0 to .8 in this set). These values were summed to produce the form metric.

Our final predictor was *grammatical gender*. This predictor was assigned a value of zero if the target–error words had different genders and a value of one if they shared gender.

In the logistic regression model, we entered the predictors of semantic and form similarity in a first step, followed by grammatical gender. As stated earlier, this allowed us to assess whether grammatical gender still exerted an effect once cross-linguistic meaning similarity and form similarity had been taken into account.

## Results

Overall, participants were highly accurate in their responses, providing 11,034 correct responses in 11,844 trials (a rate of 93.2% correct). Two hundred twenty-nine (1.9%) within-set lexical errors occurred, along with 238 (2.0%) omissions, 275 (2.3%) corrections, and 68 (0.6%) miscellaneous errors (including 8 lexical errors from outside the response set). Table 2 presents a breakdown of the lexical errors for each semantic field and each gender.

As can be seen in Table 2, grammatical gender was preserved at a rate of 46.3% across all three gender classes. This is significantly higher than the rate that would be predicted by chance (30.4%), assuming an equal probability of error occurrence between any

Table 2  
*Gender Breakdown of Errors for Words of a Given Gender: Experiment 1 (Bare Nouns)*

Target gender	Error gender					
	Feminine		Masculine		Neuter	
	No. of errors	%	No. of errors	%	No. of errors	%
Animals						
Feminine	1	5	6	28	14	67
Masculine	17	24	40	56	14	20
Neuter	9	31	4	14	16	55
Body parts						
Feminine	5	17	21	72	3	10
Masculine	12	21	32	55	14	24
Neuter	0	0	9	43	12	57
Overall						
Feminine	6	12	27	54	17	34
Masculine	29	22	72	56	28	22
Neuter	9	18	13	26	28	56

*Note.* Percentages indicate distribution of errors for targets of a particular gender and may not sum exactly to 100 because of rounding error.

two words in a given semantic field and taking into account the fact that different number of exemplars from each gender class were used in the experiment. However, as discussed earlier, we had to partial out meaning similarity before we could test whether this grammatical gender preservation effect was language specific or whether it arose as a result of conceptual correlates of gender not specific to German. The 229 within-set lexical errors were entered into confusion matrices (separately for animals and body parts) and submitted to a logistic regression analysis. In this analysis, we found that the semantic similarity predictor significantly predicted the presence of errors ( $R = -.21$ , Wald statistic [tested against a chi-square distribution] = 15.95,  $p = .0001$ ). The form predictor was only marginal in predicting errors ( $R = .06$ , Wald = 3.25,  $p = .0715$ ). Crucially, once semantic and form similarity were taken into account, gender no longer predicted the presence of errors ( $R = .0000$ , Wald = 1.40,  $p = .2374$ ).

## Discussion

The main finding of the present experiment was that grammatical gender is not a significant predictor of error pairing in bare noun naming in German once the cross-linguistic, conceptually based component of gender preservation is partialled out in the regression model. With regard to the other predictors, semantic similarity measures obtained for English significantly predicted the target–intruder pairings that actually occurred as errors in German. The form predictor, on the other hand, was only marginally significant, and we did not observe any errors in which target and intruder were related in form but not in meaning. This lack of an effect of form may be due to a number of reasons. First, it may simply reflect our item selection criteria; we preferentially selected items with little or no form similarity. Second, it may reflect task requirements; speakers in this study were pressured to name the pictures quickly. Under the proposal by Dell (1986), in which form effects on semantic errors arise as a consequence of feedback from phonological to lexical selection processes, it may be that in the

present experiment there was not enough time for feedback processes to exert any effect.

Although we observed no significant effect of language-specific grammatical gender, this is not to say that no gender preservation was present in this experiment. Gender was preserved at levels above those predicted by chance rates alone; the gender factor became nonsignificant only when English semantic similarity was taken into account. This might have occurred because a large number of observations in our study came from the animal semantic field, a field in which more agreement is observed between German gender and English speakers' intuitions than for artifacts (Boroditsky & Schmidt, 2000). Also, to our knowledge, no one has investigated cross-linguistic gender attributions for body parts, so this field might also exhibit a higher cross-linguistic correspondence than artifacts. In a follow-up study, we directly compared errors made by German speakers and errors made by English speakers to further assess language-specific gender effects beyond the contribution of cross-linguistic conceptual correspondence.

We replicated Experiment 1 with 20 monolingual English speakers from the University College London community (each paid £ 3; U.S. \$5.50), using exactly the same methodology and items. These speakers produced 226 within-set semantic substitutions (117 for animals and 109 for body parts). For each target-error pairing that occurred with a frequency of one or more in either German or English, a difference score was calculated by subtracting the number of English errors from the number of German errors for that word pair. For example, for the target word "Schaf" (sheep), German speakers erroneously produced "Ziege" (goat) 4 times, whereas English speakers made this error 3 times; the resulting difference score was 1. We then conducted an item analysis on the error difference scores, using German gender preservation between target and error (same gender vs. different gender) as a within-item factor and error difference scores as the dependent measure. The logic here was as follows. Under a semantic account, for cases in which the target and error words have the same gender, there should be more errors (positive difference score) in German than in English (in that English does not have German gender), whereas there should be fewer errors (negative difference score) in German when the target and error have different genders (a consequence of the predicted greater number of errors for same-gender word pairs). This prediction arises because, under this hypothesis, German-specific gender should contribute to the semantic similarity of German words above and beyond the contribution of any cross-linguistic, conceptual gender shared between the two languages. The results of this experiment showed otherwise; difference scores (summed across all words in a condition) were  $-1$  for same-gender pairs and  $4$  for different-gender pairs. A Mann-Whitney  $U$  test (by item pairs) revealed no significant difference between the difference scores for same-gender and different-gender word pairs ( $Z = 0.276, p = .782$ ), suggesting that English speakers made semantic substitutions that preserved (German) gender to the same extent as German speakers. Separate analyses for animals and body parts revealed no significant semantic field differences in gender effects. Thus, this follow-up study provides converging evidence against the view that gender preservation effects in semantic errors arise from language-specific properties of gender.

The lack of a language-specific gender effect in Experiment 1 suggests that, to account for the gender preservation effects ob-

served in spontaneous speech, we must consider mechanisms related to producing gender-marked frames. In Experiment 2, therefore, we assessed this possibility by presenting the same stimuli to a different group of German speakers who were asked to produce a noun phrase (definite determiner plus noun). Producing a noun phrase is more similar to connected speech, the condition under which gender preservation effects have been reported (Marx, 1999).

## Experiment 2

If the gender preservation effect observed in spontaneously occurring errors is a consequence of the fact that the error is produced in a gender-marked phrasal context, we should observe it in an experiment in which speakers are asked to produce a noun phrase. WEAVER<sup>++</sup> (Levelt et al., 1999), according to which gender-marked syntactic frames are retrieved (or built) solely after the lexical selection process has been completed, predicts no effect of phrasal production (and hence no difference between Experiments 1 and 2). A gender preservation effect is predicted by spreading activation models (Dell, 1986; Stemberger, 1985) according to which gender-marked frames can bias the lexical selection process.

## Method

*Participants.* Eighteen native German speakers ranging in age from 19 to 26 years were recruited from the Nijmegen community and from Heinrich Heine Universität Düsseldorf. Participants volunteered or received 8 DM (U.S. \$4) for taking part; all reported normal or corrected-to-normal vision. None had participated in Experiment 1.

*Materials.* Experiment 2 involved the same materials as Experiment 1.

*Procedure.* The procedure was the same as that of Experiment 1, with the following exceptions. Participants were instructed to produce a definite noun phrase in response to the picture, such as "der Leopard" (the-masculine leopard), "das Pferd" (the-neuter horse), or "die Ente" (the-feminine duck). To allow for the longer time needed to produce a determiner-plus-noun phrase, we changed the initial presentation rate (display time) slightly to 800 ms for each picture, again with a 200-ms ISI before the next picture. In this experiment, final presentation rates ranged from 600 ms to 900 ms ( $M = 778$  ms), and 2 of the 18 participants required a second practice session.

*Scoring criteria.* The scoring criteria were the same as in Experiment 1, with one additional scoring category: *determiner errors*, errors in producing the correct determiner (e.g., producing an incorrect determiner for a noun that was produced, making corrections involving the determiner, and exhibiting other miscellaneous errors that involved the determiner but not the noun).

*Data analysis.* As in Experiment 1, the presence of lexical errors across participants was recorded as entries in within-field confusion matrices, and the resulting values were submitted to a sequential logistic regression model that considered meaning similarity and form similarity as predictors in a first step and gender in the second step.

## Results

Participants' rates of correct responses were high: 10,700 correct responses in the 11,844 trials (90.4% correct). Two hundred twenty-five (1.9%) within-set lexical errors occurred, along with 242 (2.0%) omissions, 233 (2.0%) corrections, 323 (2.7%) determiner errors, and 121 (1.0%) miscellaneous errors (including 11 lexical errors from outside the response set). The 225 within-set

lexical errors were entered into confusion matrices (separately for animals and body parts) for further analysis. Table 3 presents a breakdown of lexical errors for each semantic field and each gender.

Logistic regression (as in Experiment 1) revealed, again, that semantic similarity strongly predicted the presence of errors ( $R = -.2603$ , Wald = 20.79,  $p < .0001$ ). Form similarity was also significant ( $R = .1855$ , Wald = 11.55,  $p = .0007$ ). Finally, once semantic and form similarity had been taken into account, grammatical gender significantly predicted the presence of errors ( $R = .2667$ , Wald = 21.73,  $p < .0001$ ).

As mentioned, an additional category of errors, *determiner errors*, was observed in this experiment, in 2.7% of all trials. Because such errors are extremely rare in German under normal conditions, we conducted more detailed analyses of errors of this type. Among these errors, the majority were noticeable pauses between determiner and noun (e.g., “die . . . Schalter;”  $n = 121$ ) or stutters or dysfluencies occurring during determiner production in otherwise correct utterances (e.g., “d.d.die Nase;”  $n = 84$ ). Also relatively common were corrections of various kinds involving determiners, some of which were overt and fully corrected such as “das, der Arm” ( $n = 62$ ); others involved correction of a determiner on the fly, such as “die, er Arm” ( $n = 29$ ). Of the remaining cases, 26 could be considered “true” determiner errors: utterances in which the incorrect determiner was produced along with the correct target word, such as “das Ziege” (the–neuter goat–feminine) or “die Arm” (the–feminine arm–masculine). Finally, there was only one instance of a lexical error in which a determiner error occurred—target picture: Ente (duck–feminine); response: “die, der Schwan” (the–feminine, the–masculine swan–masculine)—effectively a correction but to the wrong word.

### Discussion

A gender preservation effect was observed in this experiment, which involved the same materials and the same procedure as

Experiment 1 with only a difference in task requirements. In Experiment 1, speakers were required to produce a bare noun; in Experiment 2, they were asked to produce a definite-determiner-plus-noun phrase.

In addition, the phonological similarity predictor was significant, whereas it was not in Experiment 1. As in Experiment 1, the semantic similarity predictor was highly significant. Under the assumption of spreading activation models, in which feedback from word forms to lemmas takes time to occur (Dell, 1986), the contrasting results for form similarity in Experiments 1 and 2 might be explained in the following way: The longer picture presentation in Experiment 2 than in Experiment 1, necessary to allow phrase production, allowed enough time for feedback to occur.

In this experiment, we also observed errors in determiner production. In a number of these errors, the target word was uttered correctly but with the incorrect determiner. Responses of this kind may reflect failures in agreement processes but may also plausibly arise as self-corrections. In fact, responses of this kind were most likely to occur for those words that involved a high lexical error rate, and the erroneous determiner tended to correspond to a close competitor to the target word. For example, the neuter determiner “das” produced with the feminine noun “Ziege” might be associated with the neuter competitors “Schaf” (sheep) or “Schwein” (pig); the feminine determiner “die” produced with the masculine noun “Arm” might be related to its feminine competitor “Hand.”

The results of Experiment 2 indicate that the gender preservation effect observed in spontaneous errors is related to the processes engaged in producing gender-marked phrases. Thus, these findings argue against strict seriality between lexical selection and frame building. This is because finding an effect of the gender of the target on the intruder implies that the syntactic properties of the target must have been retrieved even if the target itself was not selected (when an error occurs), in contrast to the claim made by Levelt et al. (1999).

Does the gender preservation effect arise as a result of feedback from a syntactically (but not morphophonologically) specified frame to the lexical selection process? As depicted in Figure 1, the results of Experiment 2 are compatible with such a possibility. The gender preservation effect would arise because the activated (or built-on-the-fly) sentential frames feed back to the lexical selection process, increasing the likelihood of retrieval of a semantically related competitor sharing the same syntactic properties.

However, our results are also compatible with other mechanisms that imply a role for the morphophonological specification of gender. An influence of the morphophonological form of the determiner on differences in naming latencies has been established in previous studies. For example, Schiller and Caramazza (2003) examined gender congruency effects on naming latencies in experiments involving the picture–word interference paradigm in German. They asked speakers to produce either singular definite-determiner-plus-noun phrases (in which the determiner marks gender in the form) or plural definite-determiner-plus-noun phrases (in which the determiner does not mark gender). Gender congruency effects were found for singular (marked determiner) phrases but not for plural phrases. Schriefers, Jescheniak, and Hantsch (2002) compared the naming latencies for singular and plural bare nouns and singular and plural definite-determiner-plus-noun phrases, also in German. Whereas singular phrases involve the use of different

Table 3  
Gender Breakdown of Errors for Words of a Given Gender:  
Experiment 2 (Definite Determiner)

	Error gender					
	Feminine		Masculine		Neuter	
Target gender	No. of errors	%	No. of errors	%	No. of errors	%
Animals						
Feminine	7	54	0	0	6	46
Masculine	5	5	96	88	8	7
Neuter	7	21	2	6	25	73
Body parts						
Feminine	6	60	4	40	0	0
Masculine	2	5	36	88	3	7
Neuter	1	6	4	22	13	72
Overall						
Feminine	13	57	4	17	6	26
Masculine	7	5	132	88	11	7
Neuter	8	15	6	12	38	73

Note. Percentages indicate distribution of errors for targets of a particular gender and may not sum exactly to 100 because of rounding error.



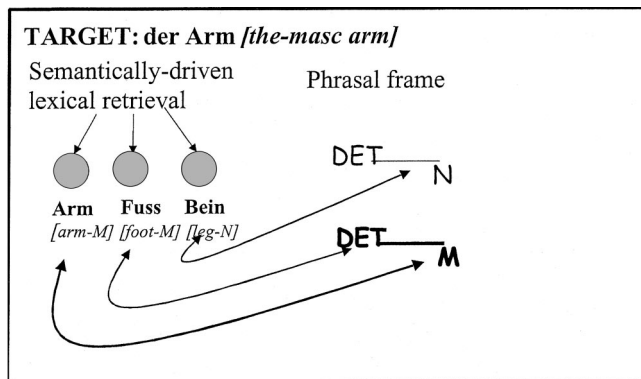


Figure 1. How the gender preservation effect might arise under conditions of feedback from a syntactically (but not morphophonologically) specified frame to lexical selection. Highly activated lexico-semantic representations feed activation to the frames, and activation would feed back from the frames to the selection process. Assuming that the target is always highly activated, its corresponding frame will also be highly activated; in turn, the frame specified for the gender of the target will send more activation back to the selection process, thus favoring intruders that share the same gender. M = masculine; N = neuter.

determiners for nouns of different gender, plural phrases involve the use of the same determiner. Furthermore, the plural determiner (in the nominative case) has the same word form as the feminine determiner. In this study, no reliable naming latency differences for words of different gender were observed when speakers produced bare nouns; however, when speakers produced phrases, an interaction between number and gender was found. Naming latencies were longer for plural masculine and neuter gendered phrases, whereas no difference between singular and plural was observed for feminine phrases.

In Experiment 3, we tested whether or not morphophonological marking of the determiner contributed to the effect observed in Experiment 2. Thus, we assessed whether the language-specific gender preservation effect on semantic errors we found when speakers produced gender-marked determiners would also occur when speakers produce determiners that are underspecified in terms of gender.

### Experiment 3

As mentioned, whereas definite determiners are always marked for gender in the nominative singular word forms we elicited in Experiment 2, indefinite determiners are not always differentially marked for gender in the nominative case. Both neuter and masculine nouns use the determiner “ein;” feminine nouns use the determiner “eine.” If the gender preservation effect arises because of direct feedback from a gender-marked syntactic frame to the lexical selection process that biases the process toward other lexical competitors fitting the same syntactic frame, we should observe a gender preservation effect regardless of whether the word form of the determiner univocally marks gender. Specifically, neuter nouns should not substitute for masculine nouns, and vice versa, when speakers are asked to produce an indefinite-determiner-plus-noun phrase.

### Method

**Participants.** Eighteen native German speakers ranging in age from 18 to 30 years were recruited from Heinrich Heine Universität Düsseldorf. Participants volunteered or received 8 DM (U.S. \$4) for taking part; all reported normal or corrected-to-normal vision. None had participated in either Experiment 1 or Experiment 2.

**Materials.** Experiment 3 used the same materials as Experiment 1.

**Procedure.** The procedure was the same as in Experiment 2, with the following exception. Participants were instructed to produce an indefinite noun phrase in response to the picture, such as “ein Leopard” (a-masculine/neuter leopard-masculine), “ein Pferd” (a-masculine/neuter horse-neuter), or “eine Ente” (a-feminine duck-feminine). As in Experiment 2, the initial presentation rate (display time) was 800 ms for each picture, again with a 200-ms ISI before the next picture. In this experiment, final presentation rates ranged from 600 ms to 1,000 ms ( $M = 783$  ms), with none of the participants requiring more than one block of practice trials.

**Scoring criteria.** Scoring criteria were the same as in Experiment 2.

**Data analysis.** The data analysis was the same as in Experiment 2. An additional analysis was carried out considering only masculine and neuter targets and errors. In this latter analysis, we focused on word pairs that shared the same determiner but differed in gender; this analysis allowed us to determine whether gender contributes to the likelihood of an error in these cases for which the determiner is the same.

### Results

Participants’ responses were again largely correct: 10,858 correct responses in the 11,844 trials (a rate of 91.6% correct). Two hundred (1.7%) within-set lexical errors occurred, along with 271 (2.3%) omissions, 162 (1.4%) corrections, 296 (2.5%) determiner errors, and 57 (0.5%) miscellaneous errors (including 5 lexical errors from outside the response set). Among determiner errors, 177 were noticeable pauses between determiner and the target noun, 70 were stutters or dysfluencies occurring during determiner production, 34 were corrections from an incorrect determiner to the correct one, and 15 were mismatches between the determiner and the target word. None of the determiner errors also involved lexical errors. The 200 within-set lexical errors were entered into confusion matrices (separately for animals and body parts) for further analysis. Table 4 presents the numbers of errors for each semantic field and each gender.

First, as in the previous two experiments, a logistic regression analysis was performed; presence or absence of errors was the response variable, and semantic and form similarity, followed by grammatical gender, were predictors. As in Experiments 1 and 2, semantic similarity proved to be a significant predictor of errors ( $R = -.2011$ , Wald = 15.37,  $p = .0001$ ). Form similarity was significant as well, as in Experiment 2 ( $R = .1424$ , Wald = 8.71,  $p = .0032$ ). Grammatical gender also predicted errors ( $R = .1340$ , Wald = 7.94,  $p = .0048$ ).

In a second analysis, feminine nouns were eliminated from the confusion matrices, and the same analyses were performed only on masculine and neuter nouns, which share the same determiner but differ in gender. As in all previous analyses, semantic similarity strongly predicted the presence of errors ( $R = -.2310$ , Wald = 10.89,  $p = .0010$ ). Form similarity was also a strong predictor ( $R = .3073$ , Wald = 17.73,  $p < .0001$ ). Grammatical gender, on the other hand, was no longer significant ( $R = .0401$ , Wald = 2.44,  $p = .1183$ ). In comparable analyses, one excluding masculine nouns and another excluding neuter nouns, significant effects of gender were observed (both  $R$ s  $> .25$ ,  $ps < .001$ ).

Table 4  
*Gender Breakdown of Errors for Words of a Given Gender:  
 Experiment 3 (Indefinite Determiner)*

Target gender	Error gender					
	Feminine		Masculine		Neuter	
	No. of errors	%	No. of errors	%	No. of errors	%
Animals						
Feminine	9	56	3	19	4	25
Masculine	6	11	35	63	15	27
Neuter	5	13	11	29	22	58
Body parts						
Feminine	14	74	4	21	1	5
Masculine	5	11	24	53	16	36
Neuter	2	8	11	42	13	50
Overall						
Feminine	23	66	7	20	5	14
Masculine	11	11	59	58	31	31
Neuter	7	11	22	34	35	55

*Note.* Percentages indicate distribution of errors for targets of a particular gender.

To minimize the possibility that the lack of a grammatical gender effect from the masculine–neuter set of words was simply due to a lack of power to detect such an effect (given that this analysis was conducted on such a limited data set), we also carried out a parallel analysis of the results from Experiment 2 (considering only masculine and neuter nouns). The pattern of results was largely the same as in the analysis of the full data set in Experiment 2. Meaning similarity ( $R = -.2435$ , Wald = 12.18,  $p = .0005$ ) and form similarity ( $R = .2508$ , Wald = 12.80,  $p = .0003$ ) were both significant predictors, and, crucially, grammatical gender still significantly predicted the presence of errors among masculine and neuter nouns ( $R = .2732$ , Wald = 14.81,  $p = .0001$ ), whereas it did not in Experiment 3.

### Discussion

The main finding of this final experiment is that gender preservation was observed only when considering target–intruder pairs for which the determiner differed (“ein” for one word, “eine” for the other); it was not observed when the gender differed but the determiner did not (masculine/neuter words, both of which use the determiner “ein”). Hence, these findings argue against an account according to which the gender preservation effect arises solely as a consequence of feedback from gender-marked syntactic frames to lexical selection processes.

### General Discussion

The results of these three experiments can be summarized as follows. For semantic substitutions, target and intruding words tend to share the same gender to a higher degree than predicted on the basis of cross-linguistic semantic similarity only when the noun is uttered in a gender-marked context. Furthermore, such a tendency to preserve the target gender depends on whether speakers are producing noun phrases in which the determiners differ between genders (Experiment 2 for all genders, Experiment 3

when one noun was feminine and the other either masculine or neuter). The tendency to preserve a language-specific target gender is not significant, however, when speakers produce bare nouns (Experiment 1) or when the form of the determiner is the same for nouns of different genders (Experiment 3 for masculine and neuter nouns).

### Gender Preservation and Models of Language Production

The vast majority of studies in the language production literature have addressed the question of how the retrieval of meaning and word form is orchestrated; only a few studies have addressed the interface between retrieving semantic and lexico-syntactic information. Our results, therefore, provide some novel constraints on the modeling of this interface. We assessed the mechanisms that give rise to syntactic preservation effects on a process (lexical selection) considered to be strictly conceptually driven.

First, the fact that these effects are present in phrasal production suggests that gender-marked frames corresponding to highly activated lemmas are also highly activated and are retrieved, regardless of whether the corresponding lemma is selected for production or not. This is because the observed gender preservation effect on semantic substitution reflects an influence of the gender of the target (a highly activated competitor never selected for production) on the selection of a lexical intruder. This finding is difficult to accommodate in WEAVER<sup>++</sup> (Levelt et al., 1999). In this theory, an active lemma coactivates its gender node (more generally, its syntactic frame) if the task requires its specification. At the same time, the theory allows only for one-way activation spreading from the lemma to its syntactic frame; thus, the gender-marked syntactic frame cannot be retrieved before lexical selection is completed. This fact excludes a simple spreading activation account of the gender preservation effect we observed in our experiments. To account for this effect, the frame must be not only activated but also selected before its lemma is ultimately selected. Namely, the target induces the frame, but an active competitor is selected and inserted in the frame, a process incompatible with the normal functioning of WEAVER<sup>++</sup>. If WEAVER<sup>++</sup> is to account for these results, the verification process from lexical concepts to lemmas must fail, leading to double lemma selection and retrieval of the corresponding gender-marked frames. Under such conditions of derailment, the discreteness assumption of WEAVER<sup>++</sup> is violated (see Vigliocco & Hartsuiker, 2002, for a discussion of the implications of double lemma selection).

Second, the results of our experiments converge with findings from studies of syntactic effects on naming latencies. It has been reported that gender congruency effects arise when phrases but not single words are produced (at least in some languages and if the determiner is gender marked; La Heij et al., 1998; Schriefers, 1993; van Berkum, 1997). An absence of grammatical class effects when speakers produce bare nouns but the presence of grammatical class effects (in this case, greater interference from a word sharing the same grammatical class) when they produce phrases has also recently been reported (Pechmann & Zerbst, 2002). These findings suggest that lexico-syntactic information is tightly linked to frame building processes and is not necessarily automatically retrieved along with the corresponding lemma. Here our study goes beyond this previous work because, first, it shows a difference between bare noun and phrasal production on errors (instead

of reaction times), thus in a different domain and with a very different methodology. Second, in contrast to previous studies, we were able to tease apart semantic and morpho-syntactic contributions to gender preservation. The studies just mentioned did not assess the role of lexico-semantics in gender and grammatical class congruency effects.

Finally, the contrasting patterns of results observed in Experiments 2 and 3 (no additional gender preservation effect was observed in Experiment 3 when the effect of determiner form was partialled out) run counter to an account in terms of feedback solely from a syntactic (i.e., not yet morphophonologically spelled out) frame to the lexical selection process. In this respect, the current results converge with the results of Vigliocco, Lauer, Damian, and Levelt (2002), who showed that semantic interference effects and gender congruency effects in adjective–noun production in Dutch do not interact. An interaction would have been expected if the process of gender-marked syntactic frame retrieval–building were to feed back to lexical selection processes (see Vigliocco, Lauer, et al., 2002, for a detailed discussion). Thus, these results speak against views that assume biasing effects on lexical selection solely from frames that are syntactically but not morphophonologically specified, along the lines of one reading of spreading activation models outlined in the introduction (Dell, 1986; Dell et al., 1997; Stemberger, 1985).

The results of Experiments 2 and 3 converge with and extend studies investigating gender congruency effects on naming latencies that have shown that gender congruency effects appear only when speakers produce a gender-marked determiner in German (Schiller & Caramazza, 2003; Schriefers et al., 2002). We extended these previous studies in that we examined a different case in which the determiner is not marked for gender in the form: indefinite determiners. This is important for the following reasons. In German, the gender of a noun is never necessary for agreement (whether within or between clauses) when nouns are plural. Therefore, it may be questioned whether the gender feature of nouns is retrieved in the plural form at all. This is not the case for singular nouns. Even when an indefinite determiner is used, the gender of the noun is still needed for agreement in some cases, such as gender marking of adjectives and pronoun selection. These results can be accounted for by allowing morphophonological information to play a role.

Taking morphophonological marking into account, gender preservation effects could arise in different manners. In an interactive activation system, they could arise if feedback to the lexical selection process is assumed to occur from morphophonological frames. However, such effects can also arise in a system that allows for activation and retrieval of more than one gender-marked frame but no feedback: Semantic substitution errors occur when lexical selection is highly competitive, and hence selection of a winning competitor is somewhat delayed. Each competitor would activate the corresponding syntactic frames, and, at least on some occasions (e.g., when lexical selection is particularly hindered by high competition), a gender-marked syntactic frame would be selected for production before lexical selection is completed. By the time the lexical selection process is finally completed, the selected lemma would be inserted into the gender-marked syntactic frame that has been independently selected (or built). In most cases, the independently retrieved frame would correspond to the syntactic frame for the target lexical entry (which we assume

would always be a highly activated competitor). In some cases, when the intruder and the frame share the same syntactic properties, the resulting morphophonological spell-out will be well formed; in other cases (when they do not share the same syntactic properties), it will not. Phrases that contain a semantic error and are also syntactically and morphophonologically ill formed would then be more likely to be identified and amended by a monitoring system than phrases that contain a semantic error but nonetheless are syntactically well formed. In this scenario, a monitoring system would be responsible for the emergence of the gender preservation effect. This alternative mechanism is illustrated in Figure 2.

Why would frames be retrieved (or built) before lexical selection is completed, if this process can lead to errors? One possible reason is that it may ensure fluency. We have characterized the lexical selection process as competitive, a claim supported by speech error data and by chronometric studies (e.g., Garrett, 1984,

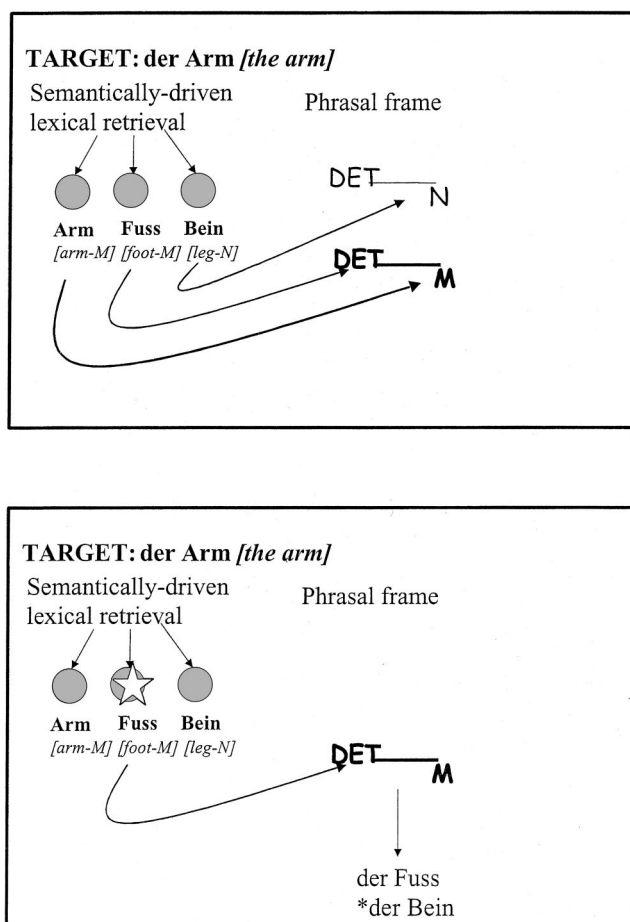


Figure 2. How the gender preservation effect might arise as a result of monitoring at the morphophonological level. There are two steps in this account. Competitor lemmas activate their corresponding gender-marked syntactic frames, and the most active frame is retrieved (top). The most active lemma is selected (indicated by a star) and inserted into the frame for morphophonological encoding (bottom). At this stage, frames that are ill formed are more likely to be spotted and aborted by a monitor (either internal or mediated through the comprehension system) than frames that instead are well formed. M = masculine; N = neuter.

1992; Levelt et al., 1999). Competition may take time, especially in cases involving a number of activated lexical candidates (as in the current experiments, in which speakers were repeatedly producing a set of semantically related words); hence, having a frame ready to be filled may prevent unwanted delays. The rarity of erroneous lexical selection (even in our experiments designed to induce errors, speakers named the pictures without errors of any kind on more than 90% of all trials) combined with monitoring processes would prevent such a mechanism from delivering erroneous utterances, except on very rare occasions.

Are the feedback and the monitoring accounts of the present data necessarily in conflict? It depends on how the monitoring processes are characterized. Our data are certainly compatible with views that do not embed feedback and in which the monitoring processes are solely carried out through the comprehension system (as originally proposed by Levelt, 1989). However, they are also compatible with models in which monitoring is production internal. In the literature on monitoring, a number of alternative proposals to comprehension-based monitoring have been advanced (see Postma, 2000, for an overview). These proposals either assume production-internal monitoring mechanisms (e.g., Berg, 1986; Laver, 1980; Postma, 2000; Postma & Kolk, 1993; Van Wijk & Kempen, 1987) or monitors based on a network partially subserving both production and comprehension (MacKay, 1992). A commonality among a number of these proposals is that the mechanism for monitoring is based on feedback (e.g., Postma & Kolk, 1993). In particular, a code that is selected at a given level will receive feedback from codes at a lower level. The amount of feedback received is indicative of whether the correct codes have been selected. This information can then be used by a (local) monitoring mechanism to detect derailments in processing and subsequently initiate corrective measures. In our specific case, feedback could occur from processes involved in developing a morphophonological representation of the to-be-uttered sentence to processes involved in building a syntactic frame (and, possibly, lexical selection). This view is not in conflict with the feedback account described here.

Finally, it is important to note that, in our description of feedback and monitoring views, we have retained the assumption that frames specified for syntactic gender but unspecified for morphophonological form are built during production. However, to account for the results of Experiment 2 and 3, we might dispense entirely with the notion of syntactically marked frames. Because gender preservation effects were observed only when gender was morphophonologically marked in the determiner, an alternative account based solely on form properties may be put forward. Namely, the gender preservation effect could be purely based on the presence of form overlap in the determiner. In this account, it would be more likely to replace “der Hund” (the dog, masculine) with “der Fuchs” (the fox, masculine) than with “das Pferd” (the horse, neuter), because the produced noun phrase would have a greater form overlap when the determiner is preserved than when it is not. In this alternative account, the results of Experiments 1–3 could be described in terms of determiner (not gender) preservation. Although such an account is sufficient for the data at hand, we prefer to maintain a view in which gender participates in syntactic-level processes, because this information must be used to compute long-distance dependencies such as noun-anaphoric pro-

noun agreement. Long-distance dependencies of this type appear to be difficult to account for purely in terms of form.

### *Semantic Representations and Grammatical Gender*

In the introduction, we discussed the implications of a semantic account of language-specific gender preservation in semantic substitutions for linguistic relativity, reviewing studies that have shown effects of grammatical gender on the mental representation of object concepts by speakers of languages such as Spanish and German (Boroditsky et al., 2003; Konishi, 1993; Sera et al., 1994). In our Experiment 1, we failed to observe language-specific gender effects, a finding that suggests that grammatical gender does not affect German speakers’ thinking for speaking. Differences in thinking for speaking are, however, a prerequisite to any stronger version of linguistic relativity according to which effects of language-specific properties can be observed in nonlinguistic tasks.

Thus, we suggest that language-specific properties such as grammatical gender are not part of the lexico-semantic representations of German speakers. This suggestion is based on our use of an on-line task, combined with analyses that allowed us to separate those effects of gender that are cross linguistic (and therefore have a conceptual foundation) from those effects of gender that are language specific. Both lexico-semantic and lexico-syntactic information are, however, automatically recruited when sentences are produced or heard.

This does not imply that gender, when conceptually grounded (i.e., biological gender), may not affect thinking for speaking in German or other languages. For example, in English the word *friend* does not carry information about whether the friend is male or female, whereas in Spanish the corresponding word is differentially inflected for a man (“amigo”) or a woman (“amiga”). In English, adjectives used as predicates (e.g., “tall” in “Luis’s friend is tall”) do not agree in gender with the noun; in Spanish, they do (e.g., “El amigo de Luis es alto” [the–masculine friend–masculine of Luis is tall–masculine] or “La amiga de Luis es alta” [the–feminine friend–feminine of Luis is tall–feminine]). Thus, in these two languages, conceptual information concerning natural gender is obligatorily conveyed by a sentence in Spanish but is optional in English, and therefore the necessity of attention to natural gender differs between the languages (see Levelt, 1989, for a similar discussion concerning deictic proximity and tense).

Finally, our findings may not generalize to other lexico-syntactic properties and to abstract words. As described in the introduction, grammatical gender is quite arbitrarily linked to the meanings of most words, but the situation is quite different for other lexico-syntactic properties that, instead, have a higher degree of conceptual motivation, such as grammatical class. As discussed earlier, grammatical class also constrains semantic substitution errors in spontaneous speech. Whereas parsimony would call for an account of grammatical class congruency along the same lines we have proposed here for the grammatical gender preservation effect, such an account is not warranted on the grounds of the available evidence. Why would grammatical class and grammatical gender behave differently? In the case of grammatical class, in contrast to grammatical gender, it seems quite plausible that children learn to distinguish among nouns, verbs, and adjectives on the basis of their appreciation of the distinction among objects, ac-

tions, and properties and later use their knowledge of grammatical class to infer semantic properties of newly learned words (Bloom, 1994; Fisher, 1994). These processes, occurring during language development, could lead to a greater degree of interdependency between semantic and syntactic information in the adult system. As argued by Boroditsky (2003), an association of masculine and feminine gender of nouns referring to objects and abstract entities with prototypical male and female properties could help children learn the arbitrary syntactic gender; however, such a process would not allow for bootstrapping from syntax to meaning (and vice versa) in the same manner as it could, for grammatical class.

### Conclusion

The present study addressed alternative accounts of the gender preservation effect on semantic substitution errors observed in spontaneous speech. First, we found that the congruency effect does not reflect greater semantic similarity among words sharing the same gender. This finding provides us with an on-line test of the idea according to which language-specific properties of nouns, such as gender, would affect people's mental representations of the corresponding entities. Second, the gender preservation effect arises in experimental conditions in which speakers are engaged in producing gender-marked phrases. This finding suggests that the retrieval–building of sentential frames does not need to wait until lexical selection is completed, in contrast to serial feed-forward models of lexical retrieval during production. Finally, we found that the gender preservation effect arises only when speakers produce phrases morphophonologically marked for gender. Thus, feedback from a syntactically but not morphophonologically specified frame to the lexical selection process is not the sole determinant of the congruency effect; instead, morphophonological information must play an important role.

### References

- Bates, E. A., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). Cambridge, England: Cambridge University Press.
- Berg, T. (1986). The aftermath of error occurrence: Psycholinguistic evidence from cutoffs. *Language and Communication*, 6, 195–213.
- Bloom, P. (1994). Possible names: The role of syntax-semantic mappings in the acquisition of nominals. *Lingua*, 92, 297–329.
- Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.
- Boroditsky, L., & Schmidt, L. (2000). Sex, syntax and semantics. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 42–46). New York: Elsevier.
- Clark, H. H. (1996). Communities, commonalities and communication. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 324–359). Cambridge, England: Cambridge University Press.
- Damian, M. F., & Martin, R. C. (1998). Is visual lexical access based on phonological codes? Evidence from a picture-word interference task. *Psychonomic Bulletin & Review*, 5, 91–95.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- 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.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 801–837.
- del Viso, S. D., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161–185.
- Fisher, C. (1994). Structure and meaning in the verb lexicon: Input for a syntax-aided verb learning procedure. *Cognitive Psychology*, 5, 473–517.
- Foundalis, H. E. (2002). Evolution of gender in Indo-European languages. In W. Gray & C. Shunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 304–306). New York: Elsevier.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). London: Academic Press.
- 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.
- 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 Gegenwartsprache. [The cognitive organization of gender assignment for one-syllable nouns in modern German]. *Zeitschrift für germanistische Linguistik*, 11, 166–182.
- La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender congruency effect in picture-word tasks. *Psychological Research*, 61, 209–219.
- Langacker, R. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Laver, J. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 287–305). New York: Academic Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- 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.
- MacKay, D. G. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition*, 1, 199–225.
- Marx, E. (1999). Gender processing in speech production: Evidence from German speech errors. *Journal of Psycholinguistic Research*, 28, 601–621.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago: University of Illinois Press.
- Pechmann, T., & Zerbst, D. (2002). The activation of word class information during speech production. *Journal of Experimental Psychology*, 28, 233–243.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97–131.
- Postma, A., & Kolk, H. H. J. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, 36, 472–487.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Sapir, E. (1921). *Language*. New York: Harcourt, Brace, & World.

- Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, 48, 169–194.
- Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841–850.
- Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 941–950.
- Sera, M., Berge, C., & del Castillo-Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, 9, 261–292.
- Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodriguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, 131, 377–397.
- Slobin, D. (1996). From “thought and language” to “thinking for speaking.” In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, England: 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.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.
- van Berkum, J. J. A. (1997). Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition*, 64, 115–152.
- Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, 19, 403–440.
- Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound and syntax in language production. *Psychological Bulletin*, 128, 442–472.
- Vigliocco, G., Lauer, M., Damian, M. F., & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 46–58.
- Vigliocco, G., Vinson, D. P., Damian, M., & Levelt, W. J. M. (2002). Semantic distance effects on object and action naming. *Cognition*, 85, B61–B69.
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (in press). *The meaning of object and action words*. *Cognitive Psychology*.
- Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of noun-verb dissociations in aphasia. *Journal of Neurolinguistics*, 15, 317–351.
- Whorf, B. (1956). *Language, thought, and reality: Selected writing of Benjamin Lee Whorf* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.

## Appendix

Words and English Translations Used  
in the Present Experiments

German	Gender	English
Katze	fem	cat
Kuh	fem	cow
Ente	fem	duck
Giraffe	fem	giraffe
Ziege	fem	goat
Maus	fem	mouse
Eule	fem	owl
Schlange	fem	snake
Bär	masc	bear
Hirsch	masc	deer
Hund	masc	dog
Esel	masc	donkey
Elefant	masc	elephant
Fuchs	masc	fox
Frosch	masc	frog
Leopard	masc	leopard
Löwe	masc	lion
Strauss	masc	ostrich
Pinguin	masc	penguin
Schwan	masc	swan
Tiger	masc	tiger
Kamel	neut	camel
Huhn	neut	chicken
Pferd	neut	horse
Känguruh	neut	kangaroo
Schwein	neut	pig
Schaf	neut	sheep
Zebra	neut	zebra
Hand	fem	hand
Lunge	fem	lungs
Nase	fem	nose
Schulter	fem	shoulder
Zunge	fem	tongue
Arm	masc	arm
Schnabel	masc	beak
Bart	masc	beard
Finger	masc	finger
Fuss	masc	foot
Mund	masc	mouth
Schwanz	masc	tail
Daumen	masc	thumb
Flügel	masc	wing
Gehirn	neut	brain
Ohr	neut	ear
Auge	neut	eye
Horn	neut	horn
Bein	neut	leg

*Note.* fem = feminine; masc = masculine; neut = neuter.

Received September 19, 2002

Revision received August 8, 2003

Accepted September 2, 2003 ■