

# How Persistent are Grammatical Gender Effects? The Case of German and Tamil

Peter Sedlmeier · Arun Tipandjan · Anastasia Jänchen

Published online: 11 February 2015  
© Springer Science+Business Media New York 2015

**Abstract** Does the language we speak shape the way we think? The present research concentrated on the impact of grammatical gender on cognition and examined the persistence of the grammatical gender effect by (a) concentrating on German, a three-gendered language, for which previous results have been inconsistent, (b) statistically controlling for common alternative explanations, (c) employing three tasks that differed in how closely they are associated with grammatical gender, and (d) using Tamil, a nongendered language, as a baseline for comparison. We found a substantial grammatical gender effect for two commonly used tasks, even when alternative explanations were statistically controlled for. However, there was basically no effect for a task that was only very loosely connected to grammatical gender (similarity rating of word pairs). In contrast to previous studies that found effects of the German and Spanish grammatical gender in English (a nongendered language), our study did not produce such effects for Tamil, again after controlling for alternative explanations, which can be taken as additional evidence for the existence of a purely linguistic grammatical gender effect. These results indicate that general grammatical gender effects exist but that the size of these effects may be limited and their range restricted.

**Keywords** Grammatical gender · German · Tamil · Extralinguistic explanations

## Introduction

Does the language we speak shape the way we think? [Whorf \(1956\)](#) indeed argued that the perception of the world is determined by language: “We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated” (p. v). Since then, this *principle of linguistic relativity*, often also termed the *Whorfian hypothesis* or the *Sapir–Whorf hypothesis*, has been the subject of hot debate.

---

P. Sedlmeier (✉) · A. Tipandjan · A. Jänchen  
Department of Psychology, Chemnitz University of Technology, 09107 Chemnitz, Germany  
e-mail: peter.sedlmeier@psychologie.tu-chemnitz.de

There exists widespread evidence that language can have a systematic effect on people's encoding and understanding of space, time, shapes, substances, numbers, and events (e.g., Boroditsky 2001; Bowerman 1996; Brown and Lenneberg 1954; Carroll and Casagrande 1958; Dehaene et al. 1999; Ervin 1962; Gelman and Gallistel 1978; Gentner and Imai 1997; Levinson 1996; Lucy 1992; Luria 1961; Martinez and Shatz 1996; Miller et al. 1995; Whorf 1940). However, the conclusion of linguistic relativity has also been regularly challenged (e.g., Barner et al. 2009; Bellugi et al. 1991; Brysbaert and Fias 1998; Chen 2007; Fodor 1983; Furth 1966; Heider 1972; January and Kako 2007; Karmiloff-Smith 1979; Li and Gleitman 2002; Malt et al. 2003; Soja et al. 1991; Takano 1989).

At the moment, it is difficult to make a broad judgment about the effects of language on thought but there seems to be a consensus that the Whorfian hypothesis cannot be upheld in its strong sense, that is, that language *controls* both thought and perception. In their comprehensive review, Hunt and Agnoli (1991) concluded that a weak version of the hypothesis—that language *influences* thought—*can* be upheld because the evidence available shows that “different languages pose different challenges for cognition and provide differential support to cognition” (p. 387). And some 12 years later, Boroditsky (2003) concluded in her summary of the literature that “linguistic processes are pervasive in most fundamental domains of thought” (p. 920). For almost all of these domains it is, however, still far from clear what exactly the effects of language on cognition are, under which conditions they arise, and how they can be explained. Such explanations might depend very much on the properties of language considered and on the specific effects examined. In the present studies, we intended to analyze one of the recently studied topics in more detail: the impact of grammatical gender on mental representations.

## The Impact of Grammatical Gender

When students learn a new language, one thing they usually note early on is that many nouns in the new language do not have the same gender as in the original language. For instance, in Spanish, the sun is masculine and the moon feminine (*el sol* and *la luna*), and in German it is the reverse (*die Sonne* and *der Mond*). And speakers of genderless languages such as English, for instance, might wonder why nouns need to have a gender at all. Let Mark Twain (1880) speak on this issue, regarding the German language:

Every noun has a gender, and there is no sense or system in the distribution; so the gender of each must be learned separately and by heart. There is no other way. To do this one has to have a memory like a memorandum-book. In German, a young lady has no sex, while a turnip has. Think what overwrought reverence that shows for the turnip, and what callous disrespect for the girl. (p. 607)

In German, Twain continued, body parts have one of three genders, and he concluded that therefore, a male, thinking about the different grammatical genders of his body parts, may find out that “in sober truth he is a most ridiculous mixture” (p. 607). Does, as Twain suggests, the grammatical gender really have implications for our understanding of the world?

As already mentioned, such effects have indeed been identified, using different languages and methods: In a number of studies participants' judgments about nouns were strongly influenced by the grammatical gender of these nouns. For instance, a recent study found that the grammatical gender of pseudonouns used to denote (unknown) musical instruments

significantly affected the mental representations about these instruments (Vuksanović et al. 2014; see below for more examples).

It is, however, still unclear how persistent grammatical gender effects are, and there remain several open questions. We were especially interested in three of them here. First, it has been suggested that grammatical gender effects might be confined to two-gendered languages (such as Italian) and be absent in languages that have three or more genders (e.g., Kousta et al. 2008). We therefore examined whether reliable grammatical gender effects could also be found for German, a language with three genders. Our second question concerned alternative explanations for the effects found. In the literature, several such explanations have been put forward but most studies did not examine the impact of the respective variables and did not control for them. In the present studies, we wanted to examine the strength of these alternative explanations by testing them on Tamil, a nongendered language, and we wanted to find out whether grammatical gender effects can still be found in a three-gendered language (German) *after* statistically controlling for alternative explanations. Finally, we were interested in examining the generality of the grammatical gender effect by using tasks that differ in respect to how closely they are associated with grammatical gender. We now briefly summarize the evidence relating to the three questions and then present the results of two studies, one with German and the other with Tamil participants.

### Can the Grammatical Gender Effect Also be Found for Three-Gendered Languages such as German?

It seems that the first systematic study on the impact of grammatical gender on cognition was conducted by Konishi (1993), who compared the reactions of speakers of German and (Mexican) Spanish. He used two lists of words, one that contained words that were masculine in German but feminine in Spanish and another list with the reverse genders. Participants were to rate every word on the two lists on several semantic differential scales. Effects were found on a “potency scale” that consisted of four dimensions—*weak–strong*, *small–big*, *dwarf–giant*, and *minor–major*—with the right sides of the dimensions (e.g., *strong*) standing for higher “potency.” Differences between the mean potency ratings for masculine and feminine words were found for speakers of both German (with three genders) and Spanish (with two genders), although the effects (calculated from the respective *F*-values; see Sedlmeier and Renkewitz 2013) were somewhat smaller for the former ( $d = .29$ ) than the latter ( $d = .41$ ; p. 529).

An impact of the grammatical gender for speakers of German was also found by Boroditsky and Schmidt (2000, Exp. 2), using a different task. In this study, native speakers of Spanish and German had to learn a set of 24 object names (half grammatically masculine in Spanish and feminine in German, and the other half the reverse) that were paired with first names (e.g., Patricia, Patrick). Participants later had to remember whether a given object had been paired with a male or a female name. They found that both Spanish and German speakers’ memories were better for gender-consistent pairings but did not report the exact size of the effects for each language. Similar results were found in a third study, again using a different method (Boroditsky et al. 2003). This time, Spanish and German speakers were to write down the first three adjectives that came to mind for 24 object names that all had inconsistent grammatical gender in Spanish and German. The study was conducted in English and native English speakers (unaware of the purpose of the study) had to rate the adjectives as masculine or feminine. Again, Spanish and German speakers differed in the associations elicited by the 24 object names. For instance, whereas German speakers described a key (masculine in German: *der Schlüssel*) as *hard*, *heavy*, *jagged*, or *metal*, Spanish speakers (feminine in Spanish: *la*

*llave*) used adjectives like *golden*, *intricate*, *little*, or *lovely*. Again, no numerical description of the effects is given.

Contrasting results were obtained by [Sera et al. \(2002\)](#) who asked their participants to assign either a male or a female voice to pictured objects. These authors found pronounced effects of grammatical gender for speakers of the two-gendered languages Spanish and French but only a small trend for speakers of the three-gendered German language. However, the German sample only consisted of children up to 9 years of age whereas the data for the speakers of other languages also included adults, for whom the effects of grammatical gender tended to be slightly higher. Note also that the task of assigning male and female voices to pictured objects does not allow a clear mapping between grammatical gender and voice for the German neuter, because there is no underlying continuous dimension as might be postulated for potency ratings or femininity–masculinity judgments. Using a different task, [Vigliocco et al. \(2005\)](#) obtained similar results. They had their participants perform “triadic similarity judgments,” in which they had to choose the two words out of three that were most similar in meaning and delete the odd one. The dependent variable was the proportion of same-gender word pairs selected as similar. They found that speakers of the two-gendered Italian language were affected by grammatical gender, although only by words with a natural gender (animal names) and not by words for artifacts. Speakers of the three-gendered German language were not affected by grammatical gender in this task. [Vigliocco et al. \(2005\)](#) explained their results by the *sex and gender hypothesis*. This hypothesis holds that associations between the gender of words and male and female features are first learned for nouns referring to humans and are then extended to other sexuated entities (animals). The sex and gender hypothesis predicts that a correspondence between gender and male or female features should be found in languages that allow for easy mapping between the gender of nouns and sex. Such a mapping is generally easier in two-gendered languages such as Italian than in languages with more than two genders, such as German. This is what [Koch et al. \(2007\)](#) examined. They compared Spanish and German speakers and used both the memory task introduced by [Boroditsky and Schmidt \(2000\)](#), see above) and the potency scale task. In the memory task, the advantage for gender-consistent pairings amounted to an effect of  $d = .67$  for speakers of Spanish and to a nonsignificant effect of  $d = .25$  for speakers of German. For the potency scale, however, even a slightly negative effect was found for the German participants (masculine words being judged as less “strong” than feminine words).

Overall, the results are not consistent. Looking only at  $p$ -levels, the evidence seems to favor a noneffect for the German gender, but an inspection of the effect sizes (calculated from the test results and reported above) indicates at least a trend effect also for the German grammatical gender. In the present studies, we conducted a strict test of the hypothesis that three-gendered languages also show a grammatical gender effect by looking at tasks of (presumably) different sensibilities to detect grammatical gender effects and by controlling for alternative explanations, described in the next section.

#### Alternative Explanations for Grammatical Gender Effects

In the literature on grammatical gender effects, several alternative explanations have been put forward for the covariation of various effects with grammatical gender. Most prominent is the hypothesis that the Spanish language (and maybe other Romance languages, as well) has a special status in that its grammatical gender might covary with features of objects that have feminine or masculine connotations (e.g., [Sera et al. 2002](#)).

### *Is the Spanish Grammatical Gender Universal?*

Sera and Berge (1994) seem to have been the first to note that speakers of English, a nongendered language, tended to attribute male and female voices to inanimate objects in a way that was consistent with Spanish grammatical gender. Could it be that the Spanish grammatical gender (or the gender in other Romance languages) is universal in the sense that Spanish grammatical gender assignments are highly correlated with features of objects that are universally perceived as male or female? In a later study, Sera et al. (2002) again found a strong impact of the Spanish gender on the assignment of male or female voices to grammatically masculine and feminine words for speakers of English and also for speakers of German, albeit to a smaller extent for the latter. Moreover, it turned out that for speakers of German, the impact of the Spanish grammatical gender on their judgments was even slightly stronger than that of the German grammatical gender (compare their Tables 12 and 11). A similar result was obtained by Koch et al. (2007). Their German participants judged the words that had masculine gender in Spanish but feminine in German as higher in potency than those that had feminine gender in Spanish but masculine in German ( $d = .51$ ). Sera et al. (2002) explained this impact of the Spanish gender as being due to the transparent mapping between grammatical gender and biological gender of human referents, that is, the special male and female features associated with the gender in Spanish.

### *Femininity Score*

A “femininity score” might at least partly explain what could make the Spanish grammatical gender universal (Leinbach et al. 1997; Mullen 1990; Sera et al. 2002): the gender in Spanish can be well predicted by such a score that can be determined from the values of four two-valued factors, with the first value referring to male and the second referring to female qualities: (a) artifact versus natural kind, (b) angular versus curved, (c) used typically by males versus by females, and (d) dense or not dense. For instance, in this study, a spoon (feminine in Spanish) was categorized as artifact (male-like), curved (female-like), used typically by females, and dense (male-like).

### *Other Potential Explanations*

The properties of the entities that nouns stand for, for instance, whether these entities are animate or whether they have more features that are generally considered masculine or feminine, might also produce effects that covary with gender effects.

### *Additional Evidence for Alternative Explanations*

Again, the evidence is not unequivocal. Boroditsky and Schmidt (2000, Exp. 1) presented their English participants a list of 50 animal names and another list of 85 names of artifacts and asked them to classify each object or animal on the list as either masculine or feminine. They found that English speakers’ judgments correlated substantially with Spanish but even higher with German grammatical gender for the animal list ( $r = .29$ , and  $r = .43$ , respectively) but less strongly for the list of artifacts ( $r = .04$ , and  $r = .11$ , respectively). So at least the results in the latter study indicate that the German language might also have some universal properties. Such properties can be studied best in nongendered languages. However, an Indo-European language with only one gender, such as English, might not be the best candidate to examine this issue because it has much in common with other Indo-European languages

such as Spanish and German. Therefore Boroditsky et al. (2003, p. 78) suggested using non-Indo-European languages to assess the generality of the findings. In the present studies, we came back to their suggestion and used Tamil, a nongendered Dravidian language, to gain insight on the generality of the alternative explanations mentioned above.

### How General Is the Effect of Grammatical Gender?

We have seen that regarding the first two questions raised above, (a) whether also in languages with more than two grammatical genders, such as German, gender has an impact on mental representations and (b) how much of the grammatical gender effect can be accounted for by alternative explanations, the results vary. This variation might be due to sampling error, but it might also be due to the quite diverse tasks used to assess the impact of grammatical gender on word connotations. Whether different tasks measure the same contents can be most easily determined if different tasks are used by the same sample. Apparently, this has been done in only one study so far. Koch et al. (2007) had their participants first perform the memory task (congruent and incongruent pairings of object names and first names) and then make judgments on the potency scale. They found a trend for the impact of the German grammatical system with the former but not the latter method. This might be seen as evidence that the methods really measured different things. However, there is a potential drawback in their procedure regarding the sequencing of their tasks. In their first task (memory task), half of the female object names were paired with female first names and half with male first names and vice versa for the male object names. In the latter task (potency scale task), the *same object words* were used. In this way, the object names might have been associated with the gender of the first names and thus the possibly quite strong effects of the first names might have had an impact on the later judgments of the potency scale, possibly leading to the null results found there. This potential problem can be overcome either by balancing out the order of the tasks or by using different sets of object names for different tasks.

There is, however, another potential problem when comparing the results of different tasks to examine the impact of three-gendered languages such as German: How can the neuter be mapped onto values of the dependent variable? This is difficult if the dependent variable can assume only dichotomous feminine and masculine values such as in a task that requires participants to associate objects with a male or female voice and in a memory task using male and female first names. Also the task of choosing the two words out of three that are most similar in meaning and deleting the odd one poses methodological problems with three genders. Better suited are tasks that allow judgments on a continuum, such as ratings on a potency scale or the maleness–femaleness ratings of adjectives that have been associated with object names. In such tasks, one would expect ratings for object names that had a neutral gender to be in between the ratings for male and female object names. Moreover, if such tasks differ in how directly they can be connected to grammatical gender, a comparison of the results might be used as an indicator of the stability of the grammatical gender effect. Whereas the potency scale task can be argued to be quite directly connected to grammatical gender, the respective relationship with the adjective rating task (having one group of participants produce adjectives that they associate with the nouns and a different group give femininity–masculinity ratings for these adjectives) is much more indirect. If then grammatical gender effects are of comparable size for these two tasks, this would lend credence to the existence of a stable grammatical gender effect.

However, both tasks might be regarded as having been more or less custom-tailored to detect an effect of grammatical gender on cognition, usually some form of associative response. What happens with less restricted tasks such as judging the overall semantic simi-

ilarity of words? Obviously, the impact of grammatical gender in this task must be less direct because the judgment involves a match between two nouns and not just associative properties of a single noun. For such a task, distinctions such as natural versus artificial or animate versus inanimate might play a stronger role than the grammatical gender. Moreover, contiguities in word use might influence the results in such a relatively unrestricted task (Elman 1990; Landauer and Dumais 1997; Wettler et al. 2005). To see how far grammatical gender effects might generalize across linguistic tasks, we had participants give similarity judgments about pairs of nouns in a list as a third task. All these tasks will be described in more detail below. Before we present the procedures and results used in our studies, we want to clarify how we dealt with some methodological issues that are specific to grammatical gender research.

## General Methodological Issues

Research on the impact of grammatical gender is plagued with several methodological problems that concern differences in languages (as spoken by participants and as used in the respective tasks), different levels of analysis, and the selection of proper stimulus material. In the following we detail how we dealt with these problems.

### Extralinguistic Differences Between Groups of Participants

Usually, the Whorfian hypothesis is studied by comparing the behavior of speakers of different languages. This, of course, does not allow for conducting proper experiments, using randomization of participants. Thus, a first problem of control in studies of this kind is to make participants in each language as comparable as possible. In our studies, we examined speakers of Tamil and German. We tried to minimize extralinguistic differences by concentrating on a group that can be regarded as relatively homogenous, both within and across language communities: students in the social sciences.

### Language Used

A second problem in studies of this kind has to do with measurement. It has been suggested that using a third language (e.g., English, for speakers of German and Spanish) might be advantageous because (a) such a design might be more suitable for studying the impact of having been raised in a particular language on language-independent thought, and because (b) instructions translated into another language might not be understood in exactly the same way (e.g., Boroditsky et al. 2003). However, if participants are not so fluent in the third language or differ in their fluency, other problems may arise. Especially if the structure of the written languages differs substantially, as is the case for the German and Tamil written language examined here, the use of English, which in its written form is close to other Indo-European languages (such as German) might bias responses for speakers of a Dravidian language (such as Tamil). Therefore, we decided to use both German and Tamil written language and to take pains to make all the expressions used have equivalent meanings.

### Two Levels of Analysis

In previous studies, grammatical gender effects have mostly been analyzed using *person-level analyses*: A given person is usually presented several nouns that fall into different grammatical categories, for example, nouns that have either masculine or feminine gender in Spanish. Because the results for single nouns lack reliability, one usually calculates summary values



over the results elicited by a certain category of words (e.g., masculine Spanish nouns) and compares that value to another summary value (e.g., that elicited by feminine Spanish nouns). In general, one calculates one or more summary values per person and usually compares these values *between* groups of persons. Such an analysis allows for statistically controlling person variables but not variables that are connected with the nature of the nouns used. To control for potential covariates of grammatical gender effects such as the “inherent” femininity or masculinity of nouns captured in the femininity score mentioned above, or the animateness and other features associated with the nouns, one needs also an analysis *within* persons. The most elegant solution to the problem seems to be a multilevel analysis that makes it possible to conduct fine-grained analysis of grammatical gender effects within individuals (Level 1) while concurrently controlling for alternative explanations and differences between individuals (Level 2; see [Cohen et al. 2003](#); [Hox 2010](#); [Raudenbush and Bryk 2002](#)). All analyses for our studies were performed with such a model using the software HLM ([Raudenbush et al. 2011](#)).

### Stimulus Selection

The impact of grammatical gender on mental representations has been found to differ dependent on the kinds of stimuli used. Effects are usually much stronger for nouns that denote animals than for those that denote inanimate things (e.g., [Boroditsky and Schmidt 2000](#); [Vigliocco et al. 2005](#)) and they also differ for natural kinds versus artifacts ([Sera and Berge 1994](#); [Sera et al. 2002](#)). Different kinds of stimuli can be taken care of by accounting for these variations in multilevel statistical analysis (see above) but in any case, stimuli have to be carefully selected and the number of words that fall into one of those categories that might influence participants’ ratings beyond the impact of grammatical gender has to be balanced out as far as possible.

For the selection of stimuli, we first collected the nouns that have been used in the previous studies outlined above. However, our stimuli had to meet several constraints. First, to avoid order effects, we needed three equivalent word lists to perform within-person comparisons for our three dependent measures: *potency scale*, *adjective productions*, and *similarity judgments*. Second, to ensure the equivalence of word lists, the joint distributions of several variables had to be as similar as possible. These variables are German gender (male, female, neuter), Spanish gender (male, female), animateness (animate, inanimate), and naturalness (natural kinds vs. artifacts). Tables 1, 2 and 3 show the final result of this complex selection process. All three lists contained eight nouns that had male gender and eight nouns that had female gender in German. Of these, half each had congruent grammatical gender in German and Spanish (e.g., *die Tür* and *la puerta*) and for the other half the grammatical gender was incongruent (e.g., *die Zeitung* und *el periódico*). Each list contained four nouns with neutral German gender. Of the 20 nouns in each list, 6 denominated animate (e.g., cow) and 14 inanimate (e.g., arrow) entities. Seven and 13 of the 20 nouns, respectively, designated artificial (e.g., door) and natural (e.g., tree) kinds. Finally, the words in each list were arranged in random order. Tables 1, 2 and 3 show the resulting lists of nouns. The tables show the words used in the two studies (German and Tamil nouns) and, for reference, the corresponding Spanish and English nouns. The tables also give the grammatical genders for the German and Spanish nouns. Finally, the tables show whether the nouns denote animate or inanimate and natural or artificial entities.



**Table 1** List A. German and Tamil noun forms, and, for reference, translations into Spanish and English

German nouns	Tamil nouns	Spanish nouns	English nouns	German gender	Spanish gender	Animate/inanimate	Natural/artificial
Tür	கதவு	la puerta	door	f	f	inanimate	artificial
Kuh	பசு	la vaca	cow	f	f	animate	natural
Baum	மரம்	el árbol	tree	m	m	animate	natural
Kranich	கொக்கு	la grulla	crane	m	f	animate	natural
Salz	உப்பு	la sal	salt	n	f	inanimate	natural
Bild	ஓவியம்	el cuadro	drawing	n	m	inanimate	artificial
Sonne	சூரியன்	el sol	sun	f	m	inanimate	natural
Ente	வாத்து	el pato	duck	f	m	animate	natural
Kopf	தலை	la cabeza	head	m	f	inanimate	natural
Insel	தீவு	la isla	island	f	f	inanimate	natural
Zeitung	செய்தித்தாள்	el periódico	newspaper	f	m	inanimate	artificial
Fahrrad	மிதிவண்டி	la bicicleta	bicycle	n	f	inanimate	artificial
Wasserfall	நீர்வீழ்ச்சி	la cascada	waterfall	m	f	inanimate	natural
Apfel	ஆப்பிள்	la manzana	apple	m	f	inanimate	natural
Fisch	மீன்	el pez	fish	m	m	animate	natural
Knochen	எலும்பு	el hueso	bone	m	m	inanimate	natural
Geige	வயலின்	el violín	violin	f	m	inanimate	artificial
Kamm	சீப்பு	el peine	comb	m	m	inanimate	artificial
Pferd	குதிரை	el caballo	horse	n	m	animate	natural
Nadel	தையல் ஊசி	la aguja	needle	f	f	inanimate	artificial

The German and Spanish grammatical gender is given for each noun (*f* feminine, *m* masculine, *n* neuter), and the last two columns indicate whether nouns denote animate/inanimate and natural/artificial entities

**Table 2** List B. German and Tamil noun forms, and, for reference, translations into Spanish and English

German nouns	Tamil nouns	Spanish nouns	English nouns	German gender	Spanish gender	Animate/inanimate	Natural/artificial
Blume	பூ	la flor	flower	f	f	animate	natural
Stern	விண்மீன்	la estrella	star	m	f	inanimate	natural
Pfeil	ஆம்பு	la flecha	arrow	m	f	inanimate	artificial
Taube	புறா	la paloma	dove	f	f	animate	natural
Tiger	புலி	el tigre	tiger	m	m	animate	natural
Sandale	மிதியடி	la sandalia	sandal	f	f	inanimate	artificial
Meer	கடல்	el mar	sea	n	m	inanimate	natural
Wind	தென்றல்	el viento	wind	m	m	inanimate	natural
Haus	வீடு	la casa	house	n	f	inanimate	artificial
Regen	மழை	la lluvia	rain	m	f	inanimate	natural
Brücke	பாலம்	el puente	bridge	f	m	inanimate	artificial
Maus	எலி	el ratón	mouse	f	m	animate	natural
Uhr	கடிகாரம்	el reloj	clock	f	m	inanimate	artificial
Giraffe	ஒட்டகச்சிவிங்கி	la jirafa	giraffe	f	f	animate	natural
Bus	பேருந்து	el autobús	bus	m	m	inanimate	artificial
Blatt	ஹூலை	la hoja	leaf	n	f	inanimate	natural
Baumwolle	பஞ்சு	el algodón	cotton	f	m	inanimate	natural
Feuer	நெருப்பு	el fuego	fire	n	m	inanimate	natural
Hund	நாய்	el perro	dog	m	m	animate	natural
Ball	பந்து	la pelota	ball	m	f	inanimate	artificial

The German and Spanish grammatical gender is given for each noun (*f* feminine, *m* masculine, *n* neuter), and the last two columns indicate whether nouns denote animate/inanimate and natural/artificial entities

**Study 1: Grammatical Gender Effects for Speakers of German**

There are three reasons why this study represents a very strict test of the existence and persistence of grammatical gender effects. First, we chose a three-gendered language, German, for which results in previous studies are not as clear-cut as for two-gendered languages, such as Italian or Spanish. Second, we statistically controlled for alternative explanations and explored how the predictive power of these alternative explanations compares to that of grammatical gender. And finally, we used three tasks that differed in their associations with grammatical gender to examine how persistent such effects are.

**Table 3** List C. German and Tamil noun forms, and, for reference, translations into Spanish and English

German nouns	Tamil nouns	Spanish nouns	English nouns	German gender	Spanish gender	Animate/inanimate	Natural/artificial
Schlüssel	சாவி	la llave	key	m	f	inanimate	artificial
Faden	நூள்	el hilo	thread	m	m	inanimate	artificial
Löwe	சிங்கம்	el león	lion	m	m	animate	natural
Badewanne	குளிக்கும் தொட்டி	la bañera	bathtub	f	f	inanimate	artificial
Luft	காற்று	el aire	air	f	m	inanimate	natural
Schaf	ஆடு	la oveja	sheep	n	f	animate	natural
Perle	முத்தும்	la perla	pearl	f	f	inanimate	natural
Berg	பலை	la montaña	mountain	m	f	inanimate	natural
Reh	யான்	el corzo, venado	deer	n	m	animate	natural
Mango	மாம்பழம்	el mango	mango	f	m	inanimate	natural
Mond	நிலவு	la luna	moon	m	f	inanimate	natural
Gabel	முள்கரண்டி	el tenedor	fork	f	m	inanimate	artificial
Wal	சிறா	la ballena	whale	m	f	animate	natural
Stift	பேனா	el aprendiz	pen	m	m	inanimate	artificial
Wüste	பாலவளம்	el desierto	desert	f	m	inanimate	natural
Fuß	கால்	el pie	foot	m	m	inanimate	natural
Buch	புத்தகம்	el libro	book	n	m	inanimate	artificial
Biene	தேள்	la abeja	bee	f	f	animate	natural
Mädchen	சிறுமி	la chica	girl	n	f	animate	natural
Brille	நீக்கூ கண்ணாடி	las gafas	glasses	f	f	inanimate	artificial

The German and Spanish grammatical gender is given for each noun (*f* feminine, *m* masculine, *n* neuter), and the last two columns indicate whether nouns denote animate/inanimate and natural/artificial entities

Method

Participants

One hundred and eight social-science students (82 female, mean age 21.6 years) at the Chemnitz University of Technology, Germany, participated for course credit in classroom settings.

Tasks

All participants performed three tasks that differed in how directly they are associated with grammatical gender: The *potency scale* task, the *adjective production* task, and the *similarity judgment* task served to find out more about the generality of the effect. These three tasks are now described in more detail. For each task, a given participant used a different word list from the three lists shown in Tables 1, 2 and 3.

The instructions for the potency scale task read as follows (all materials originally in German): “Please make a judgment about the following words on a scale from 1 = *weak* to 7 = *strong*. If in doubt just follow your feelings.” Participants then had to rate the 20 nouns in the respective list, which were all listed on one page of the questionnaire. For the adjective production task, participants read the following instructions: “Please give three adjectives that come to mind spontaneously when reading the following words. (For instance, if you see the word ‘war’ the adjectives ‘cruel,’ ‘bad,’ and ‘unpredictable’ might immediately come to mind and if you see ‘art’ you might immediately think ‘nice,’ ‘fanciful,’ and ‘fascinating’).” Then they had to produce three adjectives for each of the 20 nouns in the respective list. For the final data analysis, all adjectives were then collected and put in a list, in random order. This list was given to six independent raters who, without any knowledge about the nouns that had been originally presented, were to rate the adjectives on a femininity–masculinity scale that ranged from 1 (*feminine*) to 7 (*masculine*). From these ratings first average femininity–masculinity ratings were calculated for every adjective (averaged over six raters) and then

the average score was calculated for each originally presented noun as the average score of the three adjectives produced for a given noun, separately for every participant.

For the similarity judgment task, participants received a triangular matrix with 20 columns and 20 lines, for the 20 nouns in the respective list. Participants were to give similarity judgments for all the 190 possible word pairs  $[(20 \times 19) / 2]$ . The instructions for this task read: “In this task, we are interested in how you intuitively rate the similarity between different words. To do this you are to give a *numerical value between 1 and 10* to every word pair per column and line. ‘1’ stands for ‘not at all similar’ and ‘10’ for ‘absolutely similar.’ Please go through the task as quickly as possible and don’t think too much about your ratings.” Then, just to demonstrate what they were expected to do, participants were shown a simple example, in which the similarity between “Word A” and two other words “Word B” and “Word C” was rated as 4 and 7 (the similarity between “Word A” and “Word A” was given to be 10).

### *Alternative Explanations*

In previous studies, described above, it has been found that the kind of noun can make a huge difference. In particular, differences in response sensitivity have been found for nouns that describe animate versus inanimate entities (stronger for animate entities) and words that stand for artificial versus natural kinds (stronger for natural kinds). These two potential explanatory variables were also included in our studies. The values for these variables were determined independently by the three authors and 100% agreement was obtained for both variables (see Tables 1, 2, 3). As mentioned above, the distinction between artificial versus natural kinds has also been included in a score that correlates highly with the Spanish grammatical gender (e.g., Sera et al. 2002). Apart from the distinction (a) artifact versus natural kind (which we left in the score, for reasons of comparability), the score included three other distinctions: (b) angular versus curved, (c) used typically by males versus by females, and (d) dense or not dense (with the left expressions denoting male properties, respectively). We had six independent raters make two-valued ratings of these four items (0 if left and 1 if right expression holds). Thus, each noun received a *femininity score* that could range between 0 (artifact, angular, used typically by males, dense) and 4 (natural kind, curved, used typically by females, not dense). For the analyses, we used the mean femininity scores, averaged over the six raters.

### *Design and Procedure*

To control for order effects, the three tasks (potency scale, adjective productions, and similarity judgments) and the three word lists (see Tables 1, 2, 3) were completely counterbalanced across participants using a Latin square design. Participants were randomly assigned to one of the nine conditions of the Latin square design (12 participants in each condition) and received a booklet with the respective ordering of tasks. They were told that this was a study about the mental representation of words. Before they were to work on the three tasks, they all read the following introductory text: “We want to find out in this study how the meaning of words is mentally represented. Therefore it is important to solve the tasks quickly and intuitively and not to think too long about the answer. There are no right or wrong answers. Of course, your data will be treated anonymously.”

## Results

There was no systematic difference in the responses of male and female participants and therefore, results are combined in the following analyses. To compare the size of effects across studies, we performed multilevel analyses with standardized values (e.g., Hox 2010, p. 21) that correspond to the usual correlational effect size. Throughout, we controlled for level differences between individuals (Level 2). For the analyses of the potency scale task and the adjective production task, we entered the following Level 1 predictors: German grammatical gender, Spanish grammatical gender, femininity score, and the naturalness and animateness of the nouns used as stimuli. The values of the predictors were contrast coded for German grammatical gender (−1, 0, 1) and dummy coded for the Spanish grammatical gender, as well as for the naturalness and animateness of the nouns. For the analysis of the similarity judgment task we used the difference between femininity scores for the two nouns in each pair and the congruency of German grammatical gender, Spanish grammatical gender, and the naturalness and animateness (all dummy coded—e.g., if both nouns had the same grammatical gender or if both nouns denoted artificial entities, this was coded as “1” and if the two nouns differed in grammatical gender or one denoted a natural and the other an artificial entity, this was coded as “0”). The results of all analyses are summarized in Table 4.

### *Is There a Stable Grammatical Gender Effect for German?*

Please recall that such a strict test for the existence of grammatical gender effects has, to the best of our knowledge, never been reported before in the literature. We used a three-gendered language—a case for which previous results are mixed—and, in addition, controlled for alternative explanations. Can we still find a grammatical gender effect? The first line in Table 4 indicates that there is indeed a grammatical gender effect at work even in the German language and even if alternative explanations are statistically controlled for (see also next paragraph). However, the effects remaining are not very strong, despite the very low  $p$  values that are due to the high degrees of freedom obtained with a multilevel analysis. Note that the standardized coefficients we report in Table 4 are equivalent to correlational effect sizes. For the potency task we obtained an effect equivalent to  $r = .17$  and for the adjective ratings one of  $r = .13$ , but for the similarity judgments it was much smaller, equivalent to  $r = .03$ .

Multilevel analyses offer another way to judge effects. The maximum likelihood procedure used in these models produces a statistic called *deviance* that can be used to measure the increase in model fit if variables are added to the model (e.g., Hox 2010, p. 16). The difference in deviance of the models with and without the variables in question has a chi-square distribution with *degrees of freedom* equal to the difference in the number of variables added. For our analyses, we compared the deviance of the models with and without German grammatical gender as a predictor. All the corresponding chi-square tests (all with  $df = 1$ ) are highly significant. For the potency scale we obtained  $\chi^2 = 5,837.36 - 5,786.55 = 50.81$ , for the adjective ratings,  $\chi^2 = 5,715.51 - 5,689.84 = 25.67$ , and for the similarity judgments,  $\chi^2 = 53,917.69 - 53,901.08 = 16.61$  (the critical  $\chi^2$  with  $df = 1$  at  $\alpha = .001$  is 6.64).

### *How Strong are Alternative Explanations?*

Table 4 lists the unique results for four alternative explanations, that is, the results that remain for each of the four explanations when the three other explanations (and the impact of the German grammatical gender) are statistically controlled for. Let us first have a look at the

**Table 4** Differential impact of grammatical gender (German and Spanish) and other potential factors on German participants' responses in the potency scale, the adjective production, and similarity judgment tasks

Predictor	Task									
	Potency scale			Adjective production			Similarity judgment			
	Standardized coefficient	<i>t</i> value ( <i>df</i> = 2047)	<i>p</i> value	Standardized coefficient	<i>t</i> value ( <i>df</i> = 2026)	<i>p</i> value	Standardized coefficient	<i>t</i> value ( <i>df</i> = 20407)	<i>p</i> value	
German gender	0.168	7.698	<0.001	0.128	5.792	<0.001	-0.032	-5.175	<0.001	
Spanish gender	0.068	3.492	<0.001	0.099	5.013	<0.001	0.052	-8.319	<0.001	
Femininity score	-0.436	11.126	<0.001	-0.552	-13.930	<0.001	0.021	2.516	0.012	
Animateness	0.054	2.436	0.015	-0.056	-2.473	0.013	0.060	9.694	<0.001	
Naturalness	0.516	12.924	<0.001	0.262	6.481	<0.001	0.190	22.945	<0.001	

Given are standardized multilevel coefficients and the respective *t* and (two-sided) *p* values for five predictors. Note that for the similarity judgment task, the difference between the two respective femininity scores and the congruency of the other factors is used as predictor (see text for additional explanation)

results for the potency scale and the adjective productions. For both tasks, the Spanish grammatical gender and whether nouns denote animate or inanimate entities did not contribute much to participants' judgments. However, both the femininity scores associated with a given noun and whether these nouns denote natural versus artificial entities apparently influenced participants' judgments more than did the grammatical gender of these nouns. To predict the similarity judgments, apart from the difference in femininity scores of two given nouns, their congruency in respect to the independent variables (e.g., both nouns having the same grammatical gender or not, or both nouns denoting animate entities or not) was used. It turns out that only the congruency in respect to whether both nouns denoted natural or artificial entities made a substantial contribution to participants' similarity ratings.

### *How General is the Grammatical Gender Effect?*

In our study, we used three tasks that can be argued to systematically differ in how closely they are associated with grammatical gender. The potency scale, which can be said to have the closest connection to grammatical gender because the weak–strong continuum is strongly correlated with the feminine–masculine continuum (Konishi 1993), also yielded the largest effect ( $r = .17$ ). Even with the adjective production task, which is not associated with the feminine–masculine continuum in any obvious way, a stable small-to-medium effect size of  $r = .13$  was obtained. It seems, however, that grammatical gender has little or no impact on more unrelated verbal tasks, such as the similarity judgment task used in this study. In this task, grammatical gender cannot work directly on participants' judgments but only through the congruency of the grammatical genders of a word pair.

### Discussion

Study 1 explored the persistence and generality of grammatical gender effects by asking three partly related questions: (a) if there is also a stable grammatical gender effect for three-gendered languages such as German, even if alternative explanations are controlled for, (b) how strong these alternative explanations are compared to the impact of grammatical gender, and (c) how general grammatical gender effects are across tasks that differ in how closely they are associated with grammatical gender. The results in Study 1 can be interpreted as an indication that grammatical gender effects indeed persist even in three-gendered languages, and even if plausible alternative explanations are controlled for. However, although the results for the potency scale and adjective production tasks can be regarded as small to medium, according to Cohen's (1992) standards, the effect for the similarity judgments is very small.

The alternative explanations examined in this study that draw on extralinguistic factors, especially factors that capture the inherent femininity (or masculinity) of entities and the naturalness of entities, predict participants' responses well in tasks that have been devised to measure grammatical gender effects. Somewhat surprisingly, the animateness of entities did not explain much variance in participants' responses. The Spanish grammatical gender might capture part of these extralinguistic factors that are associated with the femininity/masculinity of entities denoted by the nouns used here. However, if these factors are statistically controlled for, as in our analysis, it does not noticeably improve predictions.

Finally, it seems that the grammatical gender effect has some generality that extends to tasks that are only indirectly related to grammatical gender (such as the adjective production task), again even if the effects of alternative explanations are statistically controlled for. However, the effect also clearly has its limits and the result in the similarity judgment task indicates that grammatical gender might not play a systematic role in verbal tasks in general.

## Study 2: Grammatical Gender Effects for Speakers of Tamil

Following Mill's (1872/1973) famous method of agreement and difference, the existence of a causal impact of grammatical gender can only be concluded if grammatical gender effects can be found in languages that have such genders and *not* in languages that do not. But, as already mentioned above, grammatical gender effects for Spanish and German genders have been found in English, a most prominent nongendered language (e.g., Boroditsky and Schmidt 2000; Sera et al. 2002). However, English contains many German words and it is very similar to Spanish in many respects because both languages contain, for instance, many words that stem from Latin. Therefore, English might not be the best candidate to test the “difference hypothesis.” In the present study, we followed Boroditsky et al.'s (2003, p. 78) suggestion to use a non-Indo-European language for this purpose. Tamil, our test-bed, is the most ancient Dravidian language, and it differs fundamentally in many respects from Indo-European languages (see Lehmann 1993; Schiffman 1999; Steever 1987). For instance, Tamil is an agglutinative language (like Turkish) with a case system that is very different from those found in Indo-European languages (Schiffman 2004). What makes it so suitable for the present research is that it (like English) does not have different grammatical genders.

The main question pursued in Study 2 was whether the effects of the Spanish and the German grammatical gender can still be found in Tamil, a nongendered language, if alternative extralinguistic explanations are statistically controlled for. In addition, we again were interested in the predictive strength of these alternative explanations across the three tasks used in Study 1.

### Method

#### *Participants and Tasks*

One hundred and twenty-one social science students (35 female, mean age 20.3 years) were recruited from five colleges in Pondicherry, India. One of the authors (AT) gave a talk on career opportunities in Germany at each college and after the talk, the respective audiences were asked to participate in the current study. Those who agreed were told the purpose of the study and given a booklet and a pen that they could keep afterward. The tasks were identical to those used in Study 1. Ratings for the femininity score were obtained from seven independent Tamil raters and the adjectives produced by the Tamil participants were rated in respect to their perceived femininity–masculinity by six additional native Tamil speakers. All materials were written in Tamil (for the Tamil nouns used see Tables 1, 2 and 3, second column).

#### *Design and Procedure*

The same Latin square design was used as in Study 1. Again, participants were randomly assigned to one of the nine conditions of the Latin squares (12 to 14 participants in each condition). Instructions were all translated from German into Tamil and were the same as in Study 1.

### Results

There were no systematic differences in any of the responses of the male and female participants and therefore, results are combined in the following analyses. The multilevel analyses



for Study 2 were conducted in the same manner as those in Study 1. The results of all analyses are summarized in Table 5.

### *Are There Grammatical Gender Effects for the Spanish and German Genders?*

Let us first have a look at the results for the potency scale and the adjective productions. Table 5 shows that for both German and Spanish grammatical genders taken together, the values of the standardized coefficients only range from -0.02 to 0.03, indicating that the grammatical gender of these two languages had no impact on Tamil speakers' judgments, if alternative explanations are statistically controlled for. This conclusion is corroborated by the fact that despite the huge values for the degrees of freedom, none of the significance tests even approached significance. For the similarity ratings, the effect sizes also are close to 0. The tests did reach significance here but with  $df = 22864$ , which is to be expected for any effect that is not strictly zero.

### *How Strong are Alternative Explanations?*

Are Tamil participants' judgments also influenced by the extralinguistic explanations examined here, that is, by the "femininity" associated with the entities denoted by the nouns used as stimuli and by the animateness and naturalness of these entities? Overall, the answer is "yes" and the pattern looks quite similar to that found for the German participants. For both the potency scale and the adjective productions, the femininity score and the naturalness of the entities in question contribute substantially to the prediction of participants' judgments. For the potency scale, but not for the adjective productions, animateness also has some predictive power. And, again, similar to the result for German participants, only the congruency concerning the naturalness of both nouns in a pair seems to have a substantial impact on participants' similarity judgments.

### Discussion

If the grammatical genders of the Spanish and German languages carry with them both intra- and extralinguistic properties, the latter might be expected to also have an impact on speakers of languages without distinct grammatical genders. If, however, the extralinguistic properties can be controlled for, no such effects should be expected. This is what we found in Study 2. For both the Spanish and the German language, the grammatical gender had no noticeable impact on Tamil participants' responses in the three tasks. In contrast, the extralinguistic properties identified here, that is, the femininity, the naturalness, and (at least in part) the animateness of entities denoted by the nouns used in our studies did influence Tamil participants' judgments. This lends credence to the assumption that the grammatical gender of a language, stripped down to its possibly random linguistic properties, indeed can have an impact on cognition.

### General Discussion

The aim of the present studies was to shed light on a central aspect of the Whorfian hypothesis: How does the grammatical gender of languages influence speakers' cognition? In particular, we wanted to find out how persistent such effects are by (a) using German, a three-gendered

**Table 5** Differential impact of grammatical gender (German and Spanish) and other potential factors on tamil participants' responses in the potency scale, the adjective production, and the similarity judgment tasks

Predictor	Task											
	Potency scale				Adjective production				Similarity judgment			
	Standardized coefficient	<i>t</i> value ( <i>df</i> = 2294)	<i>p</i> value	Standardized coefficient	<i>t</i> value ( <i>df</i> = 2294)	<i>p</i> value	Standardized coefficient	<i>t</i> value ( <i>df</i> = 22864)	<i>p</i> value			
German gender	0.001	0.047	0.963	0.032	1.543	0.123	-0.049	-7.866	<0.001			
Spanish gender	-0.023	-1.190	0.234	0.015	0.756	0.450	-0.015	-2.341	0.019			
Femininity score	-0.278	-9.472	<0.001	-0.436	-14.519	<0.001	0.088	10.616	<0.001			
Animateness	0.217	9.740	<0.001	0.045	2.006	0.045	0.004	0.710	0.478			
Naturalness	0.297	9.472	<.0001	0.262	8.087	<0.001	0.172	20.891	<0.001			

Given are standardized multilevel coefficients and the respective *t* and (two-sided) *p* values for five predictors. Note that for the similarity judgment task, the difference between the two respective femininity scores and the congruency of the other factors is used as predictor (see text for additional explanation)

language for which previous results have been mixed, (b) statistically controlling for alternative explanations, and (c) using three tasks that differed in how closely they were associated with grammatical gender. Moreover, we wanted to find out whether grammatical gender, especially the Spanish grammatical gender, carries with it some universal (extralinguistic) information that might yield “grammatical gender effects” even in genderless languages.

Our findings can be summarized as strong evidence for the existence of a small to medium-sized grammatical gender effect even for three-gendered languages and even if alternative explanations and individual-level differences are statistically controlled for. The effects for the potency scale and the adjective productions were not large but the huge number of judgments ( $108 \times 20 = 2,160$  data points per task) indicates that it is reliable. That it is a genuine effect of the grammatical gender and not an effect of information somehow attached to grammatical gender is corroborated by the null effects found for German and Spanish grammatical gender in Tamil participants’ judgments. This null effect also indicates that the femininity, naturalness, and animateness of the entities denoted by the nouns used in our studies might cover almost all extralinguistic effects attached to Spanish or German grammatical gender. Note that our results do not speak against Vigliocco et al.’s (2005) sex and gender hypothesis, but they indicate that both the Spanish and also the German grammatical gender do not have a special universal status if stripped from their extralinguistic properties.

The extralinguistic properties examined here seem to have substantial effects on verbal tasks that are on average stronger than grammatical gender effects. Even in the similarity rating task, in which the impact of grammatical gender was basically nonexistent, one of the properties, the (congruency between the) naturalness of nouns, had a sizeable impact on participants’ ratings. So, the persistence of grammatical gender effects might be quite limited. But despite these limitations, one might expect such a sensitivity to the impact of grammatical gender to influence cognition in a more general way in several respects. One might, for instance, expect such influences in the arts. As an example, Boroditsky (2009) reported that Russian painters tend to paint death as a female (death is female in Russian) whereas in German painters’ pictures, death is usually portrayed as a male (death is male in German). Such effects might also be found in music, and especially if people directly attend to associative properties of words, as when reading prose and especially poetry. Here readers might get different impressions when reading a poem in translation than when reading it in the original language (see Deutscher 2010, chapter 8). So are we sensitive to grammatical gender? Yes. Do we use our sensitivity in a general way? It seems that we do not, but we are more likely do so in areas in which verbal associations play an important role, such as in diverse fields in the arts. To examine the impact of grammatical gender in these fields should be a most rewarding task for researchers.

**Acknowledgments** We would like to thank Friederike Brockhaus, Thomas Schäfer, Marcus Schwarz, and Isabell Winkler for helpful comments on a previous version of the manuscript.

## References

- Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, *111*, 329–344.
- Bellugi, U., Bihrlé, A., & Corina, D. (1991). Linguistic and spatial development: Dissociations between cognitive domains. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological determinants of language development* (pp. 368–398). Hillsdale, NJ: Erlbaum.
- Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers’ conceptions of time. *Cognitive Psychology*, *43*, 1–22.

- Boroditsky, L. (2003). Linguistic relativity. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 917–921). London: MacMillan Press.
- Boroditsky, L. (2009). How does our language shape the way we think? In M. Brockmman (Ed.), *What's next? Dispatches on the future of science* (pp. 116–129). New York: Vintage Press.
- Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, and semantics. *Proceedings of the Cognitive Science Society*, 22, 42–47.
- Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Medow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.
- Bowerman, M. (1996). The origins of children's spatial semantic categories: Cognitive versus linguistic determinants. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge: Cambridge University Press.
- Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49, 454–462.
- Brybaert, M., Fias, W., & Marie-Pascale Noe, M.-P. (1998). The Whorfian hypothesis and numerical cognition: Is 'twenty-four' processed in the same way as 'four-and-twenty'? *Cognition*, 66, 51–77.
- Carroll, J. B., & Casagrande, J. B. (1958). The function of language classifications in behavior. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (pp. 18–31). New York: Henry Holt.
- Chen, J.-Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, 104, 427–436.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Erlbaum.
- Dehaene, S., Spelke, E., Pineda, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, 284, 970–974.
- Deutscher, G. (2010). *Through the language glass: How words colour your world*. London: William Heinemann.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–212.
- Ervin, S. M. (1962). The connotations of gender. *Word*, 18, 249–261.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Furth, H. (1966). *Thinking without language*. New York: Free Press.
- Gelman, R., & Gallistel, R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- Gentner, G., & Imai, M. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62, 169–200.
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93, 10–20.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.
- Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, 98, 377–389.
- January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, 104, 417–426.
- Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.
- Koch, S. C., Zimmermann, F., & Garcia-Retamero, R. (2007). El sol-die Sonne: Hat das grammatische Geschlecht von Objekten Implikationen für deren semantischen Gehalt? *Psychologische Rundschau*, 58, 171–182.
- Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, 22, 519–534.
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 843–858.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lehmann, T. (1993). *A grammar of modern Tamil* (2nd ed.). Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Leinbach, M. D., Hort, B. E., & Fagot, B. (1997). Bears are for boys: Metaphorical associations in young children's gender stereotypes. *Cognitive Development*, 12, 107–130.

- Levinson, S. L. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109–169). Cambridge, MA: MIT Press.
- Li, P., & Gleitman, L. (2002). Turing the tables: Language and spatial reasoning. *Cognition*, *83*, 265–294.
- Lucy, J. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*. New York: Liveright.
- Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Speaking versus thinking about objects and actions. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 81–111). Cambridge, MA: MIT Press.
- Martinez, I., & Shatz, M. (1996). Linguistic influences on categorization in preschool children: A crosslinguistic study. *Journal of Child Language*, *23*, 529–545.
- Mill, J. S. (1872/1973). System of logic. In J. M. Robson (Ed.), *Collected works of John Stuart Mill* (8th ed., Vols. 7 and 8). Toronto: University of Toronto Press. (Original edition published 1872).
- Miller, K., Smith, C., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, *6*, 56–60.
- Mullen, M. K. (1990). Children's classifications of nature and artifact pictures into female and male categories. *Sex Roles*, *23*, 577–587.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International Inc.
- Schiffman, H. F. (1999). *A reference grammar of spoken Tamil*. Cambridge: Cambridge University Press.
- Schiffman, H. F. (2004). The Tamil case system. In J.-L. Chevillard, & E. Wilden (Eds.), *South Indian horizons: felicitation volume for Francois Gros on the occasion of his 70th birthday* (pp. 293–322). Pondicherry: Institut Francaise d'Indologie de Pondichéry.
- Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (2nd ed.) [Research methods and statistics: A textbook for psychologists and social scientists]. Munich: Pearson Education.
- Sera, M. D., Berge, C., & del Castillo Pintado, J. (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development*, *9*, 261–292.
- Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., & Rodríguez, W. (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.
- Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, *45*, 101–107.
- Steever, S. B. (1987). Tamil and the Dravidian languages. In B. Comrie (Ed.), *The world's major languages* (pp. 725–746). London: Croom Helm.
- Takano, Y. (1989). Methodological problems in cross-cultural studies of linguistic relativity. *Cognition*, *31*, 141–162.
- Twain, M. (1880). *A tramp abroad*. Hartford, CT: American Publishing.
- Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*, 501–520.
- Vuksanović, J., Bjekić, J., & Radivojević, N. (2014). Grammatical gender and mental representation of object: The case of musical instruments. *Journal of Psycholinguistic Research*. doi:10.1007/s10936-014-9293-7.
- Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, *12*, 111–122.
- Whorf, B. L. (1940). Science and linguistics. *Technology Review*, *42*(229–231), 247–248.
- Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. New York: Wiley.