Cognitive Development, 9, 261–292 (1994)

Grammatical and Conceptual Forces in the Attribution of Gender by English and Spanish Speakers

Maria D. Sera
Christian A. H. Berge
University of Minnesota

Javier del Castillo Pintado
Complutense University of Madrid

We compared the assignment of gender to masculine and feminine pictured objects—as classified by the Spanish grammar—by English- and Spanish-speaking children and adults in three experiments. Across all three studies, subjects participated in one of two conditions. In one condition, pictures alone were presented; in the other condition, pictures were shown and labeled. We found that speakers of Spanish began to classify the objects according to the grammatical gender of the Spanish language in the second grade, unlike speakers of English. The effect of grammatical gender was more pronounced for speakers of Spanish when the objects were labeled, pointing specifically to the role of language in their classifications. We also found that English speakers were consistent in their judgments, often classifying artificial objects as male-like and natural objects as female-like. Spanish speakers were also sensitive to the natural-female/artificial-male conceptual division. Finally, we found that the artificial-male/natural-female link was an earlier force in classification for speakers of English than grammatical gender was for speakers of Spanish, suggesting that grammatical classifications are superimposed on conceptual ones in development.

Experiment 1 was used to partially fulfill the requirements for a master's degree by Christian A. H. Berge at the University of Minnesota. This research was supported by a McKnight Research Award from the University of Minnesota and by a research grant from the National Institute of Child Health and Human Development (NICHD; HD R0127376) to Maria D. Sera. We thank the staffs at Chicanos Latinos Unidos En Servicios (CLUES) and The Neighborhood House in St. Paul and Minneapolis for their assistance with the data collection, with special thanks to Alberto Ortiz, Roberto Aviña, Maren Swanson, Francis Mudic, Milissa Tilton, and Eric Reittinger. We also thank the teachers and children of Joaquín Costa in Madrid and extend a special thanks to Julian Diaz. A portion of this research was presented at the 60th meeting of the Society for Research in Child Development in New Orleans, LA.

Correspondence and requests for reprints should be sent to Maria D. Sera, Institute of Child Development, University of Minnesota, 51 East River Road, Minneapolis, MN 55455.

Manuscript received September 28, 1993; revision accepted May 10, 1994.

Questions of linguistic relativity, of cognition shaped and influenced by language, date back to antiquity. One has only to briefly review the writings of Aristotle, Descartes, Kant, or Locke to notice that a great deal of effort has been put forth in an attempt to understand human thought and its relation to language. Consider three claims about the relation between language and thought put forth by Edward Sapir (1912) and Benjamin Whorf (1956):

Claim 1. Structural differences between language systems will, in general, be paralleled by nonlinguistic cognitive differences, of an unspecified sort, in the native speakers of the two languages. (summarized by Brown, 1976, p. 128)

Claim 2. The structure of anyone's native language strongly influences or fully determines the world-view acquired as one learns the language. (summarized by Brown, 1976, p. 128)

Claim 3. The semantic systems of different languages vary without constraint. (summarized by Kay & Kempton, 1984, p. 66)

The current status of the Sapir–Whorf hypotheses can be briefly summarized as follows. Empirical research failed to support the third claim (Berlin & Kay, 1969), discredited strong interpretations of the second claim (Heider, 1972), and has been unable to provide solid support for the first claim. Several studies among speakers of English, however, indicate that language sometimes does affect cognition (e.g., Gelman & Gallistel, 1978; Landau & Gleitman, 1985; Markman & Hutchinson, 1984; Waxman, 1990). Consequently, we may find weaker versions of linguistic relativity to be true. In this article, we extend the evidence on linguistic relativity by asking whether the grammatical categories of a language, which appear random and chaotic, nonetheless offer speakers of that language reference points that guide their classification of objects.

We focus on the interplay between what we call natural, conceptual, and linguistic gender classifications. Natural gender is a basic dimension that is used to categorize people and animals into male and female categories. It is one of the few organizing principles present in all human societies, as it provides a salient basis for organizing information. Conceptual gender classifications are those that cannot be directly traced to linguistic or natural gender categories. For example, anthropologists such as Ortner (1974) proposed that the association of females with natural objects and males with artificial ones is a universal conceptual distinction that profoundly affects the role of men and women in society. With respect to linguistic gender, two kinds of gender classifications have been identified: *semantic* and *grammatical*. Semantic gender refers to the linguistic marking of natural gender. By traditional linguistic frameworks, languages of the world are often classified into two types: those having a natural gender and those with a grammatical gender. In languages that have a natural gender, such as

English, some words that refer to animals and humans carry a semantic value that refers directly to the gender of the referent. For example, *boy* refers to a male child and *girl* to a female child. Many other nouns such as *professor, doctor,* and *feather,* carry no explicit semantic gender markings. Thus, semantic gender in English is an integral part of some nouns. In languages with a grammatical gender system, every noun must be categorized into a gender class such as masculine, feminine, or neuter. The particular gender into which individual words fall is thought to be semantically arbitrary, as gender markings in these languages are simply believed to establish phonological agreement with other grammatical classes (e.g., articles, nouns, and adjectives). Spanish is a language with two grammatical gender classes. So, for example, the phonological ending of a word, either /a/ in the case of *niña* (girl) or /o/ in the case of *niño* (boy) marks the gender difference between a male and a female child in Spanish. Of course, the grammatical gender markings in this case are not semantically arbitrary because they mark the natural gender of the referent. However, the phonological ending of all nouns, even those that do not refer to animals or humans, also carry a gender classification in Spanish. For example, the Spanish word for feather, *pluma,* is marked as feminine but because feathers do not have a natural gender, this Spanish linguistic classification is thought to be semantically arbitrary, and thus labeled grammatical.

Many languages of the world possess a grammatical gender system, and differences across languages in the assignment of grammatical gender are well-documented. Thus, cross-linguistic differences in gender systems offer a rich domain for investigating the relationship between linguistic classification systems and conceptual ones. However, although literature exists on the classification of gender by speakers of different languages (Clarke, Losoff, MacCracken, & Rood, 1984; Clarke, Losoff, MacCracken, & Still, 1981; Ervin, 1962; Gathercole, 1989; Guiora & Acton, 1979; Mueller-Gathercole & Hasson, 1991), on the acquisition of grammatical gender by children who speak different languages (Hernández-Pina, 1984; Karmiloff-Smith, 1979; MacWhinney, 1978; Pérez-Pereira, 1991), and on cultural differences in the organization of gender concepts (Bem, 1981; Levi-Strauss, 1969; Liben & Signorella, 1993; Martin & Halverson, 1981; Mullen, 1990; Ortner, 1974; Ruble & Stangor, 1986), the degree to which linguistic gender classifications influence our thinking about objects remains an open question.

Why has there been so little experimental work on the relation between grammatical and meaningful organizations of gender? The gap in the literature may reflect the fact that many contemporary psycholinguists repeatedly point to the semantically arbitrary nature of grammatical gender (e.g., Braine, 1987; Maratsos, 1983, 1988). The following example illustrates their argument: In German, the word for knife is neuter, the word for spoon is masculine, and the word for fork is feminine. In contrast, the Spanish word for knife is masculine, the word for spoon is feminine, and the word for fork is masculine. Such

apparent inconsistencies extend beyond eating utensils. Spanish, German, and a number of other languages with a grammatical gender system disagree on the gender of many objects. The apparent disagreement of semantic gender assignments across languages with a grammatical gender system constitutes the evidence for the nonsensical, arbitrary alignment of semantic and grammatical gender.

However, just because a set of *universal* semantic features will not explain grammatical gender, does not mean that grammatical gender lacks semantic force. A close relation between grammatical, semantic, and conceptual organizations of gender may exist. Languages with a grammatical gender system seem to agree on the gender of many terms that refer to objects with a natural gender such as *man, woman,* and *king.* The assignment of grammatical gender to such objects is semantically sensible and may prove to be consistent across different languages. At least in Spanish, the grammatical gender of a noun actually marks its semantic gender as in the previous example of *niña* (girl) and *niño* (boy). Indeed, the fact that these linguistic forms convey information about natural gender is what defines the system as grammatical gender. Moreover, although the grammatical gender system of any one language may seem arbitrary and far-removed from a semantically sensible and universal classification of objects in the world to the psycholinguist, there may be no ambiguity to the monolingual speaker of that language. For example, the Spanish language requires its speakers to label objects according to gender in one particular way—oblivious to, and independently of, what German and Italian require from its speakers. So even though German classifies forks as feminine, and Spanish classifies forks as masculine, from the point of view of a monolingual speaker of Spanish (or German), there may be only one sensible semantic and conceptual classification of forks. In fact, the experimental studies that exist on the relation between grammatical and conceptual gender support just this claim (e.g., Clarke, et al., 1984; Ervin, 1962).

One of the first studies on grammatical gender was conducted by Ervin-Tripp (published as Ervin, 1962). Ervin-Tripp's article was presented as a follow-up to the classic work by Brown & Lenneberg (1954) and Carroll and Casagrande (1958), which appeared to demonstrate the influence of language on memory and categorization. Thus, Ervin-Tripp took for granted the veracity of linguistic relativity and simply sought to extend the case in favor of Whorf's hypothesis to a mandatory, grammatical, close-class aspect of the linguistic classification system. In Ervin-Tripp's study, Italian nonsense words were constructed that contained either masculine or feminine morphology. Native, adult speakers of Italian were asked to judge the meanings of the nonsense words on two scales that Osgood, Suci, & Tannenbaum (1957) identified as yielding different evaluations for men and women. She found statistically reliable differences in the semantic ratings of the masculine-sounding versus the feminine-sounding nonsense words by adult speakers of Italian. Since Ervin-Tripp's study, there

have been several studies on the perception of gender by speakers of different languages that report similar results (see Clarke et al., 1984, for a review; Mueller-Gathercole & Hasson, 1991). So despite arguments by contemporary psycholinguists that the relation between grammatical and semantic organizations of gender is arbitrary and nonsensical, the experimental evidence on the topic suggests that grammatical gender carries meaning, at least for adult speakers of languages with a grammatical gender system. Critics of linguistic relativity, however, have discredited these findings on two grounds: (a) the differences reported are caused by factors other than differences between the languages, and (b) such tasks, in which subjects are asked to classify words, directly measure a speaker's knowledge of grammatical gender and so reveal nothing beyond what is already known about the grammatical gender classifications of these languages. The design of the experiments in this article addresses these two concerns.

Developmental evidence on the acquisition of grammatical gender by children who speak German, French, Hebrew, and Spanish also sheds light on the relation between grammatical and semantic gender, although this work primarily concerns itself with the child's use of gender agreement rules, or what might be called "syntactic" knowledge of grammatical gender (Hernández-Pina, 1984; Karmiloff-Smith, 1979; MacWhinney, 1978; Pérez-Pereira, 1991). For example, in the Pérez-Pereira study, children were shown pictures of novel animate figures (some pictures depicted animates that clearly possessed a natural gender; other pictures depicted objects with ambiguous natural genders) and were told: "This is a picture of a *pifar* [an invented word]." In this way, the gender of the referent was coded both morphologically (by the article and noun ending) and semantically (by the natural gender of the object depicted in the picture). Then, a picture that was exactly the same as the original but in a different color was shown. The child was then asked to state the color of the second *pifar.* Because the gender of adjectives in Spanish agrees with the gender of the noun, the adjectival ending generated by the child reflected the child's use of gender agreement rules. The results from these studies, across different languages, are fairly consistent: Children have been shown to honor gender agreement rules by 4 years of age. However, more relevant for our purposes is the finding that semantic cues signal gender agreement rules relatively later than morphological cues. Thus, this work suggests that the correspondence between grammatical and semantic gender emerges relatively late.

The development of gender concepts—how individuals organize nonlinguistic, culturally shared information about gender—has also been widely studied. Dozens of studies exist on the cognitive structures that organize an individual's gender-related knowledge and the mechanisms that underlie the acquisition and maintenance of stereotypic beliefs (see Liben & Signorella, 1993, for a review). This literature has yielded several findings regarding the conceptual organization of gender. For example, Ruble and Stangor (1986) reported that material that

is culturally linked to one gender is better remembered by children and adults of that gender. From a different perspective, anthropologists also offer what they claim are culturally universal dimensions of gender concepts. As previously stated, Ortner (1974) argued that the association of females with natural objects and males with artificial ones is a universal conceptual distinction that profoundly affects the role of males and females in society. Although Ortner's claims have been criticized, Mullen (1990) recently reported that English-speaking children more often assign natural objects to a female category and artificial objects to a male category. In short, a close relation between culture and gender concepts is assumed to exist by researchers from a variety of perspectives. Language is unquestionably a part of culture. Ironically, although cultural effects on gender concepts are either assumed or taken for granted, the effects of language are often called into question. In this article, we begin to examine the interplay between culturally shared, conceptual organizations of gender and language-specific, grammatical ones.

We focus on English and Spanish, and begin with an outline of the Spanish gender system. In brief, the Spanish gender system works as follows: Most nouns are classified as either masculine or feminine without the possibility of being changed. There are masculine and feminine forms of articles and many adjectives. Feminine nouns typically end in /a/, /d/, /is/, or /ion/; masculine nouns typically end in /o/, /e/, or other consonants. Except for a small number of words that refer to human beings or animals, it is not possible to change the gender of one word to another. For example, *el arbol* (the tree) may not be changed to the feminine version *la arbola*. For the small group of words that may be marked with both genders, the most common masculine nouns take the /o/, /e/, /s/, /l/, or /r/ suffixes and the feminine take the /a/ suffix. For example, one may say *el americano* (the American man) or *la americana* (the American woman); *el doctor* (the male doctor) or *la doctora* (the female doctor); *el gato* (the male cat) and *la gata* (the female cat); and so on. However, both genders may not be applied to all words that refer to animals and humans: There is only one word that means giraffe (*jirafa;* it is feminine), elephant (*elefante;* it is masculine), and client (*cliente;* it is masculine). Finally, although the gender of the article agrees with the phonological ending of the noun in 99 percent of cases, a few exceptions exist (see Bull, 1965; Pérez-Pereira, 1991; or Terrell & Salgues de Cargill, 1979 for a more detailed outline of the Spanish gender system).

In the following experiments, we compared the classification of pictured objects according to the Spanish grammatical gender, and the natural-artificial division by English and Spanish speakers. By asking subjects to classify pictures, we reduced the possibility that speakers explicitly referred to language in their classifications. In all of the experiments, subjects participated in one of two conditions. In one condition, pictures alone were presented to the subjects whereas pictures were shown and labeled in the other condition. Through this

manipulation, we attempted to isolate the influence of language from nonlinguistic cultural factors. Because the Spanish labels contain the linguistic markings that we hypothesize to be affecting categorization (if any such effects exist), labeling versus not labeling the pictures constitutes a manipulation of that variable. This strategy adds an important component to our study because it would enable us to attribute the cause of differences we might find between speakers of English and Spanish specifically to language differences. In Experiments 1 and 2, we examined the effects of grammatical gender on the categorization of English- and Spanish-speaking adults across two tasks. In Experiment 1, adults classified the objects as feminine or masculine. In Experiment 2, we removed all explicit references to any aspect of language by asking subjects to attribute either a man's or a woman's voice to each object. We employed the same task in Experiment 3, in which we addressed developmental issues.

EXPERIMENT 1

In this experiment, English- and Spanish-speaking adults were asked to classify pictured objects as either masculine or feminine. Pilot research established the English and Spanish labels and the Spanish gender of each item. We also presented two kinds of stimulus items. *Control pictures* consisted of objects that possessed an obvious natural gender such as *man, woman,* and *king*. We hypothesized the classification of these objects as either masculine or feminine to be straightforward for speakers of both languages. *Test pictures* consisted of objects that we originally believed to be only grammatically gendered as masculine or feminine by the Spanish language. That is, the *test pictures* consisted of objects that lacked a natural gender (e.g., grapes, arrow, table, shoe, ice cream). We asked whether Spanish speakers would be more likely than English speakers to classify the test pictures according to the grammatical gender of the Spanish language. We also tested equal numbers of men and women in each group, because sex differences in gender concepts have often been found (e.g., Ruble & Stangor, 1986).

Method

Subjects. Thirty Spanish-speaking and 30 English-speaking adult monolinguals participated in the study (30 men and 30 women). Careful subject screening ensured the participation of native, monolingual speakers. Spanish-speaking adults came from introductory level English as a Second Language (ESL) courses taught at Hispanic community centers in St. Paul, Minnesota. Most of the Spanish speakers were from Mexico and a few were from countries in Central America. English-speaking adults were recruited at the University of Minnesota at Minneapolis and selected only if they were monolingual, native

speakers. Of the 30 Spanish-speaking adults, 24 began to learn English on the average at 24 years of age; of the remaining 6, half responded "as of today" and the rest "within the last few months." Two additional Spanish-speaking subjects failed to complete the study due to difficulties in following the experimental procedures. An equal number of male and female subjects, within each language group, participated in the study. All subjects were paid $3 for their participation.

Materials. A series of 31 laminated drawings, obtained primarily from the English and Spanish versions of the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981, 1986), were presented to all subjects. In addition, pictures of a hat and a fork were used from "Pictures, Please!" (Abbate & LaChappelle, 1979). Several drawings required slight alterations. All pictures were selected according to their grammatical gender in the Spanish language, with half receiving the feminine articles and noun endings and the other half the masculine. Pilot research with 11 subjects of Spanish and English linguistic backgrounds supported the accuracy and grammatical gender of each label used. The pictures were divided into two separate groups, control or test. The control pictures were intended to test that all subjects clearly understood the instructions of the study and that their judgments reflected their attribution of natural gender. For example, the picture of a king should have been labeled as masculine by all subjects. On the other hand, the test pictures were intended to test the effect of a grammatical gender system on the attribution of gender by Spanish speakers. Figure 1 shows examples of six pictures that were used. Table 1 shows the list of terms that we employed when the pictures were labeled. The list of terms in Table 1 also describes the full set of pictures.

Procedure. Each participant was told that about 30 pictures would be displayed, with or without a spoken label, and that their task was to judge each picture as masculine or feminine. As the researcher displayed, labeled or did not label, and numbered each picture, the subject circled either an M or F, standing for masculine or feminine on a response sheet. The exact English instructions were: "I am going to show you about 30 pictures of common, everyday objects, numbering each as I go. I want you to tell me whether each object strikes you as more masculine or feminine by circling either an M for masculine or an F for feminine on the data sheet." The exact Spanish instructions were: "Les voy a enseñar alrededor de 30 dibujos de cosas comunes, que yo voy a nombrar, una a una. En este papel quiero que pongan una M si creen que la cosa parece masculina y una F si creen que la cosa parece femenina." Subjects in the pictures + labels condition heard the name of the object pictured in addition to seeing the picture. In this experiment, subjects who received the labeled items heard only the noun. One researcher, fluent in both Spanish and English, tested all subjects in their respective native language.

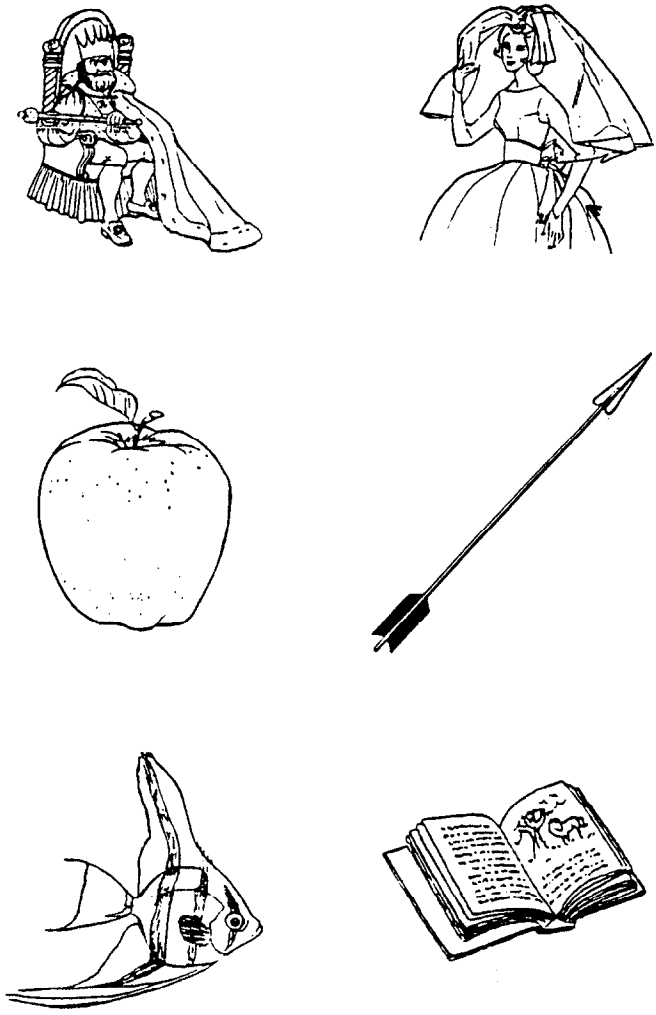

Figure 1. Examples of the pictured objects that we used as stimuli. The *king* and the *bride* are examples of two control items which possess a natural gender. The *apple* and the *arrow* are examples of grammatically feminine objects, according to Spanish. The *fish* and the *book* are examples of grammatically masculine objects.

Design. Of the 30 Spanish and 30 English speakers, 18 from each group participated in a Pictures Only condition and 12 in Pictures + Labels condition. Five randomly determined orders of pictures were used. The same number of subjects viewed each order across conditions. Subjects were tested in groups ranging in size from one to four.

Table 1. The English and Spanish Terms That Were Used When the Pictures Were Labeled Across All Three Experiments

Grammatically Masculine Terms		Gramatically Feminine Terms	
		Control	
Spanish	**English**	**Spanish**	**English**
hombre	man	mujer	woman
niño	boy	niña	girl
gigante	giant	novia	bride
rey	king	bailarina	ballerina
		Test	
Spanish	**English**	**Spanish**	**English**
helado	ice cream	uvas	grapes
fuego	fire	estrella	star
ojo	eye	pluma	feather
pez	fish	guitarra	guitar
cuadrado	square	flecha	arrow
sombrero	hat	cuchara	spoon
cuchillo	knife	araña	spider
zapato	shoe	mesa	table
tenedor	fork	rueda	wheel
libro	book	oreja	ear
helicóptero	helicopter	manzana	apple
telescopio	telescope		

Results

We calculated the percentage of times each picture was judged according to the grammatical classification of the Spanish language. We statistically examined these percentages through a five-factor analysis of variance (ANOVA) with language (English or Spanish), presentation (pictures only or pictures + labels), and gender of subject (male or female) as between-subject variables, and picture (control or test) and Spanish gender (masculine or feminine) as within-subject variables. We obtained significant main effects of language, $F(1, 52) = 42.6$, $p < .001$; Spanish gender, $F(1, 52) = 11.3$, $p < .01$; and picture, $F(1, 52) = 262.2$, $p < .001$; significant two-way interactions between language and presentation, $F(1, 52) = 7.1$, $p < .05$; between language and picture, $F(1, 52) = 78.0$, $p < .001$; and between presentation and picture, $F(1, 52) = 6.8$, $p < .05$. Finally, we observed three-way interactions between language, presentation, and picture, $F(1, 52) = 8.3$, $p < .05$, and between language, Spanish gender, and picture, $F(1, 52) = 10.3$, $p < .05$. As no significant effects of gender of subject were observed nor did gender of subject interact with any other effects, we

combined the judgments of male and female subjects in subsequent analyses and experiments. Table 2 shows the mean percentage of times test and control pictures were labeled according to the Spanish grammatical gender by English and Spanish speakers when pictures were labeled or presented alone.

Because all significant effects participated in the three-way interactions, we interpret all of our findings in the context of those interactions which we further analyzed with Tukey Tests. We begin with the three-way interaction between language, presentation, and picture (Tukey's H.S.D. = 14.09, $p < .05$). This interaction revealed that English speakers judged the control pictures according to the Spanish grammatical gender more often than the test pictures. We also found that Spanish speakers judged the test pictures reliably more often according to the Spanish grammatical gender than English speakers. Spanish speakers also judged the control pictures more often according to the Spanish grammatical gender than the test pictures, but only when the test pictures were presented alone. That is, speakers of Spanish classified the test pictures, when labeled, as systematically as they classified the control pictures. Thus, we found a reliable difference for the test pictures among speakers of Spanish such that they judged the pictures according to the Spanish grammatical gender more often when the pictures were labeled than when they were presented alone (see Table 2). However, Spanish speakers classified the test pictures according to the grammatical gender of the Spanish language more often than English speakers even when the pictures were presented alone (71% vs. 51%).

Next, we clarify the three-way interaction between language, picture, and Spanish gender (Tukey's H.S.D. = 8.0, $p < .05$). This interaction revealed no differences among the Spanish speakers in their judgments of grammatically masculine or feminine test and control pictures. We also found no differences among the English speakers in their judgments of the control pictures that were grammatically feminine or masculine. However, we found that English speakers

Table 2. The Percentage of Times English and Spanish Speakers' Classifications Honored the Grammatical Gender of the Spanish Language in Experiment 1

	Pictures Only		Pictures + Labels	
	Test	Control	Test	Control
English Speakers				
Masculine	57.7	95.8	54.7	93.7
Feminine	46.2	100	45.1	100
Spanish Speakers				
Masculine	75.9	97.2	90.2	100
Feminine	66.6	93.1	92.5	89.6

made more judgments according to the Spanish grammatical gender for the masculine test pictures than for the feminine test pictures, suggesting a nonarbitrary relation between the grammatical and conceptually masculine items. Spanish speakers also showed this trend when the pictures were not labeled (see Table 2).

We began to explore the nature of this nonarbitrary link between grammatical and conceptual classifications by examining how grammatically feminine and masculine natural and artificial pictures were judged. We classified 21 of the test items into four groups: (a) *grammatically feminine natural items* (apple, ear, feather, grapes, and spider), (b) *grammatically feminine artificial items* (arrow, guitar, spoon, table, and wheel), (c) *grammatically masculine natural items* (eye, fire, and fish), and (d) *grammatically masculine artificial items* (book, fork, hat, helicopter, ice cream, knife, shoe, and telescope). The first thing worth noting is that many more items were classified as grammatically masculine artificial items than as grammatically masculine natural items by the Spanish language. Thus, our finding of more grammatically consistent judgments among the masculine items may reflect a stronger association between the Spanish masculine items and artificial items.

We then asked whether English and Spanish speakers were sensitive to an association of artificial items as male-like and naturally occurring items as female-like. Thus, we calculated the percentage of times each subject classified the grammatically masculine natural items as feminine, the percentage of times he or she classified the grammatically masculine artificial items as feminine, and so on. Table 3 shows the percentage of times natural and artificial items were classified as feminine by each language group. We statistically examined these percentages through a Language Group (English or Spanish) × Conceptual Category (Natural or Artificial) × Spanish Gender (Masculine or Feminine) ANOVA with language as a between-subjects variable and conceptual category and Spanish gender as within-subjects variables. We observed reliable main effects of Spanish gender, $F(1, 58) = 42.6$, $p < .001$, and conceptual category, $F(1, 58) = 65.2$, $p < .001$. We also observed interactions between language group and Spanish gender, $F(1, 58) = 55.7$, $p < .001$; between language group

Table 3. The Percentage of Times Grammatically Feminine and Masculine Natural and Artificial Items Were Classified as Feminine by English and Spanish Speakers in Experiment 1

	Grammatically Feminine		Grammatically Masculine	
	Natural	Artificial	Natural	Artificial
English	61	29	61	36
Spanish	77	76	26	14

and conceptual class, $F(1, 58) = 25.3$, $p < .001$; and between language group, Spanish gender, and conceptual class, $F(1, 58) = 5.14$, $p < .05$. We examined these interactions further with analyses of simple effects.

The Language Group × Spanish Gender interaction confirms our prior finding that grammatical gender is a strong force in the Spanish speakers' but not the English speakers' judgments ($p < .001$ for the Spanish gender effect among Spanish speakers and $p = .51$ among English speakers). However, these findings also reveal that conceptual class was a force in the English speakers' judgments. English speakers judged the naturally occurring items as feminine 62% of the time whereas they judged the artificial items as feminine 32% of the time ($p < .001$). Conceptual class was also a force in the Spanish speakers' judgments, although not as strong a force. Spanish speakers judged the natural masculine items as feminine more often than they judged the artificial ones (26% vs. 14%, $p < .05$). For example, they were more likely to judge the pictures of the helicopter and telescope as masculine than the pictures of fire and fish although all four items are grammatically masculine in Spanish. However, they judged the linguistically feminine natural and artificial items as feminine equally often (77% vs. 76%). In other words, conceptual class influenced only the Spanish speakers' judgments of linguistically masculine items.

Discussion

Our most important finding is that grammatical gender can be shown to influence the attribution of gender to objects that lack a natural gender, at least among adult speakers of Spanish. The fact that the Spanish grammatical gender *often* accurately marks the object's natural gender seems to be generalized by adult speakers of Spanish to objects, that according to English speakers, do not possess the conceptual gender signaled by the Spanish grammar. The labeling of the pictures led to even more systematic judgments according to the Spanish grammatical gender system indicating that grammatical gender is a causal force in the attribution of gender to objects that lack a natural gender within Spanish speakers.

The labeling effect that we observed among speakers of Spanish provides important new evidence in support of linguistic relativity. One way that findings in support of linguistic relativity have been discredited has been by arguing that the differences reported were not caused by differences in language per se, but instead caused by an unidentified, nonlinguistic cultural variable. And in all fairness to Whorf and Sapir's critics, in studies in which language is proposed to be the underlying independent variable, it is important to try to isolate the influence of language from other cultural factors. Of course, it has been extremely difficult to tease linguistic and cultural forces apart, because language is a part of culture. In this regard, the present experiment contains a methodological advance over much of the previous work on linguistic relativity. We successfully manipulated the linguistic variable that we hypothesized to be

causing differences in the gender classifications of English and Spanish speakers. We found Spanish speakers to be sensitive to this linguistic manipulation. Thus, the possibility that a nonlinguistic cultural factor, that covaries with grammatical gender, is actually causing the difference between English and Spanish speakers is small. If a nonlinguistic cultural factor were guiding the Spanish speakers' classifications, we would not have found a *linguistic* manipulation to influence their reliance on that nonlinguistic cultural factor.

It is possible, however, that asking speakers of a language with a grammatical gender system to classify objects as masculine or feminine leads speakers of that language to explicitly refer to that language's grammatical gender system. If so, our findings reflect nothing beyond the Spanish speaker's metalinguistic knowledge of the Spanish grammatical gender system. We believe this possibility is unlikely for two reasons. First, native speakers of a language with a grammatical gender system do not receive explicit instructions regarding the masculine and feminine gender classifications of articles, nouns, and adjectives in their natural language. Second, if the Spanish speakers were explicitly referring to grammatical gender when judging the pictures, their classifications should have been perfectly correlated with the Spanish grammatical gender. However, the Spanish speakers' classifications, even when the pictures were labeled, did not perfectly correspond to the grammatical classifications. If Spanish speakers were consciously judging the items according to grammatical gender, it is also difficult to explain why they showed any sensitivity to the natural-artificial contrast. Recall that we found that Spanish speakers, like speakers of English, tended to classify grammatically masculine natural items as feminine more often than grammatically masculine artificial items. If Spanish speakers were explicitly referring to the gender system, it is difficult to explain why they did so more often for some items than others. In any case, we sought to drive the Spanish speaker's attention away from grammatical gender in Experiment 2.

EXPERIMENT 2

We undertook this experiment because we were concerned that asking speakers of Spanish to judge pictures as feminine or masculine led them to explicitly reference the Spanish gender system. In this experiment, we employed a task in which we used words that were very different from the Spanish words for feminine and masculine (*feminino/a* and *masculino/a*). We also wanted to pilot a methodology that we could employ with young children who might not know the words feminine and masculine. Our methodology involved asking subjects to attribute either a man's or woman's voice to each pictured object—a simple task that tacitly measured the conceptual attribution of gender.

Method

Subjects. Twenty native, monolingual speakers of Spanish (2 women and 18 men) and 20 native English speakers (12 women and 8 men) from the same populations as in Experiment 1 participated. Because we found no effect of gender of subject in Experiment 1, we neither controlled nor looked for effects of gender of subject in this study. All subjects were paid $3 for their participation.

Materials. We used the same pictures and labels as in Experiment 1.

Procedure. The design followed that of Experiment 1. The most important difference from Experiment 1 were the instructions. In an attempt to direct the Spanish speakers' attention away from grammatical gender, which might be activated by the words masculine and feminine, we constructed our instructions to contain no words that were similar to the Spanish words for masculine and feminine (*masculino* and *feminino*). In this experiment, subjects were told the following:

> We are thinking about making a new movie in which some everyday objects come to life. We are going to show you a series of pictures of these objects and want you to write down, on this sheet of paper, by each number, whether you think each pictured object should have a man's voice or a woman's voice. Okay, here is picture number one [the Experimenter would then show one picture to the subject, and for half of the subjects labeled the picture], should this have a woman's voice or a man's voice in the movie?

The specific Spanish instructions were:

> Estamos pensando hacer una película en la que algunos objectos comunes tienen vida. Ahora voy a enseñarles una serie de dibujos y quiero que escriban en esta hoja de papel, al lado de cada número, si piensan que la cosa en el dibujo debe tener una voz de un hombre o una voz de una mujer. Muy bien, aquí está el dibujo número uno [the Experimenter would then show one picture to the subject, and for half of the subjects labeled the picture], ¿esta cosa debe tener la voz de un hombre o la voz de una mujer en la película?

Subjects wrote their answers on numbered response sheets. A native speaker of Spanish that was also fluent in English tested all of the subjects. Of the 20 Spanish and 20 English speakers, 10 from each group participated in a Pictures Only condition and 10 participated in a Pictures + Labels condition. Three randomly determined orders of presentation were used in showing pictures to subjects. The same number of subjects viewed each order across conditions. English-speaking subjects were tested one at a time in a laboratory at the

University of Minnesota in Minneapolis. Spanish-speaking subjects were tested in groups of 10 at a Hispanic Center in Minneapolis, Minnesota.

Results

As in Experiment 1, we first calculated the percentage of times each picture was classified according to the grammatical gender of the Spanish language. Table 4 shows the mean percentage of times test and control pictures were assigned a voice consistent with the grammatical gender of the Spanish language by English and Spanish speakers when pictures were labeled or presented alone. Again we found that the control pictures were classified according to their grammatical and natural gender a high percentage of the time (99%) by all subjects in all conditions. We statistically examined the percentage of times test pictures were classified according to the Spanish language with a three-factor ANOVA with language (English or Spanish) and presentation (pictures only or pictures + labels) as between-subject variables and Spanish gender (masculine or feminine) as a within-subject variable. We observed a main effect of language, $F(1, 36) = 149.1$, $p < .001$, and a marginally significant effect of presentation, $F(1, 36) = 3.1$, $p < .085$. Although the effect of presentation was most likely due to the Spanish speakers' judging the pictures according to the Spanish grammar more often when the pictures were labeled (see Table 4), the two-way interaction between presentation and language did not reach significance, $F(1, 36) = 1.99$, $p = .169$. We found, however, the same trend as Experiment 1: Spanish speakers classified the objects according to the Spanish grammatical gender 95% of the time when the pictures were labeled and 85% of the time when the pictures were presented alone (see Tables 2 and 4).

Finally, we examined how grammatically feminine and masculine natural and artificial pictures were classified by calculating the percentage of times each subject assigned a woman's voice to each kind of item. We classified the items, as we did in Experiment 1, as grammatically feminine natural items, grammatically feminine artificial items, grammatically masculine natural items, and

Table 4. The Percentage of Times English and Spanish Speakers' Classifications Honored the Grammatical Gender of the Spanish Language in Experiment 2

	Pictures Only		Pictures + Labels	
	Test	Control	Test	Control
English Speakers				
Masculine	54.2	100.0	57.4	100.0
Feminine	51.9	100.0	50.9	100.0
Spanish Speakers				
Masculine	84.2	100.0	94.2	100.0
Feminine	86.5	100.0	95.5	97.5

Table 5. The Percentage of Times Grammatically Feminine and Masculine Natural and Artificial Items Were Assigned a Woman's Voice by English and Spanish Speakers in Experiment 2

	Grammatically Feminine		Grammatically Masculine	
	Natural	Artificial	Natural	Artificial
English	74	29	55	33
Spanish	97	90	25	4

grammatically masculine artificial items, and each subject received a score that reflected the percentage of times he or she attributed a woman's voice to items in each of the categories mentioned earlier. Table 5 shows the percentage of times a woman's voice was attributed to grammatically feminine and masculine artificial and natural objects by English and Spanish speakers. We statistically examined the classifications through a Language Group (English or Spanish) × Conceptual Class (Artificial or Natural) × Spanish Gender (Masculine or Feminine) ANOVA with language as a between-subjects variable and conceptual class and Spanish gender as within-subjects variables. We observed several reliable effects: a main effect of conceptual class, $F(1, 38) = 43.34, p < .001$; a main effect of Spanish gender, $F(1, 38) = 107.7, p < .001$; a Language × Class interaction, $F(1, 38) = 7.11, p < .01$; a Language × Gender interaction, $F(1, 38) = 74.0, p < .001$; and a three-way interaction between language, class, and gender, $F(1, 38) = 7.8, p < .01$.

We examined these effects more closely through a series of Tukey tests. The Conceptual Class × Language Group interaction indicated that English speakers honored the artificial-male/natural-female link more often than Spanish speakers, whose judgments followed a similar pattern but for whom the overall difference between artificial and natural items was not statistically reliable (H.S.D. = 14.8, $p < .05$). The Spanish Gender × Language Group interaction indicated that Spanish speakers classified grammatically feminine items as having a woman's voice more often than grammatically masculine items (H.S.D. = 17.0, $p < .05$). English speakers showed a similar trend that was not statistically reliable. The three-way interaction between language group, conceptual category, and Spanish gender clarified these effects (Tukey's H.S.D. = 13.3, $p < .05$). English speakers attributed a woman's voice to natural items that were grammatically feminine more often than to natural items that were grammatically masculine (74% vs. 55%). We observed no evidence that English speakers were sensitive to the Spanish grammatical gender classifications among the artificial items. However, Spanish speakers attributed a man's voice more often to grammatically masculine items that were artificial than to grammatically masculine items that were natural (96% vs. 75%). They showed a similar pattern for the grammatically feminine items—attributing a woman's

voice more often to the natural items than the artificial ones (97% vs. 90%). For example, the Spanish speakers attributed a woman's voice more often to a picture of an apple than to a picture of an arrow, although both *apple* and *arrow* are grammatically feminine. Thus, the Spanish speakers were also sensitive to the natural-feminine/artificial-masculine link in their classifications.

These results indicate that both conceptual category and grammatical gender are influencing the judgments of English and Spanish speakers. For example, English speakers attributed a woman's voice to 64% of all of the natural items but 74% of the time to grammatically feminine natural items and 55% of the time to grammatically masculine natural items. Similarly, Spanish speakers attributed a man's voice to 85% of the grammatically masculine items but 96% of the time to the artificial grammatically masculine items and 75% of the time to the natural grammatically masculine items. For English speakers the natural-artificial distinction yielded reliably different judgments overall. For Spanish speakers, the grammatical gender classifications yielded reliably different judgments overall. Thus, conceptual class was a stronger influence than Spanish gender among the English speakers, but Spanish gender was a stronger influence than conceptual class among the Spanish speakers.

Discussion

Using a different task, we replicated our findings from Experiment 1. Once again we observed a close relationship between the Spanish speakers' grammatical gender system and their conceptual classifications of gender. The relationship between grammatical and conceptual gender, at least among adult speakers of Spanish, is not arbitrary and nonsensical but involves a linking of language and thought. Our evidence clearly indicates that grammatical gender schemes are not encapsulated from conceptual ones. Replicating our findings with a task that should have driven subjects away from explicitly referring to grammatical gender strengthens the case in favor of linguistic relativity.

We also replicated our findings from Experiment 1 that the artificial-male/natural-female link is a force in both English and Spanish speakers' judgments, but an especially strong force for English speakers. English speakers were surprisingly consistent with their classifications, judging naturally occurring objects such as *feathers* and *apples* as female-like and artificial objects such as *helicopters* and *wheels* as male-like. Our finding that language effects in speakers of Spanish are regulated by the conceptual classification of the objects as natural or artificial also minimizes the possibility that Spanish speakers were explicitly referring to grammatical gender. If Spanish speakers had been explicitly referring to grammatical gender, why did they do so more often when grammatical and cultural classifications agreed than when they disagreed? In other words, if conscious reference to grammatical gender was the single force guiding the classifications of the Spanish speakers, we cannot account for the fact that they classified items such as *feathers* and *apples* more often according

to grammatical gender than items such as *wheels* and *arrows*. In sum, the hypothesis that Spanish speakers explicitly referred to grammatical gender in this task fails to explain the evidence.

Of course, we recognize that objects can be classified according to many dimensions, some of which are more salient than others, depending on the task. Indeed, the study of all psychological phenomena depends on the methods that are used to measure them. In this article, we were interested in the attribution of gender by English and Spanish speakers. As such, we employed tasks that measured the attribution of gender. We would not be surprised to discover that grammatically and conceptually driven gender classifications are limited to tasks in which the dimension of gender matters. In our view, tasks that measure the attribution of gender, measure a "real" aspect of normal classification. Convinced that adult speakers of English and Spanish differ in their attribution of gender on the basis of language-driven and conceptually-driven mechanisms, we sought to examine the time course of such influences in Experiment 3.

EXPERIMENT 3

Our goal in this experiment was to begin to understand the interplay between conceptual and grammatical factors in the organization of gender categories. In Experiments 1 and 2, we showed that grammatical gender was a strong influence in adult Spanish speakers' attribution of gender. Our results also indicated that the natural/artificial contrast influences both English- and Spanish-speaking adults' classifications. Mullen (1990) reported that English speakers honor the artificial-masculine/natural-feminine division by 7 years of age. Several questions thus arise concerning the relation between what is thought to be a conceptual universal (the natural-female/artificial-male link) and what we have shown to be a language-specific organization of gender: Is one kind of organization more basic and thus observed earlier in development? At what age are Spanish-speaking children conceptually sensitive to grammatical gender? What happens when the conceptual classification predicts a classification opposite that of the grammatical distinction? What happens when the conceptual classifications coincide with the grammatical ones? Although the results from a single experiment are unlikely to offer satisfactory answers to all of these questions, we began to examine the classification of the individual items by English- and Spanish-speaking children and adults in Experiment 3. By examining the classification of individual items for which grammatical and conceptual values yield competing and converging classifications, we can begin to understand the fine-grained relationship between culturally shared conceptual forces and language-specific grammatical ones in the psychological organization of gender.

Method

Subjects. Twenty kindergarten, second grade, fourth grade, and adult native speakers of English and Spanish participated (N = 160). The English-speaking kindergartners (10 boys and 10 girls) ranged in age from 5;4 (years; months) to 6;1 with a mean age of 5;6. The Spanish-speaking kindergartners (10 boys and 10 girls) ranged in age from 5;0 to 5;10 with a mean age of 5;4. The English-speaking second-graders (13 boys and 7 girls) ranged in age from 7;6 to 8;9 with a mean age of 8;0. The Spanish-speaking second-graders (13 boys and 7 girls) ranged in age from 7;0 to 7;10 with a mean age of 7;5. The English-speaking fourth-graders (10 boys and 10 girls) ranged in age from 8;11 to 10;8 with a mean age of 10;0. The Spanish-speaking fourth-graders (10 boys and 10 girls) ranged in age from 8;1 to 9;10 with a mean age of 9;4. Equal numbers of men and women from each language group also participated in the adult age group. The English-speaking adults were either students or staff members of the University of Minnesota. The Spanish-speaking adults were either students of the Complutense University or government employees in Madrid, Spain. The Spanish-speaking children were recruited from a public school in Madrid and were tested by a local, native speaker of Spanish. The English-speaking children were recruited from a public, after-school program in Hopkins, Minnesota. The English speakers were tested by a local, native speaker of English. All participants were predominantly of middle socioeconomic status.

Procedures. For the most part, the procedure and design followed those of Experiment 2 with a few modifications. As in Experiment 2, children (and adults) were told that we were thinking of making a movie in which inanimate objects come to life and we wanted them to tell us what kind of voice the item should have in the movie. In contrast to Experiment 2, the pictures were labeled with both the article and the noun (e.g., *el rey* or *the king* and *la manzana* or *the apple*) for the subjects who received the pictures with their labels, as this seemed a more natural presentation for children. Subjects in this study were also asked to *tell* the experimenter whether the pictured object should have a man's or a woman's (or a boy's or a girl's) voice in the movie because all of the children could not write their responses. All subjects were tested individually in one 10-min. session, and the experimenters recorded their answers.

Results and Discussion

We first examined performance on the control items, which had a natural gender that corresponded to the item's grammatical gender in Spanish. Subjects of all ages and in both language groups classified the items with a natural gender according to the item's natural gender 97% of the time. Thus, we can be confident that all subjects understood the task and that their classifications reflected their attribution of natural-like gender properties.

We next examined the percentage of classifications according to the Spanish language for the test items through a four-way ANOVA with grade (kindergarten, second, fourth, or adult), language (English or Spanish), condition (pictures only or pictures + labels) as between-subjects variables and Spanish gender (masculine or feminine) as a within-subject variable. This analysis yielded significant main effects of grade, $F(3, 144) = 10.3$, $p < .001$; language, $F(1, 144) = 52.3$, $p < .001$; and Spanish gender, $F(1, 144) = 53.7$, $p < .001$. We also observed a significant interaction between grade and language, $F(3, 144) = 8.5$, $p < .001$ and a marginally significant three-way interaction between grade, language, and condition, $F(3, 144) = 2.3$, $p = .084$. Figure 2 shows the percentage of classifications according to the Spanish language by English and Spanish speakers as a function of grade for the grammatically masculine and feminine items.

We interpret the main effects of age and language in the context of a two-way interaction. This interaction reflected significant simple effects between the English and Spanish speakers in the second grade and older ($p < .02$ at the second grade, and $p < .001$ at the fourth grade and among adults). However, we observed no significant differences between the English- and Spanish-speaking kindergartners ($p = .97$). Thus, according to this analysis Spanish-speaking kindergartners classify the objects in the same manner as English-speaking kindergartners. Although exact probability statements cannot be made from post hoc comparisons on an interaction that does not reach the conventionally accepted alpha level of .05, the marginally significant three-way interaction between age, language, and condition ($p = .084$) suggests a label effect for the Spanish-speaking adults ($p < .05$ by Duncan's test) such that Spanish-speaking adults were more likely to classify the test items according to grammatical gender when they were labeled (84%) than when they were shown the pictures alone (72%). Spanish-speaking fourth graders showed the same trend (80% vs. 75%). Taken together, all of these results indicate that language-driven differences between English and Spanish speakers in their classification of everyday objects begin in second grade and become stronger and more tightly linked to language with development.

As in Experiments 1 and 2, we next asked how grammatically feminine and masculine natural and artificial objects were classified. We classified the test items into grammatically feminine natural items, grammatically feminine artificial items, grammatically masculine natural items, and grammatically masculine artificial items. We then calculated the percentage of times subjects assigned a woman's (or girl's) voice to each of these four kinds of items. We statistically examined these percentages through a Language (English or Spanish) × Grade (Kindergartners, Second, Fourth, or Adult) × Conceptual Category (Natural or Artificial) × Spanish Gender (Masculine or Feminine) ANOVA with language and age as between-subjects variables and conceptual category and Spanish gender as within-subjects variables. Significant main

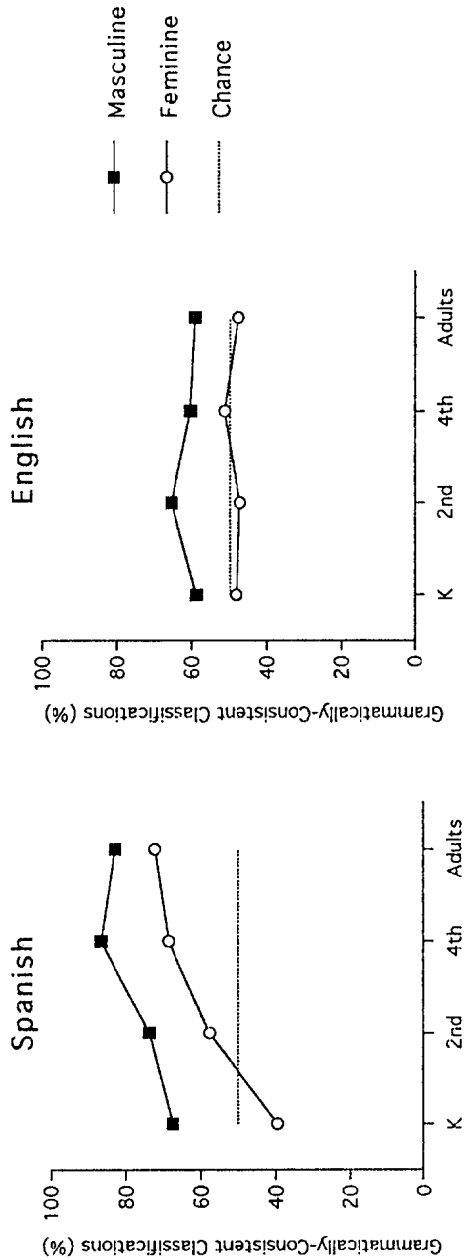

Figure 2. The percentage of times English and Spanish speakers classified the items according to the Spanish grammatical gender as a function of age in Experiment 3.

effects of language, $F(1, 152) = 4.1$, $p < .05$; conceptual category, $F(1, 152) = 11.1$, $p < .01$; and Spanish gender, $F(1, 152) = 111.2$, $p < .001$ emerged from this analysis. We also observed two-way interactions between conceptual category and Spanish gender, $F(1, 152) = 33.3$, $p < .001$; between language and conceptual category, $F(1, 152) = 4.9$, $p < .05$; between language and Spanish gender, $F(1, 152) = 41.3$, $p < .001$; between age and Spanish gender, $F(3, 152) = 8.2$, $p < .001$; and between language, age, and Spanish gender, $F(3, 152) = 8.6$, $p < .001$.

All of the effects involving language, age, and Spanish gender confirm those of our main analysis—Spanish speakers in second grade and older classified the test items according to the Spanish grammatical gender unlike speakers of English. The Language Group × Conceptual Category interaction revealed an effect of conceptual category among English speakers who classified the natural items as feminine 48% of the time, and the artificial items as masculine 60% of the time. We observed no overall effect of conceptual category among Spanish speakers of any age. Thus, although the English-speaking kindergartners differentiated among the items on the basis of the male-artificial/female-natural dimensions, Spanish-speaking kindergartners honored neither the conceptual distinction between natural and artificial kinds nor the Spanish grammatical distinction honored by older Spanish-speaking children and adults.

The Spanish Gender × Conceptual Category interaction showed an effect of conceptual category among the grammatically feminine items (the natural items were classified as female-like 60% of the time whereas the artificial items were classified as female-like 45% of the time) overall by speakers of both language groups. This interaction also revealed no effect of conceptual category for the grammatically masculine items (the natural items were classified as male-like 72% of the time and the artificial items were classified as masculine 67% of the time) by both language groups. This interaction suggests an asymmetrical relationship between the gender assignment of the Spanish language and the natural-artificial distinction: Within the grammatically feminine items, the natural/artificial distinction is observed by both language groups. However, within the grammatically masculine items, there is no such effect. English and Spanish speakers classified these items as male-like. Taken together with the finding from our main analysis earlier—that the Spanish grammatically masculine items were more likely to be judged as male-like by speakers of both groups—it may be that the grammatically masculine items capture a conceptual tendency that is more predictive of the English (and of course the Spanish) speakers' classification than the artificial-male link.

Lastly, we examined the way each particular item was classified by speakers of each language. For each item, we performed two chi-square tests—one compared the number of Spanish-speaking subjects that classified the item according to the Spanish language relative to the number that would be expected by chance; the other chi-square compared the number of English-speaking

subjects that classified the item according to the Spanish language relative to chance. The results from these tests enabled us to address the following three questions: (a) To what degree was each item classified according to chance? (b) Did the English and Spanish speakers agree on the gender of any items? If so, which ones? and (c) Did the English and Spanish speakers disagree on the gender of any items? If so, which ones and in what manner?

Several findings emerged from these item analyses. One is that nearly all of the test items (20 out of 23) were classified systematically by at least one language group. Only three objects—*guitar, spider,* and *ice cream* were classified at levels equal to chance by both language groups. As captured by our previous findings, Spanish speakers classified 16 of the 23 test items statistically different from what would be expected by chance and according to the Spanish grammatical gender. More surprising was our finding that English speakers also were consistent in their attribution of gender to 10 of the test items. Of these 10 items, they agreed with the Spanish speakers on the gender of *feather, grapes, spoon, fire, telescope,* and *helicopter*—items for which the Spanish gender system nearly always agrees with the natural-female/artificial-male distinction (*fire* was the exception as it is a naturally occurring kind but the Spanish language classifies it as male-like, as did both speakers of English and Spanish). Four of the items—*grapes, spoon, fire,* and *telescope*—were even more closely classified according to the Spanish gender by speakers of Spanish, as indicated by the different levels of alpha. Thus, the Spanish grammatical gender is driving the Spanish speakers' classifications even further in the direction that is suggested by the artificial-natural contrast when grammatical and conceptual organizations converge.

The results from the item analyses also showed that English speakers' judgments differed from chance (and disagreed with the gender prescribed by the Spanish language) on *arrow, ear,* and *wheel,* as they attributed male voices to these items that are grammatically feminine according to Spanish. Spanish speakers classified these items according to chance. Thus, for items on which the conceptual natural-artificial sorting contradicts the Spanish grammatical classification, the effects of language are observed but the language influence is not strong enough to lead the Spanish speakers to classify the objects in the opposite direction. Finally, we found that the English speakers classified 13 items—*apple, guitar, spider, star, table, book, eye, fish, fork, ice cream, knife, shoe,* and *square*—at levels equivalent to chance whereas Spanish speakers classified 10 of these items—*apple, star, table, book, eye, fish, fork, knife, shoe,* and *square*—according to the Spanish grammatical gender. Thus, for items on which clear-cut conceptual classifications do not exist (as evidenced by the English speakers' chance judgments), the Spanish grammatical gender system determines the direction of the classification for older Spanish-speaking children and adults.

In sum, four basic findings can be taken from this experiment: (a) We showed that Spanish speakers differ from English speakers in their attribution of gender by second grade, with Spanish speakers classifying the objects according to the grammatical gender of the Spanish language more often than English speakers; (b) we found that English speakers honor the artificial-male/natural-female distinction in kindergarten, earlier than Spanish-speaking children honor the grammatical contrast; (c) we found that both language groups were sensitive to the artificial-natural contrast within the Spanish feminine items, but found no evidence of an artificial-natural contrast within the Spanish masculine items—both groups tended to classify these items as male-like; and (d) we showed through our item analyses three ways that grammatical and conceptual factors interact to generate systematic classifications within Spanish speakers. Only a few objects were classified randomly by speakers of both languages. Figure 3 shows examples of the four patterns of classification that we observed. We expand our discussion of these findings later.

GENERAL DISCUSSION

We began this work by asking whether the grammatical gender system of the Spanish language offers a referential device to its speakers that leads them to systematically classify objects according to gender. Our findings offer convincing evidence that they do. Grammatical gender is not modular and encapsulated from conceptual classifications. In Experiment 1, in which adult speakers of English and Spanish classified objects that lack a natural gender as feminine or masculine, we found that the Spanish speakers, unlike the English speakers, honored the Spanish grammatical gender classifications. The effect of grammatical gender was more pronounced for the Spanish speakers when the pictures were labeled, pointing specifically to the role of language in their classifications. By manipulating the effect of grammatical gender we were able to separate linguistic from nonlinguistic cultural influences on categorization–a critical piece missing from previous work on linguistic relativity. In Experiment 2, in which English- and Spanish-speaking adults participated in a task in which they attributed either a man's or woman's voice to each pictured item—a task that should have led the Spanish speakers *away* from explicitly referring to grammatical gender—the Spanish speakers continued to classify according to the Spanish gender. Replicating the language effects in adult speakers of Spanish across a different task, together with similar findings that have been reported among speakers of Italian, Arabic, and other languages, strengthens the claim of a meaningful link between grammatical and conceptual organizations of gender. Taken together, our findings offer evidence that the relation between grammatical and conceptual classifications of gender is not arbitrary and nonsensical, but instead involves a coupling of thought to language with development.

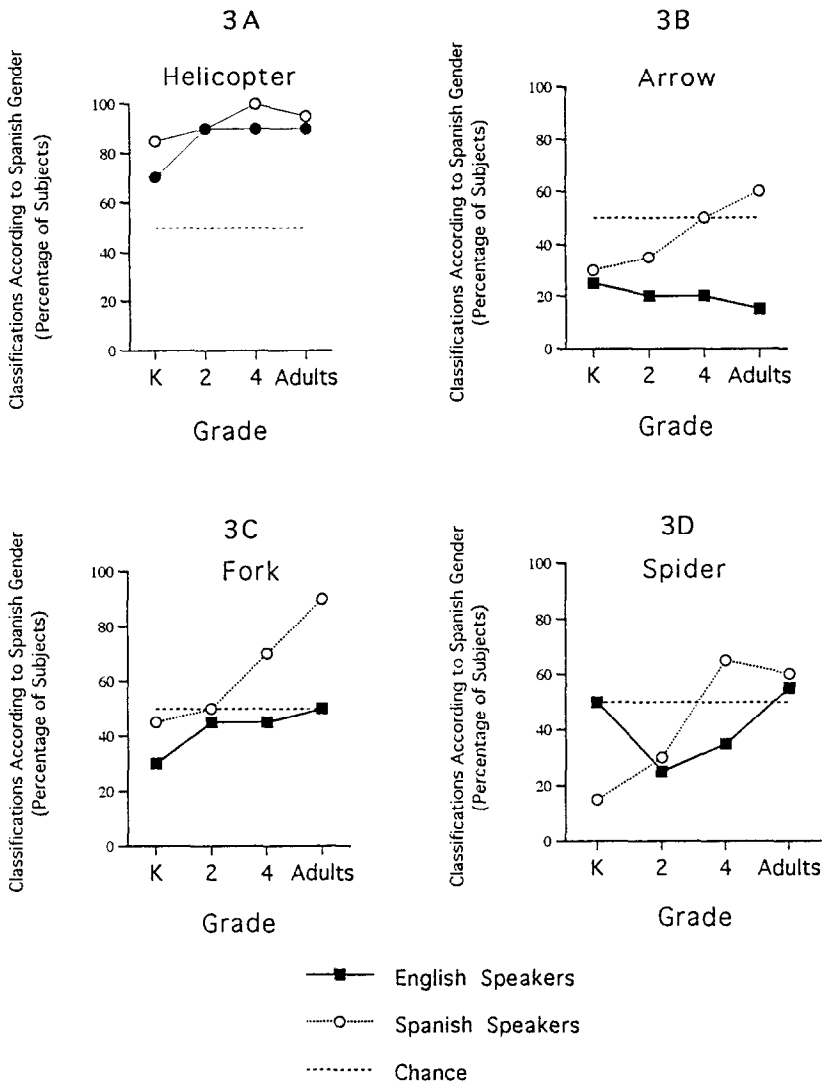

Figure 3. The percentage of English and Spanish speakers who classified the pictures of a *helicopter, arrow, fork,* and *spider* according to the Spanish grammatical gender as a function of age in Experiment 3. Each panel illustrates a pattern of performance. Panel 3A illustrates the pattern observed when grammatical and conceptual classifications by English speakers converged. Panel 3B illustrates the pattern that we found when grammatical and conceptual forces yielded opposite classifications. Panel 3C illustrates the case in which the grammatical classifications drive the judgment, as the conceptual classification by English speakers is unclear. Panel 3D illustrates the small number of cases in which we found random classifications by both language groups.

We also showed that English speakers were systematic in their classifications, honoring an artificial-male/natural-female conceptual division. It appears that even speakers of a language without a grammatical gender system, such as English, have a "folk theory" of gender that enables them to assign gender to objects that lack natural gender. For speakers of a language with a grammatical gender system, grammatical gender constitutes a reference point within that speaker's folk theory of gender, which also generates internally consistent and nonarbitrary classifications. Only when we attempt to learn a language with a gender classification system that is different from our own do we see that language's gender assignments as nonsensical and consisting of exceptions. Disagreement between languages is not evidence of an arbitrary relation between grammatical and conceptual gender organizations, but only evidence against a *universal* folk theory of gender. When grammatical classifications disagree across languages, as in the Spanish and German classifications of *fork* and *table*, we believe the folk theories of monolingual speakers of those languages will also disagree. In this sense, our findings provide the necessary foundation for many cross-linguistic questions regarding the organization of gender.

What kind of mechanism might underlie the coupling between a grammatical and a conceptually meaningful understanding of gender? One obvious vehicle is the shared morphological properties between words that refer to objects with a natural gender and those that lack one. Simply put, the classic psychological evidence on stimulus generalization may offer the proper tool for understanding the relation between grammatical, semantic, and conceptual gender. The meaningful extension of gender to objects that lack a natural gender would seem to be based on the understanding that an accurate semantic classification of natural gender is often achieved through grammatical gender markings. The attribution of semantic gender subsequently spreads to objects referred to by words with the same morphological properties. To the extent that languages of the world vary in the amount of morphological information that marks grammatical gender (e.g., agreement between articles, singular and plural nouns, and adjectives), we expect the impact of grammatical gender to vary accordingly.

This extension of semantic gender to referents that lack a natural gender would then properly place the differences between English and Spanish speakers that we found as differences in the *metaphorical* organization of gender, where the sharing of morphological properties among the set of Spanish feminine and masculine referents leads to a perceived similarity among the semantic and conceptual properties of these referents by speakers of Spanish. Now, just because we called the differences that we observed metaphorical does not mean that these differences are unimportant. On the contrary, we agree with writers such as Lakoff (1987) who claimed that metaphorical structures are central to categorization and cognition. By this view, categorization is largely a matter of making a judgment on the basis of the different similarity relations—both physical and abstract—available to individuals at any point in time. Attention to

metaphorical relations is an important aspect of categorization, not something peripheral and secondary.

Our finding that speakers of Spanish do not classify the items according to the Spanish grammatical gender system until second grade points to a late coupling of grammatical and conceptual gender. We believe that this delay reflects the young Spanish-speaking child's failure to recognize that grammatical gender often marks natural gender accurately. That is, they may not realize that for many words that refer to animates—such as the Spanish words for girl *(niña)* and boy *(niño)*—the grammatical gender cues signal the natural gender of the referent. The evidence from studies such as Pérez-Pereira's (1991) supports our claim. Pérez-Pereira found that the referent's natural gender was the latest source of information that Spanish-speaking children integrated into their knowledge of gender agreement rules. Given the young child's often-cited confusion about natural gender, it is not too surprising that morphological cues to natural gender are one of the latest sources of information that become a part of gender concepts.

Of course, because grammatical gender markings did not perfectly predict the attribution of gender by speakers of Spanish, other factors play a role in the metaphorical organization of gender concepts. We believe those forces are culturally shared, nonlinguistic, conceptual factors. Nonlinguistic conceptual factors led the English speakers, whose language lacks a grammatical gender system, to systematically classify many of the test items that we originally believed were ambiguously gendered. We found, like Mullen (1990), that the English speakers classified artificial objects more often as male-like and natural pictures more often as female-like. Apparently, the natural-female/artificial-male link is a conceptual force that organizes English speakers' gender concepts as early as 5 years of age. Does the natural-artifact distinction also account in part for the Spanish speakers' judgments? Our answer is yes. When the conceptual classification coincides with the grammatical one, the Spanish speakers' judgments were in the same direction, but more extreme, than the English speakers'. When the conceptual classification is opposite to that of the grammatical one, the grammatical classifications disrupted the conceptual classifications which resulted in classifications equivalent to chance by speakers of Spanish. When the conceptual classifications were not clearly specified, the English speakers classified the items at levels equivalent to chance, whereas grammatical gender determined the classification for speakers of Spanish. In short, the magnitude and direction of the grammatical gender effects among the Spanish speakers depended on the conceptual alignment of the items.

Another aspect of our findings that points to shared conceptual forces in both English and Spanish speakers' judgments is that speakers of both languages were more likely to attribute a man's voice to objects that are classified as masculine by the Spanish language than a woman's voice to objects classified as feminine across all three experiments (e.g., see Figure 2). A simple bias to attribute male

voices to all test items does not explain why English speakers failed to attribute male voices to the feminine items. If English speakers simply had a bias to attribute male voices to all of the objects, then they should have classified the grammatically feminine items at levels lower than would be expected by chance. However, English speakers classified the grammatically feminine objects at levels expected by chance. One possible explanation of the asymmetry could be related to linguistic markedness. Traditional accounts of marked and unmarked connotations of gender, however, do not explain our findings either. By traditional accounts, in languages that possess a grammatical gender system such as Spanish, the masculine forms are argued to be semantically unmarked because when referring to a group of people or animals of both genders, the term that carries the masculine morphological properties is used. For example, one would use the masculine form *doctores* to refer to a group of male and female doctors; use of the feminine form *doctoras* would imply a group of female doctors. Because masculine terms imply both male and female group members, they are classified as unmarked. Thus, by these accounts one would predict that linguistically masculine items possess more semantic freedom than feminine items. Yet we observed the opposite, that grammatically masculine items tended to denote masculinity more often than linguistically feminine forms denoted femininity. For these reasons, we believe that the best explanation of these findings lies in a nonarbitrary link between the Spanish grammatically masculine items and conceptually male-like dimensions. One link, suggested by our findings, is that the distribution of grammatically masculine items in the Spanish language is skewed to include more nouns that refer to artificially occurring kinds than nouns that refer to naturally occurring kinds, unlike the distribution of grammatically feminine items. However, because some natural kinds were classified as male-like by speakers of both groups (e.g., *fire*), we do not want to rule out the possibility that the grammatically masculine items capture other culturally shared, nonlinguistic conceptual or perceptual dimensions associated with males. The final explanation of this asymmetry awaits a systematic distributional analysis of grammatically masculine and feminine terms, within and across languages that possess grammatical gender systems.

In conclusion, our findings illuminate the relation between grammatical and conceptual forces in categorization. We found early effects of conceptually driven associations among speakers of English and delayed effects of grammatical gender among speakers of Spanish. The language effect among Spanish speakers was related to the direction and magnitude of the conceptual influence. The multiple relationships between grammatical and conceptual forces that we observed among speakers of Spanish, taken together with the early effects of concepts in speakers of English, and the late effects of grammar in speakers of Spanish all point to the superimposition of grammatical forces on conceptual ones in the Spanish speaker's organization of gender. Our method of isolating and studying the effects of language cross-linguistically and developmentally

provides an important piece to the puzzle on the relation between language and thought because it enables us to predict *when* and *what kind* of language effects to expect.

Results such as ours are helping to turn the tide on critics of linguistic relativity. Researchers focusing on everything from metaphor (Lakoff, 1987), and child language (Choi & Bowerman, 1991; Slobin, 1990), to cognitive development (Sera, Reittinger & del Castillo Pintado, 1991; Smith & Sera, 1992) and emotion (Lutz & White, 1986; Shweder & Bourne, 1984) are providing evidence in support of linguistic relativity. Even the traditional findings on perceptual color processing (Berlin & Kay, 1969; Heider, 1972) are being called into question (Davies et al., 1991). In an attempt to separate historical baggage from a more useful interpretation of linguistic relativity, we conclude with the modified version of the Sapir–Whorf hypotheses offered by Lucy and Wetsch (1987):

> Hypothesis 1. Language is composed not merely of forms but of meaningful forms, utilizing a finite number of devices to refer to an infinite variety of experience.
>
> Hypothesis 2. A language provides its speakers with a ready-made classification of experience which may be used as a guide for thought; furthermore, linguistic classifications vary considerably across languages.
>
> Hypothesis 3. Linguistic categories are used as analogical guides in habitual thought: a speaker in attempting to interpret an experience will use a category available in his or her language, suggesting associations not necessarily entailed by experience. (Lucy & Wertsch, 1987, p. 73)

These insights together with the evidence we offer indicate, as Sapir and Whorf suggested long ago, that language and thought are not always independent functions but closely related agents that jointly shape human experience.

REFERENCES

Abbatte, M., & LaChappelle, N. (1979). *Pictures, please! A language supplement.*

Bem, S.L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88,* 354–364.

Berlin, B., & Kay, P. (1969). *Basic colour terms: Their universality and evolution.* Baltimore: University of California Press.

Braine, M. (1987). What is learned in acquiring word classes—a step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Erlbaum.

Brown, R. (1976). Reference: In memorial tribute to Eric Lenneberg. *Cognition, 4,* 125–153.

Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology, 49,* 454–462.

Bull, W. (1965). *Spanish for teachers.* New York: Ronald Press.

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* (p. 18–31). New York: Henry Holt.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41,* 83–121.

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

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

Davies, I., Corbett, G., Laws, G., McGurk, H., Moss, A., & Smith, M. (1991). Linguistic basicness and colour information processing. *International Journal of Psychology,* 26, 311–327.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L., & Dunn, L. (1986). *Test de Vocabulario en Imagenes Peabody* [Peabody Picture Vocabulary Test]. Circle Pines, MN: American Guidance Service.

Ervin, S.M. (1962). The connotations of gender. *Word, 18,* 249–261.

Gathercole, J. (1989). The acquisition of sex-neutral uses of masculine forms in English and Spanish. *Applied Psycholinguistics, 10,* 401–427.

Gelman, R., & Gallistel, R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Guiora, A.Z., & Acton, W.R. (1979). Personality and language behavior: A restatement. *Language Learning, 29,* 193–204.

Heider, E.R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93,* 10–20

Hernández-Pina, F. (1984). *Teorías psicosociolingüísticas y su aplicación del español como lengua materna.* [Psycholinguistic theories and their application to the acquisition of Spanish as a native language]. Madrid: Silgo XXI.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference.* Cambridge: Cambridge University Press.

Kay, P., & Kempton, W. (1984). What is the Sapir–Whorf hypothesis? *American Anthropologist, 86,* 65–79.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

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

Levi-Strauss, C. (1969). *The raw and the cooked* (J. Weightman & D. Weightman, Trans.). New York: Harper & Row.

Liben, L.S., & Signorella, M.L. (1993). Gender-schematic processing in children: The role of initial interpretations of stimuli. *Developmental Psychology, 29,* 141–149.

Lucy, J., & Wertsch, J. (1987). Vygotsky and Whorf: A comparative analysis. In M. Hickmann (Ed.), *Social and functional approaches to thought and language* (pp. 67–85). New York: Academic.

Lutz, C., & White, G. (1986). The anthropology of emotions. *Annual Review of Anthropology, 15,* 405–436.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43* (1–2, Serial No. 174).

Maratsos, M. (1983). Some current issues in the study of the acquisition of grammar. In P. Mussen (Ed.), *Handbook of child psychology: Vol. 3 Congitive development* (pp. 707–786). New York: Wiley.

Maratsos, M. (1988). *The acquisition of formal word classes.* In Y. Levy, I. Schlesinger, & M. Braine (Eds.), *Categories and processes in language acquisition* (pp. 31–44). Hillsdale, NJ: Erlbaum.

Markman, E.M., & Hutchinson, J.E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16,* 1–27.

Martin, C.L., & Halverson, C.F. (1981). *A schematic-processing model of sex typing and stereotyping in children. Child Development, 54,* 563–574.

Mueller-Gathercole, V.C., & Hasson, D.J. (1991, April). *Gender marking in Spanish: Linguistic versus sociological determinants of feminine form in words for humans.* Paper presented at the Conference on Pragmatics and Language Learning, University of Illinois, Urbana.

Mullen, M.K. (1990). Children's classifications of nature and artifact pictures into female and male categories. *Sex Roles, 23,* 577–587.

Ortner, S.B. (1974). Is female to male as nature is to culture? In M. Rosaldo & L. Lamphere (Eds.), *Woman, culture and society* (pp. 67–87). Stanford, CA: Stanford Press.

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

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language, 18,* 571–590.

Ruble, D.F., & Stangor, C. (1986). Stalking the elusive schema: Insights from developmental and social psychological analyses of gender schemas. *Social Cognition, 4,* 227–261.

Sapir, E. (1912). Language and environment. *American Anthropologist,* n.s. 226–242.

Sera, M., Reittinger, E., & del Castillo Pintado, J. (1991). Developing definitions of objects and events in English and Spanish speakers. *Cognitive Development, 6,* 119–142.

Shweder, R., & Bourne, E. (1984). Does the concept of the person vary cross-culturally? In R. Shweder & R. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 158–199). New York: Cambridge University Press.

Slobin, D. (1990). The development from child speaker to native speaker. In J. Stigler, R. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 233–256). Cambridge: Cambridge University Press.

Smith, L., & Sera, M. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24,* 99–142.

Terrell, T., & Salgues de Cargill, M. (1979). *Linguística aplicada* [Applied linguistics]. New York: Wiley.

Waxman, S. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5,* 123–150.

Whorf, B.L. (1956). The relation of habitual thought and behavior to language. In J.B. Carroll (Ed.), *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (pp. 134–159). Cambridge, MA: MIT Press.