Journal of Memory and Language 66 (2012) 17–37

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Number meaning and number grammar in English and Spanish

Kathryn Bock [a,*], Manuel Carreiras [b,c], Enrique Meseguer [d]

[a] University of Illinois, Urbana, IL 61801, USA
[b] Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain
[c] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[d] Universidad de La Laguna, Tenerife, Spain

ARTICLE INFO

Article history:
Received 9 October 2010
revision received 15 July 2011
Available online 30 August 2011

Keywords:
Number agreement
Number meaning
Notional number
Grammatical number
Linguistic relativity
Number morphology

ABSTRACT

Grammatical agreement makes different demands on speakers of different languages. Being widespread in the languages of the world, the features of agreement systems offer valuable tests of how language affects deep-seated domains of human cognition and categorization. Number agreement is one such domain, with intriguing evidence that typological characteristics of number morphology are associated with differences in sensitivity to number distinctions. The evidence comes from research on language production that points to the morphological richness of languages as enhancing the expression of number distinctions. To critically test this hypothesis, native speakers of a sparse-morphology language (English) were compared with native speakers of a rich-morphology language (Spanish) in their use of semantically and grammatically motivated number agreement. With meaning-matched materials, speakers of both languages displayed significant variations in number agreement due to implicit nuances of number semantics, and the patterns and magnitudes of interaction with grammatical number were the same for both groups. In this important respect, speakers of English and Spanish appear to construe numerosity in similar ways, despite the substantial morphological and syntactic differences in their languages. The results challenge arguments that language variations can shape the apprehension of nonlinguistic number or promote differential expression of number meaning during the production of grammatical agreement.

Introduction

In English and many other languages, talking about almost anything requires a tacit evaluation of numerosity. Most English nouns come in different singular and plural forms. As a result, the use of nouns typically entails some commitment to quantity when designating a noun's referent, at least whether the referent is construed as one thing or more than one thing. So, nouns convey number meaning directly in a way that is familiar to every native English speaker. English and other languages also capitalize on number for a fundamental syntactic purpose. In grammatical number agreement, covariations in the forms of words serve the syntactic function of flagging which words or phrases modify one another: *The shape of the towers that was*… is going to say something about the shape; *The shape of the towers that were*… is going to say something about the towers. This syntactic function has nothing to do with shapes or towers or the actual topics of conversation. It nonetheless drives a need to evaluate numerosity that arises because the grammar of English demands it. The demand is unstoppable: Grammatical number variations can occur in any English sentence, requiring speakers and listeners to assess notional number (that is, the construed numerosity of an intended referent) at least once every few seconds in ordinary language use. The question we ask here is whether there are differences in how speak-

* Corresponding author. Address: Beckman Institute for Advanced Science and Technology, 405 N. Mathews Avenue, University of Illinois, Urbana, IL 61801, USA.

E-mail address: jkbock@illinois.edu (K. Bock).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2011.07.008

ers use number notionally that parallel differences in how their languages use number grammatically in agreement.

The theoretical importance of this question stems in part from its connections to linguistic determinism and relativity. Whorf (1956) contended that "the grammar ... of each language is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity" (1956, p. 212). Whorf's emphasis was on grammar, more so than single words (Lucy, 1992b), because the grammar of a language repeatedly and reliably forces speakers to use information in particular ways, and broad conceptualizations of information are likely to be shaped accordingly. Differences among grammars in how they package thoughts could cause speakers of different languages to categorize the same information differently. Likewise, speakers of dissimilar languages would be expected to vary in how adept they are in calling rapidly on concepts tailored to certain kinds of grammatical distinctions.

Take number agreement. Languages with number agreement systems force native speakers (and speakers who aspire to native-like fluency) to include information about numerosity in every nonverbal message they wish to communicate; speakers of languages without number-agreement systems face no such demand. Thus, English speakers have to represent notional numerosity virtually every time they talk, because the features associated with number agreement are essential for English utterances. There is number agreement between subjects and verbs, between nouns and determiners, and between pronouns and antecedents. In contrast, Chinese speakers are not confronted with this challenge, because number agreement is absent in their language.

Of course, along with notional number, speakers have to juggle other kinds of number information in order to implement number agreement. Linguistically, the most familiar is grammatical number, the number that covaries between linguistic elements in agreement. For example, the noun *scissors* is grammatically plural, the noun *news* is grammatically singular, the verb *were* is grammatically plural, and the verb *cuts* is grammatically singular. The usage of grammatical singulars and plurals in agreement can be independent of notional number, because the numerosity of a referent that a speaker has in mind does not always correspond to the grammatical number of a corresponding word in a referring expression. For instance, object tends to be construed as a single thing (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001), though the word that customarily refers to it (*scissors*) is grammatically plural. Conversely, what is commonly referred to as *news* tends to be evaluated as multiple things, though the word *news* is grammatically singular. Scissors *are* sometimes dull, and news *is* sometimes bad.

Languages use number grammar and number meaning in different ways and to varying degrees, making the apprehension of numerosity in the world a plausible locus of cross-linguistic, language-related variations. There is in fact tantalizing evidence that the number syntax of languages can predispose different ways of construing number notionally (Lucy, 1992a). Much of this evidence comes from research on mass and count nouns. In English, mass nouns (e.g. *toast*, *bread*) are typically treated as singulars in the grammar, despite variations in the notional-number properties of their referents. So, English speakers tend to use *toast* and *bread* to denote multiple, discrete pieces of toast and bread, even though reference to discrete objects is more often associated with nouns that alternate between singular and plural forms, count nouns like *pea* and *peas*. Yet neither mass nor count nouns have a simple link to notional number. *Pea* is a good example: In English, the fact that *pea* is a count noun (and *corn* is not) is more an accident of linguistic history than of notional number, an apparent mistaking of the old mass noun *pease* for a plural-count form.

The mass-count distinction has important syntactic and crosslinguistic implications. Syntactically, the classification of mass nouns as grammatically singular (with exceptions that convert mass to count nouns for denoting classes; e.g. *the bread***s** *of different countries*) systematically affects the grammatical number of verbs and other words that agree with mass nouns. The potential consequences of mass/count syntax for number cognition in adults and language-learning children are the focus of an extensive literature (Barner & Snedeker, 2005, 2006; Barner, Wagner, & Snedeker, 2008; McPherson, 1991; Middleton, Wisniewski, Trindel, & Imai, 2004; Soja, 1992; and many others). Crosslinguistically, the absence of mass/count syntax from some languages drives research on how speakers of such languages differ from English speakers in categorizing, individuating, and quantifying substances and objects (Barner, Inagaki, & Li, 2009; Imai & Gentner, 1997; Iwasaki, Vinson, & Vigliocco, 2010; Li, Dunham, & Carey, 2009; Lucy, 1992a). However, there are important differences between the grammatical challenges posed by mass and count nouns and the challenges posed by number agreement more broadly. The next section sketches some of these differences.

Mass/count syntax and number agreement

A principal reason for the crosslinguistic significance of mass/count syntax is that the distinction is in some ways semantically arbitrary (Quine, 1960). The grammatical treatment of mass and count nouns does not consistently capture a distinction between substances and objects, raising the possibility that the grammatical distinction on its own could affect whether referents are construed as substances or individuals. English speakers might construe a heap of small smoothish stones as gravel (an uncountable substance) or as pebbles (countable objects); the term *pea gravel* represents a compromise between these competing conceptions. The arbitrariness of grammatical number in these instances highlights the possibility that mass nouns, as a class, could be treated grammatically as invariantly plural, rather than invariantly singular. Put differently, the grammatical singularity of mass nouns is a conventional, learned property that is associated with individual words. The words themselves (not the notional number of a referent on a given occasion) determine any accompanying syntactic variations (Vigliocco, Vinson, Martin, & Garrett, 1999).

In contrast, count nouns require a tacit choice between grammatically singular (*dog*) and plural (*dogs*). The choice is often dependent on the number meaning relevant to a speaker's message. But again, once a count noun and its number are chosen, accompanying syntactic variations tend to reflect the choice. As a result, in the syntax of number agreement, when a noun is a sentence subject its verb often carries a corresponding number: The conventional rule is that singular subjects take singular verbs, and plural subjects take plural verbs. In these instances, noun number and verb number can be traced back to a notional property of the speaker's message, although the agreement patterns themselves seem to depend on specific noun choices.

However, the syntax of number agreement has wrinkles that leave specific noun properties out of the equation (Morgan, 1972, 1984). Nouns occur in noun phrases, and noun phrases are compositional and productive, just like the rest of syntax: The noun *boy* can occur in noun phrases ranging from *a boy* through *a boy who was found guilty on the basis of evidence that he had shot his parents,* and beyond. Although the noun chosen for a subject noun phrase is highly correlated with verb number, what controls agreement isn't the single noun chosen for a noun phrase, but the number chosen for the phrase as a whole. Because phrase number controls agreement, and not individual noun number, there are occasions when the association between noun number and phrase number falls apart. For example, when the noun *three* serves as a subject, it can be singular (*The three is still missing* might refer to a digit) or plural (*The three are still missing* might refer to three hostages). The variation in agreement is due to the phrasal features, roughly the features of a referring expression, that are acquired by the word *three*.

As this example illustrates, lying behind variations in phrasal number are often variations in notional number (Eberhard, Cutting, & Bock, 2005). Noun phrases are referring expressions, and the things that they refer to have notional number properties. Such properties can trump conventions of grammatical agreement and still be perfectly acceptable. Compare the conjunction *snow and rain* (both mass nouns) with the conjunction *toast and bacon* (likewise mass nouns). Conventionally, the grammatical number of a conjunction is plural regardless of whether the conjoined nouns are plural, and so the verbs that accompany them should be plural, too. Yet the treatment of phrases like *snow and rain* and *toast and bacon* in agreement relations is reliably influenced by their notionally determined grammatical number: In corpus counts of spoken and written language and in controlled testing, the phrase *snow and rain* is more likely to be treated as singular than *toast and bacon* (Lorimor, 2007; Lorimor, Middleton, & Bock, 2006). Why should this be? Snow and rain are readily construed as one weather event (a so-called "winter mixture"), yielding dispiriting forecasts like *Snow and rain is in store for tonight,* with the singular verb *is.* Toast and bacon are more likely to be regarded as two distinct entities, as in *Toast and bacon get slimy when cold*, with the plural verb *get.*

Sometimes phrasal number becomes visible only in agreement. As a result, variations in subject-phrase number may be evident only in verb number, as a consequence of the control of verb number by subject number. The diagnosis can be easy, because almost all sentences have verbs. However, the creation of the relevant number properties during language production, by individual speakers, need not be easy at all. The compositionality of phrases means that there are cases in which subject number is not a straightforward product of noun number or the form of the subject phrase. Speakers compose phrases that they have never uttered before, so number has to be calculated, not retrieved from memory. Moreover, the rate of ongoing speech means that the calculation must be possible in a few hundred milliseconds or less.

Putting notional and grammatical features together to derive subject number

In the present work, the focus was on how the demands of language production combine with those of agreement in the interaction between a speaker's evaluation of notional number and the compositional possibilities inherent in a language's grammar. To explain how speakers of different languages formulate grammatical number for novel subject noun phrases, two important considerations are (a) whether there are differences among languages in how notional number maps to the grammatical number features that are relevant to agreement; and if so, (b) whether the speakers of languages that display these differences develop different sensitivities to notional number properties as a result.

One important source of variation in notional number that has cross-linguistic consequences for agreement is *distributivity*. Distributivity is present when the same attribute or object type is distributed over multiple tokens of an object. For instance, the phrase *the picture on the postcards* can be used when referring to a picture that is displayed on several postcards. If the picture is seen as a type with several tokens, the notional number should be singular. However, because there are multiple tokens of the same picture across the postcards, the referent may be evaluated as notionally plural. This kind of referential ambiguity has a subtle effect on agreement features, disclosed in a systematic tendency for distributive subjects to be treated as plurals more often than nondistributive subjects (Eberhard, 1999). Specifically, when subject noun phrases are open to a distributive interpretation (e.g. *the picture on the flimsy postcards*) plural verb agreement is more likely than for subjects that are biased toward a nondistributive interpretation (e.g. *The picture on the flimsy hooks*). Consequently, *The picture on the flimsy postcards* ***were***... is more often produced than *The picture on the flimsy hooks* ***were***... (Bock, Eberhard, & Cutting, 2004). These are not simple agreement errors, except in a narrow prescriptive sense, because they capture a systematic semantic difference in a systematic syntactic variation.

(This kind of agreement might be dubbed "semantic," but the differences between verb agreement and clear cases of notional agreement, particularly in pronouns, imply a syntactic relationship; Eberhard et al., 2005.)

There is a provocative line of research that points to a relationship between construals of distributivity and cross-linguistic variations in number-agreement systems (Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). Although this research was neither conceived nor interpreted within a linguistic relativity framework, the findings were taken to mean that distributivity affects agreement differently as a function of the grammatical properties of languages. In their experiments, Vigliocco and colleagues elicited verb-number agreement from native speakers of Spanish, English, Dutch, and French. The task that they used required the formulation of spontaneous completions for sentence subjects (Bock & Miller, 1991). On individual trials, speakers received a noun phrase and then produced it aloud as the subject of a sentence that they created ad lib. In doing so, they produced verbs that displayed number agreement. For example, given a phrase like *the road to the mountains*, a participant might respond with "The road to the mountains was steep." The subject phrases were experimental items as well as fillers that ranged over many topics and syntactic structures, both simple and complex, but the critical phrases had features designed to create contrasting notional pressures on verb agreement. Distributive and nondistributive subjects served to manipulate notional number, using phrases equally divided between those biased toward distributive, notionally plural construals (e.g. *the label on the bottles)* and those biased toward nondistributive, notionally singular construals (*e.g. the road to the mountains*).

Because the use of verbs (and hence the production of subject–verb agreement) is an unavoidable, automatic consequence of producing full sentences, no special strategies or instructions are needed to elicit agreeing verbs in this task. As in ordinary language use, the natural focus of speakers is seldom on the verbs used, and even less often on verb number, but on the meaningful contents of what they intend to say. This yields an assessment of how number information is tacitly used in formulating and implementing subject–verb agreement.

Relying on sentence completion to examine effects of distributivity, Vigliocco and colleagues tested differences in crosslinguistic sensitivity to notional number for languages with typological differences in syntactic properties. One typological property that they assessed was the propensity in some languages to omit or delay the production of overt subjects for sentences. This property can force number-agreeing verbs to get their number features from sources other than already-formulated subject noun phrases. The second property was one that is correlated with the omission and postpositioning of sentence subjects, specifically the richness of a language's number-agreement morphology in tandem with regular morphological specifications.

The first test compared Spanish and English (Vigliocco, Butterworth, et al., 1996). In Spanish, agreement with unexpressed subjects is a common pattern, making it neither necessary nor usual to say *Yo hablo Español* [*I speak Spanish*]. *Hablo Español* is customary. Consequently, the first-person singular inflection on the verb *hablo* must be determined by something other than the grammatical number of an overt subject. In such cases, verb inflection might call directly on notional information in a speaker's message. But English reliably allows an overt subject's grammatical number to control verb number, so English inflections could be less influenced by underlying notional correlates of distributivity.

The results confirmed this prediction: The effect of distributivity on verb-number agreement was larger in Spanish than in English. However, as Vigliocco and colleagues recognized, Spanish and English differ not only in allowing pronoun subjects to be omitted, but also in the richness of their morphology. Almost all verbs in Spanish, unlike English, specify features of number agreement. To separate these possibilities, Vigliocco, Hartsuiker, et al. (1996) compared the effects of distributivity in Dutch and French, again using the sentence-completion method. Dutch is similar to Spanish in allowing verbs to be produced while subjects remain unspoken, often putting inflected verbs in early positions that can be far removed from upcoming subjects. French is not prone to omitting subjects (as Spanish does) or postposing subjects (as Dutch does), but is similar to Dutch and Spanish in being morphologically rich. Thus, if the key property behind notional sensitivity is an absent or belated subject controller for verb number, Dutch speakers should be more sensitive to notional number than French speakers. In contrast, if morphological richness is critical, Dutch and French should show similar effects of notional number. The outcome was comparable levels of notional sensitivity in Dutch and French. This led Vigliocco and colleagues to conclude that morphological richness is the critical property in promoting sensitivity to notional number.

The detailed logic behind the morphological richness hypothesis goes like this: If a speaker's choice of number inflection is under the control of information about number in the underlying message, and many words in utterances carry inflections that uniquely or selectively convey number information, each instance of inflection may entail tacit consultation with the relevant mental number representation. For example, number in Spanish is almost always specified on nouns, pronouns, verbs, and adjectives, and specifying it reliably could demand a stable, detailed representation of number meaning that remains continuously accessible to the words in an unfolding utterance. In formal linguistic terms, each inflectional morpheme may have to consult an index bearing a semantically determined number value (as in Head-Driven Phrase Structure Grammar; Pollard & Sag, 1994). A plausible consequence of this repeated interaction or consultation could be enhanced openness or sensitivity of number morphemes to notional number variations (Vigliocco & Hartsuiker, 2002).

A crucial feature of the materials in all of the languages tested was that the subject noun-phrases did not themselves instantiate the properties hypothesized to elicit differences in notional number. That is, the actual Spanish, French, Dutch and English materials were not substantially different in morphological features, nor were the sentence subjects omitted or postposed. Spanish speakers produced full noun phrases, not elided ones, and Dutch speakers placed subject noun phrases before verbs, not after them. Therefore, the notional differences that were observed between English and the other three languages were not differences due to the subjects actually spoken. Instead, the differences had to reflect predispositions in the languages or in the speakers of the languages. According to the morphological richness hypothesis, the predisposition emerges from a consistent requirement to consult notional number in order to inflect verbs and other number-specified words, regardless of whether grammatical number (as carried by subject noun phrases, for example) is present or absent.

Although the emphasis in the work by Vigliocco and colleagues was on the proclivities of languages rather than speakers, their findings are equally consistent with the predictions of linguistic relativity. Perhaps it is the speakers of rich-morphology languages who are especially sensitive to notional number, *because* of features of their language. That is, if the grammar of a language is replete with number inflections, and if the value of each inflection must be determined during ongoing language production from a representation of notional number, the *users* of rich-morphology languages may become sensitized to variations in notional number, consistent with linguistic relativity.

Slobin's thinking-for-speaking hypothesis (1996) offers a plausible mechanism for the emergence of linguistic relativity from the demands of language production, explaining how the properties of morphology could come to enhance and sharpen the ability of speakers to apprehend notional number valuations. To account for what speakers of different languages have to do in order to meet the needs of their different grammars, Slobin proposed that prelinguistic messages (the ideas or nonverbal communicative intentions behind language production) have to be packaged with grammatically relevant conceptual distinctions included. Speakers put these distinctions into messages, and do so regardless of the relevance of the distinctions to a specific intended meaning, because ever-present features of grammar alone depend on the information.

Languages with number agreement systems need notional number for agreement to work. Accordingly, speakers of English and Spanish have to tacitly represent numerosity in messages. If the richness of agreement morphology promotes the interaction between notional number and linguistic number (as it will if each number morpheme has a conceptual grounding), there will be greater pressure on speakers of Spanish than on speakers of English for a rapid and reliable evaluation of notional number. Set in light of debates about whether cognitive and linguistic mechanisms are vulnerable to feedback from subsequent processes, the question here is whether unrelenting interactions between cognition and ensuing preparations for speaking come to shape the information that cognition makes available to the language production process.

By definition, thinking-for-speaking is something that speakers do. To test its implications for linguistic relativity, the mechanisms of thinking-for-speaking (namely thinking and speaking) have to be engaged. The research on number agreement by Vigliocco and colleagues (Vigliocco, Butterworth, et al., 1996; Vigliocco, Hartsuiker, et al., 1996) did exactly that. Speakers of different languages formulated and produced sentences in a fashion that indirectly assessed the impact of variations in notional number. The assessment was covert, requiring no instructions about the production of agreement and no emphasis on number variations, because speakers produce agreement as a matter of course when they produce sentences. The unsystematic variations in grammatical number that were encountered in the experimental materials are unremarkable in language and unremarked by speakers. The languages involved in the tests differ in ways that target the grammatical dimensions hypothesized to create differences in sensitivity to notional number differences. Accordingly, viewed in terms of how language might shape thought, the research by Vigliocco and colleagues offers unusually persuasive evidence for a language-linked variation in notional number apprehension.

For the purpose of assessing linguistic relativity, however, a shortcoming in the Vigliocco, Butterworth, et al. (1996) experiment was that the English materials (adapted from Bock & Miller, 1991) were not matched to the Spanish materials in literal meaning. There were few instances in which the English and Spanish subject noun-phrases were translation equivalents, and for purposes of testing relativity, this matching is essential. For observed differences in number agreement to be attributable to sensitivity stemming from inherent properties of the speakers' languages, literal meanings have to be as similar as possible. Achieving this requires no more than a change in the experimental materials that were used by Vigliocco and colleagues, with the task and design the same. In our Experiment 1, the Spanish and English materials were matched in meaning, so as to equate the literal meanings and the implications of the meanings for notional number. This left only the speakers' inherent sensitivity to notional number to drive differences on the number agreement measure.

Our focus on English and Spanish stems from the fact that their differences constitute the critical contrast in the array of languages so far examined. Being neither morphologically rich nor pronoun-dropping nor verb-postposing, English should induce less sensitivity than Spanish to number meaning in agreement, as Vigliocco, Butterworth, et al. (1996) showed. Confirmation of this difference with matching of speakers' intended meanings would offer pivotal support for the view that differences in number grammar can steer native speakers' attention to subtle differences in notional number. Since there is ample evidence that English speakers are sensitive to *some* variations in notional number (Berg, 1998; Bock, Nicol, & Cutting, 1999; Bock et al., 2004, 2006; Eberhard, 1999; Haskell & MacDonald, 2003; Humphreys & Bock, 2005), the expectation was not that English speakers would be wholly oblivious to number meaning. Rather, the prediction from linguistic relativity, as shaped by thinking-for-speaking, is that English speakers should be less sensitive to notional plurality than Spanish speakers. The competing prediction reflects the hypothesis that the properties of the

English language do not limit the use of notional differences any more than Spanish does in the course of grammatical formulation. The morphological poverty of English may actually make the effects of notional number more visible (Bock & Middleton, in press; Foote & Bock, in press; Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008). If the latter hypothesis is right, the differences due to notional number in English should be similar to or than the differences in Spanish.

To hone the experimental tests and ensure a level playing field with respect to the number information conveyed in the stimuli, we also evaluated the materials along dimensions that have been found to be important to number agreement, including notional number, imageability, and sensibility. Identical procedures were used for testing in both languages, relying on the agreement-elicitation paradigm in which speakers complete subject noun phrases as full sentences. Most important is that the subject noun phrases had the same variations in grammatical number and the same variations in distributive properties in English and Spanish. If Spanish speakers are disposed by the morphological and syntactic features of their language toward enhanced notional number sensitivity, more plural verb agreement should occur in Spanish than in English for those sentence subjects biased toward distributive construals, regardless of grammatical number. However, if Spanish and English speakers represent and use notional number in the same ways, the same variations due to distributivity should be observed, overlaid on standard effects of grammatical number.

Experiment 1

The first experiment compared number agreement in English and Spanish using matched materials in a sentence-fragment completion task. The sentence fragments were subject noun phrases that varied in notional number due to differences in distributivity and in grammatical number properties (singularity and plurality). The experimental phrases contained two nouns that varied between singular and plural forms. The first of the nouns (e.g. *key* in *The key to the cabinet)* was the head of the subject noun phrase, which conventionally controls verb number. The second noun was the object of a prepositional phrase that modified the head (e.g. *cabinet* in *The key to the cabinet)*. The position of the second noun, dubbed the *local* noun, put it immediately before the verbs that participants produced. All combinations of singular and plural head and local nouns occurred in the experimental phrases.

Although nouns in the local position typically do not control verb number-agreement, they can affect verb number through a process of *attraction* (Bock & Miller, 1991). For instance, in a subject noun-phrase like *The key to the cabinets*, plurality in the local noun (*cabinets*) can cause the verb to become plural (*The key to the cabinets* ***were*** *lost*), displacing the singular number favored by the head noun *key*. Attraction occurs almost exclusively with grammatically plural local nouns, and as a consequence of their grammatical rather than their notional plurality (Bock et al. 2004). In the present experiment (as in Vigliocco, Butterworth, et al., 1996), this allowed the effect of grammatical number to be manipulated along with the notional number biases created by distributivity.

Participants

Native speakers of English and Spanish took part in the experiment, 32 in each group. The English speakers were recruited from an introductory psychology course at the University of Illinois in fulfillment of a course requirement. The Spanish speakers were 32 students at the Universidad de La Laguna in Spain who received course credit in return for their participation. An additional 120 English and 120 Spanish speakers served in separate norming studies. None of them took part in the main experiment or more than one norming session. Among the participants in each group, knowledge of the other language was weak to absent.

Method

The materials and procedures were the same in the English and Spanish conditions of the experiment (except, of course, that English speakers received English stimuli and produced English responses while Spanish speakers received Spanish stimuli and produced Spanish responses). Norming was also carried out with separate English and Spanish groups, employing parallel procedures. Below we describe the materials and tasks used for both groups, noting any adjustments needed to accommodate differences between the languages.

Materials

The experimental materials consisted of 64 complex subject noun phrases, here called *preambles*, such as *The label on the bottles* and *The key to the cabinets*. Every preamble contained a head noun phrase (the first noun phrase in the preamble, e.g. *the label*) followed by a prepositional phrase (e.g. *on the bottles*). The prepositional phrase included the local noun phrase (e.g. *the bottles*). As shorthand, the two noun phrases are referred to simply as the *head* and *local* nouns. Every preamble occurred in four versions made up of the four possible combinations of singular and plural head and local nouns. Table 1 shows example items for both languages in each of the four experimental conditions.

Half of the experimental preambles were semantically biased toward distributive and half toward nondistributive construals. The biases surfaced only in the versions of the preambles that had singular heads and plural local nouns, because only preambles with this property supported distributive readings. Examples of both types are shown in Table 1.

The construction of the preambles was designed to ensure that grammatical gender was balanced within the Spanish versions of the preambles. As a result, heads and local nouns in the Spanish items were equally divided in grammatical gender properties. There were 16 preambles in each of the four combinations of masculine and feminine gender. Half of the singular–plural versions of the preambles in each gender combination were distributive and half were nondistributive.

Table 1
Examples of the four versions of the experimental preambles in English and Spanish for two item types (nondistributive and distributive construals of singular–plural preambles).

Distributivity of singular/plural version	Experimental condition	Example preambles
		English examples
Nondistributive	Singular/singular	The author of the novel
	Singular/plural	The author of the novels
	Plural/plural	The authors of the novels
	Plural/singular	The authors of the novel
Distributive	Singular/singular	The label on the bottle
	Singular/plural	The label on the bottles
	Plural/plural	The labels on the bottles
	Plural/singular	The labels on the bottle
		Spanish examples
Nondistributive	Singular/singular	La autora de la novela
	Singular/plural	La autora de las novelas
	Plural/plural	Las autoras de las novelas
	Plural/singular	Las autoras de la novela
Distributive	Singular/singular	La etiqueta de la botella
	Singular/plural	La etiqueta de las botellas
	Plural/plural	Las etiquetas de las botellas
	Plural/singular	Las etiquetas de la botella

The 64 experimental preambles in Spanish were those used by Vigliocco, Butterworth, et al. (1996), apart from two single-word changes made to accommodate the dialect of the speakers. Participants in the Vigliocco, Butterworth et al. experiment spoke Mexican Spanish, where *book cover* is *la pasta* and *truck* is *el autobús*; the Spanish variety in the present study was the one spoken in the Canary Islands, where the corresponding words are *la tapa* and *el camión*. Two native Castilian Spanish speakers evaluated all of the materials for appropriateness.

The English versions of the preambles were literal or near-literal translations from Spanish except in two cases (*el responsable, la grua)* when literal translation was impossible or yielded results that strike native speakers as stilted or otherwise deviant. Minor translation variants were used in the English preambles to avoid irregular plurals (*men, women, mice, fishermen)* and awkward construction (e.g. instead of *the residence of the presidents* we used *the estate of the presidents)*. All of these changes substituted English words from the same semantic categories as the nouns in the Spanish preambles. The English preambles were evaluated by two native English speakers (one of them the age of the English participants) and checked against the Spanish materials by two Spanish–English bilinguals to ensure comparability in meaning.

An additional 64 items served as filler preambles. Fillers were made up of noun phrases with conjoined nouns (e.g. determiner-noun-*and*/*y*-determiner-noun) or determiner-adjective-noun sequences in English and determiner-noun-adjective sequences in Spanish (e.g. *the damaged elevator*/*el ascensor averiado*). There were 32 fillers with singular heads and 32 with plural heads. The fillers were composed so that in the Spanish lists, the genders of the head nouns were equally divided between masculine and feminine for each head number. Translation of the fillers followed the same steps as the experimental materials.

The experimental and filler preambles were assembled into four lists. Every list contained all of the filler preambles in the same sequential positions, with four fillers leading off the list. The four versions of every experimental preamble were assigned to the lists to create a counterbalanced arrangement in which each of the 64 experimental preambles was represented only once per list and, within lists, the number of preambles representing each of the four experimental conditions was the same, 16 per list. The order of items in each list was identical, arranged in a pseudorandom fashion so that no more than two experimental preambles occurred consecutively, and all consecutive experimental preambles represented different experimental conditions. The orders in the Spanish and English lists were the same.

Norming

The English and Spanish versions of the filler and experimental preambles were rated for notional number, imageability, and sensibility. The same lists were used for each type of norming task, in the same four forms as in the agreement task, with the same counterbalancing, presented as paper-and-pencil questionnaires. The instructions and rating scales were adopted from similar tasks reported in the literature.

The mean ratings from each task, by preamble type and distributivity, are shown in Table 2. Each set of norms was tested statistically in an omnibus analysis of variance treating items as the random factor. All analyses used distributivity as a nested factor with preamble type and language as crossed factors. To compare cell means, 95% confidence intervals for pairwise differences were calculated from the mean squared errors for the interaction among preamble type, preamble distributivity, and language. (Note that the structure of the analyses on the norming data differs from the structure that was used in the main experiments. Obviously, the main effects and interaction of distributivity and language are the same regardless of design. Treatment of preamble type as one four-level factor, rather than an orthogonal combination of head

Table 2
Mean ratings of notional number, imageability, and sensibility for experimental preambles in English and Spanish.

Distributivity of singular/plural version	Experimental condition (number of head/number of local noun)	Notional number (1 = "one thing"; 2 = "more than one thing"	Imageability (1 = not imageable; 7 = highly imageable)	Sensibility (1 = nonsense; 5 = completely sensible)
		English preamble ratings		
Nondistributive	Singular/singular	1.19	3.91	4.43
	Singular/plural	1.26	3.80	4.36
	Plural/plural	1.88	3.81	4.41
	Plural/singular	1.88	3.73	4.26
Distributive	Singular/singular	1.18	5.07	4.63
	Singular/plural	1.34	5.03	3.87
	Plural/plural	1.91	4.99	4.53
	Plural/singular	1.87	5.03	4.24
		Spanish preamble ratings		
Nondistributive	Singular/singular	1.22	4.24	3.91
	Singular/plural	1.31	4.16	3.77
	Plural/plural	1.79	4.25	3.72
	Plural/singular	1.85	4.03	3.64
Distributive	Singular/singular	1.21	5.23	3.94
	Singular/plural	1.32	5.09	3.77
	Plural/plural	1.77	5.09	3.76
	Plural/singular	1.80	4.99	3.73

number and number match between heads and local nouns, allows transparent comparisons between cell means. Since the effects of head number and number match on the ratings are of little independent interest, this simplifies the report of the analyses without obscuring important properties of the materials.)

Notional-number norming

Raters in the notional number task were 40 English and 40 Spanish speakers drawn from the same sources as the main experiment. In English, they were instructed as follows:

> For each of the phrases below, we would like you to judge whether it refers to one thing or more than one thing. So, imagine each phrase appearing in the blank in the following question:
>
> If you were thinking about ________, would you be thinking about one thing or more than one thing?
>
> Please check the box under "one thing" or "more than one thing" to indicate your answer. Sometimes both answers will seem possible. In these cases just pick the answer that makes more sense to you.

The Spanish instructions were analogous. Note that instructions indicated that both answers might seem possible, and that when this happened the answer that made more sense should be chosen. No example preambles or sample ratings were provided, in order to avoid biasing raters toward particular interpretations of what constitutes "one thing." The instructions and response options were the same as those used in Bock and Miller (1991).

Table 2 shows the mean notional ratings in English and Spanish, and Table 3 summarizes the inferential statistics. The 95% Scheffe confidence interval for pairwise contrasts between cells was .07. There was a large effect of preamble type, mainly due to the predictably higher ratings for preambles with plural head nouns. The more important difference among the preambles was the one between the singular–singular and singular–plural preambles, which was larger in the distributive (.14) than in the nondistributive (.08) condition. The effect is attributable to the influence of distributivity on notional plurality in singular–plural but not singular–singular preambles, creating a significant interaction between preamble type and distributivity. The interaction of distributivity with language was not significant.

The only other statistically significant differences were due to the English ratings being higher overall than the Spanish ratings, and chiefly when head nouns were plural. This created a main effect of language and an interaction between language and preamble type (see Table 3). Otherwise the patterns were similar for English and Spanish.

Although far from significant, there was a discrepancy in the distributivity ratings for the two languages that might be taken to indicate greater sensitivity to notional number among English than among Spanish speakers. The contrast is in the difference between the singular/plural preambles and their controls in the nondistributive condition, compared to the same difference in the distributive condition. For English the average increase from the nondistributive to the distributive conditions was .09; in Spanish it was .02. This kind of unsystematic variability is common in naive speakers' evaluations of distributivity, given the subtlety of the judgments required. Expert ratings (like those used for the same Spanish materials by Vigliocco, Butterworth, et al. (1996)) tend to be more reliable, and better predictors of notional agreement outcomes.

Imageability norming

Raters in the imageability task were 40 English and 40 Spanish speakers drawn from the same sources as the main experiment. The English instructions were as follows:

Table 3
Inferential statistics and associated probabilities from analyses of notional number, imageability, and sensibility norms for experimental items.

	F_2	Degrees of freedom (denominator)	Probability
Source	*Notional number*		
Preamble type	1169	186	<.01
Distributivity	<1	62	.97
Language	6.03	62	.02
Preamble type × distributivity	2.94	186	.03
Preamble type × language	7.60	186	<.01
Distributivity × language	1.53	62	.22
Preamble type × distributivity × language	.57	186	.63
	Imageabililty		
Preamble type	1.73	186	.16
Distributivity	19.84	62	<.01
Language	4.27	62	.04
Preamble type × distributivity	.21	186	.89
Preamble type × language	.66	186	.58
Distributivity × language	1.91	62	.17
Preamble type × distributivity × language	.28	186	.84
	Sensibility		
Preamble type	10.26	186	<.01
Distributivity	<1	62	.94
Language	151	62	<.01
Preamble type × distributivity	3.92	186	<.01
Preamble type × language	9.38	186	<.01
Distributivity × language	<1	62	.33
Preamble type × distributivity × language	7.77	186	<.01

> Please rate each of the phrases below according to the ease or difficulty with which it evokes a mental image of its referent. If an image is easily evoked (as it might be for a phrase like "the skyscraper in the city", for example), you should give the phrase a high imagery rating. Phrases that evoke images only with great difficulty or not at all (for example, a phrase like "the truth of the matter") should get low imagery ratings. Indicate your rating by circling a number on the seven-point scale beside each phrase, where 1 is lowest in imageability and 7 is highest in imageability.

The scale was the one used by Paivio, Yuille, and Madigan (1968).

The mean ratings are listed in Table 2 and the results of the statistical analysis are in Table 3. Distributive preambles were rated as more imageable overall than nondistributive preambles (5.0–3.8; a difference that should enhance distributivity effects on agreement; Eberhard, 1999), and Spanish preambles received slightly higher scale ratings than English preambles (4.6–4.4). There were no notable interactions among any of the factors, suggesting that the English and Spanish preambles were similar in relative imageability. The 95% Scheffe confidence interval for pairwise contrasts between the values shown in Table 2 was .17.

Sensibility norming

Raters in the sensibility task were 40 English and Spanish speakers from the same sources as the main experiment. Their instructions were as follows:

> For each of the phrases below, please rate how understandable the phrase seems to you. Rate each phrase on a scale from 1 (nonsense) to 5 (completely sensible). Indicate your rating by circling the corresponding number.
>
> nonsense 1 2 3 4 5 completely sensible

The instructions and scale were borrowed from Vigliocco, Hartsuiker, et al. (1996).

Table 2 presents the mean sensibility scores and Table 3 shows the results of the statistical analyses. The English versions received higher ratings than the Spanish overall (4.3–3.8) except in one case. For the singular–plural distributive preambles, the sensibility ratings were roughly the same (3.9 in English and 3.8 in Spanish). This exception created an interaction among language, preamble type, and distributivity, with a derivative main effect for preamble type and two derivative interactions (between language and preamble type, and preamble type and distributivity). The Scheffe 95% confidence interval for pairwise contrasts between cell means was .58.

Procedure

In the agreement experiment, preambles were presented to participants who repeated and completed them as sentences. Speakers were asked to use adjectives or other simple predicates as the completions, and not to think too long about what to say. This was an adaptation of the procedure used by Vigliocco, Butterworth et al. (1996, Experiment 2), who presented a specific adjective to be used as a completion on each trial. The adaptation sidestepped the problem of creating equivalently natural, matched completions for English and Spanish while minimizing the differences between the Vigliocco, Butterworth et al. preambles and the present ones.

One preamble was presented on each trial. Trials began with a display of three asterisks (***) centered on the screen for 600 ms. This served as a warning signal for the upcoming preamble. When the preamble appeared, participants were instructed to read it silently. Preambles were displayed in 18-point Courier font centered on the screen, with the first letter of the preamble capitalized and the rest in lower case. To align reading times in English and Spanish, a reading-time formula was used in both languages that allowed a constant 250 ms per word plus an additional 32 ms for each character. This gave most participants just enough time to read the preamble through once before it disappeared. When the preamble vanished, an exclamation point appeared in the middle of the screen for 1.5 s as a cue to begin speaking. At that point, participants were supposed to repeat and complete the preamble, and then press the space bar to continue to the next trial.

Three example preambles and completions were given in the instructions. The sample preambles were *The fire engine, The statue,* and *The iron* and the sample responses were *The fire engine is red, The statue was carved from wood,* and *The iron was hot*. In the United States, the preambles and cues were presented on a 19'' (18'' viewable) monitor, and experimental events were controlled by a Macintosh Quadra 650 computer running PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). In Spain, the same PsyScope script was used on a Macintosh G3 PowerBook laptop with a 15'' screen. Spoken responses were tape recorded for later transcription.

Participants were tested individually, each receiving one of the four preamble lists.

Design

The design for participants included language as a nested factor, with equal numbers of participants (32) receiving Spanish or English preambles. Preamble type was crossed with the language factor, so that each participant in each language received an equal number of items (16) representing each of the four preamble conditions (singular/singular, singular/plural, plural/plural, plural/singular). Participants received just one version of each of the 64 preambles. In the design for items, all 64 experimental items were presented to 32 participants in each cell of the design formed by crossing the factors of language and preamble condition.

Transcription and scoring

The spoken English sentences were transcribed and scored in terms of four criteria defined by the accuracy of preamble repetition and the singular or plural inflection of the verb. *Consistent* responses were those in which the preamble was correctly repeated and the verb was produced with the same number as the head of the subject noun phrase (singular or plural). *Inconsistent* responses contained correctly repeated preambles with verbs whose number differed from that of the head of the subject noun phrase. *Ambiguous* responses also had correctly repeated preambles but the verb form was the same for singulars and plurals (typically, these were regular past-tense verbs that are uninflected). *Miscellaneous* responses were those that failed to meet the criteria for the other three categories. Almost all of the miscellaneous responses were incorrect preamble repetitions. A few miscellaneous responses contained no verbs, no preamble completions, or other defects.

The scoring system for Spanish was the same as for English, except for differences created by Spanish grammar. For Spanish verb agreement, only correct preamble reproductions along with the verb number used in the completion counted toward classifying responses as corrects and errors. There were no Ambiguous responses because all the Spanish verbs were inflected for number. The miscellaneous category consisted of all other responses, with properties similar to those in the corresponding English category. Table 4 gives the numbers and percentages of responses in the major scoring categories for each language.

The Spanish scoring took account of inflections on the predicate adjective, which included inflections for gender. Predicate-adjective gender inflections were scored independently of predicate-adjective number inflections, so gender variations did not enter into the scoring of number variations, and vice versa. Consistent adjective responses contained adjectives that were the same gender as the head noun; inconsistent responses differed from the head in gender. Instances in which combinations of verb number, adjective number, and adjective gender agreed or disagreed with the head, and agreed or disagreed with the local noun, were tabulated separately. As it turned out, the incidence of gender disagreement was far too low (in all, there were nine inconsistent responses) to evaluate any relationship between gender and number inconsistency. Because of this, we omit gender from the data analyses and discussions.

Statistical analyses

Analyses of variance were performed on the proportions of inconsistent responses out of the sum of the consistent and inconsistent responses. One analysis treated

Table 4
Numbers of responses in four scoring categories in English and Spanish, Experiment 1.

Preamble type	Consistent	Inconsistent	Ambiguous	Miscellaneous
		English		
Singular/singular	352	0	122	38
Singular/plural	312 (156)	37 (29)	99 (40)	64 (31)
Plural/plural	305	18	92	97
Plural/singular	318	16	96	82
		Spanish		
Singular/singular	501	2	–	9
Singular/plural	446 (221)	38 (27)	–	28 (8)
Plural/plural	470	22	–	20
Plural/singular	471	19	–	22

Note: Numbers in parentheses for the singular/plural condition represent how many of the tabulated responses occurred for distributive preambles.

participants as the random factor and another treated items as random, and minF' was calculated from the outcomes (Clark, 1973). As in the norming analyses, 95% confidence intervals for pairwise planned comparisons were derived from the mean-squared error for the interaction among language, head number, number match, and distributivity.

Results

Fig. 1 displays the proportions of inconsistent responses (responses in which the number of the produced verb mismatched the number of the head noun) in English and Spanish. The figure shows similar patterns in the two languages, with number-inconsistent responses being much more common in the singular–plural condition than in any other. This is a typical agreement-attraction pattern (Bock & Miller, 1991). There was a slightly reduced number of consistent responses after plural heads, but the incidence did not vary much for preambles with matching or mismatching heads and local nouns.

Distributivity influenced the incidence of number-inconsistent responses in the singular–plural condition, as expected. Distributive preambles elicited more plural verbs than non-distributive preambles (.13 – .06). There was a slightly larger effect of distributivity in English than in Spanish, with distributive-nondistributive differences for plural responses in English and Spanish of .09 and .06, respectively.

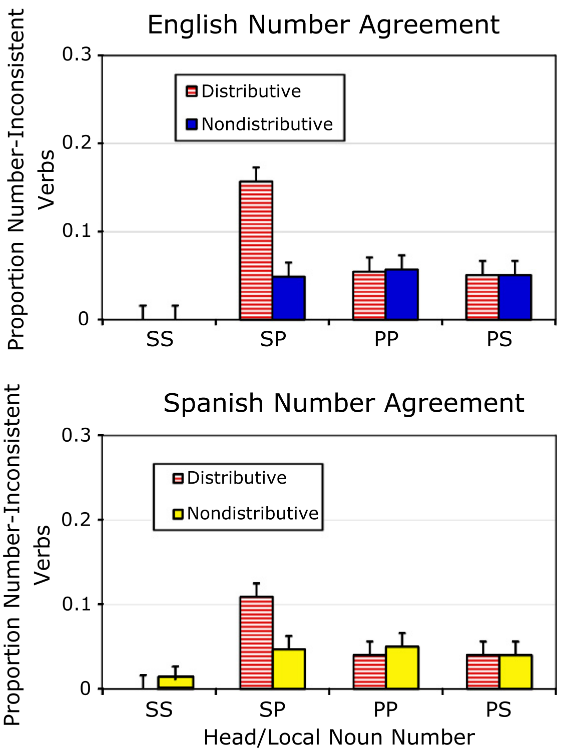

Fig. 1. Proportions of number-inconsistent verbs produced in English and Spanish sentence completions in Experiment 1. (Error bars represent halfwidths of 95% confidence intervals for pairwise contrasts between condition means.)

Statistically, as shown in Table 5, the response patterns produced the usual effects of head number, number match, and their interaction, reflecting the preponderance of number-inconsistent responses in the singular–plural condition. The interaction of these factors with distributivity was also statistically significant, reflecting the increased incidence of number-inconsistent responses after distributive preambles. None of the effects of language reached conventional levels of statistical significance, testifying to the similarity of the response patterns in English and Spanish. In particular, none of the interactions of language with distributivity approached significance, all *F*s < 2.60, all *p*s > .11.

Discussion

The variations in verb-number agreement were much the same in Spanish and English, including agreement with distributive and nondistributive subject noun-phrases. Spanish speakers were no more likely than English speakers to display sensitivity to notional number properties; indeed, they were slightly (but not significantly) less likely. These results suggest that when sentence subjects have comparable meanings, with comparable implications about distributivity, verb-number agreement is essentially the same in the two languages. The upshot is that neither morphological richness nor related typological differences between English and Spanish effectively alter speakers' use of notional number information.

It is informative to compare these results to those of Vigliocco, Butterworth et al. (1996). In Spanish (Experiment 1), there were 15.6% (40) plurals after distributive subjects and 7.4% (19) plurals after nondistributive subjects; we obtained 10.5% (27) plurals after distributive subjects and 4.3% (11) plurals after nondistributive subjects. These are similarly sized differences of 8.2% and 6.2%, respectively, in the effect of distributivity. For English, by comparison, Vigliocco, Butterworth et al. (Experiment 3) obtained 3.8% (34) and 4.0% (36) plurals after the respective distributive and nondistributive subjects; we obtained 11.3% (29) and 3.1% (8) plurals in the analogous conditions. So, for English, the respective differences between the experiments in the distributivity effect were .2% and 8.2%. (In calculating these percentages, we used the total number of responses in the relevant cells, because Vigliocco, Butterworth et al. did not break down correct responses by cell. For the Spanish comparisons the numbers of responses per cell were the same in both experiments, 256 in all cases, but for English there were 896 responses per cell in the Vigliocco, Butterworth et al. experiment and 256 in ours.)

In two crucial respects, the results replicated previous findings about number agreement, including those of Vigliocco, Butterworth, et al. (1996). First, as just noted, the results from Spanish are similar in both sets of experiments. Second, in both English and Spanish, when the head and local nouns mismatched in number, there were more number-inconsistent responses after singular than after plural heads (i.e., more agreement attraction). As a standard part of this number asymmetry, plural heads with singular local nouns elicited no more number inconsistency than plural heads with plural local nouns.

Table 5
Inferential statistics and associated probabilities for analyses of inconsistent-response proportions (Experiments 1 and 2) and consistent-response onset latencies (Experiment 2).

	Test statistics, degrees of freedom (df), and associated probabilities (p)								
Source	F_1	df	p	F_2	df	p	minF'	df	p
Experiment 1: Preamble repetition and completion in English and Spanish									
Language	.52	1,62	.47	1.08	1,61	.3	.35	1,110	.56
Number of preamble head	5462	1,62		6235	1,61		2911	1,123	
Number match of head and local noun	27.56	1,62		24.96	1,61		13.10	1,123	
Language × head number	2.06	1,62	.16	3.47	1,61	.07	1.29	1,116	.26
Head number × number match	27.8	1,62		23.32	1,61		12.68	1,122	
Language × number match	1.37	1,62	.25	2.51	1,61	.12	.89	1,114	.35
Language × head number × number match	3.12	1,62	.08	3.22	1,61	.08	1.58	1,123	.21
Distributivity	2.98	1,62	.09	3.56	1,61	.06	1.62	1,122	.21
Distributivity × language	.02	1,62	.88	.32	1,61	.57	.02	1,70	.89
Distributivity × head number	2.14	1,62	.15	4.06	1,61	.05	1.40	1,113	.24
Distributivity × number match	2.59	1,62	.11	5.93	1,61	.02	1.80	1,107	.18
Distributivity × language × head number	.16	1,62	.69	2.60	1,61	.11	.15	1,70	.70
Distributivity × language × number match	.15	1,62	.70	2.04	1,61	.16	.14	1,7	.71
Distributivity × head number × number match	5.60	1,62	.02	6.50	1,61	.01	3.01	1,122	.09
Distributivity × language × head number × number match	.07	1,62	.79	.06	1,61	.81	.03	1,122	.86
Experiment 2: Proportion inconsistent responses in preamble completions									
Number of preamble head	.71	1,63	.40	.29	1,62	.59	.21	1,106	.65
Number match of head and local noun	61.21	1,63		60.56	1,62		30.44	1,125	
Head number × number match	17.99	1,63		13.12	1,62		7.59	1,122	
Distributivity	2.54	1,63	.12	1.07	1,62	.30	.75	1,107	.39
Distributivity × head number	14.19	1,63		4.55	1,62	.04	3.45	1,98	.07
Distributivity × number match	13.71	1,63		5.59	1,62	.02	3.97	1,106	.05
Distributivity × head number × number match	14.26	1,63		3.11	1,62	.08	2.55	1,88	.11
Experiment 2: Onset latencies for consistent preamble completions									
Number of preamble head	.09	1,63	.76	.97	1,62	.84	.08	1,75	.78
Number match of head and local noun	14.35	1,63		10.55	1,62		6.08	1,122	.02
Distributivity	.29	1,63	.59	.04	1,62	.84	.04	1,79	.84
Head number × number match	17.83	1,63		12.84	1,62		7.46	1,121	.01
Head number × distributivity	.17	1,63	.68	.11	1,62	.74	.07	1,119	.79
Number match × distributivity	.33	1,63	.57	.25	1,62	.62	.14	1,122	.71
Head number × number match × distributivity	.33	1,63	.57	.63	1,62	.43	.22	1,114	.64

Conceivably, the similar rates of number-inconsistent responses for English and Spanish might be turned around and used to argue that grammatical inconsistency alone was responsible for the patterns in both languages, and not notional agreement. The argument follows from a view that blames distributivity for increased agreement attraction, not for increased notional agreement. If this is so (the argument goes), the reason that English and Spanish behaved similarly in Experiment 1 had to do merely with the factors that are responsible for attraction in both languages, not to any general similarity in the factors responsible for standard agreement. This would undercut the claim that English and Spanish speakers are similarly sensitive to the subtleties of number meaning, implying instead that the resemblance is an incidental consequence of how attraction comes about.

The weakness in this argument is that there is good evidence that attraction and notional agreement reflect different processes that sometimes have the same surface consequences. The processes can be separately manipulated (Bock et al., 2004, 2006) and better accounted for within a model in which agreement and attraction call on different sources of information (Bock & Middleton, in press; Deutsch & Dank, 2009; Eberhard et al., 2005; Staub, 2009; Staub, 2010).

Experiment 2

One explanation of the findings in Experiment 1 that would leave a relativity account intact stems from how ambiguous notional number affects agreement. Speakers could use different processes to resolve notional ambiguity, yet arrive at similar agreement outcomes. This would obscure genuine differences in language-induced sensitivity to notional number. Experiment 2 was designed to evaluate this possibility.

Notional number ambiguity comes about because most distributive preambles have potential nondistributive interpretations, and vice versa for nondistributive preambles. In the face of this ambiguity or complexity, ability to formulate agreement and complete the preambles could be compromised and slowed. However, in ordinary speaking, notional number must be rapidly apprehended, at the same pace as speech itself. These ordinary rates, roughly two words per second and 120 per minute, can allow less than a second to construe intended referents as one thing or more than one thing. To achieve this, ambiguity must remain latent, with initial biases dominating construal of notional number in messages. For speakers whose languages sensitize them to variations in notional number, automatic thinking-for-speaking procedures should capitalize rapidly on the number biases present in specific circumstances, rarely impeded by the potential ambiguity. Speakers without this language-induced facility in the apprehension of notional number could still use it in agreement, but only when they apprehend the notional ambiguity and try to resolve it. The resolution is typically difficult, and likely to take time. So, Spanish speakers could use notional number tacitly in agreement when English speakers use it more deliberately, yet arrive at agreement outcomes that exhibit similar effects of distributivity.

In the first experiment, English speakers' awareness of distributive ambiguity could arise from repeated exposure to the contrasting biases of the preambles occurring in relative proximity. These biases were stronger than those present in the English materials used by Vigliocco, Butterworth, et al. (1996) (Eberhard, 1999). Moreover, there was no need for speed in resolving the ambiguity: English speakers could have interpreted the distributive and nondistributive preambles in strategic, nonautomatic ways. In fact, all experiments that have obtained distributivity effects in English (e.g. Eberhard, 1999; Humphreys & Bock, 2005) used untimed procedures. If the English speakers in Experiment 1 produced sentences after experiencing more difficulty than is typical for number agreement, their apparent sensitivity to distributivity could be an artifactual product of out-of-the-ordinary inferences about the number properties of referents. If so, plural verb usage could increase for reasons other than an automatic evaluation of notional number during spontaneous speech.

This explanation makes a simple prediction: If English speakers became sensitized to the ambiguity of notional number, resolution of the resulting ambiguity and potential conflicts between notional and grammatical number should create a time penalty for distributive preambles. The time penalty would show up for distributed subjects even when speakers produce the conventional, grammatically expected singular verb, since competition could be present whenever agreement with a distributed subject is needed. Experiment 2 tested this prediction with measurements of the latency of verb production.

The procedure from Experiment 1 is not optimal for measuring response times, because speech began with preamble repetitions. The onset timing of preamble repetitions is unlikely to be informative with respect to competition between notional and grammatical number, which could have its impact later. To sidestep this problem, speakers in Experiment 2 completed the presented preambles without repeating them, and the verb onset latencies were timed from the offset of preamble presentation. Since the first word in the completion was almost always a number-inflected verb, its onset time could capture competition in the computation of verb number. Presentation was auditory, to control the duration of the preambles.

This procedure has its own drawback, though: It reduces the incentive for participants to listen carefully to the preambles and provides no way to assess the level of accurate interpretation. To remedy this, we included catch trials in each experimental session to obtain a measure of comprehension accuracy. On catch trials, participants were cued to repeat the just-heard preamble before completing it. This yielded a task comparable to the one in Experiment 1, allowing us to compare repetition success and agreement across the experiments.

Another modification in the procedure was aimed at reducing the variability in completion latencies due to difficulty in formulating a sensible completion. It also served to make the task more similar to one used in Spanish by Vigliocco et al. (1996, Experiment 2). Participants received instructions to complete the preambles with one of a small set of predicate adjectives, which encourages the use of number-inflected verbs and increases the number of trials available for analysis. Together, these tactics were designed to create a more stable measure of verb onset latency that was less likely to be overshadowed by interpretation, message creation, and syntactic formulation processes.

Method

Participants

The participants were 64 undergraduates from the same source as the English speakers in Experiment 1.

Materials

The experimental and filler preambles were the same as those used in Experiment 1. Three pairs of polar-opposite adjectives were selected to serve as preamble completions: *good/bad*, *big/little*, and *pretty/ugly*. Adjective selection was constrained to ensure that at least three of the adjectives formed sensible completions for each preamble and the overall number of sensible completions was comparable for distributive and nondistributive preambles (roughly 250, respectively). The number of adjectives was limited in order to minimize the difficulty of formulating completions.

The same four lists were used as in Experiment 1. Eight versions of each list were generated to create a counterbalanced set of catch trials, yielding 32 lists in all. Every list version contained eight catch trials, two in each of the four head/local number combinations. One of the catch-trial preambles in each head/local condition in every list version represented a distributive item and the second represented a nondistributive item. Across the 32 list versions, each of the four preambles derived from an individual item occurred eight times, seven on complete-only trials and once on a catch (repeat-and-complete) trial. Eight fillers also served as catch trials to create an average rate of one catch trial for every 16 trials (12.5%) per session. The same eight fillers served as repeat-and-complete trials on every list.

The preamble list orders were the same as in Experiment 1, with representatives of the same items occurring in the same list positions. All experimental catch trials were separated by 14 other trials. On every list, the first, third, and fourth trials were complete-only fillers and the second trial was a filler catch-trial.

The preambles were digitally recorded and then edited by a female native English speaker. During recording, uniformity of the preamble intonation patterns was encouraged in two ways. First, each preamble followed a carrier phrase ("The next phrase is…) which was later stripped from the recording. Second, the intonation of the preambles themselves was the flat-to-rising pattern appropriate for the beginning of a declarative sentence, without the contour that typically accompanies the ends of isolated phrases. The preambles were edited with speech-editing software to remove short pauses and shorten continuants to speed up the speech rate without compromising clarity or naturalness. The end of each edited preamble accompanied the perceived offset of speech, so that the end of the file corresponded to the end of the preamble.

Procedure

The procedure was designed to allow the collection of completion onset latencies while ensuring that participants attended to and understood the preambles. On most (87.5%) of the trials, participants heard and completed the preambles without repeating them. Immediately after the offset of the preamble, an exclamation point cued speakers to complete the preamble as quickly as possible, with the first sensible conclusion that came to mind. However, on the catch trials, the word REPEAT signaled speakers to repeat the entire preamble before completing it.

Trials began with three asterisks centered on the computer monitor for 500 ms. Then the preamble was presented and, at offset, the exclamation point or the word REPEAT appeared. Voice onset latency was measured from preamble offset. When participants finished speaking, they pressed the space bar to continue to the next trial.

The instructions were the same as in Experiment 1, except that two trial examples were provided, one for each trial type, with the same preamble (*the fire engine*) serving as the example for both. One practice trial of each type was also included in the instructions. The adjectives to be used in the completions were presented near the beginning and again near the end of the instructions. Participants were instructed to complete the preambles as quickly as possible with an adjective that made sense, but not to think too long about the completion. At the conclusion of the instructions, participants were asked to repeat the adjectives once, to ensure that they remembered them.

Design

Each of the 64 participants received 16 items in each of the four conditions formed by combining singular and plural head and local noun number. The item design was fully within items, so that each item was seen by 16 participants in the four head/local number combinations.

Scoring

Responses on all trials were scored as in Experiment 1 for the four categories of consistent, inconsistent, and miscellaneous responses. Ambiguous responses did not occur, because the requirement to use predicate adjectives elicited forms of *to be* that were always inflected for number. Tables 6 and 7 give the numbers of responses of each type, by condition, for catch-trial and completion-only trials, respectively.

Statistical analyses

Voice onset latencies for number-consistent responses with singular (*is/was*) or plural *(are/were)* copula verbs constituted the primary dependent variable. Voice onset latencies less than 200 and more than 2000 ms were removed from the analyses, as were trials with invalid voice-key registrations. These constraints eliminated 10.7% of the trials, distributed similarly over conditions (2.8% in the singular–singular condition, 2.6% in the singular–plural condition, 2.5% in the plural–plural condition, and 2.8% in the plural–singular condition). The total numbers of consistent and inconsistent responses on which the mean latencies are based are given in Table 7. Because inconsistent responses were rare, with a majority of the cells empty, complete statistical analyses were not carried out on the inconsistent-response latencies.

The latencies for consistent responses were analyzed for participants and items in each condition. The factors in the analyses of variance were head-noun number, number match, and distributivity. For participants, these factors were completely crossed; for items, distributivity was nested. To verify the replication of Experiment 1, the proportions of inconsistent responses were also analyzed in the same design. MinF' was calculated from the results of the participant and item analyses. The inferential statistics from these analyses are shown in Table 5. Confidence

Table 6
Verb number produced in responses on catch trials in English, Experiment 2.

Preamble type	Singular	Plural	Ambiguous	Miscellaneous
Singular/singular				
Nondistributive	52	0	0	12
Distributive	50	0	0	14
Singular/plural				
Nondistributive	46	5	0	13
Distributive	44	9	0	11
Plural/plural				
Nondistributive	2	54	0	8
Distributive	1	49	0	14
Plural/singular				
Nondistributive	1	47	0	16
Distributive	3	50	0	11

Table 7
Voice onset latencies, number of responses for consistent and inconsistent preamble completions, and proportions of inconsistent responses in English, Experiment 2.

	Consistent		Inconsistent		
Preamble type	Voice onset latency	Number of responses	Voice onset latency	Number of responses	Proportion inconsistent
Singular/singular					
Nondistributive	868	479	941	12	.02
Distributive	862	480	718	6	.01
Singular/plural					
Nondistributive	856	438	893	45	.09
Distributive	864	388	941	92	.19
Plural/plural					
Nondistributive	809	467	992	28	.06
Distributive	817	466	867	23	.05
Plural/singular					
Nondistributive	901	438	986	47	.10
Distributive	878	449	948	41	.08

intervals for pairwise planned comparisons were derived from the mean-squared error in the participant and item analyses for the interaction among head number, number match, and distributivity.

Results

Fig. 2 displays the proportions of inconsistent responses for the catch-trial and completion-only trials. For the completion-only trials, analyses of variance confirmed the same set of differences as in Experiment 1 (see Table 5). Most important, distributive and nondistributive trials showed the expected differences in the proportions of number-inconsistent responses: There were .19 inconsistent responses for distributive and .09 for nondistributive trials. This .10 difference is comparable to the .09 distributive/nondistributive difference obtained in Experiment 1. Responses on catch trials resembled those on the completion trials (see Table 6), although the number of responses was too sparse for meaningful analysis.

To show how distributivity affected the timing of consistent responses on the completion trials, Fig. 3 displays the voice-onset latencies for each condition. Distributivity had little impact, producing no statistically significant main effects or interactions. With singular heads, latencies were almost identical regardless of local noun number, averaging 865 ms for singular local nouns and 860 ms for plural local nouns. Head and local noun number in the preambles mattered only when preambles had plural heads, creating a significant interaction between head number and number match. With plural heads, the mean voice-onset latency was 889 ms after singular local nouns and 813 ms after plural local nouns. This 76 ms difference exceeded the confidence intervals for pairwise contrasts for participants (44 ms) and items (31 ms) calculated from the interaction between head number and head-local match.

Inconsistent responses (responses in which the elicited verb number mismatched the head number; see Table 7) were rare, precluding a complete statistical evaluation of the latencies. In any event, the only informative results are in the singular–plural condition, where the preambles invited distributive or nondistributive interpretations. For the distributive preambles that elicited plural responses, consistent with the number meaning but inconsistent with the head number, latencies were about 48 ms slower than for the nondistributive preambles (which averaged 893 ms). A paired-sample two-tailed *t*-test on the participant data yielded an associated probability of .16.

Discussion

Measurements of the time required to produce number-consistent verbs after number-varying subject noun-phrases revealed few differences due to the distributivity of the noun phrases. The absence of variation cannot be ascribed to speakers' insensitivity to distributivity, because the verbs produced on experimental and catch trials alike revealed the same distribution of plural responses as in Experiment 1. That is, speakers produced more plural verbs when singular–plural preambles had distributive than nondistributive construals. The latency results of Experiment 2 suggest that this difference cannot be explained by protracted resolution of conflicts between distributive and nondistributive readings or between notional and grammatical agreement, since distributive and nondistrib-

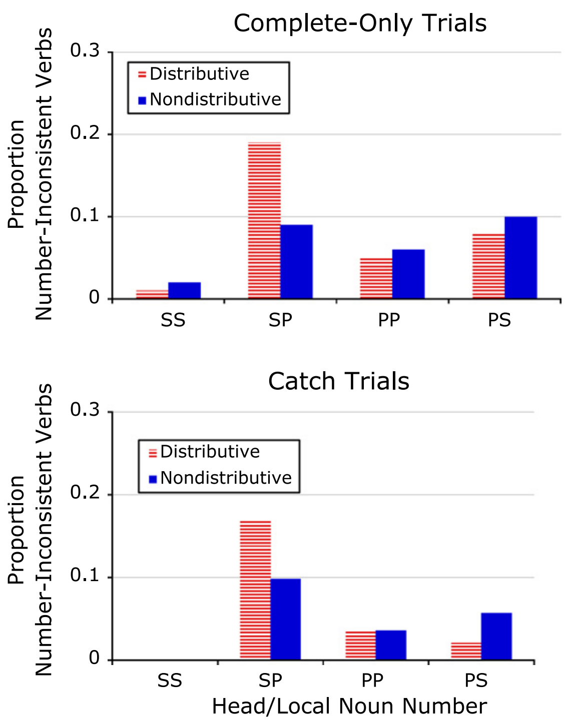

Fig. 2. Proportions of number-inconsistent verbs produced on catch trials in English (Experiment 2).

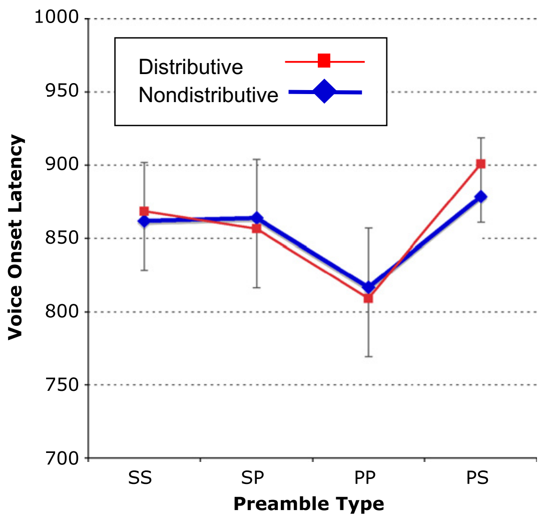

Fig. 3. Voice-onset latencies for number-consistent verbs after distributive and nondistributive subjects in Experiment 2, by preamble type. (Error bars represent the 95% confidence interval halfwidths for pairwise contrasts between condition means, drawn upward from Distributive points and downward from Nondistributive points.)

utive continuations were initiated equally rapidly in the singular–plural condition.

The only subject noun-phrases that clearly mattered to verb production latencies were those with plural heads and plural local nouns. Sentence subjects of this kind were completed faster than those in all other conditions, eliciting number-consistent plural verbs very rapidly. Since plural head nouns control verb agreement more reliably than singulars do, and plural local nouns trigger attraction more often, the combination of head and local plurality is the most plausible explanation for the fast plural responses in this condition.

Latencies for inconsistent responses were not tested statistically across all conditions, because the responses were rare and erratically distributed over participants and items. However, when distributive interpretations were likely, there was a nonsignificant trend toward slower inconsistent (i.e., plural) responses to distributive preambles. Such a trend could indicate that speakers rarely confronted the number ambiguity of distributive subjects, and when they did, opted for plural rather than singular agreement. This scenario is implausible, though. There was no tendency toward slowing when speakers produced singular verbs, despite the availability of a contradictory notional number. This would be expected only if speakers never produced singular verbs after encountering and resolving the distributive ambiguity. The absence of interference for singular responses in the singular/plural distributive condition thus suggests that there was no general agreement difficulty due to distributivity.

In short, the observed sensitivity to distributivity among English speakers in Experiment 1 is unlikely to be the product of experimentally elicited attention to number meaning or the agreement options it presents. Of course, it could be argued that the Spanish speakers in Experiment 1 may have been *less* sensitive to distributivity than the speakers in Vigliocco, Butterworth, et al. (1996), pleading for a comparison of voice-onset latencies in English and Spanish. Although such a comparison would fill out the empirical picture, its contribution would be negligible. It was Spanish that was predicted to exhibit notional sensitivity because of thinking-for-speaking or rich morphology. Experiment 1 revealed this sensitivity, replicating the Spanish results from Vigliocco, Butterworth et al. in detail, with the same materials and the same outcomes. The problem presented in Experiment 1 is not why the Spanish findings from the two experiments differed – they did not. The problem is why the English findings differed. By ruling out the contribution of an artifactual, task-induced sensitivity, Experiment 2 helps to establish that typological differences between English and Spanish are not a source of differences in how the nuances of number semantics are inferred or used.

General discussion

The present experiments offer a simple conclusion about the comparative sensitivity of speakers of different languages to number meaning. The first experiment replicated findings obtained by Vigliocco, Butterworth, et al. (1996), which showed that Spanish number agreement varied depending on the notional number of a subject noun phrase. Specifically, Spanish speakers were likely to produce plural verbs after grammatically singular subject noun phrases that could be construed as referring to multiple things. So, after distributive subjects like *The picture on the postcards*, speakers were more likely to produce plural verb agreement than after nondistributive subjects like *The key to the cabinets*. In the Vigliocco, Butterworth et al. experiments and in Experiment 1 here, Spanish showed an increased trend toward plural agreement on occasions when singular subject noun phrases had a plausible plural interpretation. This indicates a sensitivity to number meaning that influences number agreement.

Where does this sensitivity come from? It could stem from the mechanisms of language production, as hypothesized in theories that attribute differences in how number meaning is exploited to language-specific properties of syntactic and morphosyntactic processing (Vigliocco & Hartsuiker, 2002). Alternatively or additionally, there may be differences among speakers of those languages in how they recruit or categorize dimensions of nonlinguistic number meaning, molded in turn by the grammars of the languages themselves (Slobin, 1996). For Spanish and other languages, a grammatical characteristic that could shape the use or representation of number meaning is the richness of its morphology and the properties correlated with it.

Countering both views, our results showed that English speakers were just as responsive as Spanish speakers to notional number variations, despite the barrenness of English morphology. When English and Spanish speakers produced sentences designed to expose differences in number sensitivity, the levels of notionally driven agreement were much the same. This similarity between English and Spanish speakers, coupled with their similar susceptibility to other determinants of grammatical number agreement, suggests that variations in number morphology create neither enhanced use of number meaning within the mechanisms of language use nor special number sensitivity in speakers of such languages. The conclusion, then, is that lifelong exposure to the demands of rich number-agreement morphology did not predispose the Spanish speakers toward enhanced appreciation of subtle number meanings.

Experiment 2 was aimed at an alternative account of the English results. The thinking-for-speaking view of language differences suggests that Spanish speakers should take automatic advantage of distributive and nondistributive notional biases, implementing agreement in the same effortless way regardless of potential ambiguity. However, English speakers who grasp distributivity may have to resolve the accompanying ambiguity, taking more time to implement agreement. The unpressured time available in Experiment 1 would mask such a disparity in how English and Spanish speakers reached the same agreement values. Put under time pressure, though, English speakers should take longer to calculate agreement, increasing the latencies of verb production.

Experiment 2 found no support for this account. The difference in completion latencies for distributive and nondistributive subjects was only 8 ms, yet inconsistent responses revealed the same variations in verb number as in the first experiment. Responses on catch trials, when speakers performed the same task as in Experiment 1, were likewise the same. The implication is that the impact of distributivity on English agreement in the first experiment did not depend on unusually thoughtful evaluations of notional number.

The stability of the experimental outcomes is evident in the fact that the key results for each language have been obtained twice. The Spanish differences found by Vigliocco, Butterworth, et al. (1996) emerged in the same form and magnitude in Experiment 1, with different speakers from a different part of the Spanish-speaking world. The English results from Experiment 1, obtained with materials matched to those in Spanish, were replicated in Experiment 2. The findings thus put claims about language effects on number categorization into another light. Despite deep-seated contrasts between the grammars of Spanish and English in how agreement is represented linguistically, Spanish and English speakers appeared to produce agreement in similar ways, using similar language processes, and calling on similar conceptions of notional number (Eberhard et al., 2005).

What makes these similarities important is that the linguistic differences between English and Spanish are credible sources of variation in the kind or grain of number information that speakers draw on for implementing agreement. From a thinking-for-speaking perspective, it would come as no surprise to find that Spanish number morphology encourages speakers to tap and maintain detailed construals of numerosity for use in grammatical agreement. The discovery of differences between English and Spanish in the sensitive use of number information (Vigliocco, Butterworth, et al., 1996) thus seemed plausible and compelling. Turned around, if the differences disappear under strict testing, the plausibility of syntactically driven linguistic relativity is diminished.

The rich-morphology account of language differences in the expression of notional number is likewise problematic. Paradoxically, the effect of rich morphology seems to be the opposite of what the rich-morphology view predicts. Rather than creating variations in reliance on or sensitivity to number meaning, richer number morphology may in fact weaken the impact of number meaning on agreement. In a comparison of a morphologically rich language (Russian) with English, Lorimor et al. (2008) showed that the morphology-poor language, English, was more rather than less affected by number-meaning variations. A meta-analysis of cross-linguistic experiments suggested that this is a general tendency: The richer the morphology, the less likely speakers are to succumb to the impact of number meaning when producing agreement. A likely reason is that rich number morphology enforces systematic covariations in the syntax of agreement, across and within languages (Berg, 1998; Eberhard et al., 2005; Foote & Bock, in press) rather than opening more avenues for incursions from meaning. The small difference between Spanish and English in Experiment 1 can be interpreted – cautiously – in this light. If anything, rich morphology works against the expression of variations in notional number.

In the remainder of this discussion, we consider some further implications of our findings for existing perspectives on the links between syntax and number cognition. We also examine how the methods of the present experiment stack up again the kinds of considerations that emerge from classic tests of linguistic determinism and relativity.

The lens between language and thought

Just as for number agreement, other hypothesized links between language variations and number cognition can be interpreted in ways that discount a direct effect of language on how people think. Regarding the mass/count distinction, Barner, Li, and Snedeker (2010) argued that mass and count nouns provide different ways of capturing universal conceptions of objects, and not different ways of conceiving or representing them. That is, the knowledge of objects is not altered by a specific language's proprietary devices for expressing object properties. On this view, devices like mass/count syntax do not shape cognition, but serve a language's speakers by zooming in on detailed construals that might otherwise remain vague or ambiguous. For example, a language that includes mass/count syntax makes it easier for speakers to focus on individuals over species (e.g. "hunting for bears" emphasizes tokens of a type of prey) or on species over individuals (e.g. "hunting for bear" emphasizes the type of prey over the tokens; Allan, 1980). The details as well as the general outlines of knowledge that are captured in grammar will not be the same from one language to another, making the perspectives from which ideas are conveyed different. Yet the representations themselves may be fundamentally equivalent. The *what* that speakers aim to talk about is the same; the *how* differs as a function of the language used.

Our results align with this perspective and extend it to the productive, compositional processes of construal that support grammatical number agreement. Knowledge of quantities and quantification is not custom-tailored to fit the substantial typological differences in the properties of agreement systems crosslinguistically. Different agreement systems nonetheless allow speakers to specify or profile things differently. The contrast between British and American English in verb agreement with collective nouns offers an illustration. When a British speaker begins a sentence with the word *England*, the following verb can reveal whether the referent is a sports team or a country: *England* ***are*** refers to a team, such as the national cricket team; *England* ***is*** refers to the sovereign state. American English lacks this disambiguating device. Yet British and American speakers appear to have the same notional conceptions of numerosity for teams and countries. These conceptions come through in choices of pronoun number. With *team* or other collectives as antecedents, pronoun choices reflect a conception of *team* that is just as notionally plural for Americans as for Britons: American and British speakers are equally likely to use plural pronouns in reference to sports teams (Bock et al., 2006). So, for an American, after Pittsburgh *wins* (note the singular verb), *they* won (note the plural pronoun).

Although the similarities in conceptual representation that accompany linguistic contrasts can be striking, the differences are what fire perennial interest in linguistic determinism and relativity. Whorf offered his memorable examples (1956) of sentences from Hopi and other Native American languages to argue for striking divergences from ways of thinking in English. The well-known problem is that without evidence about the contents of thought, such examples need not reflect different ways of thinking at all, but only different ways of talking. A similar argument might be leveled at an interpretation of the present results in terms of nonlinguistic processes, since our findings capture ways of talking, too. However, there is a crucial difference between Whorf's evidence and ours. The weakness in Whorf's argument was that he had no handle on the nature of the thought behind the speech. In contrast, the differences in distributivity that we manipulated create notional differences that fill the gap in Whorf's logic.

Whorf's inference that what people say reflects what they think is surely correct; if it were not, human communication would fail. Where Whorf went wrong was not in using how people talk to infer how they think, but in using how people talk to infer how they think without knowing what they were thinking in the first place. When one knows what people are thinking, it can turn out that how they talk is not so different, after all (Barner et al., 2010).

Syntactic and lexical forces in linguistic relativity

Despite their relevance to the language-and-thought debate, investigations of the link between productive syntax and cognition are surprisingly scarce (among them are studies by Au, 1983, 1984; Bloom, 1981; Carroll & Casagrande, 1958). More common are studies of lexical domains, where relevant distinctions are associated with specific word knowledge, including color words and grammatical-number-bearing words like mass and count nouns. Along with Lucy (1992b), the present work treated the productivity of language and of syntax in particular as touchstones of the Whorf hypothesis. Grammatical number agreement does not depend on the features of words alone, but on structural relationships that are driven by the meanings behind phrases. These relationships can take countless forms. Such novel grammatical patterns could force language users to observe correlated patterns in meaning, and to do so habitually rather than occasionally (Slobin, 1996). This facet of linguistic relativity implies that the more often one's language requires number information in order to use the language, the more speakers should attend to notional number.

From this standpoint, it is illuminating to consider how often the nonlinguistic meanings relevant to number agreement must be evaluated. Estimates based on word frequencies imply that even in English, speakers confront the requirements of number agreement once every 16 words or so. Compare this to the frequency of words in the basic color vocabulary, one long-standing target of research on links between thought and language (e.g. Berlin & Kay, 1969; Brown, 1976; Brown & Lenneberg, 1954; Gilbert, Regier, Kay, & Ivry, 2006; Heider & Olivier, 1972; Kay & Regier, 2006; Roberson, Davies, & Davidoff, 2000). Basic color words in English occur once every 8850 words or so (calculated from the CELEX database; Baayen, Piepenbrock, & van Rijn, 1993). The upshot is that speakers make more than 550 notional number distinctions for each basic color distinction. As Lucy (1992a, 1992b) argued, if Whorfian effects are to be found, they are more likely to emerge in the distinctions that the grammars of language force speakers to make efficiently. Our results provide evidence that, at least in one important cognitive and linguistic domain, they do not.

The number knowledge that speakers of all languages have grows from deep ontogenetic and phylogenetic roots. Infants are capable of perceiving the variations in nonlinguistic number that are relevant to the grammars of language (Feigenson, Dehaene, & Spelke, 2004); for number agreement systems in particular, the small-number distinctions that are needed are accessible in a nonlinguistic format at the outset of language acquisition. This makes small-number discrimination a perceptual and cognitive ability that language can build on, rather than an ability that language creates or modifies (Bock & Kahn, 2009).

Separating thinking from speaking in research on language production

In the present experiments, the method that we used to pin down what people were thinking may seem peculiar in the context of other research on linguistic relativity. In such research, it is important to ensure that people do not use or rely on language in experimental tests. If speakers call on language to guide what they do in a categorization task, for example, the task may be little more than another way of observing differences in what things are called. The goal, instead, is to banish language as much as possible. In our experiments, though, speakers of English and Spanish heard snippets of English and Spanish, respectively, and then produced sentences in English and Spanish. It can be hard to see where there is any role for thinking divorced from language. This particular kind of worry reflects a common conception of language production as speech production, a mere conversion from inner speech to outer speech. The flaws in this "mind in the mouth" perspective are many (Bock, 1996; Oppenheim & Dell, 2008), but what is especially pernicious for psycholinguistic research on language production is the idea that when people hear speech and then produce it, the verbal chain from listening to speaking remains unbroken by nonlinguistic cognition.

To understand how nonverbal thinking comes into this sequence of communicative events, it has to be kept in mind that people are astonishingly poor at retaining the wording of what they hear or read (Potter & Lombardi, 1990), even over intervals as brief as those involved in speech shadowing (Marslen-Wilson, 1973). Yet they are very good at retaining the gist, the abstract ideas behind the language they experience. When ideas are channeled back into language, they bear the marks of thought processes shorn of linguistic form (Bransford & Franks, 1971; Bransford & Johnson, 1972), including striking similarities to the kinds of ideas inspired by the verbally unmediated perception of scenes and events (Lichtenstein & Brewer, 1980).

Research on grammatical agreement reveals the same untethering of language and thought, reflected in dissociations of the grammatical and notional number features of noun phrases (e.g. differences in verb and pronoun number agreement in English; Bock et al. 2004; and dissociations of grammatical and notional gender agreement in Hebrew; Deutsch & Dank, 2009) A mediating influence of nonverbal cognition between language comprehension and production is also needed to explain why agreement elicited by pictured objects patterns in the same way as agreement elicited by verbally described objects (Eberhard, 1999; Foote & Bock, in press). If speech gives rise to thoughts, and thoughts are uniquely shaped by characteristics of a specific language (characteristics absent from the contents of the speech itself), or if thoughts provoke the production of phrases that are uniquely colored by characteristics of a specific language (characteristics irrelevant or unnecessary for the objective expression of ideas), the presence or absence of these shades of meaning bears on whether the thoughts behind the language are shaped by the language spoken. Bluntly, the fact that a "manipulation of thought" is set in motion by language makes it no less a manipulation of thought, given the workings of everyday language comprehension and memory.

Conclusion

In summary, the present experiments challenge a plausible and provocative claim about the effects of cross-linguistic differences on the apprehension and construal of numerosity. If the grammars of languages are capable of molding nonlinguistic processes in any domain, number is a prime candidate. There is evidence in favor of this hypothesis from studies of number agreement in Spanish and English, languages in which grammatical differences could yield differences in the apprehension of number information. However, a tightly controlled comparison uncovered no contrasts between the languages when Spanish and English speakers used notional number to compose agreement features for novel phrases. The implication is that Spanish does not promote richer or finer-grained notional number representations than English, regardless of differences in grammatically relevant morphology. In short, the findings are consistent with the absence of deep-seated differences in number semantics due to crosslinguistic differences in the requirements of number grammar.

Acknowledgments

This work was supported in part by research and training grants from the National Science Foundation (SBR 98-73450, BCS 02-14270, BCS 08-43866), the National Institutes of Health (R01-MH66089, T32-MH18990), the Spanish Ministry of Education and Science (PSI_2009-08889), and CONSOLIDER-INGENIO 2010 (CSD2008-00048). A preliminary report of the first experiment was presented at the meeting of the Psychonomic Society in 2004. We thank Elizabeth Octigan and Matthew Rambert for their help with the experiments and data analyses, and Victor Ferreira, Gerry Altmann and four anonymous reviewers for comments on previous versions of the manuscript. Correspondence should be directed to Kathryn Bock, Beckman Institute, University of Illinois, 405 North Mathews, Urbana, IL 61801, USA or jkbock@illinois.edu.

Appendix A

See Table A1.

Table A1
Experimental preamble materials in English and Spanish (with token classifications from Vigliocco, Butterworth, & Garrett, 1996).

English	Spanish (singular–plural version)	Gender of head and local nouns in Spanish (f = feminine; m = masculine)
Nondistributive (Single token)		
the complaint(s) of the student(s)	La queja de las estudiantes	ff
the prize(s) for the winner(s)	La recompensa a las ganadoras	ff
the author(s) of the novel(s)	La autora de las novelas	ff
the model(s) in the photograph(s)	La chica de las fotografías	ff
the light(s) above the table(s)	La luz sobre las mesas	ff
the offense(s) against the secretary(ies)	La ofensa a las mujeres	ff
the mother(s) of the kid(s)	La madre de las niñas	ff
the house(s) beyond the hill(s)	La casa de las colinas	ff
the home(s) of my cousin(s)	La casa de mis primos	fm
the illness(es) of his brother (s)	La enfermedad de los hombres	fm
the trap(s) for the rat(s)	La trampa para los ratones	fm
the snapshot(s) of the tourists	La fotografía de los turistas	fm
the demand(s) from the owner(s)	La demanda contra los proprietarios	fm
the suggestion(s) to the producer(s)	La sugerencia a los directores	fm
the song(s) of the tenor(s)	La canción de los cantantes	fm
the theory(ies) of the graduate(s)	La teoría de los licenciados	fm
the director(s) of the movie(s)	El director de las películas	mf
the doctor(s) for the invalid(s)	El médico de las enfermas	mf
the mechanic(s) for the motorcycle(s)	El mecánico de las motocicletas	mf
the teacher(s) for the girl(s)	El maestro de las chicas	mf
the trapper(s) with the net(s)	El pescador con las redes	mf
the boy(s) with the crutch(es)	El niño con las muletas	mf

(continued on next page)

Table A1 (*continued*)

English	Spanish (singular–plural version)	Gender of head and local nouns in Spanish (f = feminine; m = masculine)
the grandfather(s) of the kid(s)	El abuelo de las niñas	mf
the debate(s) about the drug(s)	El debate sobre las drogas	mf
the road(s) to the lake(s)	El paseo por los lagos	mm
the gift(s) for the baby(ies)	El regalo para los bebes	mm
the attack(s) against the minister(s)	El atentado contra los ministros	mm
the witness(es) for the lawyer(s)	El testigo de los abogados	mm
the warning(s) of the expert(s)	El aviso de los expertos	mm
the right(s) of the worker(s)	El derecho de los trabajadores	mm
the warden(s) at the fire(s)	El responsable de los incendios	mm
the odor(s) of the almond(s)	El olor de los almendros	mm
Distributive (Multiple token)		
the crane(s) in the stone quarry(ies)	La grua de las canteras	ff
the medal(s) for the female athlete(s)	La medalla de las niñas	ff
the computer(s) in the office(s)	La computadora de las oficinas	ff
the skirt(s) for the actress(es)	La falda de las mujeres	ff
the wake(s) from the ship(s)	La estela de las embarcaciones	ff
the label(s) on the bottle(s)	La etiqueta de las botellas	ff
the chimney(s) of the house(s)	La chimenea de las casas	ff
the entrance(s) to the apartment(s)	La puerta de las casas	ff
the cover(s) of the book(s)	La tapa de los libros	fm
the breed(s) of the dog(s)	La raza de los perros	fm
the schedule(s) of the professor(s)	La agenda de los profesores	fm
the note(s) for the third grader(s)	La nota de los estudiantes	fm
the bank(s) of the river(s)	La orilla de los ríos	fm
the bow-tie(s) of the clown(s)	La corbata de los payasos	fm
the estate(s) of the president(s)	La residencia de los presidentes	fm
the defect(s) in the airplane(s)	La avería de los aviónes	fm
the coat(s) on the lady(ies)	El abrigo de las señoras	mf
the pregnancy(ies) of the teenager(s)	El embarazo de las mujeres	mf
the number(s) on the license plate(s)	El número de las tarjetas	mf
the purse(s) for the girl (s)	El bolso de las chicas	mf
the mayor(s) of the city(ies)	El alcalde de las ciudades	mf
the stamp(s) on the letter(s)	El sello de las cartas	mf
the color(s) of the flower(s)	El color de las flores	mf
the villain(s) in the movie(s)	El malo de las peliculas	mf
the bell(s) on the cat(s)	El cascabel de los gatos	mm
the baseball cap(s) for the guy(s)	El gorro de los hombres	mm
the insurance policy(ies) for the car(s)	El seguro de los coches	mm
the name(s) of the boy(s)	El nombre de los niños	mm
the slogan(s) on the poster(s)	El dibujo de los carteles	mm
the plaque(s) on the doorway(s)	El timbre de los portales	mm
the driver(s) of the truck (s)	El conductor de los camiónes	mm
the uniform(s) for the soldier(s)	El uniforme de los soldados	mm

References

Allan, K. (1980). Nouns and countability. *Language, 56*, 541–567.

Au, T. K.-F. (1983). Chinese and English counterfactuals: The Sapir–Whorf hypothesis revisited. *Cognition, 15*, 155–187.

Au, T. K.-F. (1984). Counterfactuals: In reply to Alfred Bloom. *Cognition, 17*, 288–302.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia: Linguistic Data Consortium.

Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition, 111*, 329–344.

Barner, D., Li, P., & Snedeker, J. (2010). Words as windows to thought: The case of object representation. *Current Directions in Psychological Science, 19*, 195–200.

Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition, 97*, 41–66.

Barner, D., & Snedeker, J. (2006). Children's early understanding of mass-count syntax: Individuation, lexical content, and the number asymmetry. *Language Learning and Development, 2*, 163–194.

Barner, D., Wagner, L., & Snedeker, J. (2008). Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns. *Cognition, 106*, 805–832.

Berg, T. (1998). The resolution of number conflicts in English and German agreement patterns. *Linguistics, 36*, 41–70.

Berlin, B., & Kay, P. (1969). *Basic color items: Their universality and evolution*. Berkeley: University of California Press.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Hillsdale, NJ: Erlbaum.

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421.

Bock, J. K., Butterfield, S., Cutler, A., Cutting, J. C., Eberhard, K. M., & Humphreys, K. R. (2006). Number agreement in British and American English: Disagreeing to agree collectively. *Language, 82*, 64–113.

Bock, J. K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251–278.

Bock, J. K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*, 83–128.

Bock, J. K., & Kahn, J. (2009). *Number sense and number syntax: The magical* * * *. Paper presented at the meeting of the Psychonomic Society, Boston, MA.

Bock, J. K., & Middleton, E. L. (in press). Reaching agreement. *Natural Language & Linguistic Theory*.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*, 330–346.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717–726.

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

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

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: Holt, Rinehart & Winston.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25*, 257–271.

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language, 60*, 112–143.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language, 41*, 560–578.

Eberhard, K. M., Cutting, J. C., & Bock, J. K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review, 112*, 531–559.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science, 8*, 307–314.

Foote, R., & Bock, J. K. (in press). The role of morphology in subject–verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS, 103*, 489–494.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language, 48*, 760–778.

Heider, E. R., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology, 3*, 337–354.

Humphreys, K. R., & Bock, J. K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review, 12*, 689–695.

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

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2010). Does the grammatical count/mass distinction affect semantic representations? Evidence from experiments in English and Japanese. *Language and Cognitive Processes, 25*, 189–223.

Kay, P., & Regier, T. (2006). Language, thought and color: Recent developments. *Trends in Cognitive Sciences, 10*, 51–54.

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology, 58*, 487–524.

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology, 12*, 412–445.

Lorimor, H. (2007). *Conjunction and grammatical agreement.* Unpublished doctoral dissertation, University of Illinois, Urbana, Illinois.

Lorimor, H., Bock, J. K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes, 23*, 769–799.

Lorimor, H., Middleton, E., & Bock, J. K. (2006). *One and one makes singular agreement*. Poster presented at the CUNY, New York, New York.

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

Lucy, J. A. (1992b). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, England: Cambridge University Press.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*, 522–523.

McPherson, L. (1991). A little goes a long way: Evidence for a perceptual basis of learning for the noun categories *count* and *mass*. *Journal of Child Language, 18*, 315–338.

Middleton, E. L., Wisniewski, E. J., Trindel, K. A., & Imai, M. (2004). Separating the chaff from the oats: Evidence for a conceptual distinction between count noun and mass noun aggregates. *Journal of Memory and Language, 50*, 371–394.

Morgan, J. L. (1972). Verb agreement as a rule of English. In P. M. Peranteau, J. N. Levi, & G. C. Phares (Eds.), *Papers from the eighth regional meeting, Chicago Linguistic Society* (pp. 278–286). Chicago, IL: Chicago Linguistic Society.

Morgan, J. L. (1984). Some problems of determination in English number agreement. In G. Alvarez, B. Brodie, & T. McCoy (Eds.), *Proceedings of the Eastern States Conference on Linguistics* (pp. 69–78). Columbus, OH: Ohio State University.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528–537.

Paivio, A., Yuille, J. C., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement, 76*.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633–654.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

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

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

Soja, N. N. (1992). Inferences about the meanings of nouns: The relationship between perception and syntax. *Cognitive Development, 7*, 29–45.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language, 60*, 308–327.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition, 114*, 447–454.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject–verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language, 34*, 186–215.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*, 261–298.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in language production. *Psychological Bulletin, 128*, 442–472.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*, 407–442.

Vigliocco, G., Vinson, D. P., Martin, R. C., & Garrett, M. F. (1999). Is "count" and "mass" information available when the noun is not? An investigation of tip-of-the-tongue states and anomia. *Journal of Memory and Language, 40*, 534–558.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.