

Zuzanna Fuchs\*, Maria Polinsky, and Gregory Scontras

# The Differential Representation of Number and Gender in Spanish

**Abstract:** This paper investigates the geometry of phi-features, with a special emphasis on number and gender in Spanish. We address two sets of questions: (i) are number and gender bundled together or do they constitute separate categories, and (ii) does the internal feature composition of number and gender follow a single- or a multi-valued system? Given the lack of consensus on these issues based on primary data, we approach these questions experimentally, using the phenomenon of agreement attraction: a situation in which ungrammatical sequences are perceived as grammatical when one of the NPs is erroneously identified as determining agreement. Our results offer novel support in favor of an agreement model in which number and gender are in separate projections and are valued independently. In addition, our results indicate that number but not gender in Spanish is multi-valued.

**Keywords:** agreement, phi-features, number, gender, feature geometry, small clauses, Spanish, agreement attraction, comprehension

---

**\*Corresponding Author: Zuzanna Fuchs:** Department of Linguistics, Harvard University.  
E-mail: [zuzannafuchs@fas.harvard.edu](mailto:zuzannafuchs@fas.harvard.edu).

**Maria Polinsky:** Department of Linguistics, University of Maryland, College Park.

**Gregory Scontras:** Department of Psychology, Stanford University.

## 1 Introduction

### 1.1 Bundling and splitting in feature geometry

Agreement is everywhere, which is probably why it is so hard to determine where exactly it is and what exactly is the nature and representation of the features involved in it, that is, phi-features. The classes of phi-features are well known; despite the apparent variability of agreement, the recurrent features that it may track are constant: person, number, and gender. What remains less clear is the relationship of these features to each other and their hierarchical arrangement in agreement systems. These are the relationships that are often subsumed under the rubric of feature geometry.

This paper addresses feature geometry with respect to two out of three features, namely, number and gender. The choice of these two features is not arbitrary. The hierarchical position of person with respect to the other features is relatively well understood (cf. Harley and Ritter, 2002); it is also known that person features do not appear on non-verbal probes, which again separates it from the other features (Baker, 2008). Meanwhile, the relationship between gender and number is less clear. Assuming that both features are represented in syntax, there are two analytical possibilities, both proposed in the literature. According to one scenario, gender and number are always bundled together (cf. Ritter, 1993; Carstens, 2000, 2003); we will be referring to this approach as the bundling model. Under that model, all gender morphology is either hosted on the number head, as shown in (1), or is expressed on the specifier within a number phrase, as shown in (2). Either way, gender features do not project independently of number, and the valuation of gender presupposes a valuation of number.

(1) [NumP ... [Num<sub>{Number, Gender}</sub>]]

(2) [NumP XP<sub>{Gender}</sub> [Num<sub>{Number, Gender}</sub>]]

The bundling model draws its empirical inspiration from the fact that languages regularly combine gender and number information; one rarely finds systems where the two features participate in agreement and yet are independent of each other. Furthermore, gender is lexically specified; a given noun belongs to gender X regardless of its syntactic position. The noun leaves the lexicon with a gender, and this gender persists throughout its use. Here gender stands apart from number, which is specified within a given eventuality: the number feature of a noun depends on its intended referent in a given use. Thus, number is tightly linked to event structure, the way that case is. Since gender is not directly linked to the event or argument structure, bundling models argue that it is desirable to have its representation in syntax mediated by another grammatical feature which is directly mapped into syntax. Such mediation is akin to representing animacy only indirectly via some other feature, for example via case. And finally, a strong argument for the bundling model is based on the gender of inanimate nouns. The gender of such nouns is uninterpretable, which in turn means that the gender projection cannot always have consistent semantic content. Elimination of semantically inconsistent projections is an important theoretical goal within the minimalist program (Chomsky, 1995), so eliminating an independent and semantically heterogeneous gender projection would result in a more parsimonious theory.

In the alternative, split model (Picallo, 1991; Carminati, 2005; Antón-Méndez et al., 2002, a.o.), gender morphology hosted on a nominal stem heads its own projection (GenP). GenP is dominated by NumP (i.e., the source of number features/morphology), as schematized below:

$$(3) \quad [\text{NumP} [\text{GenP} \dots]]$$

One of the major arguments in favor of the split model comes from the order of morphemes in nominal derivations. In those languages where number and gender morphology can be descriptively separated, the order is Stem-Gender-Number, as in the following Spanish examples:

- (4)    a.    [[libr]-[GenP o-] [NumP s]]    ‘books’  
          b.    [[libr]-[GenP o-] [NumP ∅]]    ‘book’

Because it levels the hierarchical distinction between number and gender, the bundling model does not have a straightforward way of predicting the ordering shown here. That the split model derives such an order is a side effect of the simple feature geometry: number dominates gender. Furthermore, as syntactic theory has been moving away from the division of features into uninterpretable and interpretable, and toward giving more weight to feature valuation itself (Pesetsky and Torrego, 2007; Preminger, 2014), one of the otherwise strong theoretical arguments in favor of the bundling model (i.e., eliminating an uninterpretable gender projection) may be losing some of its heft.

Further support for a split model of number and gender comes from cases of nominal ellipsis. In Spanish (Despiante and Masullo, 2001; Kornfeld and Saab, 2002; Saab, 2008, 2010)<sup>1</sup> and in Greek (Merchant, 2014), the number features of an elided nominal phrase do not need to match those of its antecedent, but the gender features do. The relevant contrast is illustrated for Spanish in (5). (5-a) is grammatical despite the number mismatch between the masculine plural elided nominal and the masculine singular antecedent, whereas (5-b) is ungrammatical due to the gender mismatch between the feminine singular elided nominal and the masculine singular antecedent. Under a bundling hypothesis, it is unclear why one type of feature mismatch is allowed while the other is not. This asymmetry can be accounted for under a model in which gender and number are split, with gender inside the ellipsis domain but number outside of it. However, examples such as (5) can be constructed

---

<sup>1</sup> We are grateful to an anonymous reviewer for bringing the Spanish facts to our attention.

only for animate nouns where gender is interpretable, whereas this paper is concerned with lexical gender on inanimate nouns. In fact, the approach detailed in this paper can indeed be extended to animate nouns, as discussed in section 6 below.<sup>2</sup>

- (5) a. Juan visitó a su tío, y Pedro visitó  
 Juan visit.3SG.PST to 3SG.POSS uncle, and Pedro visit.3SG.PST  
 a los suyos.  
 to DET 3.POSS.M.PL  
 ‘Juan visited his uncle, and Pedro visited his (uncles).’  
 b. \*Juan visitó a su tío, y Pedro visitó  
 Juan visit.3SG.PST to 3SG.POSS uncle, and Pedro visit.3SG.PST  
 a la suya.  
 to DET 3.POSS.F.SG  
 ‘Juan visited his uncle, and Pedro visited his (aunt).’

In what follows, we present an attempt to take the debate between bundling vs. splitting out of purely theoretical considerations and ground it in the psychological reality of agreement itself, as applied to Spanish. The logic is as follows. Assuming that grammar and its parser are in an isomorphic relation (e.g., Phillips, 2013), observing the parser allows for the observation of the grammar. In other words, speakers’ behavior in the parsing of agreement phenomena should stand as the proving ground for the theories that underlie the phenomena. Much of the work in such an approach centers around developing testable hypotheses about behavior on the basis of articulated theories of grammar. To this end, we turn now to further background on the theory of agreement features and their geometry.

## 1.2 Internal structuring of agreement features

An evaluation of the two approaches to feature geometry sketched above presupposes knowledge of the way a feature is internally structured in a given language. Let us clarify this notion by analogy with the inflectional phrase (IP). In some languages, IP is represented by a tense phrase, in others, by an aspect phrase, and its language-specific organization can have morphosyntactic

---

<sup>2</sup> We use the following abbreviations in our glosses: 3 = third person; COMP = complementizer; DET = determiner; F = feminine; INF = infinitive; M = masculine; PL = plural; POSS = possessive; PRS = present; PST = past; SG = singular.

repercussions for clause structure. Likewise, the morphosyntactic structuring of gender and number cannot *a priori* be assumed to be the same across-the-board; it has to be evaluated in the context of a given language before the evaluation of bundling vs. split models in that language.

The internal structuring of number and gender can be represented in terms of feature specification. For example, in a gender system that divides all nouns into masculine and feminine, are both genders specified morphologically? And are they both specified semantically? Morphological specification can be directly observed; if one of the genders requires extra marking and the other does not, we can state that the former is morphologically specified and the latter, unspecified. Determining semantic specification

For number, one way evaluating semantic specification is by observing whether the singular or plural is applied in borderline cases. Consider decimals in English: *0.1 miles* refers to an distance less than one whole mile. It is therefore not quite singular, but certainly it does not reference more than one mile. Nevertheless, the measure term *miles* appears in the plural. This pattern and others like it suggest that the plural is a kind of default, whereas the singular can only be applied in specific cases where a clear singular referent is intended, as in *1 mile*. Semantic specification can also be operationally determined using the Greenbergian Taghlib test (Greenberg, 1966; Sauerland et al., 2005)<sup>3</sup>: given a plurality of individuals, where some but not all have the specified (marked) property, only the unspecified form in a pair of features can be used to refer to that plurality. Let us illustrate the notion with an informal example. Should you receive an invitation to a party stating that you are welcome to bring *guests*, singular reference is included in the use of the plural. In other words, bringing only one guest would be in compliance with the invitation. However, an invitation stating that one is welcome to bring a *guest* does not include the plural reference with the use of the singular. Showing up with more than one friend would violate the letter of the invitation. Thus, in English, converging evidence suggests that singular number is semantically specified, as it restricts nominal reference to just singular individuals.<sup>4</sup> The plural, however, is semantically unspecified, permitting reference to both sin-

---

<sup>3</sup> Greenberg (1966) proposes this test and attributes it to Arab grammarians, giving it the Arabic name *taghlib* 'dominance'.

<sup>4</sup> Even Spector (2007), who proposes that the singular *morpheme* has no semantic import, agrees that "there is a sense in which the singular morpheme represents the marked item in the singular/plural opposition: namely, singular nouns are semantically *more specific* than plural nouns."

gulars and plurals. But when it comes to the actual feature content (i.e., the possible values for number), we face competing pressures from semantics and morphology. The Taghlib Test suggests a singular feature, but the morphological exponence of the plural suffix *-s* suggests a plural feature. In fact, most theories of number posit that both features get specified in the determination of nominal semantics and agreement, at least in English (cf. Sauerland, 2003; Scontras, 2013a,b).

When it comes to the specification of gender, we need to expand our sights beyond English. Many languages have gender systems, so there is no shortage of choices. Our choice of Spanish, the language we will focus on here, is motivated by several considerations. First, the structure of Spanish number is sufficiently similar to that of English, which allows for an effectual comparison between the two languages. Second, as we show below, the two genders of Spanish are distributed roughly equally, which is important in minimizing the potential imbalance between the genders due to statistical preferences. Third, there is debate surrounding the internal structure of Spanish gender (cf. Harris, 1991). Some more practical considerations include the fact that there is already some experimental work on Spanish number and gender, which allows us to build upon existing findings. And finally, Spanish is widely spoken, which facilitates access to a large pool of experimental participants.

Since most of the issues surrounding number and gender geometry arise from the consideration of inanimate nouns, our work will be limited to these nouns. A number of researchers have advocated separating the gender of animates and the gender of inanimates, distinguishing them as high and low gender, respectively (cf. Kramer, 2009, 2013; Bobaljik and Zocca, 2011, as well as the ellipsis facts discussed above). Assuming this division is on the right track, our focus is only on the low, lexical gender. Still, we anticipate that future work could include high gender; we return to this point in section 6.

The remainder of the paper is structured as follows. In section 2 we present a brief overview of the relevant empirical data in Spanish; we address the nature of number and gender oppositions, and in particular present a summary of the seminal paper by Harris (1991), who argues that Spanish gender is a single-valued opposition. In section 3 we introduce the experimental work applied in this study and present its main results. Section 4 establishes the internal structuring of Spanish number and gender in light of the experimental results. Based on this structuring, we compare the bundling and the split models of Spanish phi-features in section 5. Section 6 concludes with a discussion of our results in relation to languages other than Spanish, as well as to issues surrounding low vs. high gender.

## 2 Background: Spanish number and gender

Agreement in Spanish is rampant: determiners, adjectives, and participles must all agree with a head noun in both number and gender. In addition, this agreement in number and gender is maintained in anaphors. To better understand the features involved in this agreement, we start with a fuller discussion of Spanish number, then turn to gender.

### 2.1 Number

Spanish has two numbers: singular and plural. Much like in English, the plural is morphologically marked with the suffix *-s*, while the singular appears morphologically bare.

- (6) a. la manzana ‘the apple’  
       las manzanas ‘the apples’  
       b. el plátano ‘the banana’  
       los plátanos ‘the bananas’

In terms of the semantic specification of number, Spanish is again similar to English; consider the examples in (7).<sup>5</sup>

- (7) a. Si gana los puntos, puede traer  
       If win.3SG.PRS the point.M.PL, can.3SG.PRS bring.INF  
       invitados/niños de cualquier edad.  
       guest.M.PL/child.M.PL of any age  
       If you earn the points, you can bring guests/children of any age.  
       b. Si gana los puntos, puede traer un  
       If win.3SG.PRS the point.M.PL, can.3SG.PRS bring.INF a  
       invitado/niño de cualquier edad.  
       guest.M.SG/child.M.SG of any age  
       If you earn the points, you can bring a guest/child of any age.

Example (7-a) implies that one may bring several guests, or just one; as in English, singular reference is included in the use of the plural. On the other hand, (7-b) restricts the recipient of the invitation to bringing no more than

---

<sup>5</sup> These examples are adapted from a Scentsy.com rewards brochure outlining entrance qualifications to the Disney World park in Florida.

one guest; plural reference is not included in the use of the singular. These facts suggest that in Spanish, the singular is semantically specified, while the plural is semantically unspecified. Thus, with the internal composition of number, Spanish faces competing pressures from semantics and morphology. The Taghlib Test suggests that the singular is semantically specified, but morphology specifies the plural with *-s*. As in English, these competing pressures suggest that both singular and plural number features are equally available.

## 2.2 Gender

Spanish has two genders, masculine and feminine, which are distributed approximately equally in the nominal lexicon: masculine 52%, feminine 45% (Bull, 1965).<sup>6</sup> With respect to semantic specification, only the feminine is semantically specified. Consider *los padres*, which is ambiguous between ‘parents’ and ‘fathers’; on the first reading, feminine reference is included in the use of the masculine (‘parents’ often references classes of mixed gender). In contrast, *las madres* ‘mothers’ can only refer to a plurality of women, but not to parents of both genders. In a different realization of the same pattern, if two or more nouns of both genders are coordinated, the agreeing adjective has to be the masculine:

- (8) El libro y la pintura son caros / \*caras  
 the book.M and the painting.F are expensive.M.PL / expensive.F.PL  
 ‘The book and the painting are expensive.’

Whether or not we posit an alignment between the morphological and semantic specification of Spanish gender depends on the analysis of nominal morphology that gets assumed. For example, if we take *-o* and *-a* to be the morphological realization of masculine and feminine gender, respectively, then both genders are equally specified morphologically. However, this morphological analysis of gender has been questioned, most notably by Harris (1991). He notes that neither word marker (i.e., *-o* and *-a*) is exclusively associated with a particular gender. The typically masculine word marker *-o* can also appear on feminine nouns (e.g., *la mano* ‘the hand’), nouns that can be masculine or feminine (e.g., *el/la testigo* ‘the witness’), as well as words not associated with a gender at all (e.g., the preposition *dentro* ‘inside’). Similarly, the

<sup>6</sup> The three percent discrepancy in the numbers is due to the presence of epicene, or gender-neutral nouns.



word marker *-a*, which is typically associated with feminine nouns, can also mark masculine nouns (e.g., *el día* ‘the day’), nouns that can be masculine or feminine (e.g., *el/la turista* ‘the tourist’), and words associated with neither gender (e.g., the preposition *afuera* ‘outside’).<sup>7</sup> In addition, coordinate noun phrases with ostensibly different gender values on the constituents determine masculine agreement, as in (8) above, and neologisms are assigned masculine gender regardless of their morphological shape – for instance, the preposition *para* ‘for’ is inherently genderless, but when mentioned metalinguistically it will also determine masculine agreement.<sup>8</sup> Given the lack of a consistent mapping between specific endings and genders, we lack clear arguments concerning the morphological specification of Spanish gender. Based on the facts just sketched, Harris (1991) argues that only the feminine in Spanish gets specified morphologically, which aligns the morphological and semantic specification of gender in Spanish and provides support for his claim that only the feminine feature gets specified: in the absence of feminine, nouns appear with default (i.e., masculine) morphology and receive default semantics. In other words, what is traditionally described as masculine gender is simply the lack of a feminine feature; there is no masculine gender feature in Spanish.

The opposite view, according to which Spanish gender is multi-valued, has also received support. According to Roca (1989), Domínguez et al. (1999), and Alarcón (2006), the surface facts of Spanish indicate equal presence of a feminine and a masculine gender features in Spanish. The multi-valued account is designed to reflect the morphology of determiners, where each of the genders enjoys equal representation, and the marking on determiners is subtractive for the masculine only in the indefinite singular (*un* vs. *una*); in all other subsystems, both genders have equally visible morphology on their determiners: *el/la*, *los/las*, *unos/unas*. Thus, under the multi-valued approach, both genders are specified morphologically as a result of the availability of two gender features, masculine and feminine.

---

<sup>7</sup> Harris (1991) does note an imbalance in these facts. While many masculine nouns bear the feminine marker *-a*, only a few feminine nouns bear the masculine marker *-o*.

<sup>8</sup> But see Tiogang (2015) for instances in which neologisms are assigned feminine gender, possibly on the basis of their morphological shape.

## 2.3 Previous experimental work

To date, the work of many of our predecessors in the empirical investigation of the Spanish agreement system leaves a number of unanswered questions regarding feature specification, and as a consequence the results concerning the relationship between number and gender are inconclusive. We review this work below.

In a production study of the behavior of Spanish number and gender, which builds on Antón-Méndez (1999), Antón-Méndez et al. (2002) presented speakers with a predicative adjective as well as a complex noun, designing the complex nouns such that each contained a distractor noun phrase that could erroneously control agreement (more on this in our discussion of agreement attraction below). They then asked the speakers to reproduce the complex noun and complete the sentence using the adjective they had been shown. The researchers measured the frequency of number errors and gender errors made in agreement between the head noun and the adjective, as well as the frequency of number errors in agreement between the head noun and verb.

The authors focused on conditions in which the distractor noun mismatched the head noun in both number and gender. If number and gender are dependent (i.e., bundled together), they claim, all errors in agreement between the adjective and the head noun that involve number should also involve gender, and vice versa (what Antón-Méndez et al. refer to as “combined errors”). If they are independent, however, then laws of probability suggest we should observe far fewer combined errors than either number errors or gender errors in agreement.

The results of the experiment conducted by Antón-Méndez et al. showed that the frequency of combined errors was much smaller than the frequency of just number errors or just gender errors. In fact, a comparison of the observed probabilities of number, gender, and combined errors was consistent with a model in which the probability distributions of number errors and gender errors were independent of each other. The authors took the evidence as support for a model in which number valuation and gender valuation are independent, which in turn argues in favor of an analysis in which number and gender are projected independently, a split model.

However, Antón-Méndez et al. found that, while gender and number may not be dependent, they are sensitive to each other. Their results show that speakers were more likely to make a gender agreement error between the head noun and the adjective if the head noun and distractor noun were both singular. Similarly, speakers were more likely to make a error in number agreement if the head noun and distractor noun were of the same gender. The result is

an apparent contradiction: Valuation of number and valuation of gender are independent, yet seemingly sensitive to one another. The authors discuss this paradox, stating that the observed correlation between number and gender is inconsistent with the independence of the two features. They also suggest that these effects may be the result of a monitoring system (Levelt, 1989) that is more likely to allow an error to go unnoticed if two expressions are more similar to each other in agreement features.

The work of Antón-Méndez et al. is complemented by other experimental work on the interaction of number and gender agreement errors in Spanish. Vigliocco et al. (1996) conducted a similar production study to the one described above, but added a semantic variable for consideration. The goal of their experiment was to determine whether distributivity effects can play a role in number agreement errors. The preambles they presented to the participants of their completion study also had a head noun and a distractor noun nested within a modifier, but the plurality of one of these nouns was designed to have a possible effect on the conceptual plurality of the other. Take, for instance, *the label on the bottles*. While *the label* is grammatically singular, the plural noun in the modifier *on the bottles* may lead speakers to understand *the label* as referring to a plurality of labels (the authors call this the ‘distributivity’ effect). Vigliocco et al. found that semantics, specifically distributivity effects, does play a role in number agreement errors: Spanish speakers were more likely to make an agreement error if they perceived the head noun as possibly semantically plural based on the plurality of the distractor noun.

Alcocer and Phillips (2009) contribute to the discussion of subject-verb agreement in Spanish by making a distinction between agreement errors in sentences in which the agreeing verb is an auxiliary (9) and sentences in which the agreeing verb is a main verb (10). As in both examples below, the agreeing noun and verb were both located in a relative clause, with a non-local distractor noun located outside of the relative clause (‘the note’ in both examples below).

- (9)    la    nota    que    la    chica    va        a    escribir...  
       the note COMP the girl    go.3SG to write.INF  
       ‘the note that the girl is going to write’
- (10)    la    nota    que    la    chica    escribió...  
       the note COMP the girl    write.3SG.PST  
       ‘the note that the girl wrote’

In sentences with noun-auxiliary agreement, (9), they found that speakers were more likely to judge a grammatical sentence to be ungrammatical if the non-

local distractor was plural. In sentences with noun-main verb agreement, (10), they found no such effect. Based on these results, Alcocer and Phillips suggest that there is a difference between noun-auxiliary agreement and noun-main verb agreement. They conducted a similar experiment on English and found opposite results: English speakers were more likely to judge an ungrammatical sentence to be grammatical if the distractor noun was plural. By comparing results across the two languages, Alcocer and Phillips concluded that less marked verb forms are more susceptible to interference from a distractor noun. For instance, the English *see* is unmarked as compared to the third-person singular *sees* and is therefore more susceptible to an agreement error if a plural distractor noun is present. In Spanish, the third-person singular *va* ‘go’ is the less marked form (as compared to the third-person plural *van* ‘go’) and is more susceptible to interference from a plural distractor noun. Alcocer and Phillips thus demonstrate that properties of the agreeing verb may also have an effect on the likelihood of an agreement error in the presence of distractors.

What is missing from this not insubstantial body of experimental work on agreement in number and gender in Spanish is a systematic approach to observing the interaction between number and gender, one that tracks the feature specification while investigating its geometry. Such an approach would recognize that understanding the relationship between these two features first requires an understanding of their respective feature content. The experiment we present next aims to do just that.

## 3 Experiment

### 3.1 Goals

The goal of the experimental work presented below is to determine whether there is evidence for the bundling vs. split model of gender and number as applied to Spanish. To gain this kind of evidence we need to explore agreement violations in each of the features and more specifically to put number and gender in competition with each other. If Spanish resolves this competition differently for gender than for number, then we have evidence in favor of the split model. If gender and number are resolved together, we have evidence for bundling.

By holding number agreement constant and manipulating gender agreement in a context in which there exists a potential for conflict between features, we can observe how the two potential gender values in Spanish interact with

each other and thus gain insight into the feature content of Spanish gender. By holding gender agreement constant but creating conflict in number agreement, we can do the same for the feature content of number. Simultaneous manipulation of number and gender will similarly give insight into the interaction between number and gender and therefore into the relationship between the two features.

While ours is not the first study investigating the relationship between number and gender features, as well as the features themselves, it is the first to investigate these issues simultaneously and within a comprehension setting. Previous studies have either focused on only one of the features that we investigate and how this feature behaves in agreement attraction contexts, or they have investigated the potential differences between noun-verb agreement or noun-predicative adjective agreement. Our study also increases the syntactic distance between the source noun and agreement target in order to tax the memory and thus explore more subtle effects in agreement, which may go unnoticed when the agreeing element and the element bearing phi-features are adjacent.

### 3.2 An excursus on agreement attraction

A context where agreement features are forced to compete with each other is available when they are represented on different goals which are, licitly or illicitly, accessible to the same probe. Agreement attraction is one such context. In cases of agreement attraction, a feature-seeking probe fails to match the grammatical features of the appropriate goal and instead realizes the grammatical features of a distractor, an expression that is usually closer to the probe in linear order. Consider the following example from English:

- (11) The book that was left on the table by the girls have a blue cover.

As a speaker produces this sentence, s/he stores *the book* in memory as the subject of the sentence, looking for a verb with which it must agree in person and number. As the sentence proceeds, the memory is taxed by the introduction of several other DPs (*the table*, *the girls*). The agreement features of the original subject enter into conflict with the other information that the speaker must store in memory, including the features of these interfering DPs, even if they are embedded under a prepositional head. This may result in incorrect agreement where the features of *the book* have been replaced in memory by the features of *the girls*. Agreement attraction is particularly powerful in

single-valued systems; if only one feature value is visible and the other is simply an absence of a feature value, the visible feature should strongly drive attraction (see Bock and Cutting 1992; Franck et al. 2002, 2006 on number attraction). For this reason, it is crucial to understand the content of feature values in Spanish gender and number before we move into the comprehension of attraction errors.

Cases of agreement attraction have been experimentally studied in various languages (English: Bock and Miller, 1991; Bock and Eberhard, 1993; Vigliocco et al., 1996; Vigliocco and Nicol, 1998; Bock et al., 2012; Spanish: Vigliocco et al., 1996; Antón-Méndez, 1999; Antón-Méndez et al., 2002; Alcocer and Phillips, 2009; Italian: Vigliocco et al., 1995; Vigliocco and Franck, 1999; French: Vigliocco et al., 1996; Dutch: Bock et al., 2001; Hartsuiker et al., 2003; German: Hartsuiker et al., 2003; Russian: Lorimor et al., 2008, a.o.), testing whether there is an asymmetry between different values of features in triggering agreement errors. In particular, previous studies looked at errors involving number (Bock and Miller, 1991; Bock and Eberhard, 1993; Vigliocco et al., 1996; Bock et al., 2012; Alcocer and Phillips, 2009; Vigliocco et al., 1995; Hartsuiker et al., 2003; Bock et al., 2001; Vigliocco and Nicol, 1998), gender (Vigliocco and Franck, 1999), or both gender and number (Antón-Méndez, 1999; Antón-Méndez et al., 2002; Lorimor et al., 2008). We can build on this substantial body of work and put the phenomenon of attraction to use in exploring the difference between bundling and split approaches. We make minimal assumptions concerning the actual mechanism of attraction, noting merely the diverging predictions of bundling vs. split approaches, which we turn to next.

### 3.3 Predictions

Based on the discussion of their respective feature content above, we can make certain predictions concerning the feature content of Spanish number and gender. If a given feature value is specified, then when placed into conflict with another feature value, we should see grammaticality effects. That is, speakers will distinguish between sentences in which there are no agreement errors pertaining to the specified feature value, and ones in which there are agreement errors pertaining to the specified feature value – they should rate the former significantly higher than the latter. Consider an example:

- (12) a. El niño considera la noticia en las  
 the boy consider.3SG the news item.F.SG in the  
 revistas terriblemente aburrida.  
 magazines.F.PL terribly boring.F.SG  
 ‘The boy considers the news item in the magazines to be terribly boring.’
- b. \*El niño considera la noticia en las  
 the boy consider.3SG the news item.F.SG in the  
 revistas terriblemente aburridas.  
 magazines.F.PL terribly boring.F.PL

The first of the two sentences is grammatical, but the second has an error in number agreement – *aburridas* ‘boring’ is plural, but should be singular in agreement with *la noticia* ‘the news item’. If singular is a specified feature value in Spanish, then grammatically correct sentences like (12-a) should consistently be rated higher than sentences with an error in agreement in the singular, like the sentence in (12-b).

It follows that if a feature is multi-valued, that is, has two specified values, both of them should exhibit grammaticality effects. If, however, a feature is single-valued, only one of the two potential values will exhibit grammaticality effects. This leads us to make the following predictions:

**Prediction 1:** If number is multi-valued, both singular and plural will show grammaticality effects.

**Prediction 2:** If number is single-valued, either singular or plural (but not both) will show grammaticality effects.

**Prediction 3:** If gender is multi-valued, both masculine and feminine will show grammaticality effects.

**Prediction 4:** If gender is single-valued, either masculine or feminine (but not both) will show grammaticality effects.

Having settled the feature content of number and gender, we can then consider whether the data indicates the bundling or the split model of number and gender feature classes. Recall that if number and gender are bundled, then they ought to be valued simultaneously. This suggests the number and gender features of a noun should determine agreement together, at the same time. When an incorrect noun enters into agreement with an adjective, both its number and gender features should agree with the number and gender fea-

tures of the adjective. Building on our example in (12), consider the following ungrammatical sentences:

- (13) a. \*El niño considera la noticia en los  
           the boy consider.3SG the news.item.F.SG in the  
           periódicos terriblemente aburridos.  
           magazines.M.PL terribly boring.M.PL
- b. \*El niño considera la noticia en las  
           the boy consider.3SG the news.item.F.SG in the  
           revistas terriblemente aburridas.  
           magazines.F.PL terribly boring.F.PL
- Intended: ‘The boy considers the news item in the magazines to be terribly boring.’

Both of the above sentences are grammatically incorrect, but in each of them the local noun has entered into agreement with the adjective, which may lead to an illusion of grammaticality (i.e., attraction). If number and gender are projected and valued together, as in a bundling approach, then when the probe (incorrectly) gets a feature (e.g., number) from the local noun, it should get the other feature (e.g., gender) as well. In other words, agreement attraction in one feature ought to precipitate agreement attraction in the other feature, with the result that both of the above sentences should be rated equally high (or equally low).

If, however, number and gender are split, then they are projected and valued independently, and agreement attraction in number can proceed independently of agreement attraction in gender. This means that, all factors being equal, a violation in gender agreement may be judged higher or lower than a violation in number agreement. Of course, each type of violation may also be judged roughly the same, but crucially, they are evaluated on their own, without a connection to the other type of violations. Furthermore, if the two features are independent of each other, we can expect that a violation in both of them would be more offensive to a comprehender than a violation in just one feature. This expectation is based on the observation that the more grammatical constraints violated, the higher the degree of degradation (cf. Kluender, 2004).<sup>9</sup> Applying that logic, we expect that the violation in (13-a),

---

<sup>9</sup> The issue of cumulative effects of violations is not open-and-shut, and needs to be evaluated on a case by case basis. It is possible that some violations are more “costly” than others, and their hierarchical arrangement may call for an Optimality Theory-type ranking (see Warren and Gibson, 2002, for a discussion of such an approach to processing).



where both the gender and the number of the head noun are mismatched, should be rated lower than (13-b), where only the number is mismatched. Thus, under a split model, (13-a) should receive a lower rating.

To summarize the bundling vs. split predictions:

**Prediction 5:** If number and gender are bundled, agreement attraction in one feature should lead to agreement attraction in the other feature.

**Prediction 6:** If number and gender are split, agreement attraction in one feature need not lead to agreement attraction in the other feature.

We can now turn to the design of the experiment used to test these predictions.

### 3.4 Participants

We recruited 126 participants through Amazon.com's Mechanical Turk crowd-sourcing service. Subjects were compensated for their participation. We restricted IP addresses of subjects to just those within the United States in an attempt to control for language background, specifically familiarity with European vs. Latin American Spanish vocabulary. On the basis of a short demographic questionnaire that preceded the experiment, we determined whether subjects were native speakers. Moreover, the instructions to the experiment were given in Spanish, which allowed us additional control over the subject pool.

### 3.5 Design

Our goal of simultaneously evaluating number and gender agreement errors, with intervening material between the goal and the probe, constrained the possible syntactic frames we could use in the design of our stimuli. We needed a configuration in which a noun intervenes between the source and target of agreement, and in which all of the agreeing elements inflect for both gender and number. Such a configuration is observed in argument small clauses such as the one below:

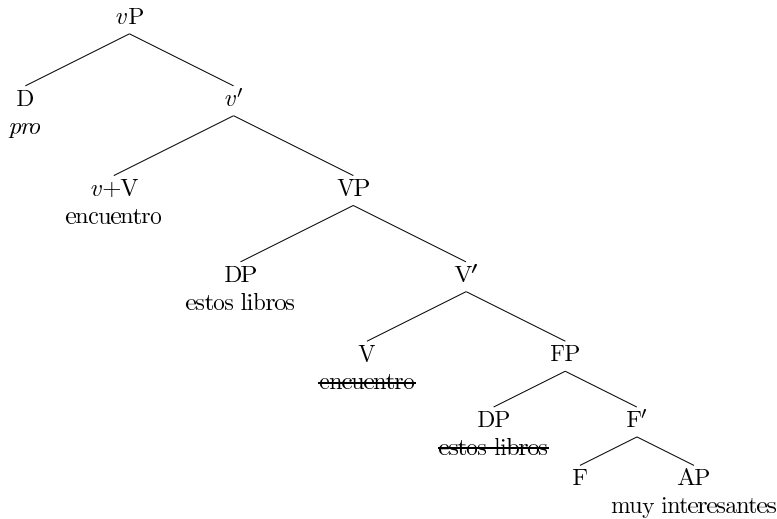
---

However, in comparing phi-feature violations, we can reasonably assume that they are comparable in visibility, which would entail a cumulative effect when both features are mismatched.

- (14) Encuentro estos libros muy interesantes.  
find.PRS.1SG these books.M.PL very interesting.M.PL  
'I find these books very interesting.'

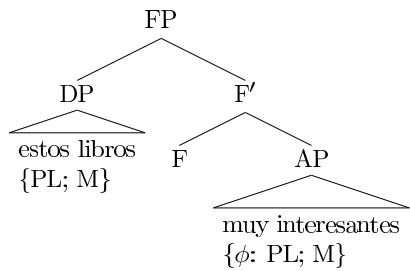
We rely on the existing syntactic analysis of such small clauses (Contreras, 1987; Jiménez-Fernández and Spyropoulos, 2013), and assume that they have the following structure, representing the sentence in (14):

- (15) *Small clause structure:*



Zooming in on the small clause portion of the example in question, we observe number and gender agreement on the predicative adjective, thus:

- (16) *Small clause feature valuation:*



The goal *DP* can be made heavier with the introduction of a prepositional phrase modifying it, thus:

- (17) Encuentro estos libros sobre la guerra muy  
 find.PRS.1SG these books.M.PL on the war.F.SG very  
 interesantes  
 interesting.M.PL  
 ‘I find these books on the war very interesting.’

This is the precisely the configuration which would allow us to investigate agreement errors and attraction by systematically manipulating the features of the two nouns and the adjective; schematically it can be represented as follows:

- (18) (Subject) Verb NP1 Prep NP2 Adv ADJ ...

Only a handful of verbs in Spanish embed small clauses. In addition to *encontrar* ‘find’, our stimuli featured *considerar* ‘consider’ and *ver* ‘see’ as the matrix verb. We made a distinction between matrix predicates which subcategorize for small clauses, and those that can co-occur with non-subcategorized small clauses or secondary predicates (i.e., depictives), such as *dejar* ‘leave’ (Contreras, 1987; Demonte, 1988; Schultze-Berndt and Himmelmann, 2004). In the experiment reported below, we included both subtypes to test if any differences between subcategorized and non-subcategorized small clauses would appear in agreement. In this paper, we will not discuss the comparison between the two types because there were no significant differences between them with respect to agreement.

Stimuli were recorded by an adult male native speaker of Spanish from Colombia. To avoid possible phonetic cues to ungrammaticality, only the grammatical versions of our items were recorded. We then split the sentence recordings at the adverb and spliced each onset ((Subject) Verb NP1 Prep NP2...) with the four possible adjective completions (ADJ.M.SG, ADJ.M.PL, ADJ.F.SG, ADJ.F.PL). All stimuli were spliced together in this manner, even the fully grammatical variants.

Fifteen items were created, with 64 versions of each. Within each item, we manipulated the number (SG vs. PL) and gender (M vs. F) of NP1, NP2, and ADJ; this manipulation yields 64 sentences:  $2_{NP1-NUM} \times 2_{NP1-GEN} \times 2_{NP2-NUM} \times 2_{NP2-GEN} \times 2_{ADJ-NUM} \times 2_{ADJ-GEN} = 64$ .

Given that we are interested in the behavior of grammatical gender, that is, the gender a noun leaves the lexicon specified for, our gender manipulation on nouns required the use of different lexical items for masculine vs. feminine values. Within an item, we matched the meaning of these nouns as closely as possible (as with *el armario* ‘the cupboard’ and *la alacena* ‘the cupboard’

in (19)); a full list of our experimental stimuli can be found online (Scontras et al., 2013).

A crucial aspect of the design of our stimuli concerns the way in which we effect agreement errors: only when NP1 and ADJ agree in both number and gender will the sentence be grammatical; all other versions of an item will involve a mismatch in the features of NP1 and ADJ and therefore be ungrammatical. In (19), we give two grammatical versions of one of our items.

- (19) a. Maria considera **el platillo** en el armario  
 Maria consider.PRS.3SG the plate.M.SG in the cupboard.M.SG  
 elegantemente *decorado* pero prefiere recoger las  
 elegantly decorated.M.SG but prefer.PRS.3SG collect.INF the  
 tetetas.  
 teapots.
- b. Maria considera **la bandeja** en la alacena  
 Maria consider.PRS.3SG the tray.F.SG in the cupboard.F.SG  
 elegantemente *decorada* pero prefiere recoger las  
 elegantly decorated.F.SG but prefer.PRS.3SG collect.INF the  
 tetetas.  
 teapots  
 ‘Mary considers the tray in the cupboard to be elegantly deco-  
 rated, but she prefers collecting teapots.’

We have assumed that NP1 (in **bold**) must agree with ADJ (in *italics*); when the features of the two do not match, ungrammaticality results. Note, however, that correctly identifying grammaticality on this basis within our items requires that there is no parse of the sentence under which ADJ intentionally agrees with and modifies NP2 (underlined) and not NP1. Were NP2 to agree with ADJ, its features (an not those of NP1) would determine grammaticality.

To rule out this unintended parse of our stimuli (under which ADJ modifies NP2), we ran a separate norming study on all of our items. 60 participants who did not take part in the critical study rated the likelihood that NP2 (and not NP1) gets modified by ADJ on a five-point scale. Subjects consistently rated this parse low (average rating: 2 out of 5; for the full set of results see Scontras et al., 2013); on the basis of these low ratings we conclude that the unintended parse is highly unlikely, if not unavailable. We furthermore conclude that our agreement error manipulation is sound: the features of NP1 determine the grammaticality of the sentence.

Subjects took the experiment online using the web-based experiment platform ExperigenRT (Becker and Levine, 2010; Pillot et al., 2012), which allowed

for the measurement of response times to our auditory stimuli. After filling out a short demographic questionnaire, subjects were presented with one version of each of the fifteen items, together with thirty fillers; stimuli were presented in a random order. Subjects provided acceptability ratings for the sentences they heard on a scale from 1 (*completamente inaceptable* ‘completely unacceptable’) to 5 (*completamente aceptable* ‘completely acceptable’).

Only responses from native speakers of Spanish were analyzed. We evaluated nativeness on the basis of two demographic questions: *What was the first language you learned?* and *What is the language you speak most at home?* Subjects who provided Spanish as answers to both questions were classified as native; data from 50 subjects was included in the analysis.

3.6 Results

Given the many factors at play in the design of our experiment, analyzing their contributions all at once would yield uninterpretable results. We therefore split the analyses into five parts: we begin by analyzing number features, looking at their effect on singular and then plural head nouns; we then shift to gender features, looking at masculine and then feminine head nouns. Finally, we compare potential attraction conditions.

SINGULAR HEAD NOUN. To avoid possible effects of gender mismatches on our investigation of number, we hold gender constant across all three of NP1, NP2, and ADJ. That is, either all three are masculine or all three are feminine. Keeping NP1 singular, we have four possibilities for the number features of NP2 and ADJ as in Fig. 1, which reports average ratings for the four resulting conditions.

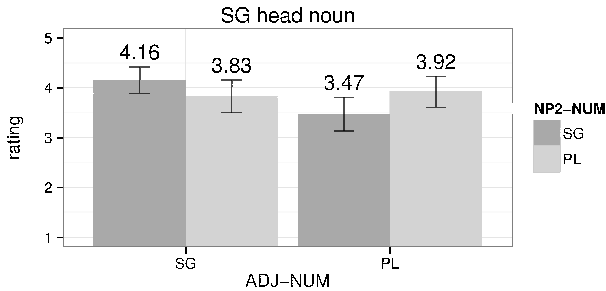


Fig. 1. Average ratings given to singular head nouns.

We fit a mixed logit model predicting response by NP2-NUM and ADJ-NUM, as well as their interaction (for full model details, see Table 1 in the Appendix). Here and in all of the analyses that follow, we trimmed the data on the basis of response time to within two standard deviations of the mean. The model finds a significant effect of ADJ-NUM ( $\beta = -0.63$ ,  $t = -3.04$ ,  $p < 0.01$ ): grammatical sentences in which ADJ agrees with NP1 in the singular are rated higher than ungrammatical sentences with a plural ADJ. We also find a significant interaction between NP2-NUM and ADJ-NUM ( $\beta = 0.74$ ,  $t = 2.61$ ,  $p < 0.01$ ): sentences in which both NP2 and ADJ are plural are rated higher.

PLURAL HEAD NOUN. As in our analysis of singular head nouns, we keep gender features constant across NP1, NP2, and ADJ: all are either masculine or feminine. Average ratings for the four conditions with a plural head noun are given in Fig. 2.

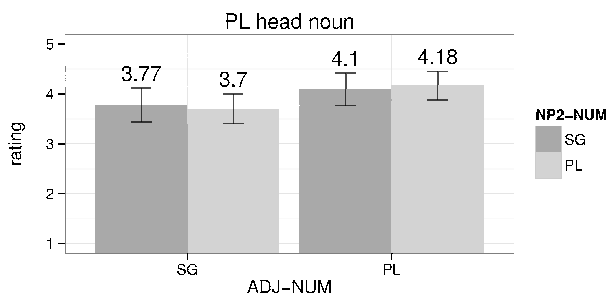


Fig. 2. Average ratings given to plural head nouns.

We fit a mixed logit model predicting response by NP2-NUM and ADJ-NUM, as well as their interaction (for full model details, see Table 2 in the Appendix). Here we find only a significant effect of ADJ-NUM ( $\beta = -0.41$ ,  $t = -1.93$ ,  $p < 0.06$ ): grammatical sentences in which ADJ agrees with NP1 in the plural are rated higher than ungrammatical sentences with a singular ADJ.

MASCULINE HEAD NOUN. Shifting our focus to the effects of gender agreement errors, we begin by analyzing responses to masculine head nouns. To avoid possible effects of number mismatch, now we keep the values for number constant across NP1, NP2, and ADJ: all three are either singular or plural. The

four possible values for NP2 and ADJ gender are given in Fig. 3, in which we report average ratings for masculine head nouns.

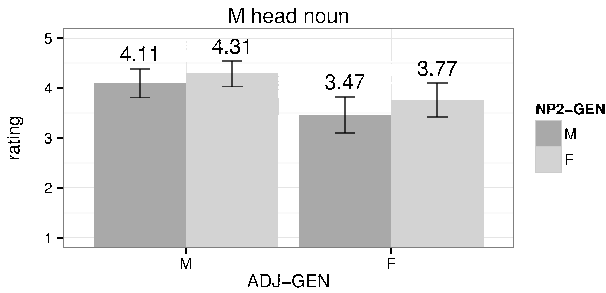


Fig. 3. Average ratings given to masculine head nouns.

We fit a mixed logit model predicting response by NP2-GEN and ADJ-GEN, as well as their interaction (for model details, see Table 3 in the Appendix). We find a only a significant effect of ADJ-GEN ( $\beta = -0.59$ ,  $t = -2.96$ ,  $p < 0.01$ ): grammatical sentences in which ADJ agrees with NP1 in the masculine are rated higher than ungrammatical sentences with a feminine ADJ.

FEMININE HEAD NOUN. In Fig. 4 we present average ratings for the four conditions in which the NP1 is feminine and all three of NP1, NP2, and ADJ match for number.

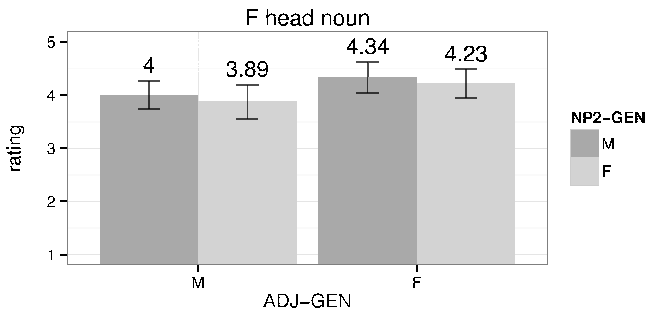
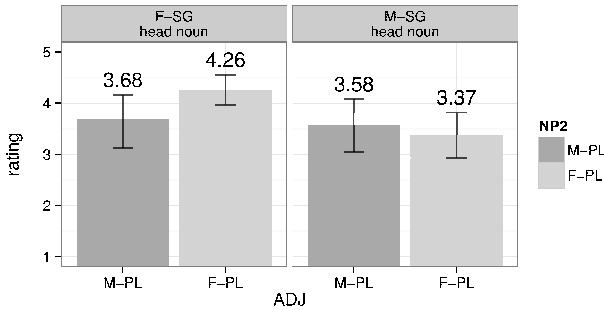


Fig. 4. Average ratings given to feminine head nouns.

We fit a mixed logit model predicting response by NP2-GEN and ADJ-GEN, as well as their interaction (for model details, see Table 4 in the Appendix). None of the effects reached significance.

ATTRACTION COMPARISONS. To evaluate possible additive grammaticality effects, and thus test bundling vs. split approaches to feature geometry, we further consider conditions with either one or two possible agreement errors, as in Fig. 5.



**Fig. 5.** Potential attraction conditions with feminine (left) and masculine (right) singular head nouns.

For feminine head nouns, the sequence in which gender is held constant and only number enters into agreement attraction, F.SG F.PL F.PL, was rated significantly higher than the sequence in which both number and gender enter into agreement attraction: F.SG M.PL M.PL ( $\beta = -0.56$ ,  $t = -1.91$ ,  $p < 0.07$ ; Table 5). A Wilcoxon test found no significant difference in the processing times for these two conditions (F.SG F.PL F.PL reaction time: 1905 ms; F.SG M.PL M.PL reaction time: 2007 ms). For masculine head nouns, the difference between ratings given for single-error attraction conditions (i.e., M.SG M.PL M.PL) and double-error attraction conditions (i.e., M.SG F.PL F.PL) was not significant ( $\beta = -0.20$ ,  $t = -0.74$ ,  $p = 0.45$ ; Table 6). A Wilcoxon test also found no significant difference in the processing times for these two conditions (M.SG M.PL M.PL reaction time: 2014 ms; M.SG F.PL F.PL: 2440 ms.).



## 4 Internal structure of Spanish number and gender

In the initial predictions, we stated that if a feature value is specified, then we should see relevant grammaticality effects. In other words, speakers should be able to distinguish between sentences in which agreement between the head noun and adjective in this feature value is successful, and ones in which there is an agreement error pertaining to the given feature value. It follows that if a feature is multi-valued, both of its values will exhibit grammaticality effects. If, however, the feature is single-valued, only one of the features might exhibit grammaticality effects.

The results presented in section 3 show that both the singular feature value and the plural feature value for number yield grammaticality effects. Errors in number between the head noun and adjective were rightly perceived as incorrect, and the corresponding sentences were rated significantly lower than sentences in which there was no error in number agreement. Recall the predictions made in section 3.3:

**Prediction 1:** If number is multi-valued, both singular and plural will show grammaticality effects.

**Prediction 2:** If number is single-valued, either singular or plural (but not both) will show grammaticality effects.

Further, recall our discussion of the possible feature content of Spanish number in section 2.1, which gave conflicting evidence as to whether the singular or plural is specified in Spanish. As in English, the Spanish plural is morphologically specified with *-s*, whereas the singular is morphologically unspecified. However, the Taghlib Test and coordination facts showed the singular to be semantically specified, and the plural to be semantically unspecified (again just like in English). Based on our results, we conclude that number in Spanish is multi-valued: both singular and plural features are available.

The picture for gender is different. Starting with the gender of the adjective, participants were not sensitive to gender errors when the adjective was masculine. If we now approach the results from the standpoint of the gender of the head noun, participants were sensitive to gender mismatches only with masculine head nouns. Recall the predictions made in section 3.3 (repeated below).

**Prediction 3:** If gender is multi-valued, both masculine and feminine will show grammaticality effects.

**Prediction 4:** If gender is single-valued, either masculine or feminine (but not both) will show grammaticality effects.

In order to determine whether feminine or masculine is the specified value in Spanish gender, we consider how masculine and feminine adjectives enter into agreement with nouns of different gender. In our experiment, sentences in which the adjective inside the small clause was masculine were all rated approximately equally high, regardless of whether the head noun was masculine or feminine. In other words, there was an absence of grammaticality effects with feminine head nouns, as the masculine adjective agreeing with a feminine noun was not perceived ungrammatical. Where feminine adjectives incorrectly agreed with masculine head nouns, the ungrammaticality was readily detected. This result indicates that masculine adjectives appear to be less offending if they do not match the gender of the head noun, thus confirming that the masculine can be evaluated as the absence of a specified gender, not a gender in its own right.

Feminine adjectives, on the other hand, evoked strong grammaticality effects. Conditions in which feminine adjectives agreed with feminine head nouns (F-F-F and F-M-F) were rated significantly higher than ungrammatical conditions in which the feminine adjective was intended to agree with a masculine head noun (M-F-F and M-M-F). This confirms that the feminine gender categorically excludes masculine from its specification.

Given the lack of grammaticality effects with masculine predicates, we conclude that Spanish gender is single-valued, and that the feminine is specified while the masculine is the realization of the absence of gender specification. This result is in full support of Harris's analysis according to which the masculine gender in Spanish is the absence of gender specification.

## 5 Spanish number and gender: Bundling or splitting?

Having settled the feature content of number and gender, we can now consider whether the two features are bundled together or split. In the experiment, we systematically manipulated number and gender in agreement attraction conditions in order to create conflict between features and observe how speakers perceive the resulting sentences.

We can compare ratings assigned to sentences in which agreement attraction occurs in both the number and gender with ratings given to sentences in which attraction occurs with just one feature. These conditions are illustrated in (20-a) and (20-b), respectively.

- (20) a. \*El niño considera la noticia en los  
           the boy consider.3SG the news.item.F.SG in the  
           periódicos terriblemente aburridos.  
           magazines.M.PL terribly boring.M.PL
- b. \*El niño considera la noticia en las  
       the boy consider.3SG the news.item.F.SG in the  
       revistas terriblemente aburridas.  
       magazines.F.PL terribly boring.F.PL  
       Intended: ‘The boy considers the news item in the magazines to  
       be terribly boring.’

Recall the predictions made in section 3.3:

**Prediction 5:** If number and gender are bundled, agreement attraction in one feature should lead to agreement attraction in the other feature.

**Prediction 6:** If number and gender are split, agreement attraction in one feature need not lead to agreement attraction in the other feature.

If number and gender are bundled, and therefore valued, together, both of the above sentences should be rated equally high (perhaps mistakenly so, as in cases of attraction). On the other hand, if number and gender are split, then speakers may perceive sentences in which there are two agreement errors as less grammatical than ones in which there is only one agreement error.

As we saw, speakers rated these two attraction cases differently: sentences with just number entering into attraction were rated higher than sentences with both number and gender entering into attraction (cf. Fig. 5). Likewise, we find a trend such that sentences where agreement attraction was caused by gender features alone were rated higher than those with both number and gender causing attraction effects. This suggests that the violation in two features creates a cumulative effect and leads to degradation in ratings, a finding that is consistent with the results of some recent work on grammatical violations in other linguistic domains. For example, Haegeman et al. (2014) present and analyze cumulative effects in what has been traditionally assumed as a homogenous phenomenon of subject island violations. It remains to be

seen what other phenomena could be accounted for in terms of cumulative violation effects.

The presence of cumulative errors (or lack of bundled attraction) may not be sufficient to discount the bundling model. To give this model yet another chance, let us consider the following metaphor. A florist is putting together a set of arrangements, where all the roses are red. She accidentally leaves a couple of white roses in one of the arrangements, and the person inspecting the arrangements notices that. The flowers are still all bundled together, but that does not prevent noticing a violation in the color scheme. One could take our gender and number results and interpret them the same way; every time there is a violation, the comprehenders rate the sentences lower, yet that does not mean they access number and gender separately. On this account, however, we would expect that violations in multi-valued features would be rated the same; to continue with our metaphor, it does not matter if our florist forgot to remove a white rose or a pink rose from the bundle, all that matters is that the red color scheme is disrupted. Recall that we determined that number is multi-valued, so it provides us with a useful test case. If all that matters is the presence or absence of a violation, regardless of the color or flavor of that violation, we would expect mismatches in number agreement to be rated the same whether it is a mismatch between the singular probe and the plural goal, or the other way around. However, the results point to a different picture; there is at least a trending difference in ratings of the singular goal and plural probe (3.77, see Fig. 2), and the plural goal and singular probe (3.92, see Fig. 1). This indicates that our flower arrangement model, which would allow us to keep all the features still bundled together, is unlikely.

The end result, then, is that the asymmetry in the ratings of agreement mismatches suggests that number and gender are valued separately; were they valued together, we should have found no difference between the conditions in which only one feature determined attraction effects and the conditions where both features caused attraction.

## 6 Conclusions and outstanding questions

The goal of this paper was to test the hierarchical organization of number and gender in Spanish. Both primary data and the experimental work presented here indicate that Spanish number is represented by a binary (multi-valued) opposition, with singular and plural both specified and equally visible in the feature specification space. With respect to gender, we find that only the

feminine is specified, and masculine is simply the absence of gender. This result, supported by our experimental investigation, is consistent with earlier experimental work on Spanish (Vigliocco et al., 1996; Antón-Méndez, 1999; Antón-Méndez et al., 2002; Alcocer and Phillips, 2009) and Italian (Carminati, 2005) and lends novel support to the conception from Harris (1991) that Spanish gender is organized as a privative opposition.

Based on the established feature specifications of number and gender, we have used an attraction experiment to determine whether number and gender belong to the same category space in the phi-feature hierarchy, or whether one feature dominates the other. These two analytical possibilities correspond to the bundling model of gender and number on the one hand, and the split model on the other. Our results show that Spanish number and gender reveal different patterns with respect to agreement attraction, which suggests that they are not equal in the phi-feature hierarchy. On the assumption that agreement in an XP occurs with the highest phi-feature (Matushansky, 2013; Preminger, 2014), we see stronger effects with number agreement, which indicates that number outranks gender on the phi-feature hierarchy. The two features are accessed differently, and the violations in gender are treated differently than violations in number. Furthermore, the presence of violations in both number and gender results in a cumulative effect; comprehenders assess such violations as significantly more offensive than violations in one feature. All factors being equal, this result constitutes an argument in favor of the split model.

Throughout this paper, we presented the two models as mutually exclusive, and we have attempted to show that the split model is superior in handling the agreement facts in Spanish. However, its success in Spanish does not yet mean that the bundling model is not applicable outside of Spanish. It may well be that both models have citizenship status in linguistic theory, and that they may apply in different languages, as a reflection of parametric variation. Before closing, we would like to offer some preliminary considerations concerning such parametric variation and the corresponding division of labor between the models.

In Spanish, as we saw, number and gender are constructed and accessed independently in the morphology, which is consistent with the split model. However, it is also possible to imagine a language where the access to all gender features or a subset of those features is via number; in such a language, it is conceivable that gender and number would be bundled together. Romanian may be an example of such a language (Bateman and Polinsky, 2010). In the singular, Romanian distinguishes two noun classes, feminine and the rest (traditional accounts posit three genders, feminine, neuter, and masculine, but neuter and masculine are indistinguishable in terms of their endings

and determine the same agreement). In the plural, there are also two classes. However, short of the small, semantically determined number of items whose natural gender is reflected in their grammatical gender, the two-way gender distinction in the plural is predicted from the form of the plural, not from the form of the singular or from independent gender information (Bateman and Polinsky, 2010, pp. 63–64). A Romanian speaker needs the rules of plural formation independent of gender, and can then use the information on the plural to predict agreement class, that is gender. Given that gender appears to get accessed via number, it is conceivable that such a system would be best captured by the bundling model, although it remains to be seen if such a model can be motivated for the singular as well.

The following conclusions follow if one agrees with the assumption that the grammar and parser are isomorphic, and that therefore observing the parser allows us to observe the grammar, as in Phillips (2013) and as assumed by this paper. For those who choose to not adopt this assumption, the challenge is to explain behavioral facts on the basis of distributional evidence – such research would need to be able to predict the results discussed in this paper based on factors such as frequency or salience of certain surface properties. At first blush, this seems plausible for number. Because plurality is expressed through *-s* whereas the singular has no phoneme associated with it, the plural is indeed more salient. This might explain why ungrammatical configurations such as SG-PL-PL were considered grammatical by the participants of our study. The salient *-s* plural marker on the distractor noun stands out to the listener, whereas the singular noun has no such salient property. This may make the listener more likely to focus on the match in plurality and therefore to rate the condition as grammatical. The challenge in this kind of approach comes from gender. Whereas the singular and plural are distinguished by the presence or absence of some salient surface property, the same cannot be said for masculine and feminine gender in Spanish. Recall that the experimental design of the study required all masculine nouns to end in the masculine marker *-o* and all feminine nouns to end in the feminine marker *-a*. The results of the study suggest that masculine and feminine behave differently with respect to grammaticality: Speakers notice grammaticality errors in conditions with masculine head nouns, but not in ones with feminine head nouns. Research that follows the assumption that the grammar and parser are not isomorphic would have to explain such a difference based on salience or other surface factors. It may prove difficult to show a significant difference in salience between the masculine *-o* and the feminine *-a*. The challenges associated with this approach suggest that assuming isomorphism between the grammar and

the parser may be a more successful research approach to the questions of number and gender specification.

A final note: We considered only the gender on inanimate nouns. Our results therefore point to the structural representation of categories within a noun phrase in (21), with gender arguably represented on the nominal root or on  $n$ . Such a representation effectively reflects the observation that nouns are lexically specified for gender in the lexicon, either at the root level, as shown in (21-a), or as property of  $n$ , as shown in (21-b). The data that we have been using do not allow us to distinguish between the two models, and we remain agnostic as to which of them is correct.

- (21) a.  $[DP \dots [NumP \dots [NP \dots [nP \dots [\sqrt{P}_{\{Gender\}}]]]]]$   
 b.  $[DP \dots [NumP \dots [NP \dots [nP \{Gender\} [\sqrt{P}]]]]]$

If we now add animate and human nouns, whose gender is often specified, evidence emerges in support of what has been known as the Distributed Gender Hypothesis (Kramer, 2013; Steriopolo and Wiltschko, 2008): there are at least two gender features, a feature for natural gender, projected at the periphery of DP, and grammatical gender, projected below the N-level (either as a property of  $n$  or as property of roots). An obvious extension of the study presented here is to test whether natural gender, which is arguably projected at the DP periphery, causes stronger attraction effects as compared to (a) lexical gender and (b) number (see Antón-Méndez et al. 2002 for a discussion of natural vs. lexical gender in relation to number). It is not inconceivable that the projection of natural gender could dominate the projection of number; if so, we would expect to see reflexes of that arrangement in an experimental setting.

**Acknowledgement:** Work on this project was supported in part by grants from the Center for Advanced Study of Language (CASL) at the University of Maryland, from the Observatory of the Spanish Language in the U.S. at Harvard University, and from the NSF (SMA-1429961) to Maria Polinsky. Any opinion, findings, conclusions or recommendations expressed in this material are those of the authors. We would like to thank Giuli Dussias, Ruth Kramer, Bradley Larson, Nicholas Longenbaugh, Nina Radkevich, Rodrigo Ranero, Allison Ray, Andrés Sanín, an anonymous reviewer, and the audiences at GLOW 37 and The Polinsky Language Sciences Lab for helpful comments on this project.

## Appendix: Full regression models

Tables 1–6 present model coefficients for the full mixed effects linear regression models used in the analyses of sentence ratings reported in Section 3. All models included the maximal random effects structure supported by the data.

**Table 1.** Full logistic regression model from the singular head noun analysis (cf. Fig. 1)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	4.10	0.18	22.32	<0.001
NP2-NUM	-0.29	0.25	-1.16	0.25
ADJ-NUM	-0.63	0.21	-3.04	<0.01
NP2:ADJ	0.74	0.28	2.61	<0.01

**Table 2.** Full logistic regression model from the plural head noun analysis (cf. Fig. 2)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	4.17	0.17	20.53	<0.001
NP2-NUM	-0.09	0.22	-0.39	0.70
ADJ-NUM	-0.41	0.22	-1.93	<0.06
NP2:ADJ	0.07	0.30	0.22	0.82

**Table 3.** Full logistic regression model from the masculine head noun analysis (cf. Fig. 3)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	4.12	0.18	22.49	<0.001
NP2-GEN	0.24	0.20	1.24	0.22
ADJ-GEN	-0.59	0.20	-2.96	<0.01
NP2:ADJ	-0.07	0.28	-0.26	0.79



**Table 4.** Full logistic regression model from the feminine head noun analysis (cf. Fig. 4)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	4.22	0.19	23.81	<0.001
NP2-GEN	0.13	0.17	0.80	0.43
ADJ-GEN	-0.29	0.21	-1.38	0.17
NP2:ADJ	-0.05	0.23	-0.21	0.84

**Table 5.** Full logistic regression model from the feminine head noun potential attraction condition analysis (cf. Fig. 5, left)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	4.27	0.24	17.99	<0.001
GEN	-0.56	0.29	-1.91	<0.07

**Table 6.** Full logistic regression model from the masculine head noun potential attraction condition analysis (cf. Fig. 5, right)

	Coef $\beta$	SE( $\beta$ )	t	p
Intercept	3.60	0.25	14.16	<0.001
GEN	-0.20	0.28	-0.74	0.45

References

Alarcón, I. (2006). *The Second Language Acquisition of Spanish Gender Agreement: The Effects of Linguistic Variables on Accuracy*. Munich: Lincom Europa.

Alcocer, P. and C. Phillips (2009). A cross-language reversal in illusory agreement licensing. Poster presented at the 22nd annual CUNY conference on human sentence processing, University of California, Davis.

Antón-Méndez, I., J. L. Nicol, and M. F. Garrett (2002). The relation between gender and number agreement processing. *Syntax* 5, 1–25.

Antón-Méndez, M. I. (1999). *Gender and Number: Agreement Processing in Spanish*. Ph. D. thesis, University of Arizona.

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.

Bateman, N. and M. Polinsky (2010). Romanian as a two-gender language. In D. Gerdts, J. Moore, and M. Polinsky (Eds.), *Hypothesis A/Hypothesis B*, pp. 41–77. Cambridge, MA: MIT Press.

Becker, M. and J. Levine (2010). Experigen: An online experiment platform. Available at <https://github.com/tlozoot/experigen>.

- Bobaljik, J. and C. Zocca (2011). Gender markedness: The anatomy of a counter-example. *Morphology* 21, 141–166.
- Bock, K., M. Carreiras, and E. Meseguer (2012). Number meaning and number agreement in English and Spanish. *Journal of Memory and Language* 66, 17–37.
- Bock, K. and J. C. Cutting (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language* 31, 99–127.
- Bock, K. and K. Eberhard (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes* 8, 57–99.
- Bock, K., K. Eberhard, J. C. Cutting, A. Meyer, and H. Schriefers (2001). Some attractions of verb agreement. *Cognitive Psychology* 43, 83–128.
- Bock, K. and C. A. Miller (1991). Broken agreement. *Cognitive Psychology* 23, 45–93.
- Bull, W. (1965). *Spanish for Teachers: Applied Linguistics*. New York: Ronald Press.
- Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy (person > number > gender) and implications for linguistic theory. *Lingua* 115, 259–285.
- Carstens, V. (2000). Concord in minimalist theory. *Linguistic Inquiry* 31, 319–355.
- Carstens, V. (2003). Rethinking complementizer agreement: Agree with a case-checked goal. *Linguistic Inquiry* 34, 393–412.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Contreras, H. (1987). Small clauses in Spanish and English. *Natural Language & Linguistic Theory* 5, 225–243.
- Demonte, V. (1988). Remarks on secondary predicates: C-command, extraction and reanalysis. *The Linguistic Review* 6, 1–39.
- Despiante, M. and P. Masullo (2001). Género y número en la elipsis nominal: Consecuencias para la hipótesis lexicalista. Paper presented at the *I Encuentro de Gramática Generativa*, Universidad Nacional del Comahue.
- Domínguez, A., F. Cuetos, and J. Segui (1999). The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistic Research* 28, 485–498.
- Franck, J., G. Lassi, U. H. Frauenfelder, and L. Rizzi (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition* 101, 173–216.
- Franck, J., G. Vigliocco, and J. Nicol (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes* 17, 371–404.
- Greenberg, J. (1966). *Language Universals*. The Hague: Mouton.
- Haegeman, L., Ángel. L. Jiménez-Fernández, and A. Radford (2014). Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31, 73–150.
- Harley, H. and E. Ritter (2002). Person and number in pronouns: A feature-geometric analysis. *Language* 78, 482–526.
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry* 22, 27–62.
- Hartsuiker, R., H. Schriefers, K. Bock, and G. Kikstra (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory and Cognition* 31, 1316–1326.
- Jiménez-Fernández, Ángel. and V. Spyropoulos (2013). Feature inheritance, vP phrases and the information structure of small clauses. *Studia Linguistica* 67, 185–224.
- Kluender, R. (2004). Are subject islands subject to a processing constraint? In V. Chand, A. Kelleher, A. J. Rodriguez, and B. Schmeiser (Eds.), *West Coast Conference on Formal Linguistics (WCCFL)* 23, Somerville, MA, pp. 101–125. Cascadia Press.

- Kornfeld, L. and A. Saab (2002). Nominal ellipsis and morphological structure in Spanish. In R. Bok-Benneman (Ed.), *Romance Languages and Linguistic Theory 2002: Selected Papers from Going Romance*, pp. 183–199. John Benjamins.
- Kramer, R. (2009). *Definite Markers, Phi-features, and Agreement: A Morphosyntactic Investigation of the Amharic DP*. Ph. D. thesis, University of California, Santa Cruz.
- Kramer, R. (2013). Gender in Amharic: A morphosyntactic approach to natural and grammatical gender. *Language Sciences* 43, 102–115.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Lorimor, H., K. Bock, E. Zalkind, A. Sheyman, and R. Beard (2008). Agreement and attraction in Russian. *Language and Cognitive Processes* 23, 769–799.
- Matushansky, O. (2013). Gender confusion. In L.-S. Cheng and N. Corver (Eds.), *Diagnosing Syntax*, pp. 271–294. Oxford: Oxford University Press.
- Merchant, J. (2014). Gender mismatches under nominal ellipsis. *Lingua* 151, 9–32.
- Pesetsky, D. and E. Torrego (2007). The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian, and W. K. Wilkins (Eds.), *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation*, pp. 262–294. Amsterdam: John Benjamins.
- Phillips, C. (2013). Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28, 156–187.
- Picallo, C. (1991). Nominals and nominalization in Catalan. *Probus* 3, 279–316.
- Pillot, C., G. Scontras, and L. Clemens (2012). ExperigenRT: Measure reaction times in web-based auditory experiments. Available online at <http://pollab.fas.harvard.edu/?q=node/212>.
- Preminger, O. (2014). *Agreement and Its Failures*. Cambridge, MA: MIT Press.
- Ritter, E. (1993). Where's gender? *Linguistic Inquiry* 24, 795–803.
- Roca, I. M. (1989). The organisation of grammatical gender. *Transactions of the Philological Society* 87, 1–32.
- Saab, A. (2008). *Hacia una teoría de la identidad parcial en la elipsis*. Ph. D. thesis, University of Buenos Aires.
- Saab, A. (2010). (im)possible deletions in the Spanish DP. *Iberia* 2, 45–83.
- Sauerland, U. (2003). A new semantics for number. In R. B. Young and Y. Zhou (Eds.), *Semantics and Linguistic Theory (SALT) 13*, Ithaca, NY, pp. 258–275. CLC Publications.
- Sauerland, U., J. Anderssen, and K. Yatsushiro (2005). The plural is semantically unmarked. In S. Kesper and M. Reis (Eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Berlin: Mouton de Gruyter.
- Schultze-Berndt, E. and N. P. Himmelfmann (2004). Depictive secondary predicates in crosslinguistic perspective. *Linguistic Typology* 8, 59–131.
- Scontras, G. (2013a). Accounting for counting: A unified semantics for measure terms and classifiers. In T. Snider (Ed.), *Proceedings of SALT 23*, pp. 549–569.
- Scontras, G. (2013b). A unified semantics for number marking, numerals, and nominal structure. In E. Chemla, V. Homer, and G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung 17*, pp. 545–562.
- Scontras, G., M. Polinsky, and Z. Fuchs (2013). Agreement mismatches in Spanish. <http://hdl.handle.net/1902.1/21954>; Harvard Dataverse Network V3.

- Spector, B. (2007). Aspects of the pragmatics of plural morphology: On higher-order implicatures. In U. Sauerland and P. Stateva (Eds.), *Presuppositions and Implicatures in Compositional Semantics*, pp. 243–281. Palgrave-Macmillan.
- Steriopolo, O. and M. Wiltschko (2008). Distributed GENDER hypothesis. Paper presented at FDSL 7.5, Independent University of Moscow.
- Tiogang, I. N. (2015). Las cuestiones del género y del número de los neologismos léxicos en el español de Guinea Ecuatorial. *Tonos Digital* 28.
- Vigliocco, G., B. Butterworth, and M. F. Garrett (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition* 61, 261–298.
- Vigliocco, G., B. Butterworth, and C. Semenza (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language* 34, 186–215.
- Vigliocco, G. and J. Franck (1999). When sex and syntax go hand in hand: Gender agreement and language production. *Journal of Memory and Language* 40, 455–478.
- Vigliocco, G. and J. Nicol (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition* 68, B13–B29.
- Warren, T. and E. Gibson (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112.