



## Romanian as a Two-Gender Language.

The Harvard community has made this article openly available. [Please share](#) how this access benefits you. Your story matters

Citation	Bateman, N., and M. Polinsky. 2010. Romanian as a two-gender language. In Hypothesis A/Hypothesis B, ed. D. Gerdt, J. Moore, and M. Polinsky, 41-77. Cambridge, MA: MIT Press.
Citable link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:37108753">http://nrs.harvard.edu/urn-3:HUL.InstRepos:37108753</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a>

## ROMANIAN AS A TWO-GENDER LANGUAGE

Nicoleta Bateman, UC San Diego

Maria Polinsky, UC San Diego and Harvard University

### 1. Introduction<sup>1</sup>

The goal of this paper is to argue that Romanian has two genders, rather than three as traditionally proposed, and in doing so to provide a comprehensive synchronic account of gender assignment in Romanian. The main argument is that gender categories can be predicted in Romanian based on semantic and formal features, and therefore nominal classes need not be specified in the lexicon. Rather, within each number there is a binary distinction of gender classes which, once determined, lead to straightforward categorization of nouns.

Following Charles Hockett, “genders are classes of nouns [systematically] reflected in the behavior of associated words” (Hockett 1958:231)<sup>2</sup>. This “behavior” is manifested in *agreement*, which we define as co-variation between the form of the trigger (noun) and the form of the target (such as adjectives and articles). Thus, particular noun forms will co-occur with particular attributive and predicate adjective forms in the singular and in the plural.

Gender categorization and assignment is a fascinating phenomenon which brings together morphology, phonology, syntax, and simple semantic structures,

so understanding categorization in a particular language offers us a glimpse into several levels of linguistic representation. Gender assignment provides a window into lexical access (which is one of the primary motivations for categorization— cf. Levelt 1989, 1993) and morphosyntactic integration, where the knowledge of a relevant gender contributes to reference identification and tracking. Romanian is particularly intriguing because of its complicated gender system which stands out among the systems of the other Romance languages. Be it the result of the conservative preservation of the Latin three-gender system or the innovation of a third gender under heavy Slavic influence, Romanian is often cited as the unique three-gender language of the Romance group.

This paper investigates this uniqueness further and brings Romanian more in line with the other, more mundane two-gender languages of its group. Specifically, we propose that Romanian has two noun classes (genders) in the singular and in the plural, but the actual division of nouns into classes in the singular is different from their division into classes in the plural. This lack of class isomorphism between the singular and the plural is the main reason why many researchers have analyzed Romanian as a three-gender system. Once we can get past the assumption that such an isomorphism is necessary, the two-gender composition of Romanian becomes much more apparent. As in many other Indo-European languages, Romance languages in particular, gender assignment is determined semantically for a small subset of nouns and by formal properties of

the nouns themselves, namely noun endings, for the majority of the nominal lexicon. Since our analysis is synchronic in nature and addresses the current state of Romanian, we will not offer any new insights into the preservation of the Latin gender system or the role of the Slavic superstrate (beyond a short discussion of the existing analyses). These issues are beyond the scope of this paper and must be addressed independently.

The paper is organized as follows. In section 2 we introduce the relevant data which lead to the main questions concerning the analysis of Romanian gender addressed in this paper. In section 3 we present and analyze the principal existing analyses of Romanian gender. While we disagree with these analyses, each offers important insights, and our own proposal builds on those insights. Section 4 outlines our proposal for analyzing Romanian as a two-gender system, showing that such a system can account for the Romanian patterns in a more straightforward manner. Section 5 provides an evaluation metric comparing our analysis with the other analyses of Romanian gender, demonstrating that our proposal fares better on virtually all criteria. We provide conclusions and identify areas for further research in section 6.

## 2. The problem

### 2.1 Data

Traditional analyses of Romanian recognize three genders: masculine, feminine, and neuter (Graur 1966; Mallinson 1986; Rosetti 1965, 1973; Corbett 1991; Chitoran 1992, 2002, among others). Gender is expressed through agreement on attributive adjectives, predicate adjectives, demonstratives, articles and other determiners, and the numerals ‘one’ and ‘two’. For the sake of simplicity, we use only adjectives to illustrate the agreement patterns. As shown in Table 1, there is significant syncretism in agreement: masculine and neuter nouns take identical agreeing forms in the singular, and feminine and neuter nouns take identical agreeing forms in the plural.

**Table 1.** Noun-Adjective Agreement in Romanian<sup>3</sup>

	SINGULAR		PLURAL	
<b>M</b>	trandafir rose. <b>M</b>	<u>frumos</u> beautiful. <b>M</b>	trandafiri rose. <b>M</b>	frumoși beautiful. <b>M</b>
<b>N</b>	palton coat. <b>N</b>	<u>frumos</u> beautiful. <b>M</b>	paltoane coat. <b>N</b>	<u>frumoase</u> beautiful. <b>F</b>
<b>F</b>	casă house. <b>F</b>	frumoasă beautiful. <b>F</b>	case house. <b>F</b>	<u>frumoase</u> beautiful. <b>F</b>

This syncretism in agreement is matched by syncretism in the number paradigm: masculine and neuter nouns are indistinguishable in the singular, illustrated in (1) as the neutralization of the masculine/neuter distinction, while neuter and feminine nouns are indistinguishable in the plural, shown in Table 2. These patterns are consistent across cases, with masculine and neuter nouns taking the same case markers in the singular, and neuter and feminine nouns taking the same case markers in the plural (see Appendix A for examples).

(1) Neutralization of masculine/neuter distinction (nominative singular definite)

a.	sabogul	[sabot- <b>ul</b> ]	‘the clog’ (shoe)	}	MASCULINE
b.	cartoful	[kartof- <b>ul</b> ]	‘the potato’		
c.	blocuk	[blok- <b>ul</b> ]	‘the block’	}	NEUTER
d.	norocul	[norok- <b>ul</b> ]	‘the (good) luck’		
e.	gardul	[gard- <b>ul</b> ]	‘the fence’		

**Table 2.** Syncretism of plural forms

	Singular	Plural		Gloss
Masculine	copac	copac- <b>i</b>	[kopatʃ <sup>j</sup> ]	tree
	sabot	sabot- <b>i</b>	[sabots <sup>j</sup> ]	clog (shoe)
	colac	colac- <b>i</b>	[kolatʃ <sup>j</sup> ]	bread roll
	codru	codr- <b>i</b>	[kodri]	field
Neuter	teatru	teatr- <b>e</b>	[teatre]	theater
	clopot	clopot- <b>e</b>	[klopote]	bell
	dulap	dulap- <b>uri</b>	[dulapur <sup>j</sup> ]	cabinet
	acvariu	acvari- <b>i</b>	[akvari <sup>j</sup> ]	aquarium
Feminine	kasă	kas- <b>e</b>	[kase]	house
	para	pará- <b>le</b>	[parále]	money
	blană	blăn- <b>uri</b>	[blənur <sup>j</sup> ]	fur
	inimă	inim- <b>i</b>	[inim <sup>j</sup> ]	heart

As Table 2 shows, there are three main plural markers, *-i*, *-e* and *-uri*,<sup>4</sup> with an additional marker *-le* which occurs on a small set of feminine nouns that end in a stressed *-á* or *-eá* (*cafea* ~ *cafele* ‘coffee’, *basma* ~ *basmale* ‘(head)scarf’, *stea* ~ *stele* ‘star’). The *-le* marker is fully predictable, and we will

therefore focus on the three main plural markers, as these are shared among the genders. *-e* and *-uri* do not appear on masculine nouns, but *-i* appears on nouns from all three genders. We will show, however, that there are actually two separate *-i* markers, one that marks traditional masculine nouns, and another which marks traditional feminine and neuter nouns.

Given these facts, the challenge of Romanian gender can be articulated as follows: Romanian seems to have *three* genders in the lexicon: masculine, feminine, and neuter. However, there are only *two* agreement patterns in the singular and the plural: masculine and feminine. Neuter nouns do not have their own dedicated marking, and they do not have their own agreement pattern. The mapping from singular to plural is not one-to-one: neuter nouns follow the masculine pattern in the singular and the feminine in the plural. Given that gender is expressed through agreement, this begs the question of whether Romanian has three genders, or just two.

## **2.2 Gender assignment: two or three genders?**

As is the case with gender in most languages, Romanian gender has a semantic core that accounts for the assignment of animate nouns to the masculine and feminine genders based on natural gender (Graur 1966; Mallinson 1986). Inanimate nouns however, are distributed among all three genders. While there is little disagreement with respect to the semantic basis for gender assignment of



animate nouns in Romanian, the factors determining gender assignment for inanimate nouns remain unclear.

The syncretism in the agreement and number paradigm described in section 1 is a vexing problem that many linguists and grammarians have grappled with since the 18<sup>th</sup> century. Syncretism is not uncommon in the world's languages (for instance, Corbett 1991 describes several languages which show a mismatch between the number of *controller* (lexical) genders and *target* (grammatical) genders). However, the syncretism exhibited in Romanian is of a different nature: it applies to an entire class, not a subset within a particular gender. Within a typical syncretism, which covers just a subset within a class (see Corbett and Fraser 1993, Baerman 2004, Stump 2001, among many others),<sup>5</sup> a learner has independent evidence that the relevant class stands on its own. In the case of full class syncretism no such evidence is available to a language learner and it is impossible to identify the criteria that separate class X from the (syncretic) class Y.

There are at least two possible approaches to the Romanian facts outlined above:

- i. Three-gender system (traditional analysis):<sup>6</sup> Romanian nouns are lexically classified into three genders (or on a modification of such analysis, three *controller* genders) that map onto two agreement

patterns (*target* genders). One agreement pattern is used in the singular, and the other in the plural.

- ii. Two-gender system (as proposed here): Romanian nouns are not lexically specified for gender. There are two genders and two agreement patterns in the singular and the plural. Class membership is determined by formal cues and a small semantic core. Agreement is straightforward once class membership is determined, by mapping noun class directly to a set of agreeing forms.

Our proposal relies on formal features of Romanian nouns, specifically the singular and plural noun endings. Recall that we define agreement as co-variation between the form of the noun (trigger) and that of the adjective (target). We capitalize on the fact that particular noun forms co-occur with particular adjective forms, thus gender specification on the noun is not necessary. In the singular these are the endings of the nominative indefinite form, and in the plural these are the plural markers *-i*, *-e*, and *-uri*. This hypothesis maintains a close relationship between the rules of plural formation and those of gender assignment and agreement. This is a welcome result given that speakers must know how to form the plural regardless of the division of the nominal lexicon into genders. Thus the same factors that are relevant for plural formation are indirectly relevant for predicting gender assignment and agreement in the plural. As a result, this analysis is more parsimonious than a three-gender analysis which does not

capitalize on the forms of the plural to determine agreement. A three-gender analysis has a more complex nominal lexicon, and needs to be supplemented with complex gender mapping rules between the singular and the plural.

In the next section we review three prior three-gender analyses of Romanian, and two prior two-gender analyses. The first two-gender proposal is by Hall (1965), who argues that “neuter” nouns are not a separate grammatical gender, but rather belong to different inflectional classes in the singular and the plural. His arguments, although not fully developed, lend themselves to the same type of analysis as the one proposed in this paper. The second proposal is a development of Farkas’s (1990) analysis into a two-gender account. We show that although it captures the attested agreement patterns, it raises several learnability questions and is in some ways very similar to three-gender analyses.

### **3. Previous analyses of Romanian genders**

#### **3.1 Three-gender analyses**

In this section we present the main three-gender analyses of Romanian. The treatment of Romanian as a three-gender language is motivated on the one hand by historical considerations, and on the other by the need to establish a systematic mapping between genders and declensional classes (Graur 1966), or between *controller* and *target* genders (Corbett 1991).

### 3.1.1 Origins of the Romanian neuter

Much has been written about the source of the Romanian “neuter”. The two main possibilities are that the Romanian three-gender system was inherited from Latin or was reintroduced after the loss of the Latin neuter gender. The resolution of this debate is beyond the scope of this paper, but we will provide synopses of the opposing views below.

Although Romanian developed from Latin, some scholars have disputed the idea that the neuter gender in Romanian has continued naturally from Latin. Since the Classical Latin neuter class became smaller in Vulgar Latin, which eventually gave rise to the two-gender modern Romance languages (French, Italian, Spanish, etc.), these scholars believe that the Romanian “neuter” does not simply continue from Latin (Mallinson 1986:246). Rather, a “reinvention” or “rebirth” is proposed for this “gender”, either due to a desire to express a distinction between “animate” and “inanimate” (Rosetti 1965:84-88), or due to contact with the South Slavic three-gender superstrate (Rosetti 1965:88, Petrucci 1993). According to Rosetti, the desire to distinguish animacy from inanimacy acted as a force that drew from the resources already available in the language (i.e. masculine singular and feminine plural endings) to create the neuter gender. Thus, under this account, the neuter was created to be *the gender for inanimates*; although not all inanimates were drawn into the neuter, there are *no animates* which are neuter<sup>7</sup> (Ibid.).

Other scholars have proposed that the Romanian gender system is continued directly from the Latin system, possibly due to contact with the three-gendered Slavic languages (Petrucci 1993:174). Petrucci refutes the reinvention of the neuter based on Slavic influence, and finds no evidence indicating that contact with the Slavic languages affected the development of the Romanian gender system. Since there is no evidence of “ambigeneric” nouns--nouns which exhibit masculine morphology in the singular and feminine morphology in the plural-- in the history of Slavic, Petrucci argues that the Romanian “neuter” could not have been borrowed from Slavic. Due to evidence of ambigeneric nouns in other Romance languages, including Italian, Dalmatian, early French and Provençal, Petrucci claims that this is a Romance-internal phenomenon (Petrucci 1993:175-6). Further, South Slavic neuter borrowings into Romanian are mostly treated as feminines. For example, the South Slavic neuter nouns [tʃudo] ‘miracle’ and [sito] ‘sieve’ were borrowed as *ciudă* [tʃudə] ‘envy’ and *sită* [sitə] ‘sieve’ respectively, both of which are feminine (Petrucci 1993:179). The facts described here strongly suggest that if there is a Romanian “neuter”, it is not a result of Slavic influence.

### 3.1.2 Three-gender analyses of Romanian

Most synchronic analyses of Romanian have relied on semantics to distinguish three genders: *masculine* for animate nouns denoting males, *feminine* for animate nouns denoting females, and *neuter* for inanimate nouns. This traditional view is that found in the Academy Grammar (Graur 1966). Corbett's (1991) and Farkas's (1990) analyses of Romanian gender each represent a step forward in the understanding of the Romanian system. Each of these analyses is discussed below.

Graur (1966) states that in Romanian grammatical gender corresponds “in principle” to natural gender, which translates primarily into the male/female/inanimate distinction, though this is “often not respected” (Graur 1966:57). In addition, Graur indicates that each noun gender corresponds to particular noun endings in the nominative singular indefinite form (Ibid. 60). These endings are later merged to describe three “traditional declensional classes” in Romanian, which suggests that declensional classes and genders should correspond to one another in some unambiguous way. As can be seen in Table 3, this is obviously not the case for Romanian: each declension contains nouns from more than one gender, showing a high degree of overlap in singular endings, particularly among neuter and masculine nouns (declension II).

**Table 3.** Romanian Declensional Classes (compiled from Graur 1966:81-82)

		<b>-ă</b>	<b>-a/-ea<sup>8</sup></b>	<b>-á/-eá</b>	<b>Special cases<sup>9</sup></b>	
<b>I</b>	F/M	<i>casă</i> ‘house’ (f)	<i>Toma</i> (m)	<i>para</i> ‘money’ (f)	<i>zi</i> ‘day’ (f)	
		<i>tată</i> ‘father’ (m)	<i>Mircea</i> (m)	<i>lulea</i> ‘pipe’ (f)	<i>tanti</i> ‘aunt’ (f)	
<b>II</b>	M	<b>-u/-w</b>	<b>-j /-C<sup>j</sup></b>	<b>-C</b>		
		<i>codru</i> ‘field’	<i>tej</i> ‘lime tree’	<i>nuc</i> ‘walnut tree’		
	<i>bóu</i> ‘ox’	<i>ochi</i> [ok <sup>j</sup> ] ‘eye’				
	N	<b>-u/-w</b>	<b>-i/-j/-C<sup>j</sup></b>	<b>-C</b>	<b>-o</b>	<b>-ú</b>
<i>lucru</i> ‘thing’		<i>alibi</i> ‘alibi’	<i>amurg</i> ‘dusk’	<i>apropo</i>	<i>atu</i>	
<i>cadou</i> ‘gift’	<i>paí</i> ‘straw’		‘by-	‘ace’		
<i>óchi</i> [ok <sup>j</sup> ] ‘fried egg’			the-	way’		
<b>III</b>	F/M/N	<b>-e</b>			<b>Special cases</b>	
		<i>sare</i> ‘salt’ (f),			weekdays in	
		<i>soare</i> ‘sun’ (m),			-C <sup>j</sup> and -j (f)	
<i>nume</i> ‘name’ (n)						

Graur (1966) notes that although Romanian nouns can be separated into the three declensional classes introduced above, “it would be more accurate to decline nouns based on their gender rather than their declensional class ... there is a declension for masculine nouns and one for feminine nouns, [with] the neuter nouns following the masculine in the singular, and the feminine in the plural” (p.

82, our translation). This statement accurately describes the division of Romanian nouns assuming the gender of a noun is clear. However, if we consider the fact that inanimate nouns are not exclusively assigned to the “neuter” gender, then this statement says nothing about how one can predict which inanimate nouns should “follow the masculine in the singular and the feminine in the plural.”

Attempts at setting up formal distinctions, such as the declensional classes described in Table 3, lead to circular arguments: the declensional classes outlined above overlap greatly with the noun genders, which is why Graur (1966) suggests that the various singular and plural nominative/accusative and genitive/dative case forms should be derived based on the gender rather than the declensional class of the noun. At the same time gender classes are defined semantically, and, as we have shown above, semantic gender distinctions do not account for all of the nouns.

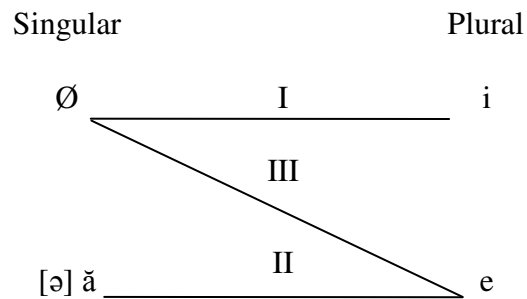
Corbett (1991) states that “gender agreement provides the basis for defining gender and establishing the number of genders in a given language” (Corbett 1991:105). He adopts Zaliznjak’s (1964:30) notion of “agreement class” in order to determine the number of genders in a language: nouns are in the same agreement class if, given the same conditions (e.g. same case/number, agreement domain and target), they take the same agreement form (Corbett 1991: 147-8). Corbett argues that although Romanian has three agreement classes, corresponding to the patterns of agreement for each of the three traditional



genders, these three agreement classes are not necessarily genders as they are in other languages, such as German or Latin. Stating that Romanian has three genders implies that agreement on targets (adjectives, demonstratives, etc.) shows a three-way distinction, which is not the case.

To solve this problem Corbett introduces the concepts of *controller* and *target* genders. Controller genders are those into which nouns are divided, and target genders are those marked on adjectives, demonstratives, numerals, etc. (Ibid.150-2). Romanian, then, has three controller genders and two target genders, corresponding to the traditional three genders and two agreement patterns, respectively. This is represented schematically in (2)<sup>10</sup>. For example, controller gender I (“masculine”) triggers agreement in *-i* on a plural adjective target, while controller gender III (“neuter”) triggers agreement in *-e* on the same target:

(2) *Gender Mapping in Romanian*



Corbett’s account is theoretically more sophisticated than the previous descriptions of the Romanian gender system, and it is superior in that it formalizes

the behavior of Romanian nouns. It posits three classes of nouns marked in the lexicon—the controller genders—and agreement mapping rules from the singular to the plural in the syntax, as represented in (2). This is attractive because it avoids the need to provide that one gender (neuter) patterns with another gender in the singular, and with a different gender in the plural. Instead, the three controller genders have mapping rules that map each of them to the same two target genders. However, this account still requires a three-to-two mapping between genders and agreement patterns, which is essentially the account provided in Graur (1966).

Finally, we turn to Farkas's (1990) analysis. Farkas's main purpose is not to decide whether Romanian has two genders or three, but rather to account for a case of apparent feature mismatch ('disagreement') in Romanian<sup>11</sup>. She provides a rule-based account for Romanian gender, assuming a three-way distinction using a single gender feature, [ $\pm$ fem]: traditional "masculine" nouns are [-fem], "feminine" nouns are [+fem], and "neuter" nouns are underspecified for gender. The patterning of the neuter nouns with masculines in the singular and feminines in the plural is obtained via feature filling rules, whereby nouns that are underspecified for the Romanian gender feature are assigned [-fem] values in the singular and [+fem] values in the plural. Farkas proposes two feature-filling rules to account for this pattern. First, she proposes the *Feature Co-occurrence*

*Restriction* rule, given below in (3), which provides that a noun that is underspecified for gender will be assigned a [+fem] value in the plural.

(3) *Feature Co-occurrence Restriction* rule (FCR):

$$[<[+N], [-V]> [+pl]] \rightarrow [+fem]$$

To account for the patterning of underspecified nouns in the singular, Farkas proposes a feature filling rule, the *Feature Specification Default* rule, given below in (4), which states that the default gender specification is [-fem]:

(4) *Feature Specification Default* rule (FSD):

$$[ \quad ] \rightarrow [-fem]$$

Farkas's analysis differs significantly from the other two described above. In particular, it avoids the syncretism in gender classification by positing an underspecified noun class for traditional neuters. This noun class, together with the feature filling rules FCR and FSD predict that neuter nouns will be masculine in the singular, feminine in the plural. This analysis maintains the stability of traditional masculine and feminine nouns which do not change gender class between the singular and the plural. However, the major drawback of this account is that it still does not predict what determines gender classification in the first place, namely which noun will be [-fem], [+fem], or underspecified for [ $\pm$ fem]. As a consequence, much of the classifying work is accomplished through the use of diacritics, whose overabundance in the system poses a serious learnability challenge—the learner is expected to do a lot of rote memorization.

In summary, the three-gender analyses in Graur (1966) and Corbett (1991) make the right generalizations with respect to agreement patterns; however, they do so in a purely descriptive fashion. The main problem with these accounts is that they lack predictive power with respect to agreement. They simply stipulate that mapping between three genders and two agreement patterns involves either “neuter” nouns patterning with “masculines” in the singular and with “feminines” in the plural (Graur 1966), or one of the controller genders mapping on to one target gender in the singular, and to another in the plural (Corbett 1991). Farkas’s (1990) analysis makes an important contribution regarding the predictability of the *behavior* of nouns, namely that masculines and feminines stay as such, but that neuters pattern with the masculines in the singular and with the feminines in the plural. Even so, this analysis does not provide a way of predicting noun classification, which is ultimately what we want to know. Given the lack of predictive power at different levels of these analyses, we now turn to alternative two-gender accounts of Romanian.

### **3.2 Two-gender Analyses**

In this section we discuss the precursors to the two-gender analysis of Romanian that we develop in section 4. Although the proposals discussed here have not been completely worked out by their respective proponents, they maintain that the nominal lexicon of Romanian can be described as a two-gender system. Hall

(1965) bases his argument for two genders on the syncretism in the number paradigm of Romanian, and Farkas (1990) on the syncretism in agreement patterns. Both accounts seem to be more economical than the three gender accounts described above. Economy is achieved through a simpler nominal lexicon, as well as more straightforward agreement rules. There are two classes of nouns mapping onto two agreement patterns in both the singular and the plural. We show, however, that while Hall's account is compatible with our own two-gender analysis, an account based on Farkas (1990) is more compatible with a three- rather than a two-gender analysis.

One of Hall's (1965) arguments against the "neuter" as a synchronic grammatical gender is that it lacks its own morphological markers in both the singular and the plural, therefore the "neuter" is lost syntactically (this argument has been mentioned by several others, including Corbett 1991, Cobet 1983-84:93). We understand this argument to mean that if gender is relevant for agreement, which is part of syntax, then there should be as many genders as there are agreement patterns. In addition, Hall argues that there is no semantic basis for this "gender": "neuter" nouns are defined semantically as inanimate, but inanimates are distributed across the feminine and the masculine genders as well.

Hall's main argument for a two-gender system of Romanian is based on the number syncretism given in Table 2 above. He proposes that "neuter nouns" are better described as *heteroclites*, "in that their chief characteristic is that they

always belong, not necessarily to different grammatical genders, but to different inflectional classes in the plural as opposed to the singular” (Hall 1965:427). This observation will be crucial for our own proposal, since we will argue that the declensional class of nouns in the singular and in the plural, and thus the actual singular and plural forms of nouns, determines agreement.

As already discussed, Farkas (1990) provides an alternative three-gender account of Romanian nouns. As Farkas herself mentions in a footnote, her analysis lends itself to a two-gender interpretation of Romanian by assuming the privative opposition [+fem] vs. underspecified [0fem] (Farkas 1990:543, fn. 9). With this contrast, only feminine nouns are specified for gender (as [+fem]), while all other nouns are underspecified with respect to gender, providing a two-way underlying distinction. Using this as a starting point we will develop Farkas’s possible two-gender analysis of Romanian.

Recall that in her three-gender analysis Farkas utilizes a *Feature Co-occurrence Restriction* rule which accounts for the behavior of some of the underspecified nouns (traditional neuters) that appear to be feminine in the plural. This is repeated here for convenience, and states that underspecified plural nouns will be feminine.

(5) *Feature Co-occurrence Restriction Rule* (FCR):

$$[<[+N], [-V]> [+pl]] \rightarrow [+fem]$$

This same rule can be applied to the possible two-gender analysis suggested, [fem]/underspecified. The key to this account is that only a subset of the underspecified nouns must undergo the FCR rule, namely all traditional “neuters”. Farkas suggests that in order for only these nouns (and not traditional “masculines”) to undergo rule (5), they would have to be marked with a diacritic that triggers the rule. Details about the exact specifications of the diacritic are not provided, except for its function as the rule trigger. With rule (5) and the diacritic in place, gender agreement follows the expected patterns. In the singular [+fem] nouns trigger agreement in *-ə* on adjectives, and underspecified nouns trigger agreement in  $\emptyset$ ; in the plural [+fem] nouns trigger agreement in *-e* on adjectives, while underspecified nouns trigger agreement in *-i* (the same patterns used by Corbett 1991:152, cf. (2)). The only underspecified nouns in the plural are traditional “masculine” nouns, since all of the “neuters” are [+fem] in the plural as a result of rule (5).

Therefore, under this account the lexicon contains nouns that are [+fem] and nouns that are underspecified for gender; among the latter there are nouns which are marked with a diacritic (traditional neuters) and nouns that are not (masculines). The problem with such an account is the use of the diacritic: it must be present on *every* traditional neuter noun. This class of nouns consists of nearly a third of all Romanian nouns, according to counts in Dimitriu (1996:129), therefore the diacritic is equivalent to maintaining a third gender. Furthermore,

the only function of the diacritic is to trigger the application of the FCR rule. This gives diacritics a greater power than they are intended to have, since they are features that attach to a particular lexical item or a *small* group of lexical items whose main function is to signal exceptional behavior with respect to some grammatical process. This is technically equivalent to rote learning of exceptional words, which means that the categorization of roughly a third of Romanian nouns would have to be rote-learned under such analysis.

#### **4. Proposed two-gender analysis**

With the foregoing background, it is time to turn to our own analysis of Romanian gender. We propose that Romanian has two noun classes in the singular and in the plural, and that this categorization is not lexically specified. The division of nouns into classes in the singular is different from their division into classes in the plural. Class assignment is determined by the ending of the noun in the singular and the plural, semantic core notwithstanding. Once class membership is established for singular and plural nouns, agreement proceeds very straightforwardly. For a few subsets of exceptional nouns, as discussed in section 3.2.1, this analysis still appeals to diacritics, but the number of diacritic-bearing nouns is kept small. Furthermore, the diacritics used in our analysis serve two functions: they determine the plural form of the noun, and they therefore predict gender agreement. In what follows we outline the details of class membership in



the singular and in the plural, and then discuss how agreement works in a two-gender system.

#### **4.1 Class membership in the singular**

It is clear in all analyses of Romanian that masculine and neuter nouns are indistinguishable in the singular, both in their form (endings) and in their agreement pattern. Our analysis capitalizes on this lack of distinction and groups masculine and neuter nouns into a single class, separate from feminine nouns. Were it not for the plural, this would be the natural categorization of Romanian nouns and no controversy with respect to number of genders would exist.

We propose that there are two noun classes in the singular, and for the sake of simplicity, we call them class A and class B. Class A includes traditional feminine nouns, and class B includes traditional masculine and neuter nouns. Class membership is determined based on a semantic core and formal cues. Animate nouns are assigned class based on natural gender, with those denoting females in class A and those denoting males in class B. Some smaller semantic subclasses include the names of trees, which are in class B, and abstract nouns, which are in class A. The formal cues that determine class membership are the final segment of the nominative indefinite form of the noun. Class A includes nouns that end in *-ə* or *-e*, and class B includes nouns that have all other endings (consonant, *-i*, *-o*, *-u*)<sup>12</sup>. As is often the case when semantic features compete with

formal features, the semantic features override the formal ones (Corbett 1991: 41); this particular ranking of features is also attested for other Romance languages (Corbett 1991:58, Tucker et al. 1977; Harris 1991). Table 4 provides examples of nouns in each class.

**Table 4.** Class A and class B nouns

CLASS A	CLASS B
boabă [boabə] ‘type of bean; grain’	buchet [buket] ‘bunch, bouquet’
umbră [umbrə] ‘shade, shadow’	borcan [borkan] ‘jar’
ghiară [gjarə] ‘claw’	buzunar [buzunar] ‘pocket’
casă [kasə] ‘house’	ficat [fikat] ‘liver’
cifră [tʃifrə] ‘number’	fior [fior] ‘shiver’
suliță [sulitsə] ‘spear; type of sword’	cartof [kartof] ‘potato’
culoare [kuloare] ‘color’	călcâi [kəlkij] ‘heel’
răbdare [rəbdare] ‘patience’	plai [plaj] ‘poetic land’
baie [baje] ‘bath’	cui [kuj] ‘nail (and hammer)’
ureche [ureke] ‘ear’	pârâu [piriɤ] ‘stream’
cheie [keje] ‘key’	panou [panow] ‘panel’
păreră [pərerə] ‘opinion’	taxi [taksi] ‘taxi’
sete [sete] ‘thirst’	alibi [alibi] ‘alibi’
	manto [manto] ‘coat’
	codru [kodru] ‘field’

## 4.2 Class membership in the plural

Traditional analyses assume that gender class determines a noun's plural form, while we take the opposite stance, namely that plural forms determine gender class. Our position is supported by the fact that in traditional three-gender analyses there is limited predictability of plural endings for nouns in the same class, clearly showing that gender specification alone does not predict plural form. For example, although the majority of traditional feminine nouns ending in *-ə* form the plural in *-e* (6a, b), quite a few form the plural in *-i*, as shown in (6c-e):

- (6)
- |    |                    |                |
|----|--------------------|----------------|
| a. | fată -- fete       | 'girl'         |
| b. | masă -- mese       | 'table'        |
| c. | bucată -- bucăți   | 'piece, chunk' |
| d. | ușă -- uși         | 'door'         |
| e. | flacăra -- flăcări | 'flame'        |

In fact, with the exception of traditional masculines, all of which take the plural marker *-i*, there are very few feminine and neuter nouns for which gender classification alone can predict plural form. For example, feminine nouns ending in *-e* take the *-i* plural marker seen above. As we mentioned previously, there are also feminine nouns ending in stressed *-á* or *-eá* which take the *-le* plural marker, and also there are neuter nouns ending in a stressed *-í* and borrowings from

French ending in *-ow* which take *-uri* in the plural. Notice that in each of these cases the plural ending is determined by the noun's ending rather than its gender class, which supports our claim that the plural forms determine class membership in the plural, rather than the other way around.

We propose that there are two noun classes in the plural: class C and class D. Class C includes traditional masculine nouns, and class D includes traditional feminine and neuter nouns. As is the case with singular nouns, class membership is determined based on semantic and formal cues. Nouns denoting males and trees are in class C, and nouns denoting females and abstract nouns are in class D. The formal cues which determine class membership are the plural noun endings, *which are the actual plural markers*. We show evidence below for Romanian possessing two plural markers in *-i*, noted here as *-i<sub>1</sub>* and *-i<sub>2</sub>*. Plural nouns ending in *-i<sub>1</sub>* are assigned to class C, while nouns taking all other plural markers (*-e*, *-uri*, *-i<sub>2</sub>*) are assigned to class D. Given the close connection between class membership and plural markers, our analysis must include rules of plural formation. We show that the form of the plural—the selection of the plural marker—is predictable from formal and semantic features, and we can immediately classify nouns into classes C and D based on the plural form. Once this classification takes place, agreement proceeds straightforwardly.

Our argument for the existence of two plural markers in *-i* is based on both diachronic and synchronic factors. First, the *-i* plural marker of traditional

masculine nouns and the *-i* plural marker of feminine and neuter nouns have different origins, as shown in Table 5<sup>13</sup>. Although the origins of the feminine *-i* plural marker are disputed, and we will not take a position here with respect to the marker's likely source, it is clear that it is not a matter of simple phonetic development from Latin. The Latin second declension nominative plural ending *-ī* produced Romanian *-i* by regular sound change, while the Latin first declension nominative plural ending *-ae* produced Romanian *-e*, which is the plural marker for the majority of traditional feminine nouns.

**Table 5.** Comparison of Latin and Romanian forms.

Latin		Development		Romanian	Gloss
	Singular	Plural		Plural	
Masculine	<i>socer</i>	<b>-ī</b>	>	socr-i [sokri]	'in-law'
	<i>oculus</i>	<b>-ī</b>	>	och-i [ok <sup>l</sup> ]	'eye'
Feminine	<i>barba</i>	<b>-ae</b>	→	bărb-i [bərb <sup>l</sup> ]	'beard'
	<i>fuga</i>	<b>-ae</b>	→	fug-i [fudʒ <sup>i</sup> ]	'run, jog'
	<i>lingua</i>	<b>-ae</b>	→	limb-i [limb <sup>l</sup> ]	'tongue, language'

Thus, one *-i* is a direct reflex of Latin *-ī* (*-i<sub>1</sub>*), and the other (*-i<sub>2</sub>*) is not.

Second, the synchronic behavior of *-i* indicates two separate markers: they combine with different noun stems in systematic ways. Speakers do not have

access to diachronic information, but they do have access to the singular form of the noun.  $-i_1$  combines with nouns that denote a male or a tree, and those that end in a consonant or  $-u$  (class C nouns).  $-i_2$  combines with nouns that denote females or abstract nouns, and those that end in  $-ə$ ,  $-e$ , or  $-ju$  (class D nouns). Given that the synchronic motivation is uncovered via morpho-phonological analysis, we use a single  $-i$  when establishing plural formation rules, to which we now turn.

#### 4.2.1 Rules of plural formation

To establish the rules of plural formation we utilized Quinlan's C4.5 Decision Tree algorithm, the details of which are not crucial here<sup>14</sup>. Let us just mention that this algorithm takes input features and categorizes data according to those features that have the highest predictive power. We found that the following elements are indicative of the plural marker selected by each noun:

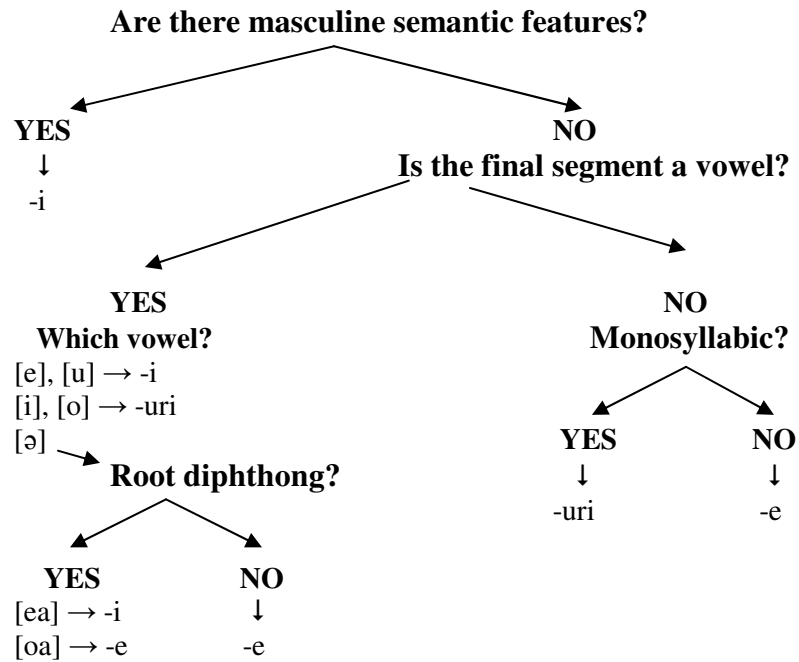
- the final segment of the nominative singular indefinite form
- the noun's semantics (masculine, tree)
- the mono- vs. polysyllabicity of the singular [indefinite] noun
- the presence and character of a root diphthong.<sup>15</sup>

The rules of plural formation are given below in (7) in the form of the decision tree obtained from the algorithm, as this is the most straightforward presentation.

We should note that this does not constitute a complete account of plural

formation rules for all nouns, as the cues determining plural marker selection for certain nouns have thus far been less transparent, as we discuss shortly.

(7) *Rules of plural formation*



The algorithm in (7) shows how the formal and semantic features rank with respect to each other in determining the choice of plural marker. Note that the first cut is based on simple semantic properties—whether the noun denotes an animate male or a tree,—thus reflecting the tendency for (typically coarse-grained) semantic features to override formal ones, as we noted in Section 4.1. Beyond this primary distinction, which is presumably subject to rote learning,

formal features predict the plural form of the noun and indirectly predict class membership. The *-i* plural markers are collapsed in (7), but recall that there are two such markers,  $-i_1$  and  $-i_2$ , according to the type of stem each attaches to. If the noun ends in a consonant or *-u* then this marker is  $-i_1$ ; otherwise it is  $-i_2$ .

The plural of a small number of class B nouns is not predicted by these rules. Some of these nouns are independent lexical items, but most can be subdivided into several small semantic categories. Under our proposal, all are marked with a diacritic specifying the plural marker they will take, but not their gender<sup>16</sup>. Since plural formation is independently needed, these subclasses of nouns have to be exceptionally marked under any analysis of Romanian and thus constitute a special case *not* just for our analysis. The following semantic categories also form the plural in  $-i_1$  (Graur 1966:58, Petrucci 1993:188).

- the names of letters of the alphabet: [doj de a] ‘two *as*’, [doj de tʃe] ‘two *cs*’.
- the names of musical notes: [doj de la] ‘two *las*’, [doj de mi] ‘two *mis*’.
- the names of months: [un januarije] ‘a (month) of January’.
- most names of numbers: [un patru] ‘a four’, [doj de zetʃe] ‘two tens’.
- (most) names of mountains and cities: *Ceahlăii* [tʃeahləjj] ‘the Ceahlăus’ (mountain, pl.), *Iașii* [jaʃij] ‘the Iașis’ (city, pl.).



- some names of plants and flowers: *trandafiri* [trandafir<sup>i</sup>] ‘roses’

*boboci* [bobot<sup>ʃ</sup>] ‘buds’.

Nouns from these semantic categories could have been included in our decision tree; however, they were left out for two reasons. First, since the initial decision relates to the presence or absence of masculine semantics, the plural forms for these nouns would have been correctly predicted, thus including them would have cluttered the algorithm needlessly. Second, these classes are very small, and most of the types of nouns they include (except for plants and flowers) do not usually lend themselves to being used in the plural. When they *are* used in the plural, they tend to form the plural in exceptional ways that do not actually change the form of the singular noun: for example, ‘two *as*’ is made plural in a construction such as *doi de a* [doj de a] ‘two of a’. Thus, our analysis does still make use of diacritics, but their use is much more limited than it would be under a proposal such as our development of Farkas’s (1990) two-gender account, and furthermore, this diacritic serves the purpose of determining plural form and indirectly agreement.

Our proposed rules of plural formation are consistent with those in Perkowski and Vrabie (1986), and Vrabie (1989; 2000), who provide a much more detailed account of plural formation for Romanian nouns. They propose additional semantic subclasses within each nominal class, and also rules based on

phonological characteristics of the nouns in the singular, in a similar vein to what we propose in this paper. Their findings support our analysis that once plural forms can be predicted, noun classification and agreement follow in a straightforward fashion.<sup>17</sup>

With the above plural formation rules in place, we can now categorize nouns into two classes in the singular and in the plural, as follows:

*Singular:*

- Class A: nouns ending in  $-\partial$  and  $-e$
- Class B: everything else

*Plural:*

- Class C: nouns ending in  $-i_l$
- Class D: everything else

Having established these noun classes, we now turn to agreement in a two-gender system.<sup>18</sup>

## **4.2 Agreement in a two-gender system**

We remind the reader that we define agreement as co-variation between the form of the trigger and the form of the target. Different agreement targets show different agreeing forms, but crucially, agreement with a particular noun class is consistent for all agreement targets (adjectives, numerals, demonstratives, etc.). The only difference among these agreement targets is the actual agreement

marker. For illustrative purposes, in (8) we show the co-variation in agreement between a noun and its attributive adjective, and we provide examples in (9). For example, when a singular noun ends in *-ə* or *-e*, an adjective modifying the noun will end in *-ə*.<sup>19</sup>

(8) *Co-variation in agreement markers*

	<i>Noun ending</i>	<i>Adjectival ending</i>
Singular	<i>-ə, -e</i>	<i>-ə</i>
	<i>-C, -u, -i, -o</i>	<i>-∅, -u</i>
Plural	<i>-e, -uri, -i<sub>2</sub></i>	<i>-e</i>
	<i>-i<sub>1</sub></i>	<i>-i</i>

- (9) a. *feli**e** bun**ă***    *feli-e bun-ə*    ‘good slice’  
    slice    good
- b. *gard bun*            *gard bun-∅*            ‘good fence’  
    fence good
- c. *mese bune*            *mes-e bun-e*            ‘good tables’  
    table    good
- d. *feli**i** bun**e***            *feli-i<sub>2</sub> bun-e*            ‘good slices’  
    slice    good



With the noun classes and the agreeing sets in place, we establish the following agreement rules, matching noun class to sets I or II:

**Table 7.** Agreement rules

Noun class	Agreeing form
A	set II, singular
B	set I, singular
C	set I, plural
D	set II, plural

The following examples illustrate how agreement proceeds straightforwardly in this two-gender system. We include details about class membership determination (noun endings).

**Table 8.** Agreement in a two-gender system

Noun form	Noun ending	Noun class	Agreeing form	N-Adjective pair		Gloss
masă table.sg	-ə	A	Set II, sg.	mas-ə table.sg	bun-ə good.sg	good table
felie slice.sg	-e	A	Set II, sg.	feli-e slice.sg	bun-ə good.sg	good slice
gard fence.sg	-C	B	Set I, sg.	gard fence.sg	bun-∅ good.sg	good fence
mese table.pl	-e	D	Set II, pl.	mes-e table.pl	bun-e good.pl	good tables
felii slice.pl	-i <sub>2</sub>	D	Set II, pl.	feli-i slice.pl	bun-e good.pl	good slices
garduri fence.pl	-uri	D	Set II, pl.	gard-uri fence.pl	bun-e good.pl	good fences
codri field.pl	-i <sub>1</sub>	C	Set I, pl.	codr-i <sub>1</sub> field.pl	bun-i good.pl	Good field

In this section we showed that with a small set of formal features and a minimal semantic core we can classify Romanian nouns into two classes in the singular and in the plural, and that once this classification is settled, agreement

proceeds very straightforwardly pursuant to agreement rules. The principal contribution of our analysis concerns the classification of nouns in the plural, as it is in this paradigm that the gender controversy resides for Romanian. Our analysis is symmetrical in that for both numbers we rely on the form of the noun to determine class membership. In the singular, the ending of the singular noun determines whether nouns will be in class A or B, and in the plural the ending of the plural noun, which happens to be the plural marker, determines whether nouns will be in class C or D. To this end we provided rules of plural formation, which are dependent on a small set of formal and semantic cues. Once we know the plural forms we can classify nouns into classes. This is the first time that such an analysis is proposed for Romanian, capitalizing on rules of plural formation to determine class membership and, indirectly, agreement in the plural. This is an important result, as speakers must know how to form the plural regardless of gender, and the fact that they can use the same information for gender agreement makes this analysis more plausible. Basically, our analysis utilizes the information that is independently available and necessary, without creating a burden for the language learner and introducing additional categories that may require more motivation.

## 5. Evaluating the analyses

It is now time to bring together the analyses considered here to determine which of them best explains the Romanian gender system. Both two- and three-gender analyses rely on the same semantic core for noun categorization: nouns denoting males, females, trees, abstract nouns, and a few others such as names of cities and mountains. Beyond this semantic core, traditional three-gender analyses do not have a principled way of categorizing nouns. Even Farkas's (1990) three-gender account, which differs from the other three-gender accounts discussed here, does not have a means to predict class membership. In all such accounts, feminine nouns are formally identified by the same features as in our proposal, namely the final vowels *-e* and *-ə* in the singular, but masculine and neuter nouns are classified arbitrarily as masculine and neuter, since they are indistinguishable from each other in the singular.<sup>20</sup> Their formal features would classify them as the same gender. The proposed two-gender analysis uses this generalization and classifies nouns into two classes in the singular and the plural, and these classes express the natural division of nouns based on their form, as well as their relationship to agreement: the same noun forms trigger the same agreement.

Our analysis is more parsimonious, as speakers need only look to a small set of semantic features and to the form of the noun in the singular and the plural in order to determine agreement. The tables below compare how agreement works in a two- versus a three-gender system. Notice that in the two-gender



system noun forms that trigger the same agreement are in the same noun class.

The behavior of the traditional neuter nouns is emergent, which is to be expected given the form of these nouns in the singular and in the plural. There is no need to mark a separate third gender. This is a generalization that cannot be captured in a three-gender analysis.

**Table 9.** Agreement in a two-gender system

	SINGULAR		PLURAL		
<b>A</b>	trandafir	<u>frumos</u>	trandafiri	frumoși	<b>C</b>
	<i>rose</i>	<i>beautiful</i>	<i>rose</i>	<i>beautiful</i>	
	palton	<u>frumos</u>	paltoane	<u>frumoase</u>	<b>D</b>
	<i>coat</i>	<i>beautiful</i>	<i>coat</i>	<i>beautiful</i>	
<b>B</b>	casă	frumoasă	case	<u>frumoase</u>	<b>D</b>
	<i>house</i>	<i>beautiful</i>	<i>house</i>	<i>beautiful</i>	

**Table 10.** Agreement in a three-gender system

	SINGULAR		PLURAL	
<b>M</b>	trandafir	<u>frumos</u>	trandafiri	frumoși
	rose.M	beautiful.M	rose.M	beautiful.M
<b>N</b>	palton	<u>frumos</u>	paltoane	<u>frumoase</u>
	coat.N	beautiful.M	coat.N	beautiful.F
<b>F</b>	casă	frumoasă	case	<u>frumoase</u>
	house.F	beautiful.F	house.F	beautiful.F

The following table allows for a simple evaluation metric of the two types of analyses. It includes the following criteria:

- Rote memorization. In any linguistic analysis, the more we can predict, the smaller the burden on the language learner. This criterion evaluates how much of nominal categorization is predictable, and how much must be memorized (i.e. via the use of diacritics).
- Semantics. Semantic distinctions in categorization are learned relatively early (e.g., Karmiloff-Smith 1979; Snyder and Senghas 1997; Suzman 1999), but these distinctions are never fine-grained—they typically cover the difference in natural gender and animacy, thus corresponding to the conceptual categories learned in early cognitive development (Mandler 2000). Beyond these coarse-grained features, overreliance on semantics in determining gender categories greatly increases the neuter gender class in some three-gender analyses (Graur 1966), as there are many non-neuter inanimate nouns (cf. discussion in section 3.1.2, where Graur (1966) acknowledges that using semantics works ‘in principle’).
- Noun forms. This criterion evaluates how much we can predict based on the formal characteristics of nouns, both singular and plural. In our account we rely heavily on form to categorize nouns, while in three-gender accounts it is unclear how much of a role noun form plays (presumably none at all in the plural, and

perhaps some in the singular, i.e. feminine nouns end in  $-ə$  or  $-e$ ). In Farkas's (1990) two-gender account we can assume that singular forms do play a role, as feminines are separated from other nouns, which are underspecified, but plural forms are not predicted and play no role.

- Agreement (mapping from trigger to target). In traditional three-gender analyses there is a complex mapping of agreement trigger to target, with neuter nouns mapping to masculine agreement in the singular and feminine in the plural. In two-gender accounts this mapping is straightforward.
- Parallelism with other Romance systems. Other Romance languages such as French and Spanish have two lexically specified nominal classes in the singular and plural. Our account brings Romanian closer to the rest of Romance at this surface level. At the lexical level, Romanian is different from other Romance languages, with no lexically determined noun classes.

**Table 11.** Comparison of the analyses

<b>Criterion</b>	<b>Proposed 2-G analysis</b>	<b>Farkas's 2-G analysis</b>	<b>Farkas's 3-G analysis</b>	<b>3-G analyses</b>
rote memorization (diacritics)	minimal	up to 30% of the lexicon ("diacritics")	up to 30% of the lexicon	up to 30% of the lexicon
contribution of semantics	minimal (small semantic core)	minimal (small semantic core)	minimal (small semantic core)	overgenerates (in some analyses)
predictive power of singular noun endings	very high	unclear	unclear	unclear
predictive power of the plural form	high	non-existent	non-existent	non-existent
mapping from trigger to target	direct	direct	direct	complex
parallelism with other Romance gender systems	yes	yes	no	no

Our proposal clearly fares better overall. It requires less rote learning and relies on fewer diacritics than any of the analyses considered here. The diacritics we

have to use are minimal and serve a dual purpose, indicating the choice of plural marker and indirectly predicting class membership and agreement. In a three-gender analysis and in Farkas's two-gender analysis, the neuter gender would have to be marked with diacritics to separate it from the masculine, and this gender comprises roughly 30% of the nominal lexicon of Romanian (Dimitriu 1996). Semantic features play a role in both types of analyses, but in some analyses (Graur 1966) semantics overgenerates. Noun endings in the singular and the plural have high predictive power in the proposed two-gender system which makes use of independently needed morphophonemic rules (plural formation). In three-gender systems such rules are not capitalized on, making these systems less parsimonious. With respect to agreement, three-gender systems, with the exception of Farkas (1990), present us with an intricate mapping from agreement trigger to target, while in the two-gender system this mapping is straightforward. And finally, on a less important dimension, our proposal brings the nominal system of Romanian closer to other Romance languages at the surface level, where nouns are categorized in only two classes.

## **6. Conclusions and outstanding questions**

This paper presented and analyzed core principles of gender assignment in Romanian, arguing that a two-gender system, like in other Romance languages, adequately accounts for the principles of gender categorization in this language.

The starting point for our investigation is the questionable status of the neuter gender in traditional analyses of Romanian. The neuter does not have its own markings or agreement pattern, being identical to the masculine in the singular and to the feminine in the plural in both these dimensions. Our analysis capitalizes on these facts and categorizes nouns into two classes in the singular and the plural. Nouns in each class share the same declension, namely noun endings (singular nominative indefinite for the singular, and plural markers for the plural). As actual plural forms determine class membership in the plural, and indirectly agreement, we provide rules of plural formation which are established based on formal features of the nouns and a small semantic core. Gender agreement is straightforwardly predictable once the noun classes are established. Agreement rules map each of the two genders in the singular and the plural to a specific set of agreeing forms.

Our proposal provides a more economical system overall. First, we claim that there are only two genders in the singular and the plural, predictable from a small semantic core and on formal properties of the nouns, namely the noun endings in singular and in the plural, as well as syllable count. Crucial to our account is that singular and plural gender assignment is established independently. Thus, unlike some other gendered languages, where the gender in the singular predicts the gender in the plural, and the plural form may not be directly relevant, in Romanian, the gender distinction in the plural is predicted

from the form of the *plural*, not from the singular. A speaker of Romanian therefore needs to know the form of the plural in order to categorize the noun as belonging to one of the two available classes. But since the plural form is needed independent of gender, the morphological features dictating plural formation have a direct bearing on syntax. To our knowledge, ours is the first proposal maintaining a tight correlation between declensional class features (specifically, features determining plural formation) and agreement. By maintaining such a connection we are able to reduce the number of diacritics introduced in the lexicon.

In addition to reducing the memory load in the gender-learning process, the proposed analysis has a number of other advantages. By showing that Romanian has a two-gender system, we can bring it closer to all the other Romance languages in which nouns divide into only two classes. As a result of the two-way distinction proposed here, agreement mapping rules from two genders to two agreement patterns become more straightforward. Finally, the prospect of such an analysis creates new analytical possibilities for other gender systems: it is conceivable that complex gender systems of other languages could be simplified if gender and number are dissociated and the issue of gender classes is raised independently for each number.

Of course, some issues remain to be dealt with in the future. Two issues particular to Romanian call for further investigation. One of these is the high

degree of variation in the choice of plural markers. For example, traditional neuter nouns *vis* ‘dream’ and *defileu* ‘gorge’ can have either the *-e* or the *-uri* plural markers, while traditional feminine nouns *monedă* ‘coin’ and *boltă* ‘arch’ can take either the *-e* or *-i<sub>2</sub>* plural markers (see also Vrabie 1989:401). There are no traditional masculine nouns which show this variation. It would be interesting to see the direction of this trend, but note that even with the variation the respective nouns remain in the same class, namely class D, so the analysis set forth in this paper would continue to apply. Second, agreement with conjoined NPs (Farkas and Zec 1995, Wechsler *to appear*, Sadler *to appear*) needs to be explored from the perspective of a two-gender system. While for combinations of male/female animate nouns there is virile agreement (agreement indexing features [+human, + male]), as in (10), agreement for different combinations of inanimate nouns shows different patterns. Only combinations of traditional masculine nouns result in a masculine agreeing form, while all other combinations result in a feminine agreeing form, as in (11).

(10) Animate: virile agreement

Pisica        și        câinele        sunt uzi.

cat.DEF[F] and dog.DEF[M] are wet.M.PL.

‘The cat and the dog are wet.’



(11) Inanimate agreement

Gardul            și    scaunul            sunt    albe.

fence.DEF[N]    and    chair.DEF[N]    are    white.F.PL.

‘The fence and the chair are white.’

This paper has concentrated on the analytical challenges particular to Romanian. However, we believe that the results achieved here, in keeping the gender system more parsimonious and in appealing to salient morphosyntactic cues readily available to young language learners, we have also touched upon the general issues of morphological relevance which now await further exploration.

## Appendix A.

**Case markers.** Romanian has five cases: nominative, accusative, genitive, dative, and vocative. The nominative and accusative cases (N/A) have the same form, as do the genitive and dative cases (G/D). In the plural, genitive, dative, and vocative forms are the same for all nouns (the suffix *-lor* attached to the nominative/accusative plural form). The vocative case is mostly used with animate nouns. We provide the definite forms in the table below.

	Singular			Plural			
	N/A	G/D	Voc	N/A	G/D	Voc	
M	brad- <u>ul</u>	bradul- <u>ui</u>	Bradul <u>e</u> !	brazi-i	brazilor	Brazilor!	fir
N	gard- <u>ul</u>	gardul- <u>ui</u>	Gardul <u>e</u> !	garduri- <u>le</u>	gardurilor	Gardurilor!	fence
F	mas-a	mes-ei	Maso!	mese- <u>le</u>	meselor	Meselor!	table

## Appendix B.

**Traditional masculine nouns ending in *-a*, *-e*** (most examples from Graur 1966:82).

1. *tată* 'father'
2. *pașă* 'pasha'
3. *popă* 'priest'
4. *vlădică* 'messenger, guard (?)'
5. *papă* 'Pope'

6. *Toma, Mina, Zaharia, Mircea, Costea* — proper names in /-a/
7. *Dănilă, Păcală, Tîndală, Nicoară* — proper names in /-ə/
8. *Gheorghişă, Peitrică, Ionică, Costică, Jenică*, etc. --proper names formed with feminine diminutive suffixes /-itsə/ or /-ikə/.

### Appendix C.

#### Traditional feminine nouns with plural in *-uri*

1. *dulceață /dulcețuri* ‘jam; types of jam’
2. *mîncare/mîncăruri* ‘food/types of food’
3. *carne/cărnuri* ‘meat/types of meat’
4. *mătase/mătăsururi* ‘silk/types of silk’
5. *marfă/mărfuri* ‘merchandise/types of merchandise’
6. *iarbă/ierburi* ‘grass/types of grass’
7. *blană/blănuri* ‘fur/types of fur’
8. *greață/grețuri* ‘nausea/repetitive episodes of nausea’
9. *otravă/otrăvuri* ‘poison/types of poison’
10. *sare/săruri* ‘salt/types of salt’
11. *lînă/lînuri* ‘wool/types of wool’ (also plural in *lîni* and *lîne*)
12. *gîlceavă/gîlcevuri* ‘bickering’
13. *leafă/lefuri* ‘wages’
14. *vreme/vremuri* ‘weather; time’

15. *gheață/ghețuri* ‘ice’  
 16. *lipsă/lipsuri* ‘lack’  
 17. *ceartă/certuri* ‘fight, quarrel’  
 18. *treabă/trebur* ‘work, task’

## Appendix D.

### Demonstratives

D.A. in this table indicates *demonstrative adjective*.

	Singular		Plural	
SET	Dem. Adj. (D.A. N)	Dem. Pn. and N D.A.	Dem. Adj. (D.A. N)	Dem. Pn and N D.A.
Set I	∅	D.A.+a	-i	D.A.+a
Set II	-ø, -a	D.A.+a	-e	D.A.+a

Demonstratives show a specific pattern of behavior. There are four types of demonstratives: of proximity (e.g. *acest* ‘this’), of proximity relative to another of the same kind (e.g. *cestălalt* ‘this other one’), of distance (e.g. *acel* ‘that’), and of distance relative to another of the same kind (e.g. *celălalt* ‘that other one’). The demonstratives relative to another of the same kind share the same behavior, and the remaining two share a different behavior. The former have the same form

both as pronouns and as demonstrative adjectives, while the latter do not.

Consider the following examples:

	SINGLE	RELATIVE TO OTHER OF SAME KIND
PROXIMITY	a. <i>acest bărbat</i> 'this man' this.M man	a. <i>cestălalt bărbat</i> 'this other man' this-other-one.M man
	b. <i>barbatul acesta</i> 'this man' man.DEF this.M	b. <i>bărbatul cestălalt</i> 'this other man' man.DEF this-other-one.M
	c. <i>acesta</i> 'this one' this.DEF.M	c. <i>cestălalt</i> 'this other one' this-other-one.M
DISTANCE	a. <i>acea femeie</i> 'that woman' that.F woman	a. <i>cealaltă femeie</i> 'that other woman' that-other-one.F woman
	b. <i>femeia aceea</i> 'that woman' woman.DEF that.F	b. <i>femeia cealaltă</i> 'that other woman' woman that-other-one.F
	c. <i>aceea</i> 'that one' that-one.F	c. <i>cealaltă</i> 'that other one' that-other-one.F

Notice that all forms of the demonstratives in the second column are identical within each gender. In addition, they have the typical endings that other adjectives have (e.g. zero for traditional masculine nouns, *-ǎ* [ə] for traditional feminine nouns). In the first column the pattern is different: the demonstrative adjective preceding the noun has the typical ending, while the demonstrative adjective and the demonstrative pronoun following the noun have the same form within each gender, and moreover they all end in *-a*. In fact, this *-a* ending appears for all such demonstratives, regardless of number/case and gender, but it is added to the regular ending corresponding to each noun class that the demonstrative modifies. Therefore, the demonstrative adjectives and pronouns are all based on the same regular form, to which the common *-a* ending is added for adjectives (when these follow the noun) and pronouns indicating a single referent (not relative to others of the same kind).

## References

- Baerman, Matthew (2004) Directionality and (Un)Natural Classes in Syncretism. *Language* 80: 807-827.
- Chitoran, Ioana (1992) Les Langues Romanes: Deux ou Trois Genres? (Le cas du roumain). *Les Langues Néo-Latines*, 4, 71-82.
- Chitoran, Ioana (2002) *The phonology of Romanian: a constraint-based approach*. Mouton de Gruyter.
- Cobeș, Doina (1983-84) Observații Privind Categoria Genului în Gramatica Românească de Pînă la Anul 1870. In *Annuaire de Linguistique et d'Histoire Littéraire XXIX*, 1-99.
- Corbett, Greville G. (1991) *Gender*, Cambridge University Press.
- Corbett, Greville G. and Norman M. Fraser (1993) Network Morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113-42.
- Dimitriu, C. (1996) *Gramatica Limbii Române Explicată*, Iași (România), Virginia Press.
- Farkas, Donka (1990) Two Cases of Underspecification in Morphology. *Linguistic Inquiry* 21, 539-550.

- Farkas, Donka and Draga Zec (1995) Agreement and Pronominal Reference. In Cinque G. and Giusti G., eds., *Advances in Roumanian Linguistics* 10, 83-101.
- Graur, Alexander, ed. (1966) *Gramatica Limbii Române*, Vol. 1 of 2, 2<sup>nd</sup> edition, Academy of the Socialist Republic of Romania, Bucharest.
- Hall, Robert (1965) The 'Neuter' in Romance: A Pseudo-Problem. *Word* 21, 421:427.
- Harris, James W. (1991) The Exponence of Gender in Spanish. *Linguistic Inquiry* 22, 27-62.
- Hockett, Charles F. (1958) *A course in Modern Linguistics*, New York, Macmillan.
- Karmiloff-Smith, Annette (1979) *A functional approach to child language*. Cambridge: Cambridge University Press.
- Levelt, Willem J.M. (1989) *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Levelt, Willem J.M., ed. (1993) *Lexical access in speech production*. Cambridge: Blackwell.
- Lumsden, John S. (1992) Underspecification in Grammatical and Natural Gender. *Linguistic Inquiry* 23, 469-486.
- Mallinson, Graham (1986) *Rumanian*, Croom Helm Descriptive Grammars.



- Mandler, Jean M. (2000) Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development* 1, 3-36.
- Perkowski, Jan L. and Emil Vrabie (1986) Covert Semantic and Morphophonemic categories in the Romanian Gender System. *The Slavic and East European Journal* 30:1, 54-67.
- Petrucci, Peter R. (1993) *Slavic Features in the History of Romanian*, USC Dissertation.
- Rosetti, Alexandru (1965) *Linguistica*, The Hague: Mouton.
- Rosetti, Alexandru (1973) *Brève Histoire de la Langue Roumaine des Origines à Nos Jours*, The Hague: Mouton.
- Sadler, Louisa (to appear). *Gender resolution in Rumanian*.
- Snyder, William, and Ann Senghas (1997) Agreement Morphology and the Acquisition of Noun-Drop in Spanish. *Proceedings of the Annual Boston University Conference on Language Development* 21, 2, 584-591.
- Stump, Gregory (2001) *Inflectional morphology*. Cambridge: Cambridge University Press.
- Suzman, Susan (1999) Learn Zulu the Way Children Do. *South African Journal of African Languages* 19, 134-147
- Tucker, G. Richard, Wallace E. Lambert, and André A. Rigault (1977) *The French Speaker's Skill with Grammatical Gender: An example of Rule-Governed Behavior*, The Hague: Mouton.

Vrabie, Emil (1989) On the Distribution of the Neuter Plural Endings in Modern Standard Romanian (MSR). *The Slavic and East European Journal* 33:3, 400-410.

Vrabie, Emil (2000) Feminine Noun Plurals in Standard Romanian. *The Slavic and East European Journal* 44:4, 537-552.

Weschler, Stephen (to appear) Elsewhere in Gender Resolution. In Kristin Hanson and Sharon Inkelas (eds.) *The Nature of the Word—Essays in Honor of Paul Kiparsky*. MIT Press.

Zaliznjak, Andrej A. (1964) K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom ruskom jazyke. *Voprosy jazykoznanija*, 4:25-40.

---

<sup>1</sup> We fondly dedicate this paper to David Perlmutter, from whom we have both learned so much, and whose unwavering confidence in this project has been a constant source of inspiration to us. In his inimitable manner, David has often told us that the two-gender analysis of Romanian is as clear as daylight. We hope that our readers will concur with his assessment.

---

For helpful discussions of this project, we are grateful to Eric Bakovic, Bernard Comrie, Grev Corbett, Donka Farkas, Jay Jasanoff, Andy Kehler, John Moore, Andrew Nevins, Keith Plaster, Sharon Rose, Steve Wechsler, and an anonymous reviewer. We regret that we were unable to take into account all of their excellent suggestions.

<sup>2</sup> Noun class and gender are different terms denoting the same concept (Corbett 1991:1); “class” and “gender” will be used interchangeably in this paper.

<sup>3</sup> Romanian nouns inflect for one of five cases: nominative, accusative, genitive, dative, and vocative. The vocative case is quickly losing ground to the nominative, and the other four cases have only two distinguishing forms: nominative/accusative and genitive/dative forms (Graur 1966:79). When inflecting for case, nouns can be singular or plural, and definite or indefinite. Definite forms have an enclitic definite suffix, while indefinite forms are preceded by a separate indefinite article.

<sup>4</sup> These have various realizations according to the morpho-phonological rules of the language.

<sup>5</sup> Syncretism has been a difficult issue for morphological theories and subject to heated debate. For our purposes, nothing hinges on a particular model of morphology with respect to syncretism—the crucial point, which no one seems to

---

dispute, is that syncretic clusters occur within paradigms but do not span the entire class of nouns/paradigm.

<sup>6</sup> Some early grammarians argued for as many as five genders (Eustatievici, Văcărescu, Golescu as cited in Cobeț 1983-84), whereas others argued for only two: *masculine* and *feminine*—either ignoring the neuters or saying that they are simultaneously masculine and feminine (Micu, Șincai and others as cited in Cobeț 1983-84). Arguments for only two genders arose in an attempt to be true to the etymological definition of “neuter” as “neither one nor the other of two”, thus also explaining the lack of correspondence in content or form between the Romanian and Latin neuter genders (Cobeț 1983-84:92). A fourth gender has also been proposed—the “personal gender”, which forms a subset of masculine and feminine (Rosetti 1965:85, Graur 1966:59-60). The “personal gender” is expressed by adding the particle *pe* before proper names and names of personified animals:

(i) *Am văzut-o pe Ioana*

have.1s see.past—3s.f.clitic on Ioana

‘I saw Ioana,’

and parallels Spanish personal *a*, cf.

(ii) *Lo vi a Juan*

‘I saw Juan’.

---

<sup>7</sup> The word *animal* ‘animal’ is neuter and it is animate. Mallinson suggests that this word could eventually be reinterpreted as masculine by a new generation of speakers (Mallinson 1986:247). There are some collective nouns denoting groups of people, but not individuals, which are also neuter, i.e. *popor* ‘people’, *tineret* ‘youth’.

<sup>8</sup> The *-a* and *-ea* endings on masculine nouns appear only on proper names.

<sup>9</sup> These are the only singular feminine nouns ending in [-i].

<sup>10</sup> After Corbett 1991:152, Fig. 6.1. (2) shows only the main agreement markers for each target gender:  $\emptyset$  and *-i* for one, and *-a* and *-e* for the other.

<sup>11</sup> The notion of ‘disagreement’ is used when the noun and the target have different genders. In Romanian when a demonstrative refers to an event it is feminine (*asta*), while the adjective describing the event is ‘masculine’ (*uluitor*): *Asta*[fem] *e uluitor*[masc]. ‘This is amazing.’ See Farkas (1990) and Lumsden (1992) for discussion.

<sup>12</sup> There is a subset of class B nouns (numbering around 80) which end in /-e/. Of these, 47 are assigned to class B via semantics, and the remaining have to be marked with a diacritic as belonging to class B.

<sup>13</sup> Many thanks to Ioana Chitoran for pointing this out and providing the examples. As used in Table 5, “>” indicates development of the Romanian plural ending via regular sound change from the corresponding Latin form, while “→”

---

indicates that the Romanian plural form was remade between Latin and Romanian.

<sup>14</sup> Visit <http://ling.ucsd.edu/~bateman/AppendixE.pdf> for decision trees we obtained in the singular and the plural, and for additional information regarding accuracy rates and features used in each case.

<sup>15</sup> We are grateful to Ioana Chitoran for discussion and comments regarding this feature. See Chitoran (2002) for further discussion.

<sup>16</sup> We thank an anonymous reviewer for pointing out examples of lexical items which would need diacritics under our analysis. These are traditional masculine inanimate nouns such as *cercel* ‘ear ring’, *chilot* ‘underpants’.

<sup>17</sup> To mention just a few of these rules, by positing 47 minor distribution rules for the *-e* plural marker in the traditional neuter class, “as many as 2,857 di- and polysyllabic nouns [are saved] from arbitrariness’ (Vrabie 1989:407). These rules include very specific endings, such as *-îst*, *-ăț*, *-cons + ru*, which are beyond what we attempted to accomplish in this paper. Our goal was to show that the plural can be predicted based on formal and semantic features, and that noun classification can be obtained based on the singular and plural forms. We believe we have accomplished that goal, and Perkowski and Vrabie’s (1986) and Vrabie’s (1989; 2000) rules for plural formation, while much more articulated, strongly support our analysis of the Romanian gender system.

---

<sup>18</sup> As an anonymous reviewer pointed out, the stability of the correspondences between class D in the plural and class A in the singular, and also between class B in the singular and class C in the plural, should be captured in the complete analysis of the operation of Romanian gender. Our proposal correctly predicts classes to which a noun will belong in the singular and the plural but does not currently attempt to formalize any correspondences between singular and plural classes.

<sup>19</sup> Some adjectives are invariable in form for all genders, thus occurring in just one form in the singular and one form in the plural, for example:

verde [verde] ‘green (all genders, sg)’

verzi [verzʲ] ‘green (all genders, pl)’

This is an example of low level syncretism, which does not pose a problem for our analysis since most adjectives distinguish gendered forms.

<sup>20</sup> Neuter nouns that have specific endings such as stressed *-i*, *-o*, or those borrowings from French which end in *-ow* constitute small exceptional classes that can be identified as different from masculine nouns.